RESEARCH ARTICLE

A Novel Approach for Predicting Remaining Useful Life and Capacity Fade in Lithium-Ion Batteries Using Hybrid Machine Learning

SADIQA JAFARI[ID][1], **YUNG-CHEOL BYUN**[ID][2], **AND SEOKJUN KO**[3]

[1]Department of Electronic Engineering, Institute of Information Science & Technology, Jeju National University, Jeju 63243, South Korea
[2]Department of Computer Engineering, Major of Electronic Engineering, Jeju National University, Institute of Information Science & Technology, Jeju 63243, South Korea
[3]Department of Electronic Engineering, College of Engineering, Jeju National University, Jeju 690-756, South Korea

Corresponding authors: Yung-Cheol Byun (ycb@jejunu.ac.kr) and Seokjun Ko (sjko@jejunu.ac.kr)

ABSTRACT Since lithium-ion batteries (LIBs) are essential to many different sectors, accurate estimates of their Remaining Useful Life (RUL) are necessary to maximize Battery Management Systems (BMS). In this study, we introduce an innovative approach that combines machine learning techniques to create a hybrid model, enhancing the precision and reliability of battery analysis. Our proposed model leverages the power of k-Nearest Neighbors (kNN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) algorithms to capture complex relationships and patterns in battery data effectively. Our major objective is to precisely estimate the residual energy and RUL of LIBs, allowing for the efficient evaluation of battery health and deterioration over time. We meticulously curate a comprehensive dataset comprising essential battery parameters, including capacity, voltage, cycle, and temperature. The proposed hybrid model achieves impressive results with an R2 value of 0.996457, a minimal RMSE of 0.016861, and a low MAE of 0.008956. Our analysis provides valuable insights for optimizing battery performance, informed maintenance planning, and enhancing energy storage system efficiency.

INDEX TERMS Charge cycles, lithium-ion batteries, RUL, capacity fade, battery performance, feature selection.

I. INTRODUCTION

Lithium-ion batteries are essential for our devices, powering various electronic and mobile devices with their remarkable features, including high energy density, rapid discharge rates, and resistance to memory development. In addition to increasing battery efficiency, accurate life prediction techniques reduce the possibility of unplanned failures. Determining these lithium-ion batteries' Remaining Useful Life (RUL) is essential for maintaining performance and enhancing battery management systems. RUL is the number of charge and discharge cycles under specific operating circumstances from new to EOL, with a typical capacity deterioration of 20% upon reaching EOL [1]. Several variables influence battery capacity aging, including charge and discharge profiles, ambient temperature, electrode materials, and capacity regeneration. Predicting the RUL accurately poses challenges due to the complexity and nonlinearity of the aging process. In order to obtain dependable RUL prediction, researchers have developed three main types of models: models that are data-driven, semi-empirical, and mechanism-based. Each category offers its unique approach to capturing the intricate dynamics of battery aging and providing insights for RUL estimation [2]. Mechanism-based models offer a quantitative approach to assessing capacity loss by simulating the battery system's electrochemical responses and aging mechanisms [3], [4]. The procedures entail dissolving transition metals from

The associate editor coordinating the review of this manuscript and approving it for publication was Mehrdad Saif[ID].

the cathode, mechanically crushing electrode materials, and creating and growing solid electrolyte interfaces [5], [6]. Nevertheless, mechanism-based models are unique to certain battery material systems and need a thorough comprehension of the intricate differential equations and underlying physical-chemical processes [7], [8]. By removing model parameters from accelerated battery aging data, semi-empirical models offer a more approachable option. These models employ linear, exponential, Eyring, and Arrhenius equations to integrate stress factors such as operating temperature, depth of discharge, charge-discharge rate, and State of Charge (SoC). However, designing optimal trials for accelerated aging research can be time-consuming and challenging, especially when exploring interactions between different aging processes [9], [10], [11], [12]. Utilizing pertinent attributes, the data-driven model makes use of past aging data. Changing the model weights to reduce the discrepancy between the expected and actual values improves prediction accuracy. This approach circumvents the need for intricate physical-chemical equations and offers greater flexibility in capturing the underlying patterns and dynamics of battery aging. Without depending on explicit mechanical insight, the data-driven model successfully learns from the data and makes correct predictions by utilizing machine learning and statistical approaches [13]. While more complex than the previously mentioned models, the intelligent algorithm model offers enhanced nonlinear fitting capabilities. Contrasting time series data characterize battery capacity degradation, and methods such as Recurrent Neural Networks (RNN) are well-suited for predicting such unknown series. RNNs excel at capturing complex temporal dependencies and can effectively model the nonlinear dynamics of battery degradation. Their ability to process sequential data and retain historical information makes them a valuable tool for accurately predicting and forecasting battery performance [14]. However, many review papers have been written on machine learning (ML)-based RUL prediction, often without delving into the foundational development of widely used ML algorithms. These evaluations, such as Gaussian Process Regression (GPR) [15] and Relevance Vector Machine (RVM) [16], offer valuable insights into the application of ML in RUL prediction. While the literature has explored various feature selection techniques that complement ML algorithms and demonstrated adequate performance in estimating battery lifespan, there remain research gaps. The predominant focus in the existing literature has been on predicting RUL, addressing the remaining operational life of batteries. Additionally, the phenomenon of Capacity Fade, reflecting the decrease in a battery's capacity due to usage and aging over time, has been explored. However, a critical research gap remains. In response to this gap, this paper introduces a novel hybrid approach, which simultaneously considers both RUL and capacity fade as dual target variables within a single predictive model. This endeavor seeks to offer a more comprehensive understanding of battery behavior and degradation patterns by jointly predicting RUL and Capacity Fade. The innovative integration of these two variables has the potential to unearth valuable insights that previous research has yet to fully explore. Our suggested prediction model makes use of state-of-the-art machine learning methods, such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), and k-Nearest Neighbors (kNN). Our main goal is to create a reliable and accurate prediction framework that can estimate capacity fading and RUL for lithium-ion batteries (LIBs) by utilizing these cutting-edge methodologies. The main contributions of this paper are outlined as follows:

- Introducing a new system by combining RUL and capacity fade as target variables in a machine learning model.
- Proposing a method to quantify capacity loss during battery cycles.
- Introducing Remaining Capacity Cycles (RCC) to analyze RUL and capacity degradation.
- Improved predictive models: developing more accurate predictive models for battery performance by incorporating RCC as a target variable.
- Comparing the performance of three model ML algorithms, RF and XGBoost, for battery RUL prediction.

The rest of the paper is organized as follows: Section II describes the brief related work of RUL battery prediction. Then, in section III, we describe experimental data and analysis. Section IV describes modeling methodology, including hybrid model structure, RUL and capacity fade estimation, combined RUL and capacity fade, and model output. Section V introduces the machine learning workflow. After the experimental setup and dataset used for assessment, section VI presents the experimental findings. In section VII, we conclude our study and outline our next steps.

II. RELATED WORK

Prior studies have made significant strides in advancing our understanding and techniques in the realm of battery RUL and capacity estimation. In this section, we review key research efforts and notable approaches that have contributed to the foundation of our investigation.

A. RUL OF PREDICTION BASED ON THE MACHINE LEARNING

The research has produced a new hybrid model that integrates the Broad Learning System (BLS) and the Long Short-Term Memory Neural Network (LSTM-NN) to predict the RUL of batteries. The proposed method aims to increase RUL prediction accuracy, which is a challenging task. The BLS algorithm extracts feature nodes from historical battery data, which have further been improved and have been used as inputs for the LSTM-NN. Experiments have been conducted using NASA and cycle battery aging data demonstrate a significant reduction in the required training data, highlighting the effectiveness of this approach [17]. Another cutting-edge

hybrid method for forecasting battery capacity and RUL integrates the Variational Modal Decomposition (VMD), Particle Filter (PF), and Gaussian Process Regression (GPR). The recorded battery capacity data has been split into distinct residual sequences and aging trend sequences by the VMD algorithm. The Posterior Feedback Confidence (PFC) approach has been utilized to ascertain the quantity of modal layers. To anticipate these sequences, prediction models based on PF and GPR algorithms have been created, providing a distinctive viewpoint on RUL prediction [18]. Furthermore, the paper has proposed a SoH estimation method for both in-service and retired batteries. This method has combined the Incremental Capacity Analysis (ICA) technique with an improved BLS network. It has been started with voltage data from constant current charging to create an IC curve with noise reduction. Correlated Health Indicators (HIs) have been extracted through Pearson correlation analysis and have been used for SoH estimation, facilitated by the Particle Swarm Optimization (PSO-BLS) network. Experiments involving various charging multipliers validate the method and have demonstrated its precision and robustness [19].

