*mathematics*

*Article*

# Lithium-Ion Battery Estimation in Online Framework Using Extreme Gradient Boosting Machine Learning Approach

**Sadiqa Jafari †,‡, Zeinab Shahbazi †, Yung-Cheol Byun * and Sang-Joon Lee**

Department of Computer Engineering, Institute of Information Science & Technology, Jeju National University, Jeju 63243, Korea; s.jafari1370@yahoo.com (S.J.); zeinab.sh@jejunu.ac.kr (Z.S.); sjlee@jejunu.ac.kr (S.-J.L.)
* Correspondence: ycb@jejunu.ac.kr
† These authors contributed equally to this work.
‡ Department of Electronic Engineering, Institute of Information Science Technology, Jeju National University, Jeju 63243, Korea.

**Abstract:** The battery management system in an electric vehicle must be reliable and durable to forecast the state of charge. Considering that battery degradation is generally nonlinear, state of charge (SOC) estimation with lower degradation can be challenging. Lithium-ion batteries are highly dependent on the knowledge of aging, which is usually costly or not available online. In this paper, we suggest the state of charge estimation of lithium-ion battery systems by using an extreme gradient boosting algorithm for electric vehicles application, which acquires the nonlinear relationship model can with offline training. The extreme gradient boosting algorithm is the tree on based learning, which effectively performs and speeds. Voltage-time data used as an input of this system from the partial constant current phase; the proposed algorithm improves the accuracy of predicting the relevant. Additionally, no initial state of charge is required in our proposed method; thus, estimating the state of charge can consider each battery state.

**Keywords:** lithium-ion battery; capacity; state of charge; extreme gradient boosting



**Citation:** Jafari, S.; Shahbazi, Z.; Byun, Y.-C.; Lee, S.-J. Lithium-Ion Battery Estimation in Online Framework Using Extreme Gradient Boosting Machine Learning Approach. *Mathematics* **2022**, *10*, 888. https://doi.org/10.3390/math10060888

Academic Editors: Ripon Kumar Chakrabortty and Ioannis G. Tsoulos

Received: 25 January 2022
Accepted: 8 March 2022
Published: 10 March 2022



## 1. Introduction

Lithium-ion batteries play a significant role in portable consumer electronics, electric vehicles, hybrid electric vehicles, and large power energy storage systems in portable consumer electronics, electric vehicles, and large power energy storage systems. Their prominent features include lightweight, high efficiency, long lifespan, and low self-discharge. Similarly, the performance of Li-ion batteries diminishes over time as their electrochemical constituents degrade, leading to a loss of capacity and power [1]. Overcharging or over-discharging for a battery management system is a crucial indicator to optimize control strategies to the batteries due to preclude to stretch their life. When a battery cell's capacity is accurately decided, the battery can be replaced before the device limitation life and extend proper battery life without compromising safety. The capacity ratio of the current state cycle of charge and discharge is used to determine the health state of cells to compare the capacity of the initial status [2]. Due to the increased safety risks associated with Li-ion batteries, developed battery management systems are needed to prevent fires and explosions caused by reactive chemicals. Batteries should be managed by a system that evaluates safety-relevant input from the battery, such as temperature, current drawn from the battery, and the state of health and charge of the battery. An essential parameter of the BMS is the measurement of SOC, which provides the key to ensuring that all cells have a constant voltage [3]. At the Lithium-ion batteries electrochemical systems, in general, cells' performance degrades during usage and storage, increasing the need to evaluate a cell's reliability and longevity. Thus Lithium-ion battery health monitoring, as a battery management system, can consider a function that is important for manufacturers and customers of electric vehicles. Additionally, lithium-ion rechargeable batteries have become increasingly popular over the past two decades in various applications, such as power backup, consumer electronics, and grid storage. For ensuring that Li-ion battery packs perform safely and reliably, battery management systems must be able to display real-time information about each cell [4]. In recent years, various estimation methods have been suggested, divided into conventional methods, adaptive filtering, and machine learning. Utilizing conventional methods at the battery can estimate the state of charge directly, such as discharge current, resistance, and impedance [5]. Researchers' modeling for online capacity estimation presented intelligent data-driven methods with attention to multiple features and battery capacity, such as neural networks [6,7]. Recurrent neural networks are a type of neural network that can use for getting the nominal capacity and as an input time sequence [8].

State of charge and state of health usually determine the battery state. Besides providing information about the charge-discharge operation, taking into account, the battery state is also beneficial for determining whether the operating environment is safe and reliable. An accurate estimation of a BMS's SOC will be difficult due to aging, varying environmental situations, and charge–discharge cycles associated with a battery, which will make SOC comparable to a fuel consumption indicator on gasoline cars. As well as the state of health-giving the percentage of remaining battery life, there is no measurement of an exact physical property for state of health because there is no consensus on its description [9]. SOC crucially influences battery state estimation since it represents the remaining capacity and determines how long a battery will last before it requires a recharge. As a general rule, $Q_a$ available is defined as the ratio in SOC. Furthermore, $Q_r$ is described as rated capacity, which is seen in the Equation (1) [10].

$$SOC = \frac{Q_a}{Q_r} \tag{1}$$

Lithium-ion batteries, however, experience electrochemical reactions over time that can change their rated capacity to some period and subsequently gradually reduce it. Thus, calculating SOC operating a fixed rated capacity remains controversial [11]. Applications based on machine learning require little effort to model with the correct test data. With the development of algorithms, it is possible to employ machine learning applications in practical applications for SOC estimation. Machine learning applications for SOC estimation are currently in usage in the literature because they are comfortable to use and provide accurate results with accurate data. SOC estimation has benefited from the development of new methods in data science. Recent approaches such as XGBoost have been created and are applied to SOC estimation. Similarly, the choice of health metrics and battery performance evaluations depends not only on the battery model content, but also on data-driven approaches. Electric vehicle battery management systems (BMS) rely heavily on the battery's state of charge (SOC) estimation. The study utilizes the XGBoost algorithm to predict the state of charge of lithium-ion batteries in electric vehicles based on the data collected from the battery management system used a comprehensive online phase procedure for online estimation of lithium-ion batteries to determine the online capacity estimation performance. XGBoost algorithm cells are used to find the remaining capacity under operation. A cell's partial are constant current charging curve is the input to this paper's network, with voltage and time samples, temperature. The information is deemed the most reliable without operational interference and does not require further feature engineering or processing. The proposed model's ability evaluates the prediction and classification accuracy.

The main contribution of this process summarized as below:

- Using machine learning techniques to estimate battery capacity online is the main focus of this work.
- We considered a dataset of aging cell experience from lithium-ion batteries.
- In this study, the main variables are voltage and temperature.

- As a random process, the suggested method shows impressive estimation performance, such as learning the relationship between the features and the state of charge.
- The paper's final portion illustrates how the XGBoost model can predict and perform aging cell batteries.

The rest of this process is as followed: we begin in Section 2 with related work, while Section 3 will focus on the method. The result is shown in Section 4. We present a discussion in Section 5. Conclusions and future work are presented in Section 6.

