

# Comparison of machine learning techniques for SoC and SoH evaluation from impedance data of an aged lithium ion battery

Davide Aloisio<sup>1</sup>, Giuseppe Campobello<sup>2</sup>, Salvatore Gianluca Leonardi<sup>1</sup>, Francesco Sergi<sup>1</sup>, Giovanni Brunaccini<sup>1</sup>, Marco Ferraro<sup>1</sup>, Vincenzo Antonucci<sup>1</sup>, Antonino Segreto<sup>2</sup>, Nicola Donato<sup>2</sup>

<sup>1</sup> Institute of Advanced Energy Technologies “Nicola Giordano”, National Research Council of Italy, Salita S. Lucia sopra Contesse, 5 - 98126, Messina, Italy

<sup>2</sup> University of Messina, Department of Engineering, C.da di Dio, Vill. S.Agata, 98166 Messina, Italy

## ABSTRACT

State of charge estimation and ageing evolution of lithium ion (Li-Ion) batteries are key points for their massive applications in the market. However, the battery behavior is very complex to understand because many parameters act in determining their ageing evolution. Therefore, traditional analytical models employed for this purpose are often affected by inaccuracy. In this context, machine learning techniques can provide a viable alternative to traditional models and a useful tool to characterize the batteries behavior.

In this work, different machine learning techniques were applied to model the impedance evolution over time of an aged cobalt based Li-Ion battery, cycled under a stationary frequency regulation profile for grid application. The different ML techniques were compared in terms of accuracy to determine the state of charge and the state of health over the battery ageing phenomena. Experimental results showed that ML based on Random Forest algorithm can be profitably used for this purpose.

Section: RESEARCH PAPER

**Keywords:** Machine Learning; Electrochemical impedance spectroscopy EIS; Lithium-ion battery; State of Charge; State of Health

**Citation:** Davide Aloisio, Giuseppe Campobello, Salvatore Gianluca Leonardi, Francesco Sergi, Giovanni Brunaccini, Marco Ferraro, Vincenzo Antonucci, Antonino Segreto, Nicola Donato, Comparison of machine learning techniques for SoC and SoH evaluation from impedance data of an aged lithium ion battery, Acta IMEKO, vol. 10, no. 2, article 12, June 2021, identifier: IMEKO-ACTA-10 (2021)-02-12

**Section Editor:** Ciro Spataro, University of Palermo, Italy

**Received** January 18, 2021; **In final form** April 29, 2021; **Published** June 2021

**Copyright:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by the Italian Ministry of Economic Development under the programme “Ricerca di Sistema”, project electrochemical storage

**Corresponding author:** Davide Aloisio, e-mail: [aloisio@itae.cnr.it](mailto:aloisio@itae.cnr.it)

## 1. INTRODUCTION

As well known, Machine Learning (ML) is a subfield of computing, an Artificial Intelligence (AI) technique that provides machines with the ability to learn from the field data without explicit programming [1]. In particular, ML can be really useful in applications that try to extract some information or unknown properties (‘features’) from the dataset (usually called ‘training set’) coming from data warehouses or data lakes. Information extracted from this kind of data analyses can be used to develop prediction models for systems behaviour (subject to certain operative conditions and under some constraints). In particular, the battery behaviour characterisation is quite complex to be described through analytical models, mainly due to many parameters act in determining the ageing evolution (e.g. charge

and discharge current rates, operative temperature, depth of discharge (DoD) reached, state of charge (SoC) during rest periods and so on. Therefore, the combination of the aforementioned parameters makes the systems hard to model via analytical equations. This is particularly evident for Li-Ion batteries, for which it is more difficult to describe electrochemical processes with analytical equations, due to the nonlinearities present in their behaviours. The analytical models require, in addition to input data of the actual working conditions (current, temperature, etc.), the knowledge of many parameters (geometry, density and porosity of materials, etc.) These data are not always available or easily measurable and can vary over time (e.g. due to ageing). Therefore, the analytical models can be affected by inaccuracy.