The article has introduced battery health monitoring for EV batteries that are sold commercially. Lithium-ion batteries are the primary power source for Electric Vehicles (EVs) and Hybrid Electric Vehicles (HEVs). These batteries are renowned for their long lifespan, high energy density, and effective charging and discharging. The study has addressed the importance of battery health monitoring and capacity estimation, particularly focusing on EVs and HEVs [20]. To evaluate battery SoH and capacity, the study has conducted a comparative analysis of various data-driven approaches, including Convolutional Neural Networks (CNN), feed-forward neural networks, and Long Short-Term Memory (LSTM) networks. These techniques have been applied to NASA's Li-Ion battery dataset, utilizing charging cycle data for voltage, current, and temperature. The study has found that LSTM-based machine learning outperforms other methods due to its inherent long-term memory capabilities, and it has been recommended to use LSTM for accurate battery health monitoring and capacity estimation [21]. The work has also concentrated on accurate State of Charge (SoC) calculation for lithium-ion batteries, which are widely utilized in electric vehicles. The study used thevenin-equivalent circuit theory to assure safe battery operation, and it evaluated SoC using MATLAB/Simulink-implemented Coulomb counting and Extended Kalman Filter (EKF) techniques [22]. Additionally, the paper has highlighted the need for proper battery maintenance and management systems that rely on metrics such as SoH and SoC for assessing battery performance and condition. It has been suggested a machine learning approach to extract key features from discharge curves for effective SoH and SoC estimation, with simulations demonstrating the method's effectiveness under various currents and temperatures [23].

TABLE 1. Typical specifications of NMC-LCO 18650 battery.

Specification	Value
Nominal Capacity	2600mAh - 3400mAh
Nominal Capacity	3000mAh (3.0 Ah)
Nominal Voltage	3.6 volts
Chemistry	Lithium-ion NMC-LCO
Maximum Continuous Discharge Current	10A - 20A
Operating Temperature Range	$-20°C$ to $60°C$ ($-4°F$ to $140°F$)
Dimensions	Diameter: 18.6mm, Length: 65mm

III. EXPERIMENTAL DATA AND ANALYSIS

Open-source data sets [24] are utilized for training and verifying the model suggested in this paper. Table 1 concisely represents the NMC-LCO 18650 battery's technical specifications. It provides information about various aspects of the battery, such as its nominal capacity, nominal voltage, chemistry, maximum continuous discharge current, operating temperature range, and dimensions. Each row in the table represents a specific specification, and the corresponding value for that specification is provided in the adjacent cell.

In general, the number of charge-discharge cycles it can undergo before its discharge capacity drops below a predefined failure threshold from the RUL of a battery refers to, typically set at 80% of the nominal capacity, marking the end of its useful life. However, the paper introduces a novel approach to defining RUL specific to our research context.

$$RUL_\alpha = M_{EOL} - M_{CC\alpha} \quad (1)$$

Equation (1) estimates the battery's RUL by subtracting the number of completed charge cycles from the estimated number of remaining effective operational cycles. Where RUL_α describes the RUL of the battery, M_{EOL} denotes the estimated number of remaining effective operational cycles, and $M_{CC\alpha}$ represents the number of completed charge cycles.

A. DATASET INFORMATION

The dataset consists of information from 14 batteries, each subjected to more than 1000 charge/discharge cycles at a temperature of 25°C, charging at a rate of C/2 and discharging at 1.5°C. The dataset contains 15064 rows, representing the results of these cycles for each battery. Each row provides essential features for analysis, including the cycle index, total discharge duration, duration of discharge from 3.6 V to 3.4 V, starting voltage during discharge, starting voltage during charging, duration of charging up to 4.15 V, duration of constant current charging, and total charging duration.

IV. MODELING METHODOLOGY

This paper presents a comprehensive hybrid model designed to predict battery capacity fade and RUL accurately. The model's structure is outlined in section (A), while section (B) delves into the intricacies of RUL and capacity fade estimation, encompassing capacity fade estimation and the

fusion of RUL and capacity fade information. Section (C) offers insights into constructing the output model and briefly introduces the overall framework. Our approach amalgamates a trio of machine learning algorithms kNN, RF, and XGBoost–primed for predicting RUL and capacity fade in LIBs. The method unfolds through a series of key steps, commencing with rigorous dataset preprocessing, including normalization, scaling, data cleaning, and feature engineering. Feature selection identifies the most pertinent battery parameters, ensuring the model focuses on the critical variables. The subsequent phase involves training the models with tailored hyperparameter optimization to accommodate the unique nuances of battery data. Ensemble techniques then amalgamate the predictions derived from these models, endowing the approach with heightened accuracy and robustness. In some instances, a meta-learner is introduced to enhance prediction refinement further. The method's resilience and generalizability make it valuable for battery management and optimization, providing actionable insights. Illustrated in Figure 1 is the proposed model's flowchart, delineating the four sequential steps: variable extraction, data preprocessing, model training, and the ultimate prediction of RUL and capacity fade for LIBs, thus facilitating effective battery management system operation and improved energy storage efficiency.

A. HYBRID MODEL STRUCTURE

In this study, we propose a hybrid model that combines three distinct machine learning algorithms: kNN, RF, and XGBoost, to create a versatile framework for battery analysis and prediction. The hybrid model begins with a comprehensive dataset preprocessing phase, involving tasks such as data normalization, scaling, cleaning, and feature engineering. Critical battery parameters are carefully selected through a feature selection process, ensuring a focus on variables crucial for predicting capacity fade and RUL. Subsequently, each of the three models (kNN, RF, and XGBoost) undergoes meticulous training with tailored hyperparameter optimization to align their parameters with the nuanced characteristics of battery data. To generate highly accurate predictions, ensemble techniques are employed to harmonize individual model predictions, thereby enhancing predictive accuracy and model robustness. In specific scenarios, the introduction of a meta-learner further refines predictions through an iterative process guided by rigorous evaluation metrics. This hybrid approach seamlessly integrates the strengths of kNN, RF, and XGBoost, capturing intricate relationships and patterns within battery data. Specifically, kNN, as a non-parametric algorithm, classifies data points based on the majority vote of their nearest neighbors, making it effective for data with clear patterns and clusters. RF, an ensemble method, combines multiple decision trees, each trained on a random subset of features and data points to reduce overfitting. High-dimensional data management, complicated interaction capture, and feature significance ranking are among RF's strong points. XGBoost,

a gradient-boosting algorithm, optimizes gradient descent and trains weak learners (decision trees) sequentially to minimize the loss function, making it well-suited for capturing complex relationships and handling large datasets. The proposed method takes input data that includes crucial parameters such as capacity, voltage, cycle, temperature, and other relevant battery information. This input data is processed through the hybrid model, allowing for the accurate diagnosis of the battery's condition and precise evaluation of RUL and capacity which Figure 2 shows a flow chart of our proposed method.

The hybrid model leverages the strengths of each base model, training them on the available data. Predictions from the base models are combined using an ensemble method, while a meta-learner generates the final prediction. This iterative refinement process aims to enhance the model's performance. The hybrid model captures complex relationships and patterns by integrating kNN, RF, and XGBoost, ensuring accurate RUL and capacity estimation within the battery domain. The hybrid model can be succinctly represented as follows in Equation (2):

$$H(x) = \frac{1}{3} \left(\alpha \cdot \text{kNN}(x) + \beta \cdot \text{RF}(x) + \gamma \cdot \text{XGBoost}(x) \right) \}$$

(2)

Here, $H(x)$ is the final prediction for input x, with α, β, and γ as weighting factors for kNN, RF, and XGBoost predictions, respectively. The averaging balances their contributions for accurate hybrid predictions. The model, denoted as $H(x)$, generates predictions for a given input x. It combines three machine-learning algorithms: kNN, RF, and XGBoost. Each algorithm has parameters, such as the number of neighbors, distance metric, and weights for kNN, or the number of trees and maximum depth for RF. Similarly, XGBoost involves parameters like the number of boosting rounds, learning rate, and selected features. We employ weighting factors, w_1, w_2, and w_3, to adjust the influence of each algorithm's prediction in our hybrid model.

