









Article

# Random Forest-Based Grouping for Accurate SOH Estimation in Second-Life Batteries

Joelton Deonei Gotz <sup>1,2,\*</sup>, José Rodolfo Galvão <sup>3</sup>, Fernanda Cristina Corrêa <sup>3</sup>, Alceu André Badin <sup>1</sup>, Hugo Valadares Siqueira <sup>3</sup>, Emilson Ribeiro Viana <sup>4</sup>, Attilio Converti <sup>5,\*</sup> and Milton Borsato <sup>2</sup>

- <sup>1</sup> Graduate Program in Electrical and Computer Engineering (CPGEE), Universidade Tecnológica Federal do Paraná (UTFPR-CT), Curitiba 80230-901, Brazil; badin@utfpr.edu.br
- <sup>2</sup> Postgraduate Program in Mechanical and Materials Engineering (PPGEM), Universidade Tecnológica Federal do Paraná (UTFPR-CT), Curitiba 81280-340, Brazil; borsato@utfpr.edu.br
- <sup>3</sup> Graduate Program in Electrical Engineering, Universidade Tecnológica Federal do Paraná (UTFPR-PG), Ponta Grossa 84017-220, Brazil; jgalvao@alunos.utfpr.edu.br (J.R.G.); fernandacorrea@utfpr.edu.br (F.C.C.); hugosiqueira@utfpr.edu.br (H.V.S.)
- <sup>4</sup> Physics Department, Universidade Tecnológica Federal do Paraná (UTFPR-CT), Curitiba 80230-901, Brazil; emilsonjunior@professores.utfpr.edu.br
- <sup>5</sup> Department of Civil, Chemical and Environmental Engineering, University of Genoa, Pole of Chemical Engineering, Via Opera Pia 15, 16145 Genoa, Italy
- \* Correspondence: gotz@alunos.utfpr.edu.br (J.D.G.); converti@unige.it (A.C.)

**Abstract:** Retired batteries pose a significant current and future challenge for electric mobility due to their high cost and the need for a state of health (SOH) above 80% to supply energy efficiently. Recycling and alternative applications are the primary options for these batteries, with recycling still undergoing research as regards more efficient and cost-effective techniques. While advancements have been made, researchers are actively seeking improved methods. Repurposing retired batteries for lower-performance applications like stationary systems or low-speed vehicles is recommended. Second-life batteries (SLB) can be directly reused or reconstructed, with the latter involving the disassembly, measurement, and separation of cells based on their characteristics. The traditional measurement process, involving full charge and discharge cycles, is time-consuming. To address this, a Machine Learning (ML)-based SOH estimator is introduced in this work, offering the instant measurement and estimation of battery health without complete discharge. The results indicate that the model can accurately identify SOH within a nominal capacity range of 1400–2300 mAh, with a resolution near 45.70 mAh, in under five minutes of discharging. This innovative technique could be instrumental in selecting and assembling SLB packs.

**Keywords:** lithium-ion batteries; second-life batteries; state of health; machine learning; SOH estimation



**Citation:** Gotz, J.D.; Galvão, J.R.; Corrêa, F.C.; Badin, A.A.; Siqueira, H.V.; Viana, E.R.; Converti, A.; Borsato, M. Random Forest-Based Grouping for Accurate SOH Estimation in Second-Life Batteries. *Vehicles* **2024**, *6*, 799–813. <https://doi.org/10.3390/vehicles6020038>

Academic Editors: Yiqun Liu and Mohammed Chadli

Received: 9 February 2024

Revised: 27 April 2024

Accepted: 29 April 2024

Published: 30 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Retired batteries have become essential for electric mobility's present and future [1]. The motivation for this is related to the fact that most types of lithium-ion batteries (LIB) are suitable for supplying enough energy for high-speed electric cars and trucks when their state of health (SOH) is above 80% [2,3].

After that, these batteries must be repurposed, either by recycling or a second application as a second-life battery (SLB) [4,5]. The recycling process is still under investigation, and despite several advancements in the last few years, new and optimal solutions must be compared to actual techniques [6]. Then, reducing costs and improving performance are necessary to maximize material recovery and eliminate environmental contamination [7]. A review of the recycling methods for retired batteries can be found in [8,9].

The other option is a second application in SLB [10,11]. An SLB can be applied to semi- or entirely stationary systems to store energy from solar and wind-solar farms [12–14].

Besides that, SLBs can supply power for low-speed vehicles such as golf carts, trucks, lifts, automated guided vehicles (AGVs), or others [10,15].

Lithium-ion batteries (LIB) are the most common type of storage system adopted these days due to the several advantages offered by this technology, such as long lifetime, high density and capacity, and others [16,17]. However, it is sensitive to failures, and unique management systems are required to maintain the cells under safe conditions during their operation [16–18]. Therefore, SLBs need more robustness control due to the unique and particular characteristics that a pack of them will show, with cells having different capacities and resistances [11,19,20].

Thus, the cells of an SLB pack must be correctly selected in the packing assembly phase so as to avoid failures and optimize the cycles of the batteries. In this way, a new group can be built via direct reuse or new pack formation [21,22].

In direct use, the pack is removed from the original vehicle and coupled into the second application. This does not require work to understand the characteristics of each cell. Consequently, a high difference can be acquired in the capacity or impedance of the cells. This action can lead to a very imbalanced system that can accelerate cell degradation and reduce cycle time [21].

On the other hand, reconfiguration consists of disassembling the pack, separating the cells, and measuring the cells according to their capacity, SOH, resistance, and other factors. Then, the cells are divided into clusters with similar characteristics and assembled with similar features. The new reconfiguration allows for a pack construction with more homogeneous cells, reducing the risk associated with a highly imbalanced system [21].

Thus, the second approach (reconfiguration) is more optimal for SLB. However, this process is prolonged and can take a much longer than re-using [21]. Among the activities, capacity measurement is one of the most time-consuming, and is not applicable to battery selection. This high time consumption arises because it is necessary to charge the cells fully and then discharge them to measure their capacity. Based on the estimated capacity, it is possible to calculate the SOH of the battery [23].

To reduce the time-consuming traditional approach to capacity measurement used for SLB selection, this work presents the application of a Machine Learning (ML) technique to estimate SOH as capacity within a few minutes without knowing the previous characteristics of the cell. In this scenario, the proposed model learns from a historical database, and can predict the final capacity of the test batteries.

The rest of the work is segmented as follows: The state-of-the-art is discussed in the second section. Then, in the third section, the proposed idea is elaborated; the results and discussion are given in the fourth section, and finally, the conclusions follow.

## 2. State-of-the-Art

SOH is one of the essential indicators of a battery's health. It indicates how long the battery will be suitable for in its first or second application. Besides this, it is one of the most critical inputs for the security control of a battery management system (BMS) [24]. In the literature, three different approaches are given to estimating the SOH: model-based (MD), knowledge-based (KB), and data-driven (DD) [24].

MD systems use mathematical calculations, circuit building, and empirical experiments to create algorithms to perform SOH estimations [17,25]. Several applications of SOH estimation using the traditional model can be found in the literature. A complete review of the SOH model-based system is presented by Cacciato et al. [26]. The authors applied the SOH model and then improved the SOH estimation, adding a PI-based observer scheme to enhance the model's accuracy and estimate SOC and SOH in real time.

In contrast, KB systems are developed by recognizing the issues under examination. This approach suits nonlinear and complex problems, such as those observed in the LIBs curves [17]. In this area, vital works such as that of Dai et al. [27] present a combination of knowledge-based neural networks to estimate SOH. According to the authors, the combination could reduce the error to less than 1.7%.

Finally, the use of DD for LIB applications has increased significantly in the last decade. That has happened mainly due to the advent of Industry 4.0, the increasing number of Internet of Things (IoT) devices, and the large datasets available worldwide [28]. The main contribution of DD is that this approach makes decisions based on the data. It enables the application of ML and Deep Learning algorithms. These learn from historical data and can perform inferences in real time [17,29,30].