In this context, ML techniques represent a viable alternative and a useful tool for modelling the battery behaviour. ML algorithms learn directly from experimental data, reducing the complexity of modelling, usually due to the high number of parameters and empirical adjustments needed. In addition, according to the recent literature, the application of ML techniques in the prediction of the ageing of Li-Ion batteries shows errors in the range between 0.5% - 5.5% [2]-[4]. This range of accuracy is considered a good compromise among algorithm complexity, effort spent on model development and reliability of results.

Many physical and electrical parameters are characteristic of the chemical reactions inside Li-ion batteries. Therefore, these relations could be used as tools for the battery state modelling [5], [6]. Typically, these features are derived from charging and discharging curves, since typical battery management systems (BMS) are able to collect current-voltage data. Hence these are the most commonly used parameters for real time battery monitoring [7], [8]. Various approaches based on the use of different parameters were proposed in the literature to train machine learning models. Among battery parameters, single-point terminal voltage, current, temperature, charge/discharge profiles [2], [9], [10] or their geometrical characteristics [11] were employed for this purpose. However, much more information about the status of the battery can be extracted from the impedance spectra recorded by means of electrochemical impedance spectroscopy (EIS) [12]. Indeed, the impedance spectrum of a lithium cell contains rich information on all materials properties, interfacial phenomena and electrochemical reactions.

From a practical point of view, many of these can be extrapolated by the Nyquist diagram in which inverse imaginary part of impedance is plotted against the real one for each investigated frequency (of solicitations). In the case of Li-Ion batteries, the Nyquist diagram consists of four distinct regions typically belonging to the frequency range between 10 mHz to 10 kHz [13]. In the low frequency region, an almost linear trend in the Nyquist plot is representative of the solid diffusion of lithium ions through the electrodes material. In the medium-high frequencies range, one or more semicircles usually represent the impedance of either charge transfer phenomena or passivation layers on the electrodes surface (solid electrolyte interphase-SEI). The intersection of the impedance spectrum with the real axis (pure ohmic impedance) represents the cell internal resistance. Finally, the high frequency region is representative of inductive phenomena. Since each one of these phenomena are strictly related to temperature, to SoC and to the state of health (SoH) of the cell, then the analysis of the impedance data can be used to monitor the status of the battery [14]. However, due to the large number of data involved in a single EIS spectrum and the amount of information it can contain, the use of conventional data analysis methods may be difficult. Also, because of the difficulties in measuring the impedance while the cells are active, EIS is not widely used [15]. To overcome this drawback, increasing attention is paid to the implementation of ML approaches, either to aid the fitting of the parameters of equivalent circuits able to describe the battery impedance [16], or by directly modelling the entire impedance spectrum [17].

In this paper, ML is addressed to identify possible methodologies to estimate the SoC and SoH of Li-Ion battery from EIS data, mainly aiming at developing a feasible model easy to be integrated in a battery management system (BMS). Implementation in BMSs of techniques able to extend batteries

useful life, estimating the possible replacement time (estimation of Remaining Useful Life, or RUL), is considered a key research activity in the field [18].

In Section 2, some state-of-the-art of ML techniques applied to SoH and RUL estimation are reviewed. Section 3 describes the experimental procedures employed to age the Li-Ion cell; the main parameters extrapolated to create the dataset for the algorithm; and the methods for their collection. Section 4 describes the methodology used to carry out the first selection of ML algorithm and the validation of the model. Section 5 presents the main results related to the use of different classifiers to model both SoC and capacity loss of the Li-Ion cell. Finally, in Section 6, the main observations are summarised.

## 2. ML ALGORITHMS FOR STATE OF HEALTH (SOH) AND REMAINING USEFUL LIFE (RUL) EVALUATION: A BRIEF REVIEW

Thanks to the remarkable computational capabilities of today's systems, learning algorithms applied to large quantities of data have often become the preferred approach in the search and identification of complex system behaviour, and therefore represent a valid tool for SoH estimation of batteries. In these techniques, a large amount of data, constituted by main battery parameters, are collected continuously up to the end of their life. The dataset analysis of the battery life, performed by learning algorithms, allows extracting non-linear relationships among the various parameters. The knowledge derived from this kind of information can allow a careful management of the battery, helping to extend the useful life and giving reliable prediction on possible replacement times, with obvious positive impact on costs and investments.