## 2. Related Work

A battery system needs to accurately estimate the state of charge to make a more suitable battery cell. For lithium-ion batteries, an accurate SOC can supply precise parameters. A stable SOC is vital for E.V. battery systems, controlling the batteries from being over-discharged and over-charged, thereby keeping the system's safety, maximizing efficiency, and growing battery life. This identification can only be achieved by remembering the parameters and estimating their state using a battery model. Li-ion batteries can be estimated based on electricity, open-circuit voltage (OCV), impedance, internal resistance, and lithium content associated with their state of charge (SOC) resolution. The state of charge could not accurately estimate the reason for battery aging. Additionally, the main causes of battery aging are the decay of internal resistance, capacitance, and accessible power; Figure 1 shows the factors and analysis of the reasons for battery aging [12].

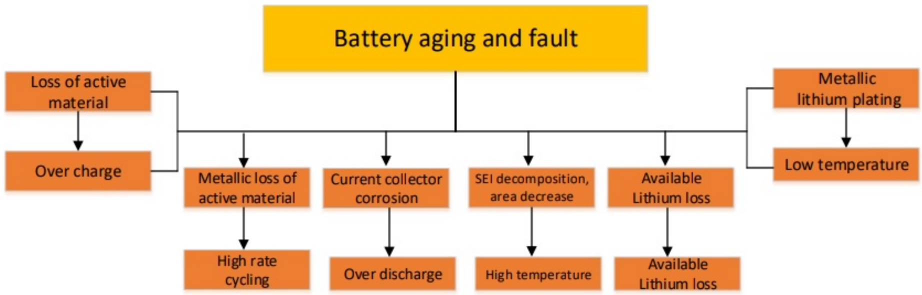


**Figure 1.** Battery aging analysis.

As renewable-energy-based technologies advance, lithium-ion batteries provide the power to power them. An overview of a lithium battery dataset in the public domain is presented. The following studied the existing public domain battery dataset and categorized the testing regimes, cell specifications, and data files [13]. Electric vehicles require lithium-ion batteries to have a predictable cycle life. Since lithium-ion batteries are comparatively safer and have more energy and power density than other commercialized batteries, they are gaining extensive attention. Lithium-ion batteries, whose degradation mechanisms have yet to be distinctly apprehended, produce complex, nonlinear behavior, thereby limiting standard predicting methods. Machine learning techniques have been attracted increasing attention for empirically learning and predicting battery behaviors. By using machine learning approaches, this analysis has been illustrated how to detect hidden features in complex, nonlinear methods that can accurately predict a system's service life in this article [14]. This work suggested an estimation of lithium-ion battery state of charge based on alternate adaptive extended Kalman filter and ampere-hour counting methods; these strategies can be challenging to apply in real applications due to the limited computational capacity of the BMS. They operated a battery module; the situation is especially problematic when every battery pack module needs to be estimated simultaneously. This article aimed to estimate the SOC per battery module in a battery pack simultaneously, handled by this balance [15,16].

*2.1. Overview Lithium-Ion Batteries with Electrochemical Model and Impedance Spectroscopy*

To estimate a cell's health status, several approaches such as controlled experimental, controlled modeling approach. This work has been presented a characterization of high-power lithium-ion batteries using electrochemical impedance spectroscopy that a commercial high-power lithium-ion cell's performance concerning the temperature and charge state [17–19]. Lithium batteries have a higher energy and power density, a higher frequency of operation, and an improved temperature range compared with other technologies. Monitoring the battery's state online is crucial to avoid dangerous operating conditions and extend battery life. In this study, a method has been described to examine the aging of lithium polymer batteries using electrochemical impedance spectroscopy based on an analysis done of the impedance spectra [20–22]. It is crucial to predict degradation in lithium-ion batteries to ensure battery safety. The first prediction of capacity fades in lithium-ion batteries using electrochemical impedance spectroscopy under overcharge conditions [23,24]. A variety of mechanisms cause Lithium-ion batteries to degrade. Conductivity loss, active material loss, and lithium supply loss to simplify aging mechanisms is degradation modes. In the battery management system, based on a decreasing capacity and an increase in resistance, the battery's state of health has typically been used to quantify battery degradation. Incremental electrochemical capacitance spectroscopy has been compared with differential voltage spectroscopy electrochemical impedance spectroscopy over the same data set [25,26]. Model electrochemical is complex because it has a complicated identification process; thus, identifying parameters relies on identifiability analysis. The article has proposed Parameter sensitivity analysis for lithium-ion batteries using electrochemical model-based. An electrochemical model under realistic charging and driving conditions have been analyzed to determine the parameter sensitivity [27,28]. An extended model is developed based on absolute nodal coordinates, which provides high accuracy while requiring little processing power. In this study, they used an observer that continuously has estimated parameters in real-time with attention to the problem of parameter identification in an electrochemical model of a Lithium-ion battery [29–31]. A composite electrode lithium-ion battery can improve energy and power density and longer cycle life than batteries made from a single active material, which presented an electrochemical model reduced from composite; they used an extended Kalman filter to estimate charge state in real-time [32–34]. For lithium-ion batteries to be reliable and efficient, the state of charge must be accurately modeled and estimated. Physics-based electrochemical models have been highly desirable to push batteries to their physical limits, presenting a trial of proportional–integral observers to estimate the state of charge, capacity simultaneously, and resistance for lithium-ion batteries [35,36].

*2.2. Lithium-Ion Batteries with Machine Learning Algorithms*

When a lithium-ion battery is in operation, the discharge current varies according to the load, making it challenging to measure capacity online using the traditional method. There are solutions for problems suggested based on the charge curve to estimate a lithium-ion battery's capacity and state of health. They used voltage variations, and charge currents have been used as the health indicators for predicting capacity [37]. A nonlinear autoregressive was proposed to estimate the state of charge and health works based on recurrent networks, and the battery's internal parameters have not been required. The long short-term memory for the lithium-Ion battery-based estimation has been used because the online state of health is the main issue for battery management due to the condition limit environment for measurement [38]. An accurate prediction in the lithium-ion battery has a significant role in the intelligent battery health management systems that use life batteries. Using an autoencoder and deep neural network to predict the remaining useful life of lithium-ion batteries has been proposed [39]. An online method for lithium-ion battery remaining practical life estimation using importance sampling and neural networks has been afforded an online strategy; battery performance would differ depending on age and condition. The voltage curves have been analyzed for each cycle number during the charging process [40]. The loss in rated capacity has been described as the health of lithium-ion batteries. The proposed support vector machine for online state health estimation has been used to analyze the extraction of three feature variables like energy signal, throughput and charge duration [38]. The state of charging and discharging happens under a constant current in the battery, which this situation might be available for hence voltage versus time measurements. Situ capacity estimation of lithium-ion batteries using the Gaussian process regression method has considered voltage measurements for battery capacity over short periods of the galvanostatic process [41]. The analysis has been focused on healthy features and long short-term memory in the battery. The battery aging practice always requires considerable time for data recovery in actual purposes and real performance in the battery, which speeds up the aging operation. The state of health estimation focusing on healthy features and long short-term memory in the battery with charging and discharging battery voltage while changing has been collected in the cycle life experiment until the healthy features have been extracted correlating to battery degradation. Then, the state of health of the debasement station of lithium-ion batteries has been determined [42]. Different lithium-ion batteries require separate aging mechanisms, depending on the primary battery scheme and applied materials. Besides, battery systems are the main critical components of battery electronic vehicles that effectively affect each charge and driving performance: a multi-island genetic algorithm and Gaussian process regression for the state of health estimation. The aim has been an efficient parameter of Incremental capacity curves has been identification [43]. The following studies illustrate, in Table 1, the charging features, discharging features, and estimation status of batteries.