Several approaches are presented in the literature to estimate SOH using the DD concept. Wang et al. [31] used Random Forest (RF) and Gated Recurrent Unit (GRU) for SOC estimations. On the other hand, Jia et al. [32] developed a Gaussian Process Regression (GPR) model to predict SOH and Remaining Useful Life (RUL) to estimate the lifetime of the battery. GPR was also used by Wang et al. [33]. The authors performed several experiments, and the proposed model showed a root-absolute-error (MAE) of just 1.7%. On the other hand, Chen et al. [34] presented an SOH estimation based on a Support Vector Machine (SVM).

Besides that, some neural networks have been used to predict the SOH of batteries. Van and Quang [35] used Long Short-Term Memory (LSTM) to indicate not only SOH but also the solid electrolyte interphase (SEI) and the internal resistances of the cell. Their results have been compared with those of a feed-forward network (FNN). Their LSTM gave better results than FNN.

In the same direction, Zhang et al. [36] used the NASA and CALCE datasets to estimate the SOH of the batteries. The authors also used LSTM and indicated that the proposed work could predict the available capacity and RUL of the batteries. Finally, Xu et al. [37] assessed a combination of convolution neural networks (CNN) and LSTM to estimate SOH. The authors added a CNN to reduce the neural network degradation generated by the multi-layer LSTM. According to the authors, the proposed method could estimate SOH with an RMSE below 0.004 when using the NASA and Oxford datasets.

Differently from the approaches in the above references, SOH is a standard approach that uses some indirect estimations. It is possible to calculate SOH based on capacity or internal resistance [30,38,39]. Calculating the SOH based on capacity is a traditional mathematical approach that takes into account the entire accumulated discharge current for a cycle. Then, the SOH is derived from the ratio of the actual capacity to the nominal capacity of the cell, as observed in Equation (1) [30,38].

$$\text{SOH} = \frac{\text{Capacity}_{\text{current}}}{\text{Capacity}_{\text{new}}} \quad (1)$$

Conversely, SOH can be estimated based on the cell's internal resistance. SOH estimation based on capacity is more accurate. On the other hand, it is faster when based on the internal resistance [39,40]. Besides this, according to Hoque et al. [40], two main characteristics make this parameter excellent for use in SOH estimation. Firstly, internal resistance is not a linear behavior, unlike the normal aging process of batteries. Therefore, this indicator can offer an excellent parameter for SOH estimation in early cycles. Secondly, the model can be applied to different scenarios and conditions because the model does not need to observe the discharge curve, which consumes a lot of time.

Equation (2) shows the SOH calculation based on resistance [38]. It is calculated by considering the actual and final resistance when SOH is 100%. Furthermore, as mentioned in the introduction, a new SLB pack can be formed from reused cells or with a new reconfiguration. Even though reuse is faster, some of the cells may show severe differences in their main features, limiting good performance. Therefore, reconfiguration is ideal when constructing a new SLB pack [21].

$$\text{SOH} = \frac{\text{Resistance}_{\text{end\_of\_life}} - \text{Resistance}_{\text{actual}}}{\text{Resistance}_{\text{end\_of\_life}} - \text{Resistance}_{\text{new}}} \quad (2)$$

In this scenario, the worst cell of a pack usually sets the limit of the charge and discharge process. This limitation is essential to keeping all cells within the safety and security limits [41]. Therefore, to optimize the cycle time, a new SLB must comprise cells with similar capacity requirements, degradation factors, and impedance in order to prolong cycle time while maintaining secure operation [21].

The traditional construction of a new SLB pack comprises several time-consuming steps. The first step is the disassembly of retired packs. As even batteries from the same group have different capacities and resistances, it is difficult to maintain and recover historical data from the cells, and select them, without a batch of experiments. The old BMS is discarded because it will not be precise in the new application. Therefore, several experiments are performed after disassembly to measure parameters such as capacity, impedance, and internal resistance [21,41]. This task consumes a significant amount of time and effort.

Capacity is one of the most time-consuming processes because it requires fully charging the cell and discharging it to measure its available energy. Then, the cells are separated into different groups. Cells with similar characteristics can be grouped into the same pack and will perform similarly for a new application. After that, cells will be assembled with a new BMS to guarantee the security and performance of the pack. A new BMS must be selected according to the characteristics of the cells in the new pack and its application [41,42].

In the literature, some works have prepared such frameworks to reduce the time taken for SLB pack formation. Meng et al. [42] critically reviewed the Artificial Intelligence (AI) and ML approaches to LIB disassembly. According to the authors, AI and ML are prominent tools used for solving the security, heterogeneity, and uncertainty problems in LIB disassembly. The work also presents applications such as SVM, RF, a combination of RFID and IoT, and others in this area.

The disassembly of the LIB pack and assembly of the SLB pack is not a new topic, but the process still takes a long time. Therefore, applying new approaches such as ML and IA can speed up the process, requiring less energy and time. Besides this, it is a more secure approach, since there is less manual interaction with the batteries during this process, which can preclude accidents [43]. In this way, works presented by Zhou et al. [41], Zhou et al. [44], and Lee and Kum [21] offer interesting frameworks for reducing the time involved in SLB cell separation and pack formation.

Zhou et al. [41] presented a framework employing an unsupervised bisecting K-means algorithm combined with a fast pulse test to segment the cells into groups with similar characteristics. According to the authors, the solution has a comparable accuracy (over 2 min) to the SLB traditional selection process, which takes, on average, 5 h. Compared to this work, our proposal uses supervised ML to classify batteries. It also finds the actual capacity of the cell within a few minutes, which will be demonstrated in the following sections.

On the other hand, Lee and Kum [21] developed a framework to select cells for SLB according to homogeneous parameters. According to the authors, their primary intention was to build a homogeneous pack that reduces the variation in cells and the loss of energy in the group. They developed a screening algorithm that could reduce 60–70% of the cell-to-cell variation in the pack.

Zhou et al. [44] presented the use of SVM for screening the batteries. The authors disassembled four packs of batteries in carrying out the model. According to the manuscript, they reached an accuracy of about 97% and reduced the disassembly time by four-fifths. In this way, our proposal can reduce the time required from 5 h to a few minutes, and offers an RMSE of about 45 mAh.

Liu et al. [45] used a data-driven approach with a convolutional auto-encoder to extract the features of batteries and fuse a feature base, and then use conditional generative adversarial networks to enhance them. Their results indicate an inconsistent rate reduction of about 96%. A previous work by Liu et al. [46] used a convolution neural network in

two-step time-series clustering and hybrid resampling. They applied the model to imbalanced data and derived an inconsistent rate of reduction of about 91%.

Gu et al. [47] presented an approach to the large-scale screening of retired batteries based on cloud-edge architecture. The proposed idea uses a Light Gradient Boosting Machine trained on the cloud and makes inferences related to the edge to be scaled. The model uses partial charge/discharge curves and can reduce the time sixfold. The results indicate an accuracy of about 97%.

However, Lu et al. [43] also demonstrated the use of AI for disassembling retired batteries, but their work is focused on the security process during the disassembly. They used a thermal camera to monitor the cell's temperature and separate the batteries according to their types and models. Similar work on AI and vision computing can be observed in Zorn et al. [48], but it focuses more on sorting the batteries.

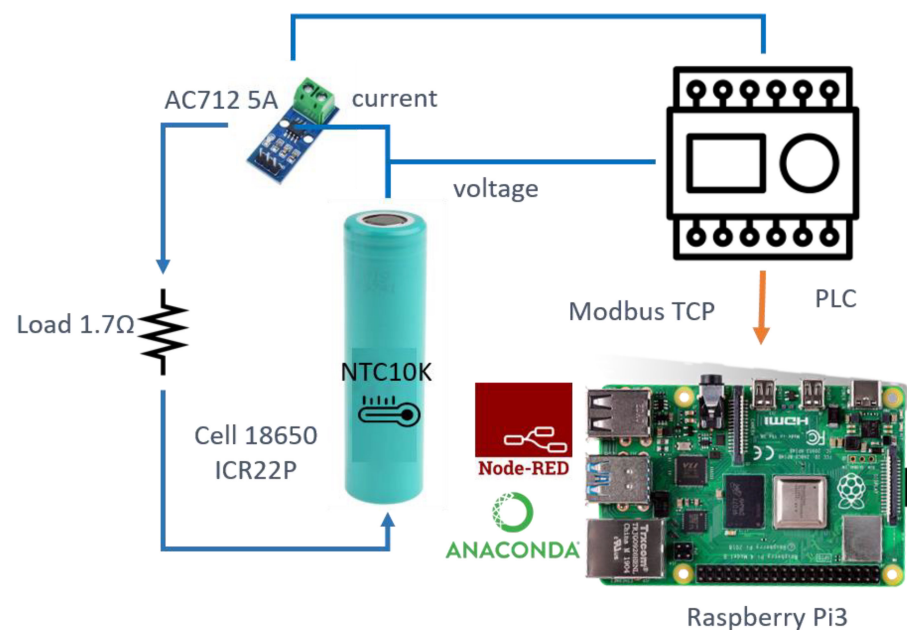
### 3. Development Description

As observed in the literature, SOH based on capacity and resistance is time-consuming and needs to be made more accurate using Coulomb counting. Despite that, ML approaches can be helpful for SOH estimation because, based on a large dataset, the models can learn the behaviors of the main parameters of the cell and then predict the SOH.