ML techniques such as Fuzzy Logic (FL), Support Vector Machine (SVM) and Artificial Neural Networks (ANN) have extensively been applied for the estimation of the health of batteries, and a brief review can be found in [3]. In most cases, SoH is estimated by determining battery capacity and internal resistance, parameters strictly related to SoH, from input variables behaviour analysis (current, temperature, voltage, etc.) An application of Fuzzy logic with a potential use in portable devices is reported in [19], where Electrochemical Impedance Spectroscopy (EIS) technique was used for the dataset creation. However, improper hypotheses in the Fuzzy rules [3] and reduced set of observations can lead to substantial errors. The Support Vector Machine is a regression algorithm which converts nonlinearities in a lower dimension space to a linear model developed in a higher-dimensional one [20]. Examples of application of this technique applied to SoH are reported in [21]-[25]. In particular, in [25] an online method for SoH estimation was developed determining support vectors by means of pieces of charging curves. SoH with less than 2% error for 80% of all the cases for commercial NMC LI-ion batteries was achieved. The accuracy of the results is strongly dependent on the noise and operational conditions; hence, other data manipulation techniques (particle filter, Bayesian technique) have to be used in conjunction with SVM to increase the robustness of the estimation [26], thus increasing the complexity of implementation. Relevance Vector Machine (RVM) is suggested as a possible improvement of this approach in [20]. Artificial Neural Networks (ANNs) are probably the most used approach, inspired by the biological functioning of the human brain, for modeling nonlinear systems. SoH estimation using an independently recurrent neural network (in RNN) was realised in

Table 1. Main characteristics of the tested Li-Ion cell.

Description	Value
Nominal voltage	3.7 V
Nominal capacity	1.1 Ah
Max charge current	4 A
Max discharge current	10 A
Maximum voltage	4.2 V
Minimum voltage	2.5 V
Discharge temperature	-30 ÷ 60 °C
Charge temperature	0 ÷ 60 °C
Chemistry	LiCoO <sub>2</sub> -LiNiCoMnO <sub>2</sub> /Graphite

[3]. Here SoH was predicted accurately with a root means square error (RMSE) of 1.33% and mean absolute error (MAE) of 1.14%. The main limitation is the need of a detailed analysis on experimental dataset. Different chemistries can require a precise identification and understanding of input parameters.

In [27], an improved neural network method based on the combination of a LSTM (long-short-term memory) and (PSO) Particle Swarm Optimisation was developed. The methodology proposed here uses some additional techniques in each part of the learning process, such as PSO for optimisation of the weights, dynamic incremental learning for SoH model updating, CEEMDAN method to denoise raw data, with the aim of increasing the accuracy of the model [27].

Another hybrid approach can be found in [28], where false nearest neighbour method was used in conjunction with a mixed LSTM and convolutional neural network (CNN) as a solution for unreliable sliding window sizes, a problem commonly present in data-driven RUL evaluation approaches. The complexity and topology of ANN used in these works is actually classified as deep learning, an evolution of the machine learning concept coined for neural networks which exploits the concept of multi-layer perceptron (MLP). A comparison of deep learning and different other common techniques showing its potential and advantages of data driven approaches was presented in [4]. The outcomes showed the goodness of deep neural networks (DNN), which are suitable when high accuracy is needed. However, also this technique is not easy to be implemented due to higher computational complexity and resources needed [4].

A lot of other techniques and approaches can be found in the literature. Although out of the scope of the present work, the goal of a possible implementation in BMS suggests the choice of low-complexity approaches to reduce computational resources needed and thus leading to lower energy consumption [29]. A possible alternative is given by Random Forest algorithms. They generally use reduced computation resources, and thus can result preferable in comparison to other techniques analysed, based on SVM and NN. In general, Linear regressors or Random Forest response is faster than complex model and is easily interpreted. However, it has to be underlined that the accuracy in Random Forest models is related to the number and size of trees and therefore to the availability of memory [1], [30].