**Table 1.** Related studies on lithium-ion batteries in machine learning.

| Reference | Model | Error Rate | Benefit |
|---|---|---|---|
| [44] | Elman neural network | MAE 1.29% | Prediction |
| [45] | Semi-supervised transfer component analysis | MAE 1.29% | Learning |
| [46] | Incremental capacity analysis technique | RMSE 2.99% | Analysis technique |
| [47] | Gaussian process regression | RMSE 3.45% | Optimize |
| [48] | Extreme learning machine | RMSE 2% | Prediction |
| [49] | Geometrical approach | RMSE 3.84% | High accuracy |
| [50] | Random forest | RMSE 3.58% | Prediction |

## 3. Methodology

In this section, the state-of-charge estimation of batteries, state-of-charge analysis of machine learning algorithms, model framework, data information, and extreme gradient boosting are presented in detail.

### 3.1. State-of-Charge Estimation of Batteries

In electric vehicles, it is not easy to predict the state of charge due to the durability of the battery management systems. Due to this, we proposed a type of ML algorithm that can be used to create an accurate state of charge (SOC). SOC estimation using ML is shown in Figure 2.

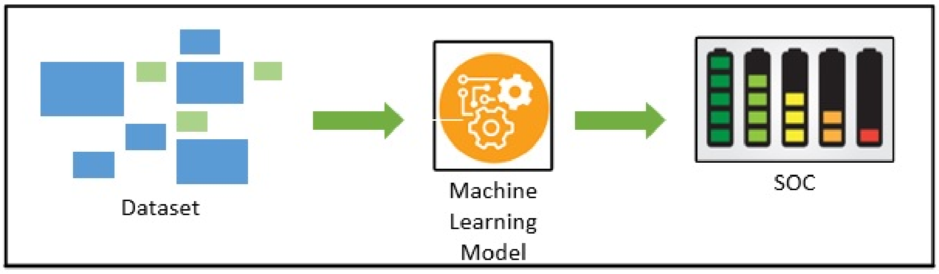


**Figure 2.** Block diagram of machine learning for SOC estimation.

*3.2. State of Charge in the Machine Learning Applications*

Several models in machine learning presented the state of charge estimation utilization and different regression methods studied nowadays to estimate the state of charge and state of health, executed with a dataset to various parameters. The absence of the necessity of parameter identification is the most significant benefit of the data-driven models since parameter identification demands comprehensive tests that can take months to complete. Applications that use battery management systems gain a great deal from this feature. Due to their variable time constants, different models have been created for different cell chemistries [51]. This disadvantage has been eliminated with data-driven models. However, data-driven models also need a considerable amount of data. Likewise, we mentioned several previous research studies, such as support vector machine [52], Gaussian process regression [53], etc., in the machine learning algorithm. The existing literature is constantly revised with new machine learning algorithms.

*3.3. Model Framework*

Overview suggested model includes a server, aging experiments, electric vehicles, and battery management system. A data-driven approach to SOC estimation is described in Figure 3. Here are two central parts in the online capacity estimation model processes with the XGBoost method. We used a dataset processed before a train to train save in the server for the training model. The main framework in the proposed model includes a training model and online estimation of the trained model to vehicles on the operation. There is a bidirectional connection in the server, and the main task is to present the best model with the devices connected to it.

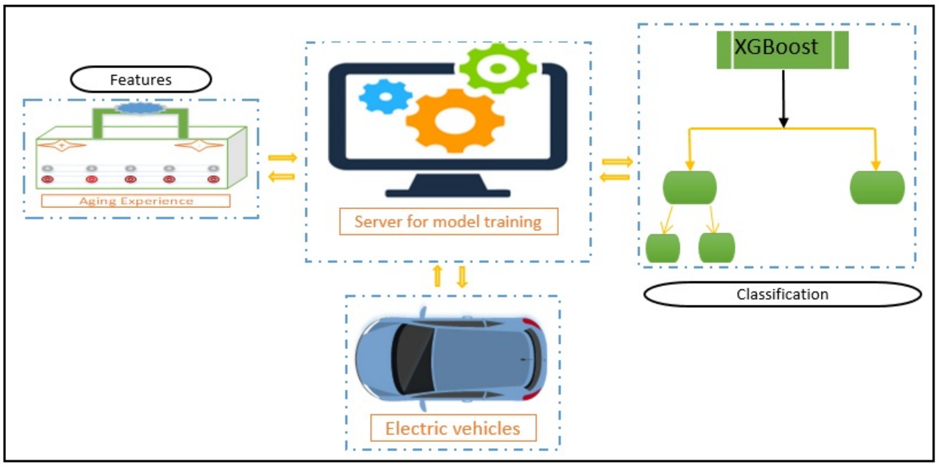


**Figure 3.** The framework of the proposed model.

*3.4. Data Information*

The dataset has been collected from NASA Ames PCoE for 18,650 lithium-ion batteries; the dataset has been used in this study based on the lithium-ion battery aging data [54]. This study has been generated a method to estimate the remaining capacity for equivalently simulated cells to determine cell-specific degradation. A cyclic aging experiment includes characterization tests that have been interspersed. The Table 2 shows the information types of cells.

**Table 2.** Cells information.

| Cell Type of Dataset | |
|---|---|
| **Specifications** | |
| Nominal voltage | 3.6 V |
| Charging method | Constant current<br>Constant voltage |
| Maximum weight | 44.5 g |
| Room temperature | 24 °C |
| The end of life criteria | 30% fade |
| Train | 80% |
| Test | 20% |

There are 48 cells in the dataset, but we show Figure 4 as a degradation trend over a single cell's lifetime. The trend isn't linear, but changes at a given point in the life of each cell.