B. REMAINING USEFUL LIFE AND CAPACITY FADE ESTIMATION

Estimating the life cycle is challenging and often dependent on either capacity or impedance. While a decline in impedance is often linked to a power fade, capacity fade is typically used in literature as a major signal for RUL. Batteries go through various cycles with various rest times at various rates of charge and discharge at various temperatures. Because its past is frequently unknown, a battery that has been used previously is considerably more complicated. Several manufacturers provide life cycle curves to indicate the number of cycles that can be anticipated for a constant depth of discharge. Life cycle testing is generally used to support these. Calendric aging, or battery resting, is sometimes divided from cyclic aging, which is thought to be more difficult. Despite this complexity, the academic community continues to work hard to develop creative and challenging
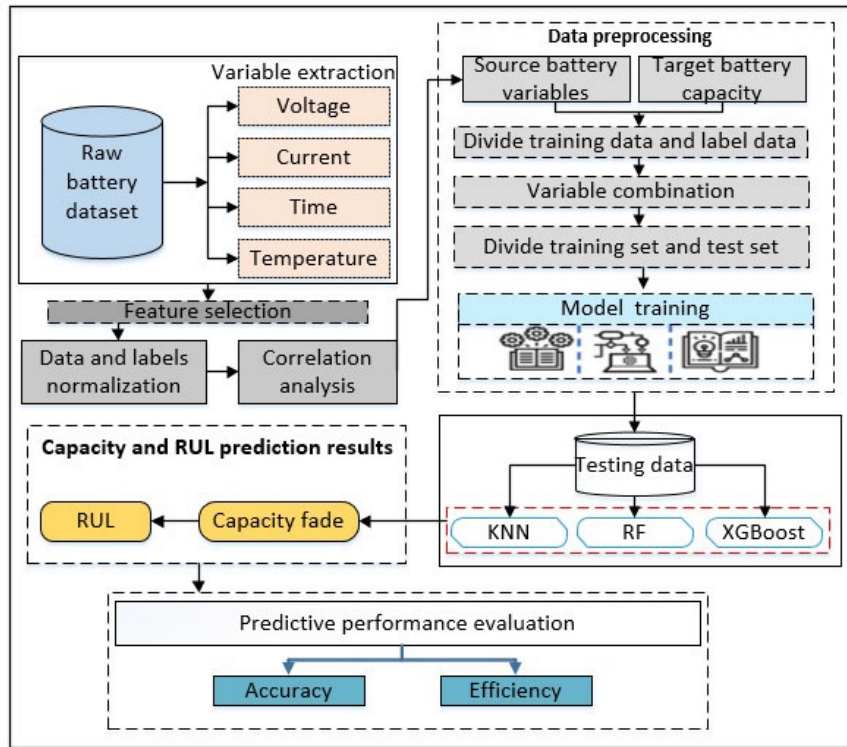
FIGURE 1. The proposed approach structure.

life cycle estimate (or RUL) computations based on a wealth of data [25], [26]. These are frequently separated into model-based approaches and data-driven methods. Model-based solutions are less dependent on past data and need an understanding of the LIBs degrading mechanism. Current research frequently makes use of the electrochemical model or comparable circuit model. In data-driven approaches, prospective rules are discovered by analyzing previous data, which are then used to forecast the course of occurrences. It is optional to have an in-depth understanding of a battery's degrading mechanism. Rules generated from past data can be used to forecast the deterioration. LIBs employed in automobiles can be in one of three states: charged, discharged, or standing. Following each charge, the battery alternates between standing and being in the discharged state. Standing will lengthen the battery life since the diffusion action causes the ion concentration to balance and the voltage to climb. We referred to it as the self-recharge phenomenon in restoration. As seen in Figure 3, vehicle LIBs often drain intermittently and (a) show continuous discharge and (b) show intermittent discharge. Occasional discharge can lengthen the life of a battery. When a pulse current is passed through a battery for an intermittent discharge, the battery relaxes for a while. As a result, the consumption and active material recovery in the diffusion process increase, which enhances battery performance. LIBs frequently require self-recharging since an automobile is typically in motion and at rest. Additionally, the RUL is impacted by the LIBs'

self-recharge strength as battery charge and discharge times alter [27].

1) CAPACITY FADE ESTIMATION

A systematic approach is employed to assess the capacity fade of LIBs, encompassing the following steps: first, the battery's initial capacity, which represents its original energy storage capability, is determined. The information can be acquired from the battery manufacturer or through characterization tests. Next, the battery's current capacity is measured at a specific time or after several cycles. Accurate capacity measurements can be obtained using discharge tests and impedance analysis counting techniques. After that, the capacity fade is computed by deducting the original capacity from the current capacity. The calculation estimates the capacity lost or faded over the defined period or cycles. Capacity fades, a vital parameter for evaluating the degradation of LIBs, is expressed as a percentage.

$$\text{Capacity Fade (\%)} = \left(\frac{\text{IC- CC}}{\text{IC}}\right) \times 100 \qquad (3)$$

The capacity fade, calculated as a percentage using Equation (3), provides a quantitative measure of capacity loss about the battery's Initial Capacity (IC) and Current Capacity (CC). The metric shows the battery's overall health and rate of degradation. Analyzing and interpreting the calculated capacity fade makes assessing the extent of energy storage capacity decline possible. Higher capacity fade values
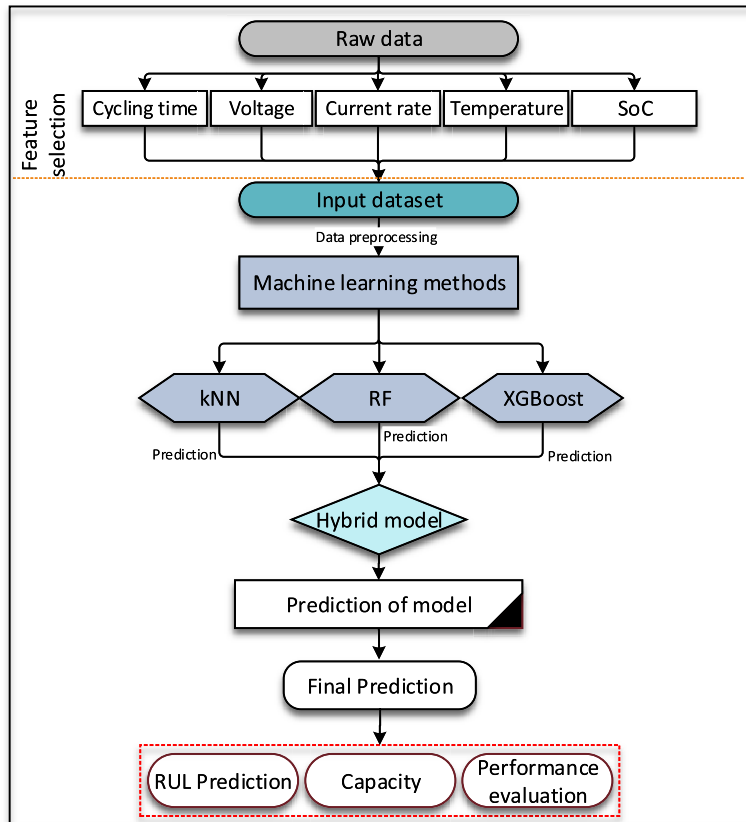
FIGURE 2. Flow chart for battery data analysis and prediction using machine learning and hybrid method.
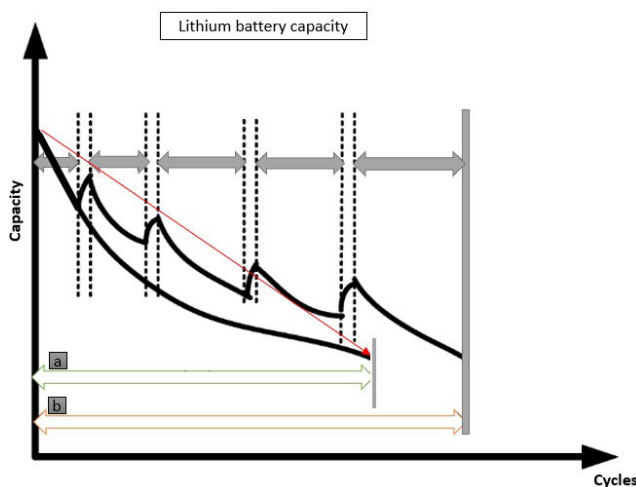

FIGURE 3. Lithium battery capacity changing with the discharge process.

signify a more substantial decrease in capacity, indicating increased degradation and potential issues with the battery's performance and longevity.

2) COMBINE RUL AND CAPACITY FADE
The combination of RUL and capacity fade in a battery analysis provides a comprehensive understanding of battery behavior and degradation. By merging these two parameters,

we can gain deeper insights into the battery's remaining operational life and capacity deterioration over time. RUL predicts the battery's remaining charge cycles or lifespan, while Capacity Fade quantifies the loss of capacity during operation. Considering both factors, the overall degradation pattern of the battery and its remaining capacity can be evaluated more effectively. One approach involves calculating the RCC for each cycle, representing the number of charge cycles required for the battery's capacity to degrade to a specified threshold. It can be achieved by dividing the capacity fade at each cycle by the average capacity fade per cycle and subtracting this value from the RUL defined as (4):

$$RCC = RUL - \frac{Capacity\_Fade}{Avg\_Capacity\_Fade\_Per\_Cycle} \quad (4)$$

The combined approach enhances battery management strategies and decision-making processes. It enables the planning of maintenance, replacement methods, and optimal battery utilization in a variety of applications, including consumer devices, renewable energy systems, and electric cars.