### 3. DATASET COLLECTION AND CREATION

The present work was aimed at the development of a method to identify the degradation level induced by the use of Li-Ion batteries in a primary frequency regulation (FR) service. More precisely, the activity was focused on the identification of the main parameters indicating the state of battery degradation. For

this purpose, cylindrical-type 18650 Li-Ion-ion cells (Table 1) were cycle aged according to a test profile extrapolated by the standard IEC 61427-2 [31].

The standard profile requires that the storage system is able to provide symmetrical charging and discharging phases at constant power of 500 kW and 1000 kW, respectively, with a voltage range of 400–600 V. Therefore, the profile was adapted to the single cell characteristics. Moreover, in order to enhance the degradation of the cell (thus limiting the overall duration required for data collection) FR ageing tests were accelerated by operating at an ambient temperature of 45 °C. In fact, the degradation processes of Li-Ion batteries are accelerated by temperature increase [32].

The ageing tests were performed by a dual-channel Bitrode FTV1 battery cyler. In addition, the cell was tested under temperature-controlled atmosphere in an Angelantoni Discovery DM 340 BT climatic chamber.

The FR ageing profile with actual power steps imposed to the cell is shown in Figure 1. The full ageing protocol consisted of a first charge of the cell up to 100% SoC and then an execution of the FR profile. Once the cell reached the lower voltage cut-off threshold (discharged), a recharge up to 100% SoC was performed and then the cycle was restarted.

The ageing level was defined in terms of residual capacity retained by the cell. This information was obtained from periodic check-ups carried out on the cell, approximately every 10 days of operation. Parametric check-ups of the cells performed the extraction of residual capacity and impedance evaluations by means of EIS technique. Both analyses were carried out through a high reliability Autolab 302N potentiostat/galvanostat (whose potential accuracy and current accuracy are both ±0.2% of the full-scale value). It is worth noting that, due to instruments calibration and performance, measurements were considered reliable and having no impact of uncertainty on the model. The robustness of the model will be investigated in a future work.

Capacity tests, constituted by a galvanostatic discharge at nominal c-rate and room temperature, allowed to extrapolate the characteristic parameters indicative of the SoH of the cell. The recorded discharge curves at the begin of life (BoL) and different SoH levels are reported in Figure 2a. In particular, Residual capacity ( $C_d$ ) and Residual energy ( $E_d$ ) were collected and used as output variables of the database. The value of  $C_d$  was obtained by integrating the actual current ( $I_d$ ) between begin of discharge ( $t_0$ ) and end of discharge ( $t_f$ ), within the upper and lower voltage cut off limit

$$C_d = \int_{t_0}^{t_f} I_d(t) dt . \quad (1)$$

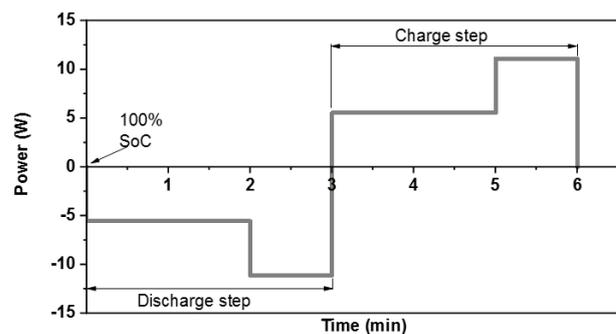


Figure 1. Power profile used to age the battery according to a frequency regulation profile extrapolated from the international standard IEC 61427-2.

Table 2. Statistical data of used dataset.

	f (Hz)	Re{Z} (Ω)	Im{Z} (Ω)	SoC (%)	C_loss (%)
Range min-max	0.01-10000	0.041-0.207	-0.072-0.067	0-100	0-8.27
Mean	797	0.0762	0.0019	50.00	4.54
STD	1951.6	0.0212	0.0175	35.36	2.69

The quantity  $Ed$  was obtained by integrating the actual power ( $i$ ) between the begin of discharge ( $t_0$ ) and the end of discharge ( $t_f$ ), within the upper and lower voltage cut off limits.

$$Ed = \int_{t_0}^{t_f} Pd(t) dt, \quad (2)$$

where  $Pd(t) = V(t) \cdot Id(t)$ , with  $V(t)$  and  $Id(t)$  representing the instantaneous values of voltage and current, respectively.