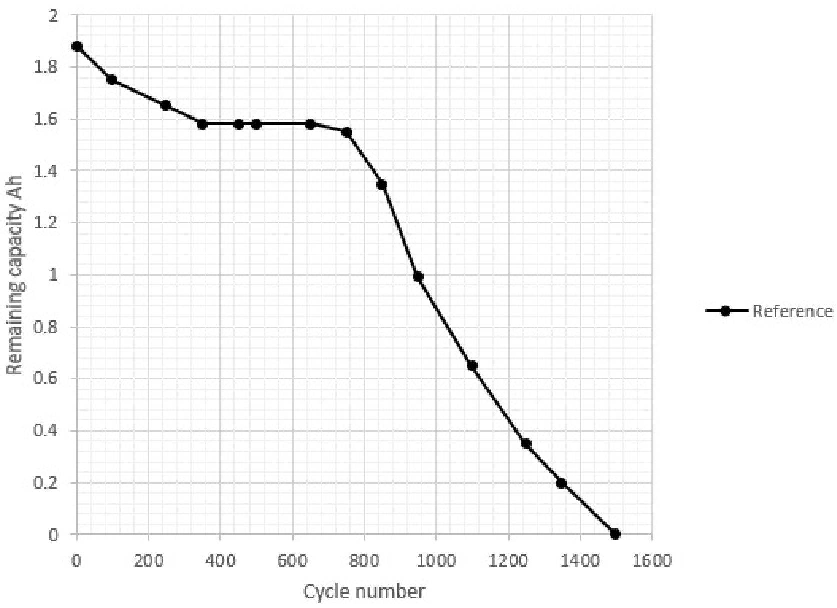


**Figure 4.** One cell's degradation over time.

Cycling of the cells is conducted at a constant temperature of 24 °C. A cycle begins with 2.7 V discharge at 1.5 A in constant current mode and finishes with 4.2 V charge at 2 A in continuous current mode, followed by a constant voltage (CV) phase for 30 min. A battery's accelerated aging is caused by repeated charge and discharge cycles. Voltage, output current, battery temperature, current, load voltage, time, battery capacity with attention to the discharge cycle collected. Thus, there is a non-uniform sampling rate used for each of them. In this work, the proposed method focuses on known partial data for predicting the complete charging curve. We considered three modules: input parameters, feature extraction, and estimates of the state of charge. In the dataset, one temperature and profile for battery aging are provided as the first proof of concept. We have divided the dataset into training, validation, and testing. There are 80% training and 20% validation in the training set, which includes 43,355 data values. The experiments are performed at 24 °C at a constant temperature. The battery's performance is affected by changes in the rising temperature and increase. Nomenclature in the last page shows the notations used in the proposed approach.

### 3.5. Extreme Gradient Boosting

An optimized distributed gradient boosting library, XGBoost is flexible, scalable, and portable. XGBoost includes clever penalization of trees, a proportional shrinking of leaf nodes, newton boosting extra randomization parameter. With XGBoost, trees can have variable nodes, and trees calculated with less evidence have smaller left weights. Boosting by Newton is an alternative to gradient descent that uses the Newton–Raphson approximation. Adding an extra randomization parameter can achieve a lower correlation between trees. A classifier ensemble with a low correlation will perform better than one with a high correlation. It has been possible to decide regression and classification issues utilizing decision trees since the 1980s. In the recent past, the XGBoost algorithm has attracted much attention. The decision tree basis of XGBoost's evaluation offers possible explanations dependent upon specific situations. Different models have been combined, and random decision trees are gathered, like bags. A model with high-performing conduct is given (boosted) more dominance to achieve error minimization. Parallel processing has been carried out in addition to regularization to control overfitting. XGBoost modeling determines a regression tree that fits the residual of the last prediction based on a generalized definition of the objective function. Although decision trees can be unstable and generally incorrect in data errors, they offer many benefits. Besides being a scalable machine learning algorithm consistent with distributed examples, XGBoost does not depend on only one machine learning model. XGBoost also utilizes parallel and dispersed computing to explore model data fast. A fast learner iterates constantly to make continuous predictions. The XGBoost algorithm achieves high error minimization and can predict high accuracy, even with very little data. Compared to other machine learning algorithms, they require reasonably little data to complete an outcome; in addition, they are pretty simple and easy to understand. The algorithms or machine learning models can also be composed of these models [55]. The machine learning (ML) algorithms can develop an accurate state of charge. Therefore, according to the offered strategy, a wide field of battery situations have been evaluated to consider the performance state of charge estimation. The lithium-ion battery's four important parameters can be used to estimate the state of charge established on the XGBoost algorithm; based on the available data, the battery current, voltage, capacity, and temperature can be determined. The XGBoost algorithm utilized in this study is elaborated in the following. Boosting based on gradients is another commonly operated machine learning algorithm. A feeble learner has been transformed into a powerful one. Thus, the final algorithm can be viewed as an orchestra comprising various tools. XGBoost is an ensemble classifier that uses gradient boosting, which the model structure to a loss function that is further expanded by adding an expansion function. In general, gradient boosting algorithms optimize different loss functions to enhance prediction. Decision trees have been used to control the complexity of trees using variations in loss functions. Understanding XGBoost requires understanding the underlying machine learning algorithms and concepts: supervised learning, decision trees, ensemble learning, and gradient boosting. Figure 5 shows a XGBoost classifier [56].

The regression tree provides a residual of a previous prediction, given a generalized description of the objective function. It can decide to predict a category or numerical value. Designing gradient boosted decision trees has been used for performance, speed, and implementation. A short explanation from XGBoost is random sample selection and column subsamples, XGBoost stochastic methods, reduce overfitting and speed up training. The XGBoost algorithm can use a compacted column-based system for minimizing the mathematical of computation by finding the best split. XGBoost has been presented in the following brief way. With attention to the data set, $m$ represents the features, and $N$ represents the capacity in the data set, which includes Equation (2).

$$E = \{(a_i, b_i)|i = 1 \ldots n\}, if \begin{cases} a_i \in R^m \\ b_i \in R \end{cases} \quad (2)$$

We have $n$ train samples and $m$ features for each sample. $b_i$ shows the state of charge in the $i$ sample. The trees boosting are grouped as a collection:

$$F = \{f_k(a) = w_q(a)\}, ifk \in R^m \quad (3)$$

Here is the formula that predicts the value of a tree boosting model, which as output $b_i$ and $f$ is a function in Equation (3). What follows is a definition of $K$ trees in the following equations:

$$\hat{b}_i = \sum_{K=1}^{k} (f_k(a_i)) | \{f_k \in F\} \quad (4)$$

Minimizing the followership regularized target function $f_k$ is demanded to determine a suitable set of functions $f_k$:

$$b_i = \sum_{k=1}^{k} \Omega D \rightarrow D = f_k(a_i) \quad (5)$$

where

$$\Omega(D) = \alpha N + \frac{\beta \|w\|^2}{2} \quad (6)$$

In Equation (6) $\Omega$ determine the complexity of the model, $\alpha$ is parameters controlling and $\beta$ the number of leaves $N$, $W$ is the magnitude of leaf weights. During each round of model training $f_k$, the XGBoost algorithm adds a new function to the model, keeping the prediction results of the last band unchanged [56].

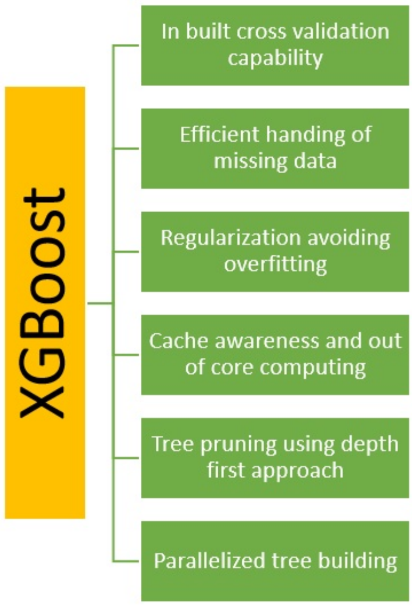


**Figure 5.** XGBoost architecture.

## 4. Results

This section details the implementation process, such as experimental setup, performance evaluation, and SOC estimation based on the extreme gradient boosting system.