C. MODEL OUTPUT
Predicted values for the RUL and battery capacity fade are part of the model output. These predictions are made using fresh input data and the trained machine-learning models.

The result includes estimated values for the RUL, which shows how many charge cycles each battery has left, and the capacity fade, which shows how much capacity has been lost compared to the battery's starting capacity.

V. MACHINE LEARNING WORKFLOW

The machine learning workflow section outlines the step-by-step process of applying machine learning algorithms, including kNN, RF, and XGBoost. This section briefly describes each model we used in our proposed method.

A. KNN REGRESSION

For the kNN algorithm, we focused on tuning parameters such as $n-neighbors$ which determines the number of nearest neighbors to consider, weights to assign different weights to neighbors, and metrics to choose an appropriate distance metric. The choice of these parameters significantly impacts the algorithm's effectiveness in capturing patterns in our battery data. kNN is an effective and widely used method in pattern recognition for classification tasks. It functions by assigning a test point to one of its k nearest neighbors in the feature space, depending on the majority class of those neighbors. These neighbors are chosen from a group of training locations with well-known classifications. Lazy learning algorithms, such as kNN, locally estimate the function and postpone calculation until needed. Notably, it only requires an explicit training step besides storing the training data set. In regression tasks, kNN estimates the response of a testing point x_t by calculating a weighted average of the responses of the k nearest training points, x(1), x(2), ..., x(k), in the neighborhood of x_t. The weights assigned to each neighbor are typically calculated using a kernel function, which takes into account the proximity of the neighbors to the test point. This approach allows kNN to approximate the target variable based on the information from its nearby training examples. The equation (5) provides the estimated response value in kNN regression for a testing point x_t.

$$\hat{y}(x_t) = \frac{1}{k} \sum_{i=1}^{k} y_i \tag{5}$$

where the average of the responses (y_i) from the k closest training points nearby (x_t) is used to estimate the predicted value (denoted as $\hat{y}(x_t)$) for a testing point (x_t), the parameter k value controls how many nearest neighbors are taken into account.

B. RF REGRESSION

An ensemble learning method, RF regression, averages and integrates the results of many decision tree models. It is essential within the framework of our investigation to identify complex correlations and patterns in the battery data. By aggregating the predictions of individual decision tree models, RF regression offers improved accuracy and robustness in capturing complex relationships within the data [28]. RF employs a clever strategy to pursue enhanced predictive accuracy. It accomplishes this by assembling a multitude of decision trees, each crafted using a distinct random subset of predictors and a randomized training data subset. This introduction of randomness and diversity in the creation of decision trees bestows upon the RF model a remarkable ability: the capacity to adeptly amalgamate the forecasts generated by each tree. This amalgamation leads to a more resilient and precise estimation of the target variable. In our study, we thoroughly fine-tuned the hyperparameters for RF, which is an ensemble method known for its versatility. The parameters we focused on include n-estimators to determine the number of decision trees in the forest, max-depth to control the depth of the trees, and $min-samples-split$ and $min-samples-leaf$ to influence the structure of the trees and prevent overfitting. Additionally, we explored $max-features$ to decide how many features should be considered when determining the optimal split. Equation (6) provides a concise expression of the mathematical representation of the RF regression approach.

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \tag{6}$$

where $\hat{y}$ represents the predicted value, N is the number of decision trees in the RF model, and $f_i(x)$ denotes the prediction of the i-th decision tree.

C. XGBOOST

XGBoost is a high-performance tree-based ensemble model known for its exceptional predictive capabilities. It employs boosting techniques to sequentially train decision trees and capture complex patterns in the data [29]. In order to create a single, more stable, strong tree, XGBoost aggregates several weak trees and combines their split features. In each phase of the XGBoost training procedure, a new simple tree is constructed to make up for the prediction residuals of earlier simple trees, reducing the loss function [30]. In the case of XGBoost, we thoroughly adjusted parameters such as $n-estimators$ for controlling the number of boosting rounds, $learning-rate$ to manage step size, $max-depth$ to limit tree depth, and subsample to introduce randomness. Additionally, we optimized $colsample-bytree$ to specify the fraction of features used for growing trees and $gamma$ to apply regularization. The XGBoost model formulation can be expressed as follows in Equation (7):

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{7}$$

where $\hat{y}_i = \phi(x_i)$ combines predictions from multiple decision trees (up to K trees) to make a final prediction $\hat{y}_i$ for a given input x_i. Each tree contributes its prediction $f_k(x_i)$.

VI. RESULTS AND DISCUSSION

This section presents a comprehensive analysis of our study, starting with the exploratory analysis. We delve into the

insights gained from the data, identifying key patterns and trends that guide our research. Subsequently, we highlight the crucial step of data processing, which involves handling missing values and outlier removal to ensure the integrity and consistency of the dataset. By computing the error-index, we assess the efficacy of our suggested model and obtain important insights into the precision and accuracy of the RUL.

A. EXPLORATORY ANALYSIS

Figure 4 presents a dataset's visual representation of data distribution after removing outlier data points. Outliers are data points that significantly deviate from most of the data and can substantially impact the overall data distribution. When outliers are removed from a dataset, the distribution of the remaining data is likely to change. The visualization helps to observe how the data points are distributed without the influence of extreme values, allowing a clearer understanding of the main direction and spread of the data. Visualizing the distributions after outlier removal allows one to assess whether the data has become more concentrated around the main values or if variations still need to be addressed.

The Figure shows 5 a heatmap representing the correlation between different variables in the dataset before outlier removal. Colors indicate the strength and direction of the correlation: yellow for positive, dark blue for negative, and green for no correlation. This visualization helped identify relationships and patterns among variables, guiding data analysis and modeling decisions.

Figure 6 represents a heatmap of the correlation matrix after removing outliers. The correlation matrix shows how each variable in the dataset is related to every other variable, with values ranging from -1 to 1. The color-coded cells in the heatmap visually represent the correlation strength. Darker shades of color indicate a higher positive correlation between variables, while lighter shades represent a weaker or negative correlation. Figure 6 indicates that outliers, which are extreme values that may distort the relationships between variables, have been removed from the dataset. Removing outliers improved the accuracy and reliability of statistical analyses and machine learning models.

B. DATA PROCESSING

In the data processing stage, we performed several essential tasks to prepare the dataset for analysis and modeling. These tasks included separating the data into training and testing sets to evaluate model performance, scaling the data to ensure that all features are on a similar scale, encoding categorical variables to make them suitable for modeling, addressing any issues related to imbalanced data, cleaning the data to remove any errors or inconsistencies, selecting relevant features that provide insights into battery behavior, and engineering new features to capture additional insights. The Table 2 provides a summary of the division of the dataset into training and testing sets. The total dataset consists of 14,845 rows and seven columns. The training set comprises 80% of the total data, corresponding to 11,876 rows, while the remaining 20%

TABLE 2. Summarize the training and testing dataset division.

Number	Data	Percentage(%)	Row
1	Total data	100	(14845, 7)
2	Traing data	%80	(11876, 7)
3	Test data	%20	(2969, 7)

forms the test set, consisting of 2,969 rows. This separation allows the model to be evaluated on a smaller test dataset for performance assessment and training on a larger training dataset.

C. FEATURE SELECTION

In our study, we harnessed the power of Sequential Feature Selection (SFS) to enhance our battery analysis. This technique, available through the SequentialFeatureSelector transformer, offers both forward and backward selection options. In the feature selection process, forward-SFS begins with an empty feature set and systematically adds the most valuable features by maximizing cross-validated scores. This ensures that only the most informative features are included in the final set. Conversely, backward-SFS starts with all available features and iteratively removes the least relevant ones, allowing us to fine-tune our feature selection approach. SFS operates exclusively on the features themselves (X), regardless of any desired output (y), making it valuable in unsupervised learning contexts. Our adoption of Sequential Feature Selection optimized our feature selection process, guaranteeing that our predictive model leverages the most relevant features for precise capacity fade and RUL predictions.