The recorded discharge curves at begin of life (BoL) and different SoH levels are reported in Figure 2a. SoH levels were defined as capacity loss of the cell identified during each parametric check-up.

As input variables of the algorithm, the complex impedance values were collected at different frequencies and SoH levels of the cell. Such information came from EIS analysis carried out in correspondence of parametric check-ups. To create the database, the impedance of the cell was recorded at different SoCs (100%, 75%, 50%, 25%, 0%) at BoL and every ten days of operation under FR cycle, until a loss of capacity (Closs) of about 8% was reached. Moreover, the loss of capacity (Closs, as effect of ageing) was used as indicative parameter of the cell SoH. Nyquist plots of the impedance recorded for different SoCs at BoL and five different SoH levels are reported in Figure 2b.

The impedance was recorded in the frequency range between 10 mHz to 10 kHz with ten points per decade, which leads to 61

values for each SoC. Finally, the data set used for the case study consists of 1830 impedance measurements. Table 2 contains some statistical information on the dataset used.

#### 4. METHODOLOGY

The above-mentioned dataset has been used to test various classification and regression techniques through the Scikit-learn tool [33], an open-source library for machine learning developed in Python. Among them, K-nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Gaussian Naive Bayes (GNB), Support Vector Classification (SVC), Decision Tree (DT), Linear Regressor, Lasso, Ridge and Random Forest were considered for performances comparison.

In order to avoid an influence on the results by the particular previous partitioning, a cross-validation technique was also used. In particular, in this phase the original data set was partitioned into 5 subsets (folds) used for tests and training. In the case of regressors, the value of the MAE and the determination coefficient ( $R^2$ ) was calculated for each round. Similarly, the accuracy (ACC) was measured for the classifiers.

The models were then compared on the basis of the average values of the aforementioned metrics obtained in the 5 validation rounds. The standard deviation (STD) was also determined from the same metrics, which provides information on the robustness of the model (in fact, lower values of STD generally correspond to more robust models).

#### 5. RESULTS

##### 5.1. Data analysis

First, correlation coefficients were analysed to investigate relationships among impedance measurements and the corresponding SoC and Closs values.

Correlation coefficients of SoC and Closs, specifically achieved for the RF cycle, are summarised in Table 3 for both rectangular and polar forms of the impedance. The analysis of the correlation coefficients shows that the highest correlation value is between the Closs measurements and the real part of the impedance ( $Re(Z)$ ), for which a correlation coefficient of 0.471 was obtained. It is also possible to observe that the correlation coefficient obtained between Closs and the impedance module ( $Abs(Z)$ ) is just smaller (0.456). The similarity between these two correlation coefficients suggests that, for the purpose of Closs modelling, it is possible to use either the module or the real part of the impedance.

In the case of SoC, the highest correlation value is obtained with the impedance phase values ( $Arg(Z)$ ) for which a correlation coefficient of 0.337 was obtained. Rectangular coordinate values, on the other hand, are uncorrelated to SoC. Therefore, it can be assumed that, for SoC modelling, the phase values of impedance are the most useful, at least for this set of data. As a consequence, it is to be expected that machine learning algorithms will perform better with the use of impedance values represented in polar coordinates rather than rectangular ones.