This work presents a method based on ML to predict the SOH of an SLB according to its capacity within a few minutes, using an RF approach composed of one hundred estimators. The basic idea is to indicate the SOH estimation of the SLB without knowing its characteristics, such as the aging process, internal resistance, internal capacity, and other features.

Therefore, it is essential to have a dataset featuring a massive variation in capacity. Then, OD abuse was applied to a LIB cell for 100 cycles. The cell was fully charged for each cycle using the current-constant and the voltage-constant method. After that, it was discharged until its voltage was below 1 V, forcing critical degradation.

Figure 1 shows the proposed monitoring architecture. A Programmable Logic Controller (PLC) LOGO from Siemens® was used to collect the main parameters from the analog inputs. Then, a couple of sensors were used to monitor the cell's voltage, current, and temperature during the SOH estimation.



**Figure 1.** The proposed architecture comprises a PLC, Raspberry, and sensors to monitor the batteries.



An ACS712 5A has been used to monitor the current of the discharge. This Hall effect sensor interrupts the flowing current in order to measure its magnitude. Its sensibility is 180 mV for each Ampere. The output is connected to the analog input from PLC. Additionally, the second analog input receives the cell's voltage, which is used to monitor the energy level during discharging.

Conversely, two NTC100K sensors monitor the environmental and cell temperature. The cell temperature is used to estimate SOH, maintains the operation under controlled conditions and avoids security and performance problems.

Then, a Raspberry Pi3 communicates with the PLC to request the incoming data. In a node-red flow written in the Pi3, a node requests data every second using the Modbus TCP protocol. After each request, the PLC sends a package containing the demanded data with a response according to the registered table protocol.

To evaluate the proposed idea, an 18650 Samsung LIB with a nominal capacity of 2200 mAh and 3.7 V (see details in Table 1) was used. The cell was charged and discharged for one hundred cycles under an OD abuse of 1.0 V to provoke loss capacity. During each charging cycle, the cell was charged in two phases. Firstly, it was charged under a current constant of 1.7 A until the voltage came to 4.2 V, and then, it was charged under a voltage constant until its current decayed to zero A.

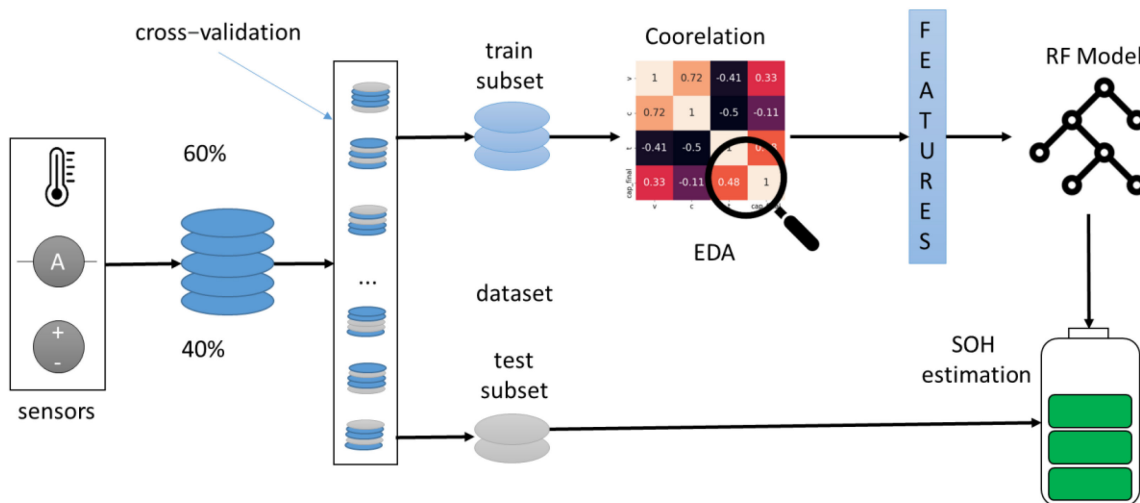
**Table 1.** Specifications of the lithium-ion ICR18650-22P battery used in the experiments.

| Item                      | Specification           |
|---------------------------|-------------------------|
| Cathode                   | Nickel-cobalt-manganese |
| Anode                     | Graphite                |
| Nominal capacity          | 2200 mAh                |
| Used capacity             | 1700 mAh                |
| Internal resistance       | 70 mΩ                   |
| Nominal voltage           | 3.7 V                   |
| Upper voltage             | 4.25 V                  |
| Lower voltage             | 2.5 V                   |
| Maximum charging current  | 1C                      |
| Maximum discharge current | 10 A                    |
| Dimensions                | Ø18.25 × 65 mm          |
| Weight                    | 42 g                    |

Conversely, the cell was discharged with a constant load of 1.7 Ω. This load maintained the discharge current rate close to 1C until the voltage reached 2.5 V. After that, the current decayed due to the low energy availability in the cell. When discharged below the safety level, low voltage (about 2.5 V) caused a loss of capacity and the internal resistance rise.

The degradation process enhanced the data's quantity and heterogeneity, which is very useful when validating the proposed idea under different battery conditions. For each experiment, the cell's ambient temperature, the current applied in the process, and the cell voltage were monitored every second and saved in the CSV file.

Then, the data were uploaded to the Anaconda Environment, where they were processed in a Jupyter Notebook using Python. Figure 2 shows the ML pipeline followed to build and validate the proposed idea. Firstly, the incoming CSV composed a dataset. According to the degradation rate, this dataset has been split into twenty ranges of capacity, which will be described in the next section. A cross-validation method was used to select some cycles of each subset to compose the training subset, and the rest was used to compose the test dataset; 60% of the data were used for training, and the rest for testing.



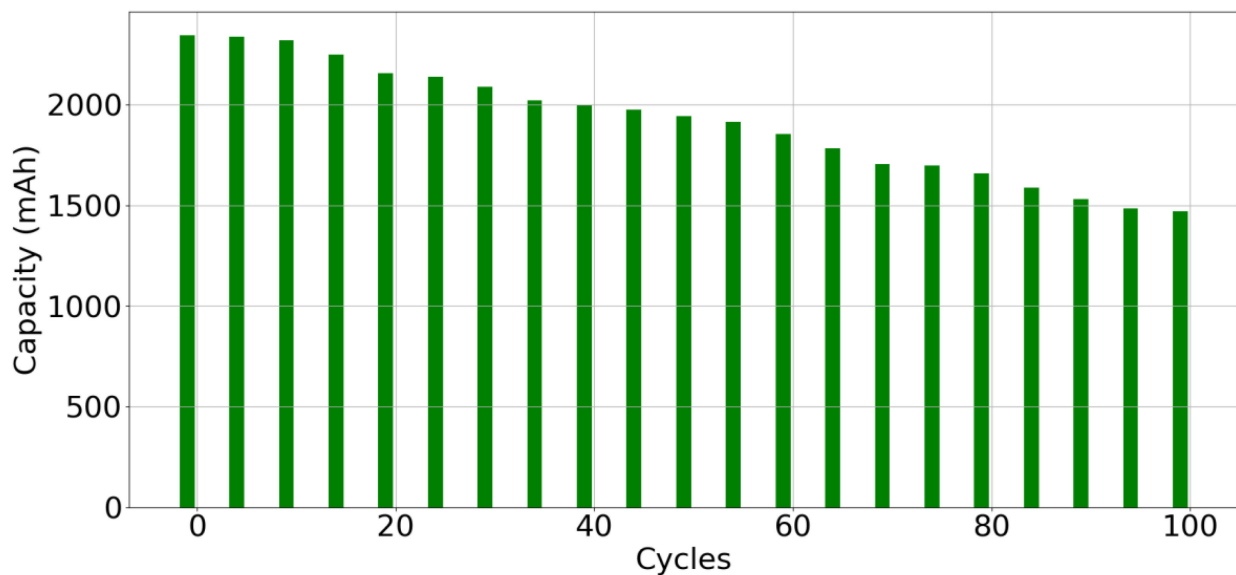
**Figure 2.** Machine Learning pipeline of the method proposed to estimate SOH of SLB.