D. ERROR INDEX

The performance of the suggested hybrid model is assessed in our research utilizing three metrics, namely Mean Absolute Error (MAE), Absolute Correlation Coefficient (R^2), and Root Mean Square Error (RMSE), for assessing the effectiveness of the prediction model. MAE can accurately measure the model's prediction error, and R^2 shows the model's fitting effect [31]. The following calculation method for RMSE measures a model's ability to predict outcomes accurately. The formula for calculating RMSE, MAE, and R^2 are as follows in the Equations (8), (9) and (10):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{\mathrm{real}} - \hat{y}_{\mathrm{pred}})^2} \quad (8)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{\mathrm{real}} - \hat{y}_{\mathrm{pred}}| \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{real} - \hat{y_{pred}})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (10)$$

In the Equations (8), (9) and (10), N represents the total number of data points or observations, y_{real} represents the real or actual values, and y_{pred} represents the predicted values. The formulas calculate the deviation or difference between the
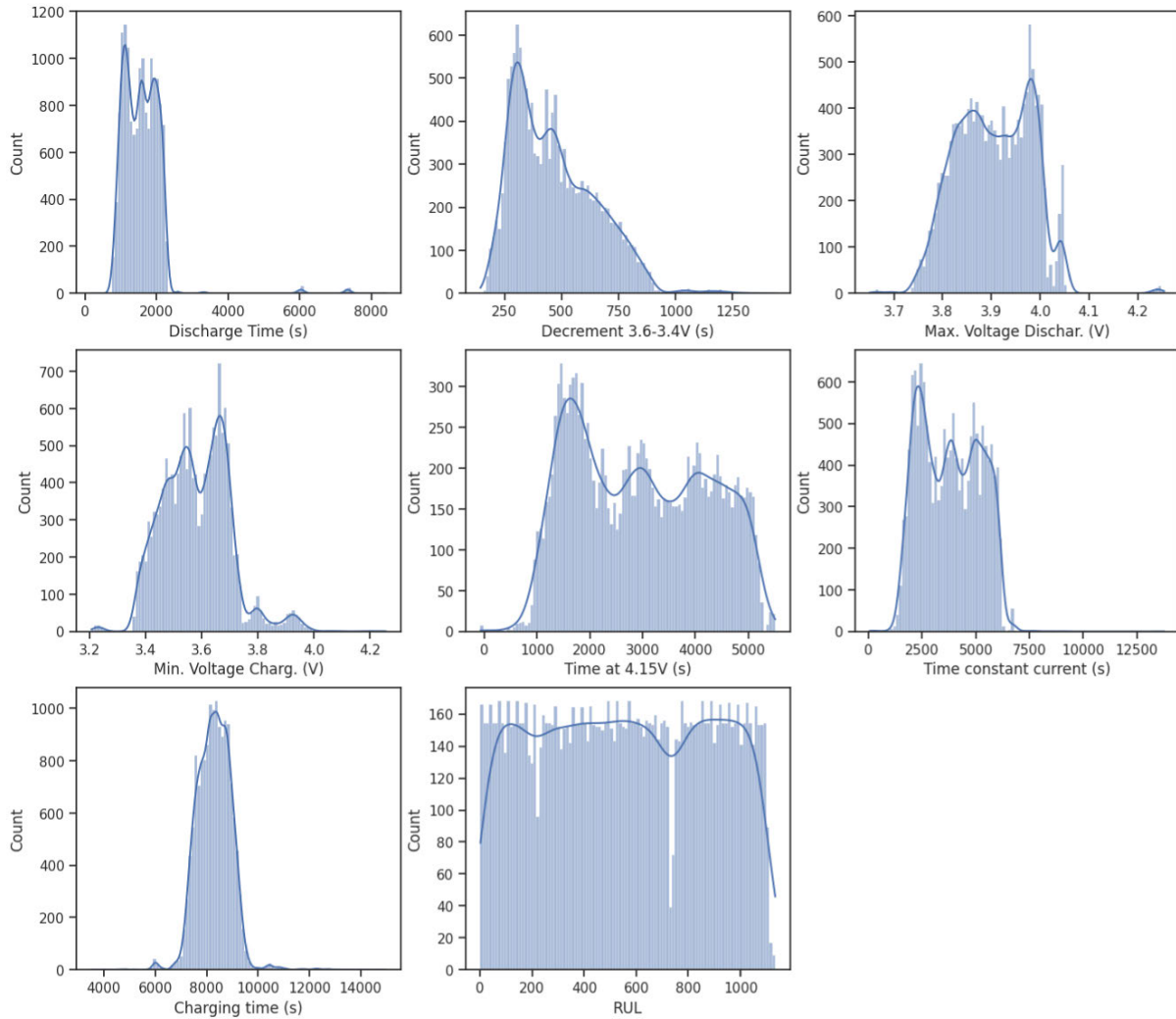
FIGURE 4. Distributions after removal of outliers.

real and predicted values by taking the squared differences for RMSE or the absolute differences for MAE. The sums and averages are then computed to obtain the final values, and R^2 indicates the proportion of the variance in the dependent variable that the independent variables can explain. In the context of battery capacity prediction, the values of y_{real} and y_{pred} correspond to the capacity data, representing the actual capacity measurements and the predicted capacity values, respectively. The RMSE and MAE provide information about the accuracy and precision of the capacity predictions. A lower RMSE and MAE indicate better performance and closer alignment between the predicted and actual capacity values.

E. PERFORMANCE EVALUATION OF THE PROPOSED MODEL IN COMPARISON TO RELATED MODELS

The performance of the suggested hybrid model is compared with three alternative models in Table 3: kNN, RF, and XGBoost, using R^2, RMSE, and MAE as

TABLE 3. Performance evaluation of the proposed model with three different models.

Model	R^2	RMSE	MAE
kNN	0.995797	0.018274	0.00762
RF	0.996041	0.017824	0.009138
XGBoost	0.995661	0.018679	0.00976
Hybrid model	0.996457	0.016861	0.008956

assessment measures to evaluate prediction accuracy. The results indicate that the hybrid model outperforms the other models. It achieved the highest R^2 value of 0.996457, demonstrating its ability to explain more variance in the target variable. Additionally, the hybrid model obtained the lowest RMSE of 0.016861, reflecting its superior prediction accuracy compared to kNN, RF, and XGBoost. The MAE value of 0.008956 confirms the hybrid model's effectiveness in accurately estimating the target variable.
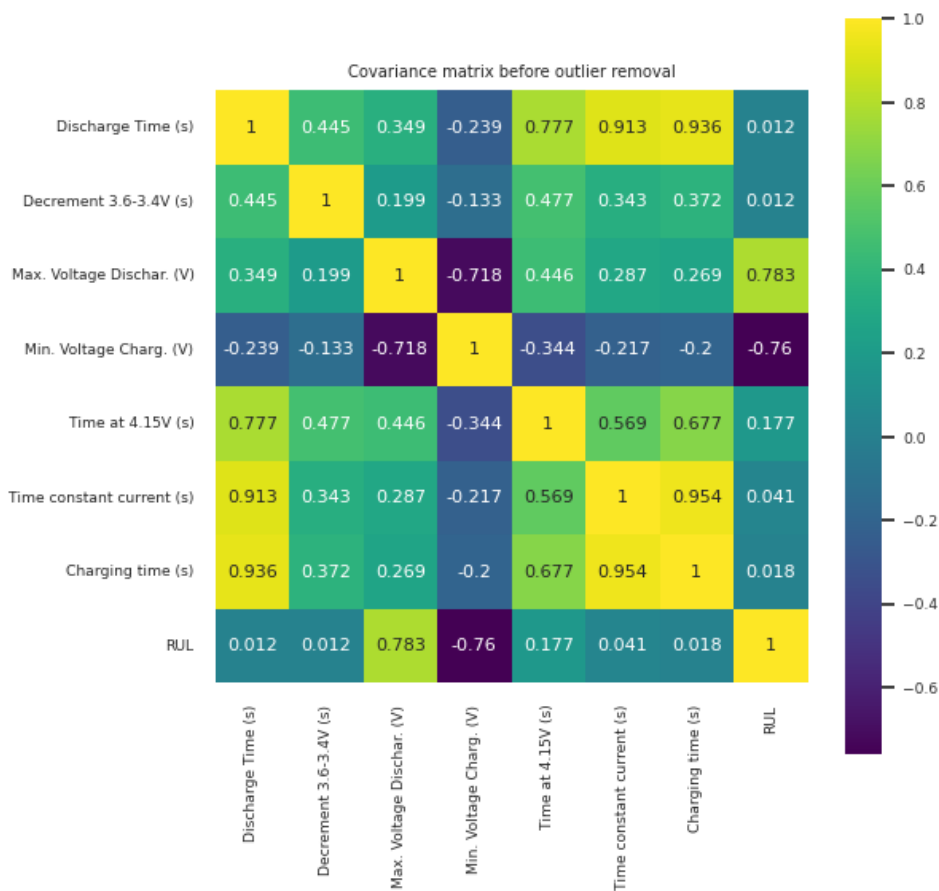
FIGURE 5. Covariance matrix before outlier removal.