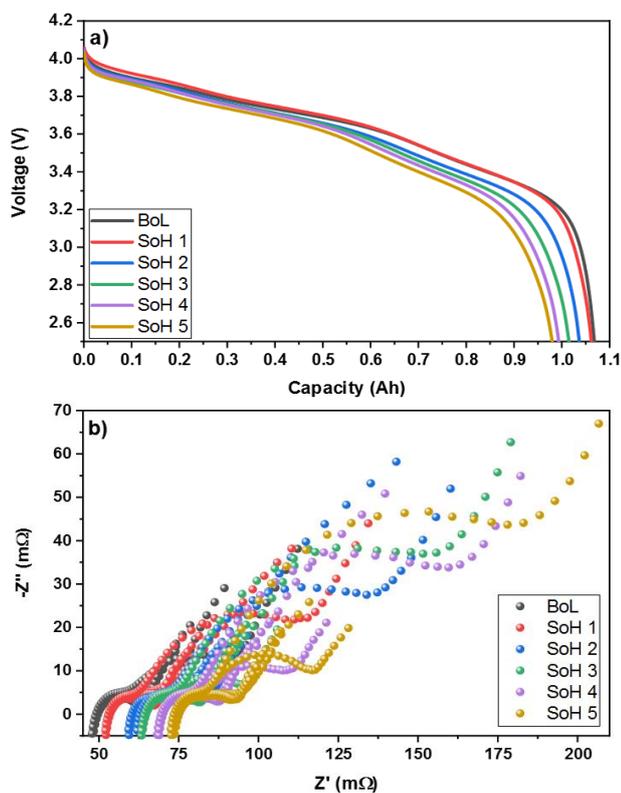


Figure 2. a) Discharge curves for extrapolation of output variables; b) Nyquist plot of the impedance used as input variables of the database.

Table 3. Correlation matrix for impedance measures evaluated on original/unfiltered data

	Re{Z} ( $\Omega$ )	Im{Z} ( $\Omega$ )	Abs{Z} ( $\Omega$ )	Arg{Z} ( $\Omega$ )
<b>Closs</b>	+0.471	+0.044	+0.456	-0.002
<b>SoC</b>	-0.166	-0.103	-0.170	-0.337

Table 4. Correlation matrix for impedance measures evaluated only on low-frequency data.

	Re{Z} ( $\Omega$ )	Im{Z} ( $\Omega$ )	Abs{Z} ( $\Omega$ )	Arg{Z} ( $\Omega$ )
<b>Closs</b>	+0.477	+0.119	+0.458	-0.001
<b>SoC</b>	-0.213	-0.1239	-0.215	-0.706

Table 5. Accuracy of some classifiers used for modeling the SoC starting for unfiltered data in rectangular and polar representation.

Classifier	Representation Z	Mean	STD
LDA	Rectangular	0.234	0.027
LDA	Polar	0.333	0.045
GNB	Rectangular	0.196	0.020
GNB	Polar	0.370	0.053
SVC	Rectangular	0.192	0.014
SVC	Polar	0.380	0.068
KNN	Rectangular	0.222	0.072
KNN	Polar	0.383	0.088
DT	Rectangular	0.329	0.020
DT	Polar	0.915	0.047

Table 6. Accuracy of some classifiers used for modeling the SoC in polar representation for filter and unfiltered data.

Classifier	Mean	STD	Mean	STD
	(original data)		(filtered data)	
LDA	0.333	0.045	0.602	0.192
GNB	0.370	0.053	0.397	0.027
SVC	0.380	0.068	0.374	0.066
KNN	0.383	0.088	0.392	0.084
DT	0.915	0.047	0.938	0.024

The above analysis was repeated considering only EIS impedance data corresponding to frequency values lower than 350 Hz. Henceforward, in this work, we will refer to this data set as filtered data. Indeed, as also reported in [34] where a similar lithium cell was used, the most important features induced by ageing on physico-chemical processes were observed only for the negative imaginary part of the impedance, which, in our case, matches with the selected filtered frequency range.

Also, it is well known that EIS at moderate and high frequency is strongly dependent on the experimental setup and cables, thus leading to measurement errors and scattered data [35].

Accordingly, the comparison of correlation coefficients reported in Table 3 and Table 4 (for original and filtered data, respectively) reveals a marked improvement in SoC correlation when only low frequency measures are considered. In particular, in the case of filtered data, a correlation coefficient of 0.706 was obtained between the SoC and the impedance phase, which is relatively higher in comparison to the value of 0.337 obtained for

the original dataset. As a consequence, machine learning algorithms are expected to provide higher performance results when trained with the filtered dataset.

## 5.2. Comparison of machine learning algorithms

Performance of several machine learning classifiers and regressors were evaluated and compared. Among them, K-nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Gaussian Naive Bayes (GNB), Support Vector Classification (SVC) and Decision Tree (DT), were considered as representative classification algorithms. The aforementioned algorithms were compared in terms of accuracy achieved on both original dataset and filtered dataset, using a cross-validation method on 5 folds.