### *4.1. Experimental Setup*

In this study, we presented a summary for experimental setup in Table 3, in which the system takes out all experiments and results by using a Windows 10. Furthermore, the device's memory is 30 GB memory with CPU Intel(R) Core(T.M.) i5-9600 CPU @ 3.20 GHz processor. The lithium-ion battery was used estimation capacity. This work performed classification and prediction using the XGBoost machine learning algorithm. In addition, Python was used as a framework and library for developing the proposed system. A version of WinPython-3.8.3 was used to design this system.

**Table 3.** System components and specification.

| Component | Description |
|---|---|
| Operating system | Windows 10 64 bit |
| Browser | Google Chrome |
| CPU | Intel(R) Core(TM) i5-9600K CPU @ 3.70 GHz |
| Memory | 30 GB |
| Programing language | Win Python 3.8.3 |
| Library and framework | Python |
| Machine learning algorithm | XGBoost |
| Battery | Lithium-Ion |

### *4.2. Performance Evaluation*

To compare our forecasted SOC to the experiment SOC to evaluate its actual performance. Therefore, we need to consider performance metrics by comparing our predicted SOC with the experiment SOC's results. As a measure of accuracy, the root means the square error is frequently the square root of the square mean of all the errors. RMSE can only be used to compare predictions from the model to actual data, not to compare comparisons between variables. A commonly used metric for measuring the difference between predictions and observations is the root mean square error (RMSE). The MAE is identical, but offers more weight to more significant absolute values, penalizing them more than MAE does. As the variance in individual errors increases, the MAE/RMSE difference increases the RMSE can be described as follows in Equation (7):

$$\text{RMSE} = \sum_{i=1}^{n} \frac{1}{n} (y_i - \hat{y}_i)^2 \tag{7}$$

In Mean Absolute Error (MAE), all the errors have the same weight. It averages the absolute differences between the tested and predicted values when Mean Absolute Error (MAE) is small, accurate to the forecast results in Equation (8) [57].

$$\text{MAE} = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n} \tag{8}$$

Below is a definition of the error metrics in Equation (9): $y$ shows the real estimated capacity, $\hat{y}$ shows, respectively, and $n$ shows the number of data points per battery cell.

$$\text{APE} = \left| \frac{1}{y} (\hat{y} - y) \right| \times 100\% \tag{9}$$

here for evaluating metric each battery cell with calculating the mean absolute percentage error (MAPE) presented in the Equation (10):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left( \left| \frac{1}{y} (\hat{y} - y) \right| \times 100\% \right)_i \tag{10}$$

The absolute percentage error (MAPE) calculates the average accuracy of the sample's estimates over the cell's entire life. In this test, data is randomly selected and dropped from the model before it is fed input so that the model can demonstrate its predictive abilities. The estimation network was used to validate the model by randomly dropping 15% of the input samples before forwarding it to simulation.

The features extraction such as terminal voltage, temperature charge, and discharge is an input put and output used as shown in the Table 4.

**Table 4.** Features extraction.

| Features | Description |
| --- | --- |
| Terminal voltage | provided voltage |
| Temperature | provided temperature |
| Charge | State of charge |
| Discharge | State of discharge |

In the following Table 5, we provide the results metrics found in the validation set.

**Table 5.** Model validation based on extreme gradient boosting.

| Definition | |
| --- | --- |
| Train | 0.96% |
| Test | 0.92% |
| Validation | 0.78% |

The error rate comparisons in Table 6 show that our suggested approach performs at least on par with existing methods. Based on the error rate comparisons, we estimated the proposed method to be at least as good as existing methods.

**Table 6.** Comparison of error rates between previous approaches and the proposed method.

| Reference | Model | Error Rate |
| --- | --- | --- |
| [58] | Unscented Kalman filter | RMSE 2.00% |
| [59] | Convolutional gated recurrent unit –recurrent neural network | MAE 3.96% |
| Proposed Method | XGBoost | RMSE 2.56<br>MSE 10.03 |

A vital reference in the battery is estimating the available battery capacity to determine the battery life before it needs to be charged. While charging from the vehicle can obtain relatively easy sensor data, charging is a consistent process during the life of a cell since it coincides with charging conditions or over a similar period. In this method, we utilized constant current-constant voltage as inputs to the model because the upper step current-constant will be available on the curve charge in the cell's lifetime. However, they can be the raw charging curve of time and voltage patterns. Discharge curves can depend on diverse vehicle behavior, attention to distance traveled, traffic trends, and driving. Additionally, the vehicle can start to charge from whatever point in the charge cycle, but usually go to 100% before stopping.

The battery's terminal voltage and temperature curves have significantly been modified during discharge. Different battery aging stages produce different terminal voltages and temperatures under constant load discharge profiles as seen in the Figure 6, along with the deepening of aging, which verifies that the presented features Terminal voltage and Temperature are age-aware.

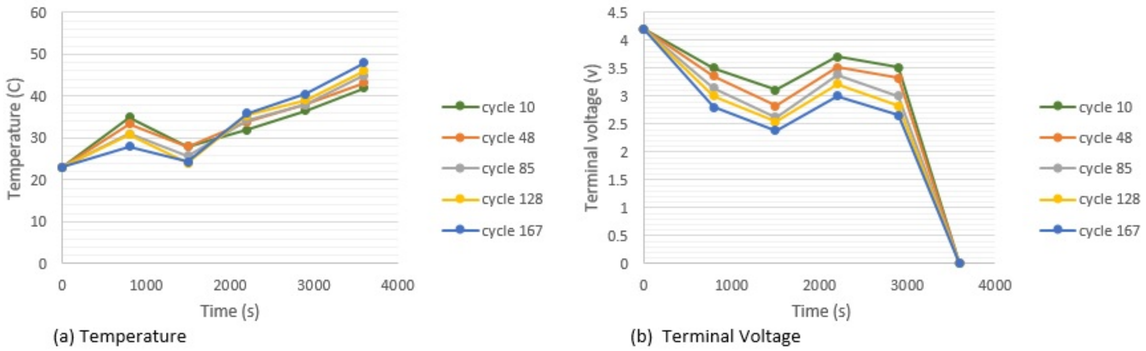


**Figure 6.** The voltage and temperature changes.

### 4.3. Soc Estimation

This section offers the capacity estimation state of charge based on the XGBoost algorithm. A separate set of training and test data are used for XGBoost, which reports the accuracy of the proposed method and the results presented. To train the machine and to obtain results, XGBoost has been used. The identical training data have been used to estimate SOC. The terminal voltage, current, and temperature are available from monitoring data at the battery management system. This section estimates the state of charge. We considered voltage and temperature from the discharge process. At the offline process, the elements are divided using XGBoost for the features-state of charge mappings training until the efficiency and accuracy improve, with attention to split based on the discharge voltage curve. Additionally, the offline trained represses are used to estimate the state of charge online.

At the different stages of battery aging, temperatures curve and produce different voltage curves during the discharge process also show significant changes in the proposed method. Then, the proposed temperature in the room is 24 °C. As shown in Figure 7, there are charge and discharge cycles in which repeated processes accelerate battery aging. Additionally, during charging, the current has been constant.