Figure 7 shows a bar plot comparing different regression models based on the R^2 scores, which the models include kNN, RF, XGBoost, and the proposed method, which measures the goodness of fit for each model. Each bar represents a regression model, and the height of the bar indicates the corresponding R^2 score. The hybrid model outperformed the other models, achieving the highest R^2 score, approximately 0.9965, as shown on the x-axis, which represents the names of the regression models. The corresponding R^2 values are displayed on the y-axis.

F. PERFORMANCE VALUATION OF THE PROPOSED MODEL IN COMPARISON TO RELATED MODELS

Table 4 offers a comprehensive comparative analysis of predicted outcomes between our proposed hybrid model and established techniques. Each row corresponds to a distinct reference or model, featuring notable approaches like LightGBM, Adaptive Unscented Kalman Filter and Genetic Algorithm optimized Support Vector Regression (AUKF-GASVR), Adaptive Feature Separable Convolution (AFSC), Deep Learning model, Autoencoder with Deep Neural Network (DNN), and GRU. The table prominently displays crucial evaluation metrics, encompassing RMSE,

TABLE 4. Comparison of predicted results between the suggested model and existing techniques.

#Ref	Model	RMSE	MAE	R^2
[32]	LightGBM	1.16	1.17	-
[33]	Adaptive Unscented Kalman Filter and Genetic Algorithm optimized Support Vector Regression (AUKF-GASVR)	0.0185	0.0091	0.95
[34]	Adaptive Feature Separable Convolution (AFSC)	0.35	0.12	-
[35]	Deep Learning model	0.0213	0.0145	-
[36]	Autoencoder with Deep Neural Network (DNN)	6.66	-	93.34
[37]	GRU	0.0389	0.0267	-
Proposed method	Hybrid model	0.0168	0.0089	0.996

MAE, and the R^2. RMSE and MAE serve as measures for predictive precision, with lower values signifying heightened accuracy. Conversely, R^2 gauges the model's fitting quality, with elevated values indicating better alignment. Notably, our proposed method represents the hybrid model introduced in this study, providing profound insights into its remarkable
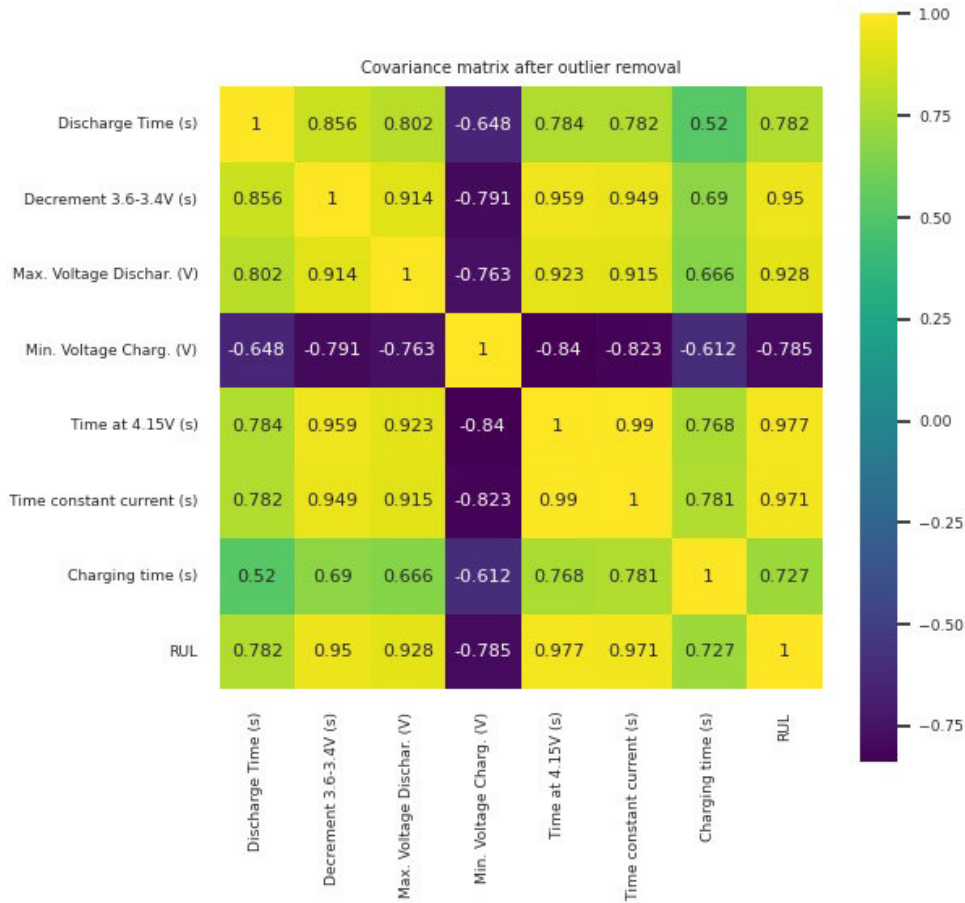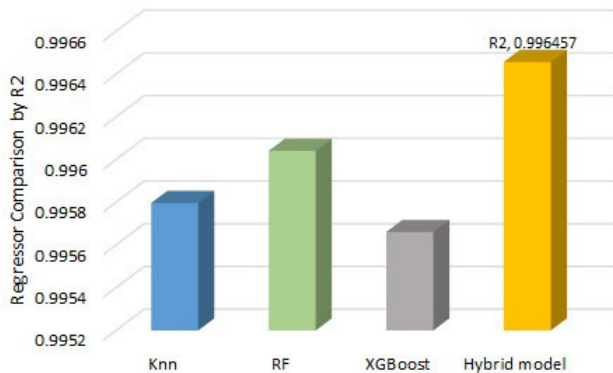
FIGURE 6. Variance matrix after outlier removal.

FIGURE 7. Comparison models by R^2.

predictive capabilities. Particularly noteworthy is the RMSE metric, where our Proposed method achieved an exceptionally low value of 0.016. This result shows our hybrid model's superior predictive performance compared to other battery RUL prediction approaches.

Figure 8 visually defines the connection between cycles and capacity, which is essential for battery behavior analysis.

The horizontal axis, labeled as cycles, quantifies the number of charging and discharging cycles endured by the battery, serving as a metric for its usage over time. The vertical axis, denoted as capacity, signifies the battery's remaining charge storage capacity relative to its original capacity. The blue line within the figure depicts the actual battery capacity data, providing insights into how the battery's capacity evolves throughout multiple cycles, as observed in real-world scenarios. The orange line represents predictions generated by our proposed model, indicating how the battery's capacity is estimated to change with increasing cycles.