Thus, an engineering data analysis (EDA) has been performed to understand the problem, clean the data, and identify the correlation between the inputs in order to choose the features of the model, as observed in Figure 2. Following EDA, features were selected to construct the input data for the RF model. Then, an RF model was built with one hundred decision tree estimators, with the voltage, current, and temperature as the model’s inputs (features), and the SOH estimation capacity as the output [49,50].

After the model’s building and training, it was tested with the test subset. The results will be demonstrated in the next section.

**4. Results and Discussion**

The data collected from one hundred cell experiments under a process of discharging to 1.0 V have been saved and uploaded to the Anaconda Environment. Figure 3 shows the loss capacity of the cell discharged to 1.0 V. According to the picture, the cell lost power, dropping from 2345 mAh to 1461 mAh, representing a loss of 37% capacity.



**Figure 3.** Variation of the capacity of the cell under one hundred cycles, when discharged to 1.0 V.

This capacity variation is useful in proving the efficacy of our model, because it indicates the actual capacities of batteries with different characteristics originating from

retired packs. Next, the dataset was split into 60% for training and 40% for testing. However, Figure 3 shows twenty different capacity ranges, each with about five cycles. As observed in Figure 2, a cross-validation approach has been used to randomly select 60% of each range to provide the training dataset, while two other cycles of each range have been used for the test subset.

Next, the EDA was performed. In this step, the noise from the data is cleaned, and a correlation is constructed to understand the relationship between variables. Figure 4 shows the Pearson and Spearman correlations. The Pearson correlation indicates a linear relationship [45]. Thus, according to the figure, voltage (v) and temperature (t) have a high linear correlation, and current has an inverse linear relationship with the capacity (final\_capacity). This proves that a linear model such as RF could be suitable for use in estimating SOH.

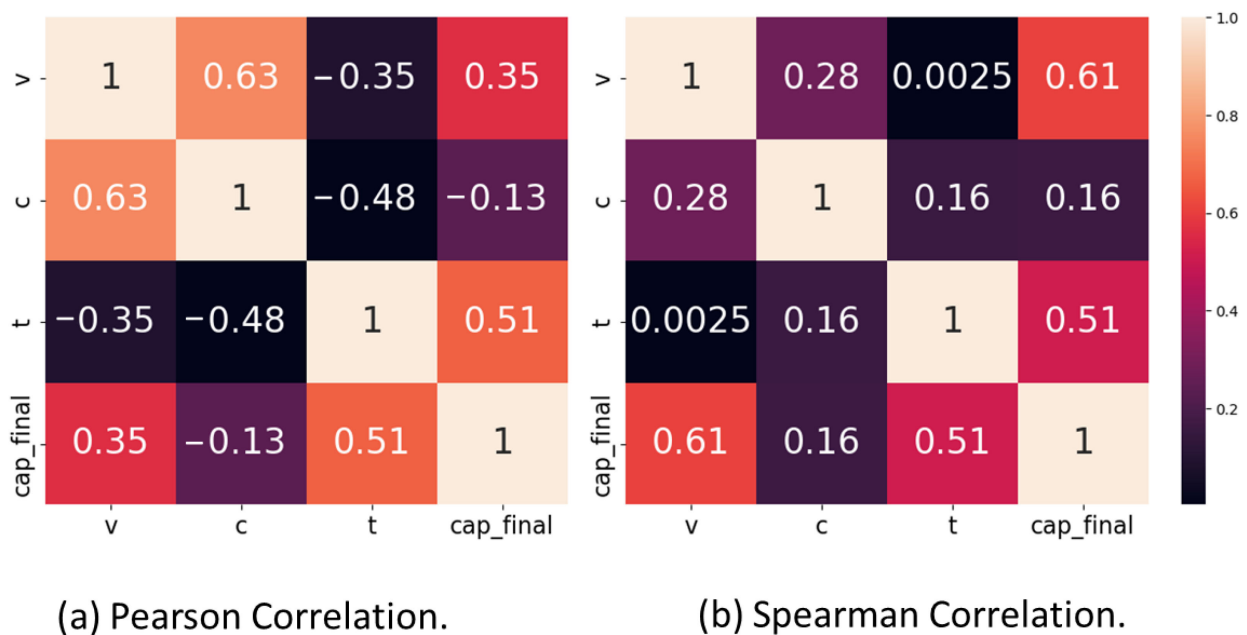


Figure 4. Pearson and Spearman correlations between voltage, current, temperature, and the cells' capacity.

Conversely, Figure 4 also shows the Spearman relationship. It indicates not only the linear but also the monotonic relation between the variables, i.e., if the relation increases, the monotonic relation suggests that it will continuously increase. The Spearman correlation also indicates a considerable relation between the variables [51].

Therefore, due to the Spearman and Pearson analysis, voltage, current, and temperature are the features that should be described by the model in Figure 2. After that, the model is chosen. RF was here selected because of its unique characteristics, making it more robust than other statistical ML models. RF comprises several decision tree nodes (named "Estimators") that endow it with ensemble learning characteristics. This means that each single estimator will deliver a "weak" answer. Conversely, if the RF model combines all "weak" answers, it can yield a "strong" answer, reducing the chance of overfitting and bias. Besides that, compared to neural network models, RF is softer and requires fewer datasets and lower-powered computing resources [52–55].

With the RF selected, the next step is the selection of hyperparameters. In the literature, it is possible to find some hyperparameters for tuning RF, such as the tuning-grid search algorithm (TGSA) or the tuning-randomized search algorithm (TRSA). Both algorithms attempt all possibilities to find the best option [55]. In this model, the primary search hyperparameters are the number of estimators, the maximum depth to which the forest