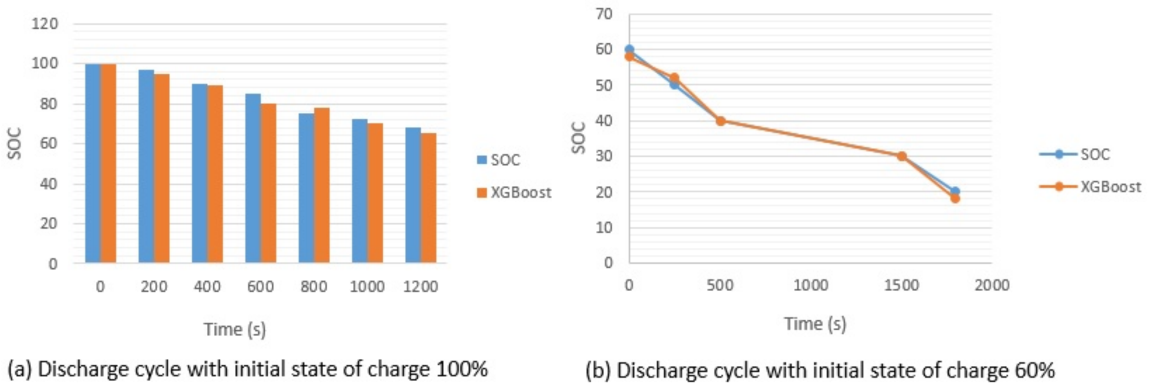


**Figure 7.** Charge cycle of the battery aging reputation.

In this part, for estimating the online state of charge, we used offline training to simulate online training. Here, data points have been randomly selected of all the life cycles of a battery, as shown in Figure 7; the estimation results in attention to different aging degrees and primary state of charge. Then, the result has been compared between the actual state of charge and the XGBoost model, in which the proposed method has been showing good performances.

Figure 8 shows the validation result for the comparison plot for estimates and reference prediction. (a) is the worst aging cell in the battery, and (b) is the best aging.

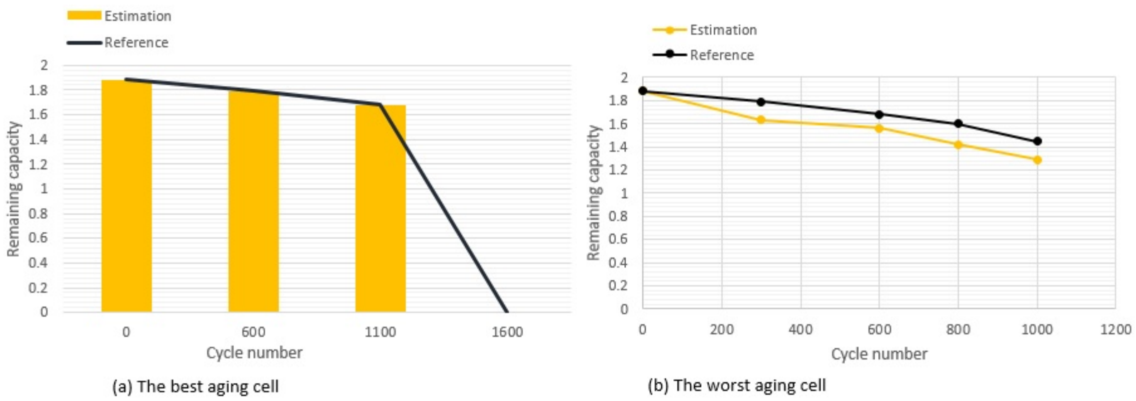


**Figure 8.** Comparison for estimate and reference.

## 5. Discussion

Parallel decision trees speed up operations by evaluating various parameters according to several data. Therefore, XGBoost provides more accurate and faster SOC estimations than other algorithms. Operating machine learning algorithms overcome the drawbacks of conventional SOC estimation techniques for test data, overfitting, and parameter tuning. The lithium-ion battery has been widely used in different applications. The state estimation has attracted increasing interest, particularly for state-of-charge and state-of-health. There are significant challenges in charge, such as accuracy and computational complexity. In Table 7, the benefits and drawbacks of ML and DA have been compared. Similarly, we have outlined some of the benefits and drawbacks of each process [5].

**Table 7.** Comparing SOC estimation techniques—benefits and drawbacks.

| Methods | Benefit | Drawback |
|---|---|---|
| Machine learning | 1- Estimation accuracy<br>of high quality.<br>2- Models based<br>on physical<br>properties are<br>not required.<br>3- Dynamic operational<br>situation | 1- Complex computations.<br>2- The modality and<br>amount of training<br>data affect estimation accuracy. |
| Differential Analysis | 1- Available<br>2- Comfortable to integrate<br>into a BMS<br>3- Computability low | 1- The variation in<br>temperature affects<br>the accuracy of estimations.<br>2- Charges and discharges<br>must be controlled. |

## 6. Conclusions and Future Direction

We present a methodology for predicting the battery SOC based on an ensemble boosting algorithm. We use the proposed machine learning model to analyze the nonlinear mapping between voltage and current values as input features into SOC estimation. The battery SOC estimation technique uses a machine-learning algorithm because it is better suited for handling nonlinear data. Due to the diversity of parameters available for evaluation and the parallel decision trees that can be constructed, also operation can be accelerated. As a result, the extreme gradient boosting algorithm can generate faster and more accurate estimates for SOC applications. As a result of the optimized features input into the machine learning model, the user and researcher will predict the battery state of charge by identifying the best battery for their specific applications. Simulations and experiments demonstrate that the extreme gradient boosting algorithm performs well in estimating SOCs. In future work, other parameters will be incorporated into the proposed estimation techniques. Research can be concentrated on the change of cells in a battery pack since batteries are widely used in many applications. The state of charge estimation and state of health in the area from the battery can be in actual performance situations. Developing the model's predictive capacity by analyzing can be additional options for raising training efficiency and looking at new network architectures in the work future. The scope of this model will be expanding in various approaches in the future. Furthermore, it can analyze additional options to increase training efficiency, which is key to improving the model's predictive capability by using another machine learning algorithm. Another can be validation, which would be a necessary step. A part of future work and the estimation accuracy be evolved by bypassing overfitting and thus lessening errors.

**Author Contributions:** Data curation, Z.S. and S.J.; funding acquisition, Y.-C.B. and S.-J.L.; Investigation, Z.S.; methodology, Z.S.; writing original draft, Z.S. and S.J.; supervision, Y.-C.B.; Project Administration Y.-C.B. and S.-J.L.; validation, Z.S. and S.J.; visualization, Y.-C.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was financially supported by the Ministry of Small and Medium-sized Enterprises(SMEs) and Startups(MSS), Korea, under the "Regional Specialized Industry Development Program(R&D, S3091627)" supervised by Korea Institute for Advancement of Technology(KIAT). This research was also financially supported by the Ministry of SMEs and Startups(MSS), Korea, under the "Startup growth technology development program(R&D, S3125114)".