Figure 9 provides a direct comparison between the actual RUL values and those predicted by our hybrid model for the last 100 test samples. The x-axis represents the indices of these last 100 test samples, allowing for a chronological understanding of the predictive performance over time. Each data point on the plot corresponds to a single test sample. On the y-axis, we quantify the corresponding RUL values, describing the RUL for each test sample. The real RUL line shows the real RUL values received from the test data. It serves as a benchmark for evaluating the model's performance in estimating the RUL of the tested systems. The blue data points in Figure 9 describe the real RUL values,
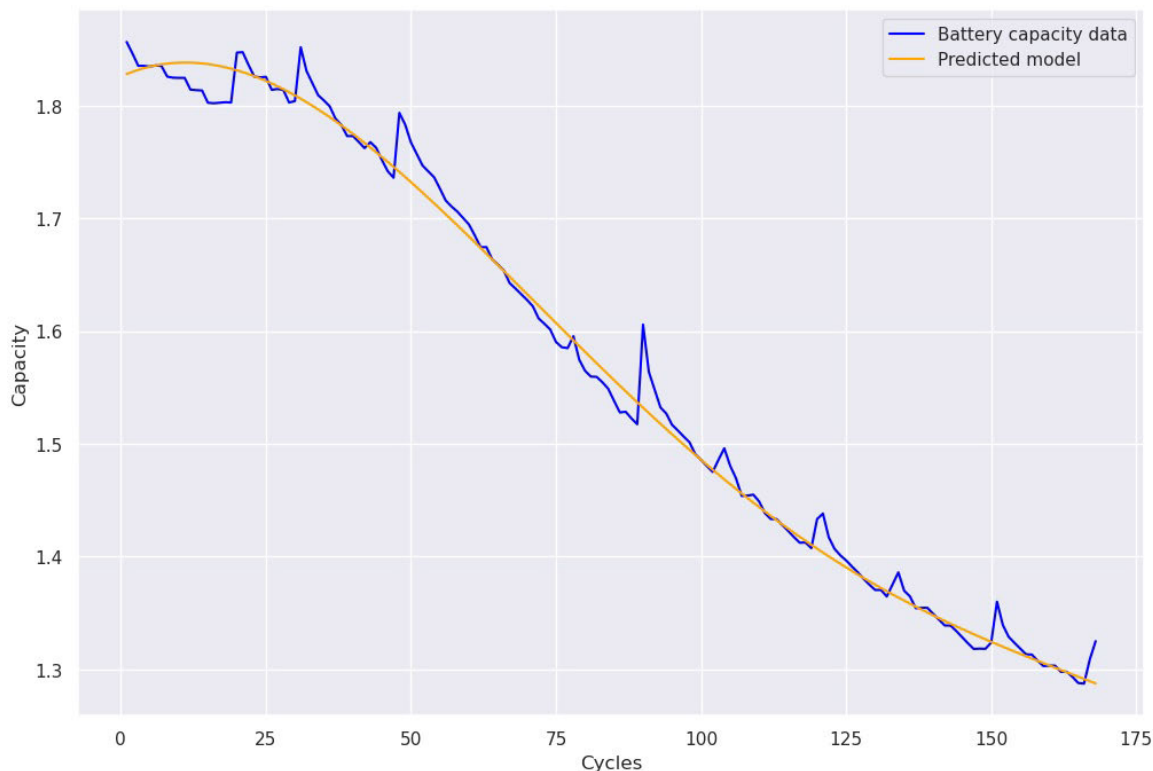
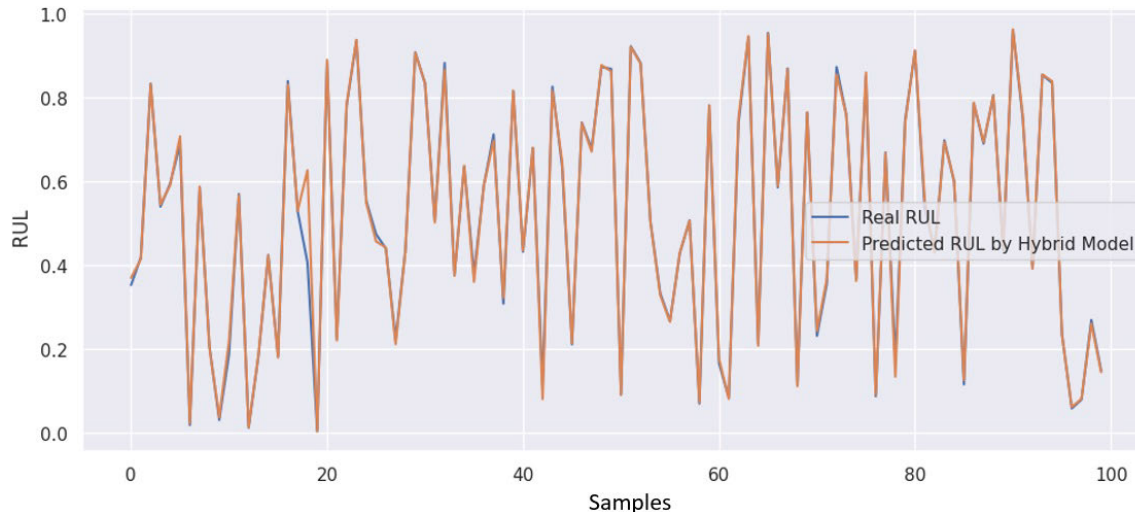**FIGURE 8.** Capacity prediction result by a hybrid model.

FIGURE 9. Comparison of RUL Values and predicted RUL by a hybrid model.

while the orange data points define our model's predictions. Notably, there is a close match between our predictions and the actual data in many points, indicating the model's effectiveness.

G. LIMITATIONS IN RUL BATTERY PREDICTION
The limitations of RUL battery prediction encompass a range of challenges encountered by researchers, practitioners, and industries in their quest for more accurate predictions. Key limitations include a reliance on historical battery data, which significantly impacts model precision. The complexity of real-world battery behavior, influenced by factors such as environmental conditions and usage patterns, poses intricate challenges in predicting RUL accurately. Moreover, there is a critical need for adaptable models capable of accommodating diverse battery types used across

TABLE 5. The list of abbreviations.

Nomenclature	Abbreviation
Lithium-ion Batteries	LIBs
Remaining Useful Life	RUL
Battery Management System	BMS
Electric Vehicles	EVs
State of Health	SoH
k-Nearest Neighbors	kNN
Random Forest	RF
Extreme Gradient Boosting	XGBoost
Remaining Capacity Cycles	RCC
State of Charge	SoC
Recurrent Neural Networks	RNN
Gaussian Process Regression	GPR
Relevance Vector Machine	RVM
Broad Learning System	BLS
Long Short-Term Memory Neural Network	LSTM-NN
Variational Modal Decomposition	VMD
Particle Filter	PF
Posterior Feedback Confidence	PFC
Incremental Capacity Analysis	ICA
Correlated Health Indicators	HIs
Particle Swarm Optimization	PSO-BLS
Hybrid Electric Vehicles	HEVs
Initial Capacity	IC
Constant Current	CC
Constant Voltage	CV
Root Mean Square Error	RMSE
Mean Absolute Error	MAE
Absolute Correlation Coefficient	R^2

various industries. Addressing uncertainties and ensuring the interpretability of predictive results adds another layer of complexity. Balancing computational demands, model generalizability, and the influence of external factors presents further intricacies in the field. The necessity for comprehensive data, adaptation to evolving battery technology, and the successful implementation of predictions into real-world decision-making processes further complicate the landscape. However, these limitations are not merely hurdles; they also present opportunities for innovation and refinement, driving the pursuit of more accurate and practical battery management practices.

VII. CONCLUSION AND FUTURE WORK

A novel prediction strategy of capacity and RUL is proposed based on machine learning using a hybrid model, which investigated the effectiveness of a hybrid model incorporating kNN, RF, and XGBoost algorithms in predicting the RUL and capacity fade of batteries. By leveraging the strengths of these models, we achieved accurate and robust predictions of battery performance. The dataset analysis revealed valuable insights into battery degradation and capacity loss. The experimental findings have demonstrated that the suggested approach can efficiently and correctly assess batteries' capacity fading and RUL. The experiment result shows the potential of the hybrid model in enhancing battery management and decision-making processes. Furthermore, integrating machine learning techniques allows for more precise RUL estimation and capacity fade prediction. This

research contributes to the field of battery analytics and provides a foundation for developing efficient battery management strategies. Future work can focus on refining the hybrid model and studying the RUL prediction of LIBs batteries under diverse operating conditions. This affects analyzing datasets that cover various factors like temperature and currents to understand their impact on battery performance. By doing more accurate and comprehensive prediction models can be developed.

Table 5 shows a comprehensive list of abbreviations and corresponding definitions for our research work.

REFERENCES

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

[2] X. Pang, S. Zhong, Y. Wang, W. Yang, W. Zheng, and G. Sun, "A review on the prediction of health state and serving life of lithium-ion batteries," *Chem. Rec.*, vol. 22, no. 10, Oct. 2022, Art. no. e202200131.

[3] J. Vetter, P. Novák, M. R. Wagner, C. Veit, K-C. Möller, J. O. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Vogler, and A. Hammouche, "Ageing mechanisms in lithium-ion batteries," *J. Power Sources*, vol. 147, nos. 1–2, pp. 269–281, 2005.

[4] W. Chen, S. Qi, L. Guan, C. Liu, S. Cui, C. Shen, and L. Mi, "Pyrite FeS_2 microspheres anchoring on reduced graphene oxide aerogel as an enhanced electrode material for sodium-ion batteries," *J. Mater. Chem. A*, vol. 5, no. 11, pp. 5332–5341, 2017.

[5] S. J. Wachs, C. Behling, J. Ranninger, J. Möller, K. J. J. Mayrhofer, and B. B. Berkes, "Online monitoring of transition-metal dissolution from a high-Ni-content cathode material," *ACS Appl. Mater. Interface*, vol. 13, no. 28, pp. 33075–33082, Jul. 2021.

[6] X.-Q. Zhang, X.-M. Wang, B.-Q. Li, P. Shi, J.-Q. Huang, A. Chen, and Q. Zhang, "Crosstalk shielding of transition metal ions for long cycling lithium–metal batteries," *J. Mater. Chem. A*, vol. 8, no. 8, pp. 4283–4289, 2020.