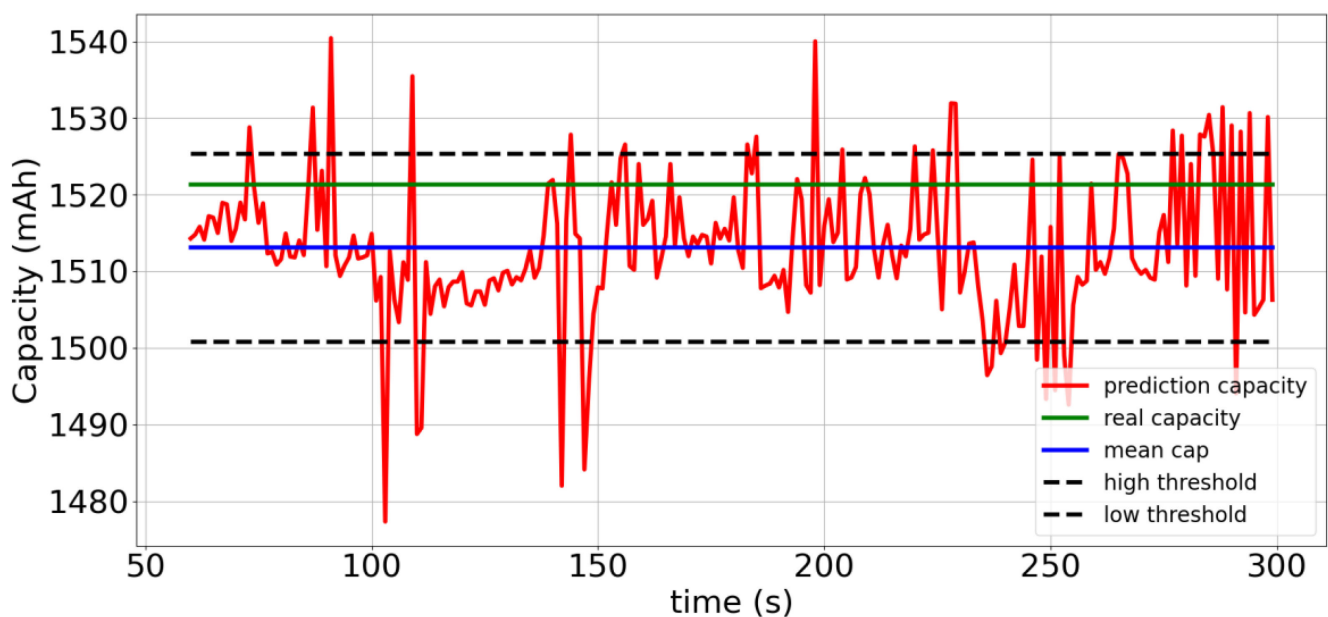


grows, the minimum number of samples for each estimator (node), and the maximum number of features [56,57].

Despite all this, we built a simple search hyperparameter with some singular characteristics. By default, this algorithm maintains the number of features, the maximum depth of the forest, and the minimum number of samples for each estimator. It has a range of estimators: 20, 40, 60, 80, 100, 120, 140 and 160. The process results show that the best performance was achieved when using 100 estimators.

Next, as the model seeks to reduce the capacity measurement time of the traditional process of testing batteries, both training and testing RF models have been employed using the initial data from each cycle. Firstly, the model was trained and applied using data from 50 to 300 s. As samples were collected every second, each cycle's dataset yielded 250 samples related to the three features.

Figure 5 shows the result of the RF applied in cycle number #89. The model received data from 50 to 300 s to make inferences. According to Figure 5, the actual capacity in this cycle was 1521 mAh. The RF model yields a value of predicted capacity every second as well as an error related to the real capacity. However, when the mean was calculated within 250 s, the mean value was a capacity of 1512 mAh, indicating an error of 0.5%. This result could reduce the testing time for SLB pack formation to 250 s and demonstrate a battery capacity that is almost the same as the actual value.



**Figure 5.** Results of the application of RF for cycle number #89.

Besides this, Figure 5 shows one high threshold composed of the mean estimation plus the root-mean-squared error (RMSE), as well as a low point comprising the mean estimation minus the RMSE. The high threshold for this case was 1525 mAh, and the low threshold was 1500 mAh. The predicted values are maintained under the thresholds. According to the picture, most of the predicted points are inside of the thresholds, indicating the strong performance of the model.

On the other hand, Figure 6 shows the actual versus predicted capacity based on the RF model trained and tested using just 250 s of each cycle (from 50 to 300 s of the experiment). According to the figure, the RF could perform better for all forty experiments. This error is maintained in Figure 7, demonstrating the RMSE of each cycle.

According to Figure 7, the average RMSE of the model was just 45.7 mAh, representing a minor error. However, in some cycles, the model had a more significant error. Cycle #1 represents the highest RMSE, with about 130 mAh. Besides this, cycles #5 and #39 had an

absolute error of 100 mAh. Finally, cycles #9 and #39 had an error above 100 mAh above 100 mAh. Conversely, most cycles had an error below 60 mAh, which could reduce the mean RMSE by about 45.7 mAh.

A complementary result can be observed in Figure 8, where the actual error percentage of the RF model is visible. The figure shows a comparison between the mean estimated capacity and the actual capacity of the cell. In this image, it is possible to see that while the capacity decays over the cycles, the error in capacity prediction using RF is constantly maintained, with a maximum error of about 7%, but on average 2.86%, compared to the actual capacity of the cycles.

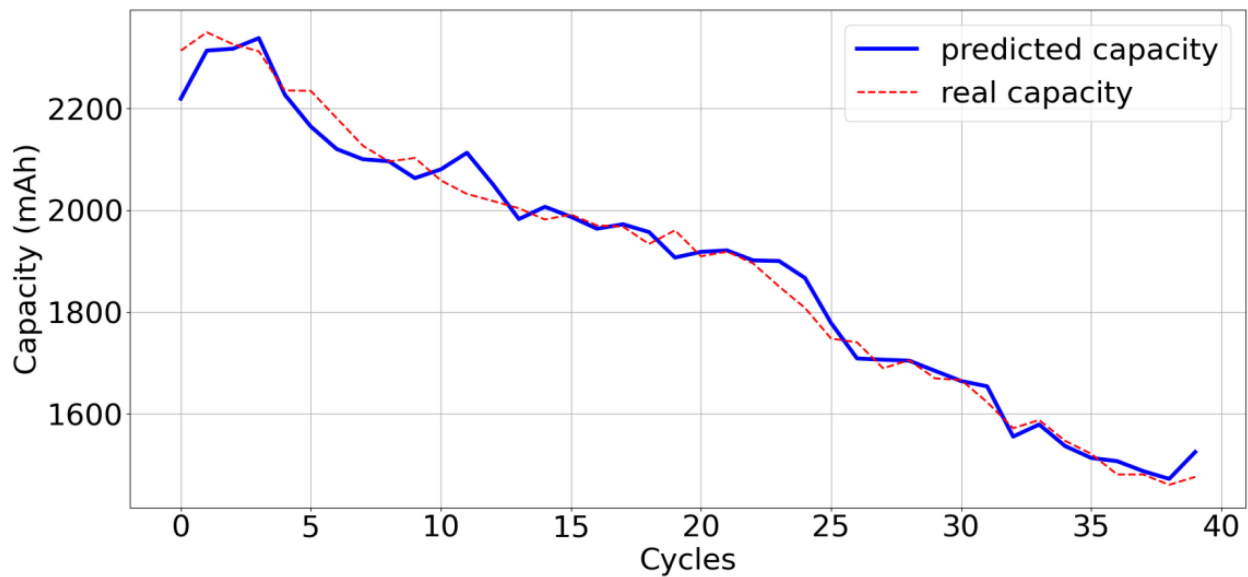


Figure 6. Real x predicted capacity of RF SOH estimated for eighty-two cycles.