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

The following abbreviations are used in this manuscript:

| | |
|---|---|
| Symbols | |
| I | Current |
| t | Time |
| $\Omega$ | Penalizes the complexity of the model |
| $\alpha$,$\beta$ | Parameter Control |
| N | The number of leaves |
| w | Weight |
| $b_i$ | Output |
| $a_i$ | Input |
| m | Features |
| n | Train sample |
| $f_k$ | Function |
| K | trees |
| $Q_{max}$ | Represents the maximum Capacity |
| Acronyms | |
| SOC | State of charge |
| EV | Electrical Vehicle |
| Ah | Ampere hour |
| OCV | Open Circuit Voltage |
| SOH | State of Health |
| BMS | Battery Management System |
| XGBoost | Extreme gradient boosting |
| ML | Machine learning |
| DA | Differential Analysis |
| Subscript | |
| min | Minimum value |
| max | Maximum value |

## References


1. Vetter, J.; Novák, P.; Wagner, M.R.; Veit, C.; Möller, K.C.; Besenhard, J.; Winter, M.; Wohlfahrt-Mehrens, M.; Vogler, C.; Hammouche, A. Ageing mechanisms in lithium-ion batteries. *J. Power Sources* **2005**, *147*, 269–281. [CrossRef]
2. Wang, Z.; Feng, G.; Zhen, D.; Gu, F.; Ball, A. A review on online state of charge and state of health estimation for lithium-ion batteries in electric vehicles. *Energy Rep.* **2021**, *7*, 5141–5161. [CrossRef]
3. Rivera-Barrera, J.P.; Muñoz-Galeano, N.; Sarmiento-Maldonado, H.O. SoC estimation for lithium-ion batteries: Review and future challenges. *Electronics* **2017**, *6*, 102. [CrossRef]
4. Shen, S.; Sadoughi, M.; Chen, X.; Hong, M.; Hu, C. A deep learning method for online capacity estimation of lithium-ion batteries. *J. Energy Storage* **2019**, *25*, 100817. [CrossRef]
5. Hannan, M.A.; Lipu, M.H.; Hussain, A.; Mohamed, A. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations. *Renew. Sustain. Energy Rev.* **2017**, *78*, 834–854. [CrossRef]
6. Li, Y.; Dong, B.; Zerrin, T.; Jauregui, E.; Wang, X.; Hua, X.; Ravichandran, D.; Shang, R.; Xie, J.; Ozkan, M.; et al. State-of-health prediction for lithium-ion batteries via electrochemical impedance spectroscopy and artificial neural networks. *Energy Storage* **2020**, *2*, e186. [CrossRef]
7. Eddahech, A.; Briat, O.; Bertrand, N.; Delétage, J.Y.; Vinassa, J.M. Behavior and state-of-health monitoring of Li-ion batteries using impedance spectroscopy and recurrent neural networks. *Int. J. Electr. Power Energy Syst.* **2012**, *42*, 487–494. [CrossRef]
8. Chaoui, H.; Ibe-Ekeocha, C.C. State of charge and state of health estimation for lithium batteries using recurrent neural networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 8773–8783. [CrossRef]
9. Kozlowski, J.D. Electrochemical cell prognostics using online impedance measurements and model-based data fusion techniques. In Proceedings of the 2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652), Big Sky, MT, USA, 8–15 March 2003; Volume 7, pp. 3257–3270.
10. Piller, S.; Perrin, M.; Jossen, A. Methods for state-of-charge determination and their applications. *J. Power Sources* **2001**, *96*, 113–120. [CrossRef]
11. Zhang, J.; Lee, J. A review on prognostics and health monitoring of Li-ion battery. *J. Power Sources* **2011**, *196*, 6007–6014. [CrossRef]
12. Wu, C.; Zhu, C.; Ge, Y.; Zhao, Y. A review on fault mechanism and diagnosis approach for Li-ion batteries. *J. Nanomater.* **2015**, *2015*, 8. [CrossRef]
13. Dos Reis, G.; Strange, C.; Yadav, M.; Li, S. Lithium-ion battery data and where to find it. *Energy AI* **2021**, *5*, 100081. [CrossRef]
14. Su, L.; Wu, M.; Li, Z.; Zhang, J. Cycle life prediction of lithium-ion batteries based on data-driven methods. *eTransportation* **2021**, *10*, 100137. [CrossRef]
15. Liu, Z.; Li, Z.; Zhang, J.; Su, L.; Ge, H. Accurate and efficient estimation of lithium-ion battery state of charge with alternate adaptive extended Kalman filter and ampere-hour counting methods. *Energies* **2019**, *12*, 757. [CrossRef]
16. Jamil, F. Energy Service Architecture Based on SECaaS and AIoT for Energy Trading and Resource Management in Smart Nanogrid. Ph.D. Thesis, Jeju National University, Jeju, Korea, 2021.
17. Andre, D.; Meiler, M.; Steiner, K.; Walz, H.; Soczka-Guth, T.; Sauer, D. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. II: Modelling. *J. Power Sources* **2011**, *196*, 5349–5356. [CrossRef]
18. Andre, D.; Meiler, M.; Steiner, K.; Wimmer, C.; Soczka-Guth, T.; Sauer, D. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation. *J. Power Sources* **2011**, *196*, 5334–5341. [CrossRef]
19. Shahbazi, Z.; Byun, Y.C. Knowledge Discovery on Cryptocurrency Exchange Rate Prediction Using Machine Learning Pipelines. *Sensors* **2022**, *22*, 1740. [CrossRef]
20. Galeotti, M.; Cinà, L.; Giammanco, C.; Cordiner, S.; Di Carlo, A. Performance analysis and SOH (state of health) evaluation of lithium polymer batteries through electrochemical impedance spectroscopy. *Energy* **2015**, *89*, 678–686. [CrossRef]
21. Jafari, S.; Shahbazi, Z.; Byun, Y.C. Designing the Controller-Based Urban Traffic Evaluation and Prediction Using Model Predictive Approach. *Appl. Sci.* **2022**, *12*, 1992. [CrossRef]
22. Jafari, S.; Shahbazi, Z.; Byun, Y.C. Traffic Control Prediction Design Based on Fuzzy Logic and Lyapunov Approaches to Improve the Performance of Road Intersection. *Processes* **2021**, *9*, 2205. [CrossRef]
23. Togasaki, N.; Yokoshima, T.; Oguma, Y.; Osaka, T. Prediction of overcharge-induced serious capacity fading in nickel cobalt aluminum oxide lithium-ion batteries using electrochemical impedance spectroscopy. *J. Power Sources* **2020**, *461*, 228168. [CrossRef]
24. Shahbazi, Z.; Byun, Y.C. Analyzing the Performance of User Generated Contents in B2B Firms Based on Big Data and Machine Learning. Available online: https://www.scmij.com/articles.html#tabs1-1k (accessed on 24 January 2022).
25. Pastor-Fernández, C.; Uddin, K.; Chouchelamane, G.H.; Widanage, W.D.; Marco, J. A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. *J. Power Sources* **2017**, *360*, 301–318. [CrossRef]
26. Jamil, F.; Ibrahim, M.; Ullah, I.; Kim, S.; Kahng, H.K.; Kim, D.H. Optimal smart contract for autonomous greenhouse environment based on IoT blockchain network in agriculture. *Comput. Electron. Agric.* **2022**, *192*, 106573. [CrossRef]
27. Li, W.; Cao, D.; Jöst, D.; Ringbeck, F.; Kuipers, M.; Frie, F.; Sauer, D.U. Parameter sensitivity analysis of electrochemical model-based battery management systems for lithium-ion batteries. *Appl. Energy* **2020**, *269*, 115104. [CrossRef]
28. Shahbazi, Z.; Byun, Y.C.; Kwak, H.Y. Smart Home Gateway Based on Integration of Deep Reinforcement Learning and Blockchain Framework. *Processes* **2021**, *9*, 1593. [CrossRef]