[7] T. R. Ashwin, A. Barai, K. Uddin, L. Somerville, A. McGordon, and J. Marco, "Prediction of battery storage ageing and solid electrolyte interphase property estimation using an electrochemical model," *J. Power Sources*, vol. 385, pp. 141–147, May 2018.

[8] P. Ramadass, B. Haran, P. M. Gomadam, R. White, and B. N. Popov, "Development of first principles capacity fade model for Li-ion cells," *J. Electrochem. Soc.*, vol. 151, no. 2, p. A196, 2004.

[9] S. L. Hahn, M. Storch, R. Swaminathan, B. Obry, J. Bandlow, and K. P. Birke, "Quantitative validation of calendar aging models for lithium-ion batteries," *J. Power Sources*, vol. 400, pp. 402–414, Oct. 2018.

[10] X. Han, Z. Wang, and Z. Wei, "A novel approach for health management online-monitoring of lithium-ion batteries based on model-data fusion," *Appl. Energy*, vol. 302, Nov. 2021, Art. no. 117511.

[11] F. Cadini, C. Sbarufatti, F. Cancelliere, and M. Giglio, "State-of-life prognosis and diagnosis of lithium-ion batteries by data-driven particle filters," *Appl. Energy*, vol. 235, pp. 661–672, Feb. 2019.

[12] Y. Li, H. Sheng, Y. Cheng, D.-I. Stroe, and R. Teodorescu, "State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis," *Appl. Energy*, vol. 277, Nov. 2020, Art. no. 115504.

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

[14] M. Catelani, L. Ciani, R. Fantacci, G. Patrizi, and B. Picano, "Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[15] D. Yuanchang, P. Xiaoqiong, J. Jianfang, S. Yuanhao, W. Jie, and L. Xiao, "Remaining useful life prediction of lithium-ion batteries based on SVD-SAE-GPR," *Energy Storage Sci. Technol.*, vol. 12, no. 4, p. 1257, 2023.

[16] D. Liu, J. Zhou, D. Pan, Y. Peng, and X. Peng, "Lithium-ion battery remaining useful life estimation with an optimized relevance vector machine algorithm with incremental learning," *Measurement*, vol. 63, pp. 143–151, Mar. 2015.

[17] S. Zhao, C. Zhang, and Y. Wang, "Lithium-ion battery capacity and remaining useful life prediction using board learning system and long short-term memory neural network," *J. Energy Storage*, vol. 52, Aug. 2022, Art. no. 104901.

[18] C. Zhang, S. Zhao, and Y. He, "An integrated method of the future capacity and RUL prediction for lithium-ion battery pack," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2601–2613, Mar. 2022.

[19] C. Zhang, S. Zhao, Z. Yang, and Y. Chen, "A reliable data-driven state-of-health estimation model for lithium-ion batteries in electric vehicles," *Frontiers Energy Res.*, vol. 10, Sep. 2022, Art. no. 1013800.

[20] F. A. Shah, S. S. Sheikh, U. I. Mir, and S. O. Athar, "Battery health monitoring for commercialized electric vehicle batteries: Lithium-ion," in *Proc. Int. Conf. Power Gener. Syst. Renew. Energy Technol. (PGSRET)*, Aug. 2019, pp. 1–6.

[21] S. S. Sheikh, F. A. Shah, S. O. Athar, and H. A. Khalid, "A data-driven comparative analysis of lithium-ion battery state of health and capacity estimation," *Electr. Power Compon. Syst.*, vol. 51, no. 1, pp. 1–11, Jan. 2023.

[22] M. Khalid, S. S. Sheikh, A. K. Janjua, and H. A. Khalid, "Performance validation of electric vehicle's battery management system under state of charge estimation for lithium-ion battery," in *Proc. Int. Conf. Comput., Electron. Electr. Eng. (ICE Cube)*, Nov. 2018, pp. 1–5.

[23] S. S. Sheikh, M. Anjum, M. A. Khan, S. A. Hassan, H. A. Khalid, A. Gastli, and L. Ben-Brahim, "A battery health monitoring method using machine learning: A data-driven approach," *Energies*, vol. 13, no. 14, p. 3658, Jul. 2020.

[24] G. dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100081.

[25] S. Wang, S. Jin, D. Deng, and C. Fernandez, "A critical review of online battery remaining useful lifetime prediction methods," *Frontiers Mech. Eng.*, vol. 7, Aug. 2021, Art. no. 719718.

[26] L. Wu, X. Fu, and Y. Guan, "Review of the remaining useful life prognostics of vehicle lithium-ion batteries using data-driven methodologies," *Appl. Sci.*, vol. 6, no. 6, p. 166, May 2016.

[27] X. Yu, "A car-following model with stochastically considering the relative velocity in a traffic flow," *Acta Phys. Sinica*, vol. 52, no. 11, p. 2750, 2003.

[28] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C.-W. Chan, P. van den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 232, pp. 197–210, Dec. 2018.

[29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[30] F. Jiang, J. Yang, Y. Cheng, X. Zhang, Y. Yang, K. Gao, J. Peng, and Z. Huang, "An aging-aware SOC estimation method for lithium-ion batteries using XGBoost algorithm," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2019, pp. 1–8.

[31] C. Wang, N. Lu, S. Wang, Y. Cheng, and B. Jiang, "Dynamic long short-term memory neural-network-based indirect remaining-useful-life prognosis for satellite lithium-ion battery," *Appl. Sci.*, vol. 8, no. 11, p. 2078, Oct. 2018.

[32] Z. Jiao, H. Wang, J. Xing, Q. Yang, M. Yang, Y. Zhou, and J. Zhao, "A LightGBM based framework for lithium-ion battery remaining useful life prediction under driving conditions," *IEEE Trans. Ind. Informat.*, vol. 19, no. 11, pp. 11353–11362, Nov. 2023.

[33] Z. Xue, Y. Zhang, C. Cheng, and G. Ma, "Remaining useful life prediction of lithium-ion batteries with adaptive unscented Kalman filter and optimized support vector regression," *Neurocomputing*, vol. 376, pp. 95–102, Feb. 2020.

[34] X. Pang, W. Yang, C. Wang, H. Fan, L. Wang, J. Li, S. Zhong, W. Zheng, H. Zou, S. Chen, and Q. Liu, "A novel hybrid model for lithium-ion batteries lifespan prediction with high accuracy and interpretability," *J. Energy Storage*, vol. 61, May 2023, Art. no. 106728.

[35] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, and L. Zhang, "Remaining useful life prediction for lithium-ion battery: A deep learning approach," *IEEE Access*, vol. 6, pp. 50587–50598, 2018.

[36] C. Chen, J. Wei, and Z. Li, "Remaining useful life prediction for lithium-ion batteries based on a hybrid deep learning model," *Processes*, vol. 11, no. 8, p. 2333, Aug. 2023.

[37] R. Rouhi Ardeshiri and C. Ma, "Multivariate gated recurrent unit for battery remaining useful life prediction: A deep learning approach," *Int. J. Energy Res.*, vol. 45, no. 11, pp. 16633–16648, Sep. 2021.

SADIQA JAFARI received the B.S. degree in electrical engineering from Pooyesh University, Iran, and the M.S. degree in control engineering from the University of Qom, Qom, Iran, in 2019. She is currently pursuing the Ph.D. degree with the Machine Learning Laboratory (MLL), Jeju National University, Jeju, Republic of Korea. Her research interests include energy systems, model predictive control, fuzzy control, multi-agent systems, artificial intelligence, machine learning, natural language processing, and data mining.

YUNG-CHEOL BYUN received the B.S. degree from Jeju National University, in 1993, and the M.S. and Ph.D. degrees from Yonsei University, in 1995 and 2001, respectively. From 1998 to 2001, he was a Special Lecturer with Samsung Electronics and SDS. From 2001 to 2003, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI). In 2003, he was promoted to an Assistant Professor with Jeju National University, where he is currently a Full Professor with the Department of Computer Engineering. His research interests include AI and machine learning, pattern recognition, blockchain and deep learning-based applications, big data and knowledge discovery, time-series data analysis and prediction, image processing, medical image applications, and recommendation systems.

SEOKJUN KO received the B.S., M.S., and Ph.D. degrees from Sungkyunkwan University, in 1996, 1998, and 2001, respectively. From 2002 to 2004, he was a Senior Researcher with Samsung Electronics. In 2005, he was promoted to an Assistant Professor with Jeju National University, where he is currently a Full Professor with the Department of Electronic Engineering. His research interests include signal processing for radar and communication systems, radar sensors for automotive applications, radar systems for space rockets, and communication systems in vehicle-to-vehicle.

● ● ●