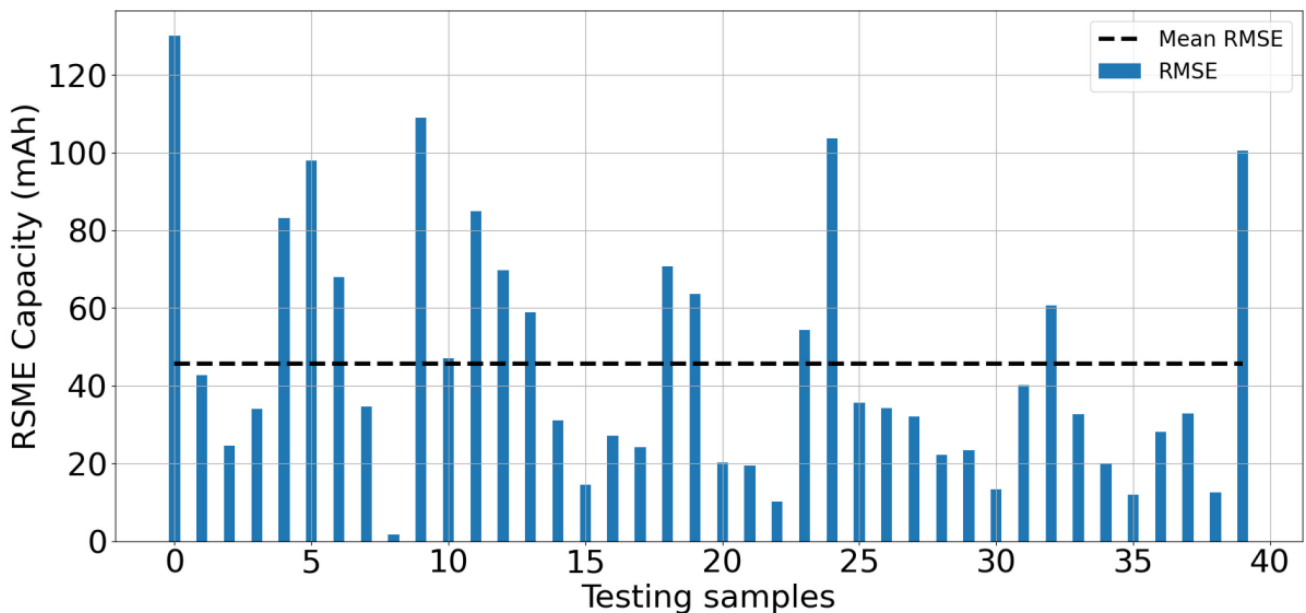
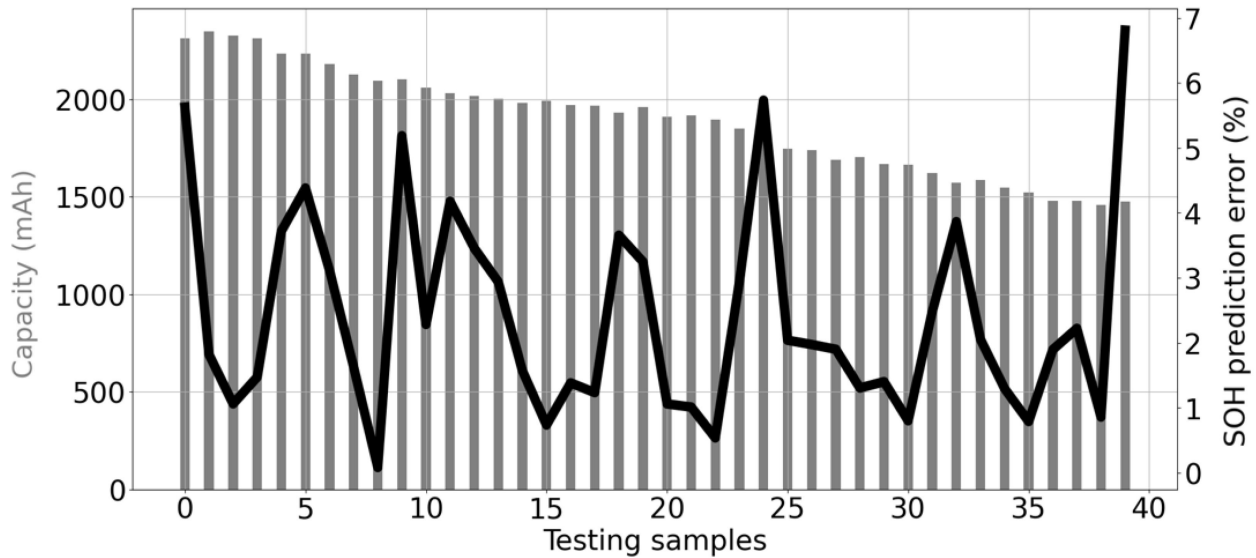


Figure 7. SOH estimation via RMSE for the RF application with data between 50 and 300 s.

The results indicate the model’s excellent performance in estimating SOH for SLB based on incoming data ranging from 50 to 300 s. Despite this, it is essential to analyze and determine whether 250 s is the optimal testing time for SOH estimations of SLB. Then, the

following experiments will demonstrate the impact on SOH predictions, with variations in the time of the available data. Therefore, new models have been trained and tested with different input window times, as shown in Table 2. After re-testing, Table 2 indicates the RMSE (mAh) results for different RF model databases.



**Figure 8.** Percentage error of the RF model for each cycle (black) compared with the real capacity of the cycle (in gray).

**Table 2.** Results of the variation on RF testing.

| Prediction Time(s) | RMSE (mAh) |
|--------------------|------------|
| 50–100             | 48.08      |
| 50–200             | 47.85      |
| 50–300             | 45.70      |
| 50–400             | 46.23      |
| 50–500             | 43.34      |
| 50–600             | 47.51      |
| 50–700             | 51.55      |
| 50–800             | 51.83      |
| 50–900             | 51.10      |
| 50–1000            | 53.74      |
| 50–2000            | 46.10      |
| 50–3000            | 43.70      |

This work aims to estimate a real SOH without knowing the characteristics of the cycle in the first stage. However, when the model fits better, the problem must be managed. On the one hand, the best result was given by the 50 to 3000 s dataset, for which the RMSE was about 43.70 mAh. However, it takes a long time, and the goal of the work is to reduce the time spent on battery characterization. Therefore, the best option would be to test with a database for 50–100 s. However, the error here will be more significant and can result in incorrect conclusions.

Thus, according to Table 1, the best options seem to be variations of 50–300 and 50–400, for which the RMSE values were 45.70 and 46.23 mAh, respectively. As both RMSEs are similar, the best variation for OD degradation is 50–300, with which the model can perform well and quickly. Despite the results, it is essential to highlight that the model performs well with different time window ranges. This is affirmed by the fact that the highest RMSE mean error was just 53.74 mAh, which was observed for the 50–1000 s window.

Therefore, this model's performance depends on the amount of data and can be used for SLB pack formation.

Compared to works presented in the literature, our idea has several advantages. Firstly, it can identify the capacity within a few minutes with a low RMSE, as shown in Table 2. Besides that, it requires minimal hardware and component resources, which reduces the cost of the experiments. Finally, the proposed model can be evaluated and improved with greater data availability, i.e., the more tests it does, the more assertiveness the model develops.

## 5. Conclusions

SLBs are becoming an increasingly suitable solution for the use of retired batteries from electric vehicles. In this area, two options are available regarding new pack formation: reuse and reconfiguration. Reusing the batteries consumes less time, but the new pack arrangement can contain batteries with different characteristics. Therefore, reconfiguration is more appropriate for second applications.

Despite this, reconfiguring a new pack involves disassembling old groups, testing the batteries, measuring the capacities, resistances, and impedances, and separating the cells to assemble a new pack. This workload is very time-consuming, and reducing the time spent on some tasks can accelerate the process.

Among the activities, measurement is one of the most time-consuming tasks because it is necessary to charge the cell and discharge it in order to calculate the capacity and estimate some vital indicators, such as SOH.

Therefore, this work represents a novelty in applying an SOH estimator composed of an RF model to estimate the SOH of the cells within just a few minutes of the discharging process. In this way, an RF model has been built, trained, and applied to a database constructed after 90 cycles of the discharging of an SBL with a range capacity between 2200 and 1400 mAh.

The proposed SOH estimator could estimate the capacities of cells with an average error of less than 45.70 mAh in less than five minutes. Further, in just 100 s, the SOH of the cells could be estimated with an error below 49 mAh, which indicates that the model is suitable and helpful for SLB formation.

Finally, this work has presented a technique that can accelerate the classification of SLBs. However, disassembling cells in the pack is still necessary in order to apply the algorithm. Thus, in future studies, the authors intend to apply this method directly to a retired pack without disassembling the cells. This could reduce the time-consuming process of SLB preparation.

**Author Contributions:** Conceptualization, J.D.G. and M.B.; methodology, J.D.G.; software, J.D.G.; validation, J.D.G., J.R.G. and M.B.; formal analysis, J.R.G., F.C.C. and E.R.V.; investigation, J.D.G., H.V.S. and A.C.; resources, F.C.C. and M.B.; data curation, J.D.G. and J.R.G.; writing—original draft preparation, J.D.G., J.R.G., F.C.C. and M.B.; writing—review and editing, J.R.G., A.A.B., A.C. and E.R.V.; visualization, J.D.G., H.V.S. and A.C.; supervision, M.B.; project administration, A.C. and M.B.; funding acquisition, F.C.C., H.V.S. and A.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is funded by FUNDEP Rota 2030 Public Call \#01/2020, Agreement #271 92.03.01/2020.16-00.

**Data Availability Statement:** The data are confidential.

**Acknowledgments:** Grateful acknowledgment is extended to FUNDEP (Rota 2030) for its invaluable financial support of our research project.