29. Limoge, D.W.; Annaswamy, A.M. An adaptive observer design for real-time parameter estimation in lithium-ion batteries. *IEEE Trans. Control Syst. Technol.* **2018**, *28*, 505–520. [CrossRef]
30. Shahbazi, Z.; Byun, Y.C. Twitter Sentiment Analysis Using Natural Language Processing and Machine Learning Techniques. In Proceedings of the KIIT Conference 2021; pp. 42–44. Available online: https://www.dbpia.co.kr/journal/publicationDetail?publicationId=PLCT00000960 (accessed on 24 January 2022).
31. Shahbazi, Z.; Byun, Y.C. Improving the Product Recommendation System based-on Customer Interest for Online Shopping Using Deep Reinforcement Learning. *Soft Comput. Mach. Intell.* **2021**, *1*, 31–35.
32. Bartlett, A.; Marcicki, J.; Onori, S.; Rizzoni, G.; Yang, X.G.; Miller, T. Model-based state of charge estimation and observability analysis of a composite electrode lithium-ion battery. In Proceedings of the 52nd IEEE Conference on Decision and Control, Firenze, Italy, 10–13 December 2013; pp. 7791–7796.
33. Shahbazi, Z.; Byun, Y.C. Deep Learning Method to Estimate the Focus Time of Paragraph. *Int. J. Mach. Learn. Comput.* **2020**, *10*. [CrossRef]
34. Shahbazi, Z.; Byun, Y.C. LDA Topic Generalization on Museum Collections. In *Smart Technologies in Data Science and Communication*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 91–98.
35. Zheng, L.; Zhang, L.; Zhu, J.; Wang, G.; Jiang, J. Co-estimation of state-of-charge, capacity and resistance for lithium-ion batteries based on a high-fidelity electrochemical model. *Appl. Energy* **2016**, *180*, 424–434. [CrossRef]
36. Shahbazi, Z.; Byun, Y.C. Computing focus time of paragraph using deep learning. In Proceedings of the 2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), Seogwipo, Korea, 8–10 May 2019; pp. 1–4.
37. Yang, N.; Feng, J.; Sun, Q.; Liu, T.; Zhong, D. Online estimation of state-of-health for lithium ion batteries based on charge curves. In Proceedings of the 2016 11th International Conference on Reliability, Maintainability and Safety (ICRMS), Hangzhou, China, 26–28 October 2016; pp. 1–8.
38. Zhang, J.; Hou, J.; Zhang, Z. Online state-of-health estimation for the lithium-ion battery based on an LSTM neural network with attention mechanism. In Proceedings of the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 1334–1339.
39. Ren, L.; Zhao, L.; Hong, S.; Zhao, S.; Wang, H.; Zhang, L. Remaining useful life prediction for lithium-ion battery: A deep learning approach. *IEEE Access* **2018**, *6*, 50587–50598. [CrossRef]
40. Wu, J.; Zhang, C.; Chen, Z. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Appl. Energy* **2016**, *173*, 134–140. [CrossRef]
41. Richardson, R.R.; Birkl, C.R.; Osborne, M.A.; Howey, D.A. Gaussian process regression for in situ capacity estimation of lithium-ion batteries. *IEEE Trans. Ind. Inform.* **2018**, *15*, 127–138. [CrossRef]
42. Wu, Y.; Xue, Q.; Shen, J.; Lei, Z.; Chen, Z.; Liu, Y. State of health estimation for lithium-ion batteries based on healthy features and long short-term memory. *IEEE Access* **2020**, *8*, 28533–28547. [CrossRef]
43. Wang, Z.; Ma, J.; Zhang, L. State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the Gaussian process regression. *IEEE Access* **2017**, *5*, 21286–21295. [CrossRef]
44. Li, W.; Jiao, Z.; Du, L.; Fan, W.; Zhu, Y. An indirect RUL prognosis for lithium-ion battery under vibration stress using Elman neural network. *Int. J. Hydrogen Energy* **2019**, *44*, 12270–12276. [CrossRef]
45. Li, Y.; Sheng, H.; Cheng, Y.; Stroe, D.I.; Teodorescu, R. State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis. *Appl. Energy* **2020**, *277*, 115504. [CrossRef]
46. Stroe, D.I.; Schaltz, E. Lithium-ion battery state-of-health estimation using the incremental capacity analysis technique. *IEEE Trans. Ind. Appl.* **2019**, *56*, 678–685. [CrossRef]
47. Yang, D.; Zhang, X.; Pan, R.; Wang, Y.; Chen, Z. A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. *J. Power Sources* **2018**, *384*, 387–395. [CrossRef]
48. Shu, X.; Li, G.; Zhang, Y.; Shen, J.; Chen, Z.; Liu, Y. Online diagnosis of state of health for lithium-ion batteries based on short-term charging profiles. *J. Power Sources* **2020**, *471*, 228478. [CrossRef]
49. Lu, C.; Tao, L.; Fan, H. Li-ion battery capacity estimation: A geometrical approach. *J. Power Sources* **2014**, *261*, 141–147. [CrossRef]
50. Mawonou, K.S.; Eddahech, A.; Dumur, D.; Beauvois, D.; Godoy, E. State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. *J. Power Sources* **2021**, *484*, 229154. [CrossRef]
51. Ardeshiri, R.R.; Ma, C. State of Charge Estimation of Lithium-ion Battery Using Deep Convolutional Stacked Bidirectional LSTM. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6.
52. Klass, V.; Behm, M.; Lindbergh, G. A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. *J. Power Sources* **2014**, *270*, 262–272. [CrossRef]
53. Xiong, R.; Zhang, Y.; Wang, J.; He, H.; Peng, S.; Pecht, M. Lithium-ion battery health prognosis based on a real battery management system used in electric vehicles. *IEEE Trans. Veh. Technol.* **2018**, *68*, 4110–4121. [CrossRef]
54. Saha, B.; Goebel, K. Battery data set. In *NASA AMES Prognostics Data Repository*; NASA Ames Research Center: Moffett Field, CA, USA, 2007.
55. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
56. Shahbazi, Z.; Byun, Y.C. Product recommendation based on content-based filtering using XGBoost classifier. *Int. J. Adv. Sci. Technol.* **2019**, *29*, 6979–6988.

57. Li, Y.; Zou, C.; Berecibar, M.; Nanini-Maury, E.; Chan, J.C.W.; Van den Bossche, P.; Van Mierlo, J.; Omar, N. Random forest regression for online capacity estimation of lithium-ion batteries. *Appl. Energy* **2018**, *232*, 197–210. [CrossRef]
58. Chen, Y.; Huang, D.; Zhu, Q.; Liu, W.; Liu, C.; Xiong, N. A new state of charge estimation algorithm for lithium-ion batteries based on the fractional unscented kalman filter. *Energies* **2017**, *10*, 1313. [CrossRef]
59. Huang, Z.; Yang, F.; Xu, F.; Song, X.; Tsui, K.L. Convolutional gated recurrent unit–recurrent neural network for state-of-charge estimation of lithium-ion batteries. *IEEE Access* **2019**, *7*, 93139–93149. [CrossRef]