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

## References

1. Shahjalal, M.; Roy, P.K.; Shams, T.; Fly, A.; Chowdhury, J.I.; Ahmed, M.R.; Liu, K. A review on second-life of Li-ion batteries: Prospects, challenges, and issues. *Energy* **2022**, *241*, 122881. [[CrossRef](#)]
2. Rimpas, D.; Kaminaris, S.D.; Piromalis, D.D.; Vokas, G.; Orfanos, V.A. Impact of lithium battery recycling and second-life application on minimizing environmental waste. *Environ. Sci. Proc.* **2023**, *26*, 41. [[CrossRef](#)]
3. Saxena, S.; Le Floch, C.; MacDonald, J.; Moura, S. Quantifying EV battery end-of-life through analysis of travel needs with vehicle powertrain models. *J. Power Sources* **2015**, *282*, 265–276. [[CrossRef](#)]
4. Hu, X.; Deng, X.; Wang, F.; Deng, Z.; Lin, X.; Teodorescu, R.; Pecht, M.G. A review of second-life lithium-ion batteries for stationary energy storage applications. *Proc. IEEE* **2022**, *110*, 735–753. [[CrossRef](#)]
5. Hua, Y.; Liu, X.; Zhou, S.; Huang, Y.; Ling, H.; Yang, S. Toward sustainable reuse of retired lithium-ion batteries from electric vehicles. *Resour. Conserv. Recycl.* **2021**, *168*, 105249. [[CrossRef](#)]
6. Rufino Júnior, C.A.; Riva Sanseverino, E.; Gallo, P.; Koch, D.; Kotak, Y.; Schweiger, H.G.; Zanin, H. Towards a business model for second-life batteries: Barriers, opportunities, uncertainties, and technologies. *J. Energy Chem.* **2023**, *78*, 507–525. [[CrossRef](#)]
7. Chen, M.; Ma, X.; Chen, B.; Arsenault, R.; Karlson, P.; Simon, N.; Wang, Y. Recycling end-of-life electric vehicle lithium-ion batteries. *Joule* **2019**, *3*, 2622–2646. [[CrossRef](#)]
8. Arshad, F.; Li, L.; Amin, K.; Fan, E.; Manurkar, N.; Ahmad, A.; Yang, J.; Wu, F. A comprehensive review of advancement in recycling anode and electrolyte from spent lithium ion batteries. *ACS Sustain. Chem. Eng.* **2020**, *8*, 36. [[CrossRef](#)]
9. Faessler, B. Stationary, second use battery energy storage systems and their applications: A research review. *Energies* **2021**, *14*, 2335. [[CrossRef](#)]
10. Michellini, E.; Höschele, P.; Ratz, F.; Stadlbauer, M.; Rom, W.; Ellersdorfer, C.; Moser, J. Potential and most promising second-life applications for automotive lithium-ion batteries considering technical, economic and legal aspects. *Energies* **2023**, *16*, 2830. [[CrossRef](#)]
11. Kampker, A.; Heimes, H.H.; Offermanns, C.; Vienenkötter, J.; Frank, M.; Holz, D. Identification of challenges for second-life battery systems—A literature review. *World Electr. Veh. J.* **2023**, *14*, 80. [[CrossRef](#)]
12. Dong, Q.; Liang, S.; Li, J.; Kim, H.C.; Shen, W.; Wallington, T.J. Cost, energy, and carbon footprint benefits of second-life electric vehicle battery use. *iScience* **2023**, *26*, 107195. [[CrossRef](#)] [[PubMed](#)]
13. Kebir, N.; Leonard, A.; Downey, M.; Jones, B.; Rabie, K.; Bhagavathy, S.; Hirmer, S. Second-life battery systems for affordable energy access in Kenyan primary schools. *Sci. Rep.* **2023**, *13*, 1374. [[CrossRef](#)]
14. Falk, J.; Nedjalkov, A.; Angelmahr, M.; Schade, W. Applying lithium-ion second life batteries for off-grid solar powered system—A socio-economic case study for rural development. *Z. Energiewirtschaft.* **2020**, *44*, 47–60. [[CrossRef](#)]
15. Colarullo, L.; Thakur, J. Second-life EV batteries for stationary storage applications in Local Energy Communities. *Renew. Sust. Energ. Rev.* **2022**, *169*, 112913. [[CrossRef](#)]
16. Chen, Y.; Kang, Y.; Zhao, Y.; Wang, L.; Liu, J.; Li, Y.; Liang, Z.; He, X.; Li, X.; Tavajohi, N.; et al. A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards. *J. Energy Chem.* **2021**, *59*, 83–99. [[CrossRef](#)]
17. Hu, X.; Zhang, K.; Liu, K.; Lin, X.; Dey, S.; Onori, S. Advanced fault diagnosis for lithium-ion battery systems: A review of fault mechanisms, fault features, and diagnosis procedures. *IEEE Ind. Electron. Mag.* **2020**, *14*, 65–91. [[CrossRef](#)]
18. Gotz, J.D.; Galvão, J.R.; Werlich, S.H.; Silveira, A.M.D.; Corrêa, F.C.; Borsato, M. Reducing the capacity loss of lithium-ion batteries with machine learning in real-time—A study case. *Machines* **2022**, *10*, 1114. [[CrossRef](#)]
19. Gotz, J.D.; Galvão, J.R.; Silveira, A.; Viana, E.R.; Correa, F.C.; Borsato, M. Intelligent management for second-life lithium-ion batteries with backup cells. In *Flexible Automation and Intelligent Manufacturing: Establishing Bridges for More Sustainable Manufacturing Systems*; Silva, F.J.G., Ferreira, L.P., Sá, J.C., Pereira, M.T., Pinto, C.M.A., Eds.; Springer: Cham, Switzerland, 2024; pp. 1011–1018.
20. Börner, M.F.; Frieges, M.H.; Späth, B.; Spütz, K.; Heimes, H.H.; Sauer, D.U.; Li, W. Challenges of second-life concepts for retired electric vehicle batteries. *Cell Rep. Phys. Sci.* **2022**, *3*, 101095. [[CrossRef](#)]
21. Lee, K.; Kum, D. Development of cell selection framework for second-life cells with homogeneous properties. *Int. J. Electr. Power Energy Syst.* **2019**, *105*, 429–439. [[CrossRef](#)]
22. Canals Casals, L.; García, B.; González Benítez, M. A cost analysis of electric vehicle batteries second life businesses. In *Project Management and Engineering Research, Proceedings of the XVIII Congreso Internacional de Dirección e Ingeniería de Proyectos, Alcañiz, Spain, 16–18 July 2014*; Springer: Cham, Switzerland, 2016. [[CrossRef](#)]
23. Teng, J.H.; Chen, R.J.; Lee, P.T.; Hsu, C.W. Accurate and efficient SOH estimation for retired batteries. *Energies* **2023**, *16*, 1240. [[CrossRef](#)]
24. Goh, H.H.; Lan, Z.; Zhang, D.; Dai, W.; Kurniawan, T.A.; Goh, K.C. Estimation of the state of health (SOH) of batteries using discrete curvature feature extraction. *J. Energy Storage* **2022**, *50*, 104646. [[CrossRef](#)]
25. Sun, S.; Zhang, H.; Ge, J.; Che, L. State-of-health estimation for lithium-ion battery using model-based feature optimization and deep extreme learning machine. *J. Energy Storage* **2023**, *72*, 108732. [[CrossRef](#)]
26. Cacciato, M.; Nobile, G.; Scarcella, G.; Scelba, G. Real-time model-based estimation of SOC and SOH for energy storage systems. *IEEE Trans. Power Electron.* **2016**, *32*, 794–803. [[CrossRef](#)]
27. Dai, H.; Zhao, G.; Lin, M.; Wu, J.; Zheng, G. A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and Markov chain. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7706–7716. [[CrossRef](#)]



28. Reis, M.S.; Saraiva, P.M. Data-driven process system engineering—Contributions to its consolidation following the path laid down by George Stephanopoulos. *Comput. Chem. Eng.* **2022**, *159*, 107675. [[CrossRef](#)]
29. Sheng, H.; Zhou, Y.; Bai, L.; Shi, L. Transfer state of health estimation based on cross-manifold embedding. *J. Energy Storage* **2022**, *47*, 103555. [[CrossRef](#)]
30. Zhang, M.; Yang, D.; Du, J.; Sun, H.; Li, L.; Wang, L.; Wang, K. A review of SOH prediction of Li-ion batteries based on data-driven algorithms. *Energies* **2023**, *16*, 3167. [[CrossRef](#)]
31. Wang, X.; Hu, B.; Su, X.; Xu, L.; Zhu, D. State of health estimation for lithium-ion batteries using Random Forest and Gated Recurrent Unit. *J. Energy Storage* **2024**, *76*, 109796. [[CrossRef](#)]
32. Jia, J.; Liang, J.; Shi, Y.; Wen, J.; Pang, X.; Zeng, J. SOH and RUL prediction of lithium-ion batteries based on Gaussian process regression with indirect health indicators. *Energies* **2020**, *13*, 375. [[CrossRef](#)]
33. Wang, J.; Deng, Z.; Yu, T.; Yoshida, A.; Xu, L.; Guan, G.; Abudula, A. State of health estimation based on modified Gaussian process regression for lithium-ion batteries. *J. Energy Storage* **2022**, *51*, 104512.
34. Chen, Z.; Sun, M.; Shu, X.; Xiao, R.; Shen, J. Online state of health estimation for lithium-ion batteries based on support vector machine. *Appl. Sci.* **2018**, *8*, 925. [[CrossRef](#)]
35. Nguyen Van, C.; Quang, D.T. Estimation of SoH and internal resistances of Lithium ion battery based on LSTM network. *Int. J. Electrochem. Sci.* **2023**, *18*, 100166. [[CrossRef](#)]
36. Zhang, L.; Ji, T.; Yu, S.; Liu, G. Accurate prediction approach of SOH for lithium-ion batteries based on LSTM method. *Batteries* **2023**, *9*, 177. [[CrossRef](#)]
37. Xu, H.; Wu, L.; Xiong, S.; Li, W.; Garg, A.; Gao, L. An improved CNN-LSTM model-based state-of-health estimation approach for lithium-ion batteries. *Energy* **2023**, *276*, 127585. [[CrossRef](#)]
38. Li, Y.; Luo, L.; Zhang, C.; Liu, H. State of health assessment for lithium-ion batteries using incremental energy analysis and bidirectional long short-term memory. *World Electr. Veh. J.* **2023**, *14*, 188. [[CrossRef](#)]
39. Ramadan, M.; Pramana, B.; Widayat, S.; Amifia, L.; Cahyadi, A.; Wahyunggoro, O. Comparative study between internal ohmic resistance and capacity for battery state of health estimation. *J. Mechatron. Electr. Power Veh. Technol.* **2015**, *6*, 113. [[CrossRef](#)]
40. Hoque, M.A.; Nurmi, P.; Kumar, A.; Varjonen, S.; Song, J.; Pecht, M.G.; Tarkoma, S. Data driven analysis of lithium-ion battery internal resistance towards reliable state of health prediction. *J. Power Sources* **2021**, *513*, 230519. [[CrossRef](#)]
41. Zhou, Z.; Ran, A.; Chen, S.; Zhang, X.; Wei, G.; Li, B.; Kang, F.; Zhou, X.; Sun, H. A fast screening framework for second-life batteries based on an improved bisecting K-means algorithm combined with fast pulse test. *J. Energy Storage* **2020**, *31*, 101739. [[CrossRef](#)]
42. Meng, K.; Xu, G.; Peng, X.; Youcef-Toumi, K.; Li, J. Intelligent disassembly of electric-vehicle batteries: A forward-looking overview. *Resour. Conserv. Recycl.* **2022**, *182*, 106207. [[CrossRef](#)]
43. Lu, Y.; Maftouni, M.; Yang, T.; Zheng, P.; Young, D.; Kong, Z.J.; Li, Z. A novel disassembly process of end-of-life lithium-ion batteries enhanced by online sensing and machine learning techniques. *J. Intell. Manuf.* **2023**, *34*, 2463–2475. [[CrossRef](#)] [[PubMed](#)]
44. Zhou, Z.; Duan, B.; Kang, Y.; Shang, Y.; Cui, N.; Chang, L.; Zhang, C. An efficient screening method for retired lithium-ion batteries based on support vector machine. *J. Clean. Prod.* **2020**, *267*, 121882. [[CrossRef](#)]
45. Liu, C.; Tan, J.; Wang, X. A data-driven decision-making optimization approach for inconsistent lithium-ion cell screening. *J. Intell. Manuf.* **2020**, *31*, 833–845. [[CrossRef](#)]
46. Liu, C.; Tan, J.; Shi, H.; Wang, X. Lithium-ion cell screening with convolutional Neural Networks based on two-step time-series clustering and hybrid resampling for imbalanced data. *IEEE Access* **2018**, *6*, 59001–59014. [[CrossRef](#)]
47. Gu, X.; Li, J.; Zhu, Y.; Wang, Y.; Mao, Z.; Shang, Y. A quick and intelligent screening method for large-scale retired batteries based on cloud-edge collaborative architecture. *Energy* **2023**, *285*, 129342. [[CrossRef](#)]
48. Zorn, M.; Ionescu, C.; Klohs, D.; Zähl, K.; Kisseler, N.; Daldrup, A.; Hams, S.; Zheng, Y.; Offermanns, C.; Flamme, S.; et al. An approach for automated disassembly of lithium-ion battery packs and high-quality recycling using computer vision, labeling, and material characterization. *Recycling* **2022**, *7*, 48. [[CrossRef](#)]
49. Siqueira, H.; Santana, C.; Macedo, M.; Figueiredo, E.; Gokhale, A.; Bastos-Filho, C. Simplified binary cat swarm optimization. *Integr. Comput. Aided Eng.* **2021**, *28*, 35–50. [[CrossRef](#)]
50. Belotti, J.; Siqueira, H.; Araujo, L.; Stevan, S.L., Jr.; de Mattos Neto, P.S.; Marinho, M.H.; de Oliveira, J.F.L.; Usberti, F.; de Almeida, L.F.M.; Converti, A.; et al. Neural-based ensembles and unorganized machines to predict streamflow series from hydroelectric plants. *Energies* **2020**, *13*, 4769. [[CrossRef](#)]
51. Gotz, J.; Guerrero, G.; De Queiroz, J.; Viana, E.; Borsato, M. Diagnosing failures in lithium-ion batteries with Machine Learning techniques. *Eng. Fail. Anal.* **2023**, *150*, 107309. [[CrossRef](#)]
52. Belouadah, E.; Popescu, A.; Kanellos, I. A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Netw.* **2020**, *135*, 38–54. [[CrossRef](#)]
53. Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random forests and decision trees. *Int. J. Comput. Sci. Issues* **2012**, *9*, 272–278.
54. Gotz, J.D.; de Souza, P.H.G.; Galvão, J.R.; Corrêa, F.C.; Badin, A.A.; Viana, E.R.; Borsato, M. Enhancing SOC estimation accuracy via incremental learning techniques for second-life batteries. *Sustain. Energy Technol. Assess.* **2024**, *65*, 103753. [[CrossRef](#)]
55. Zhu, N.; Zhu, C.; Zhou, L.; Zhu, Y.; Zhang, X. Optimization of the Random Forest hyperparameters for power industrial control systems intrusion detection using an improved grid search algorithm. *Appl. Sci.* **2022**, *12*, 10456. [[CrossRef](#)]

- 
56. Le, H.-L.; Le, T.-T.; Vu, T.-T.-H.; Tran, D.-H.; Chau, D.V.; Ngo, T.-T.-T. A survey on the impact of hyperparameters on Random Forest performance using multiple accelerometer datasets. *Int. J. Comput. Their Appl.* **2023**, *30*, 351–361.
  57. de O. Santos, D.S., Jr.; de Mattos Neto, P.S.G.; de Oliveira, J.F.L.; Siqueira, H.V.; Barchi, T.M.; Lima, A.R.; Madeiro, F.; Dantas, D.A.P.; Converti, A.; Pereira, A.C.; et al. Solar Irradiance Forecasting Using Dynamic Ensemble Selection. *Appl. Sci.* **2022**, *12*, 3510. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.