Table 5 shows the average values and the standard deviation of the accuracy obtained for the above classifiers in the case of SoC prediction obtained by training the algorithms with the original dataset.

It is possible to observe how the use of polar representation leads to an improvement in accuracy for all classifiers with an increase between 40% and 270%, depending on the classifier. For both representations (rectangular/polar), the Decision Tree (DT) exhibits a better performance, obtaining an average accuracy equal to 0.915 with the polar representation.

This analysis was repeated considering the filtered dataset, i.e. removing high frequency points from the original dataset. As can be seen from Table 6, filtering improves accuracy of almost all classifiers (for the sake of clarity, in Table 6 only results for polar coordinates are reported).

For better comparison, in Figure 3, a box plot shows the median value (orange line), quartiles, and range of accuracy values (minimum and maximum values) in the case of algorithm trained considering filtered (Figure 3b) and original (Figure 3a) datasets. From the comparison between Figure 3a and Figure 3b it can be observed that for the LDA classifier the use of filtered values leads to a marked improvement in performance. Moreover, in the case of DT, in addition to an increase of the average accuracy value, there is also a significant reduction of data dispersion that justifies the reduction of the standard deviation in Table 6, obtained in the case of filtered data.

Similar considerations can be carried out using for comparison purpose the F1 metric [36]. In fact, as shown in Figure 4 where the macro-average F1 score obtained for the same classifiers (for both filtered and un-filtered datasets) is reported, DT classifier achieves better results even considering the macro-averaged F1 metric instead of the accuracy.

Finally, Figure 5 shows the confusion matrix obtained for the DT classifier in the case of the filtered dataset for an 80/20 distribution, i.e. with 80% of the data used for training and 20% for testing. A total of 272 SoC values were tested and only 16 of them were wrongly classified, thus obtaining an accuracy on the specific test set of 94.31%. Therefore, the achieved classification can be effectively used to evaluate the state of charge of the battery starting from the impedance and, in particular, to predict when the state of charge is below 50%. It is worth mentioning that, the choice of using classifiers instead of regressors is related to the specific application. In a few cases, in fact, classifiers able to simply detect discrete values of SoC can be useful for detecting when specific critical threshold levels have been reached, i.e. the 20% of capacity reduction commonly used for automotive applications.

It is worth noting that, in the previous analysis, the default values of Scikit-learn was used for all classifiers, i.e. all classifiers

were applied without any optimisation. This fact partially justifies why most classifiers exhibit poor performances. In addition, it is well known that LDA, like other linear classifiers and regressors such as Ridge and Lasso, adapts well to linear models while the dependence of SoC on impedance curves does not. Nevertheless, it is generally better to test and compare them due to their lower computational complexity.

Therefore, a similar analysis was carried out for Linear, Lasso, Elastic, Ridge, Gradient Boosting, Ada Boost and Random Forest regressors with the main difference that, in the case of regressors, performance was measured in terms of MAE and the determination coefficient ( $R^2$ ).

The regressor with the best performance in terms of both  $R^2$  and MAE was the Random Forest.

The distributions of the predicted values by Random forest regressor, when trained with the filtered data, are reported in Figure 6a and Figure 6b for the modeling of SoC and Closs, respectively. Figure 6 also reports the average values and standard deviations of  $R^2$  and MAE. In particular, in the case of SoC, an average value of  $R^2$  of 0.98 was achieved (see Figure 6a). In comparison to [37], which considered unfiltered data, a significant reduction was obtained in the MAE, from 2.65 to 1.87. In addition, as discussed in the following subsection, the use of filtered data leads to models with lower complexity.

### 5.3. Analysis of the Random Forest parameters

Different tradeoffs between performance and complexity of machine learning algorithms can be obtained by a proper tuning of related parameters. In the specific case of the Random Forest, the most important parameters that impact on both performance and overall complexity are the number of trees ( $n\_estimators$ ) and the maximum depth of trees ( $max\_depth$ ). Generally,

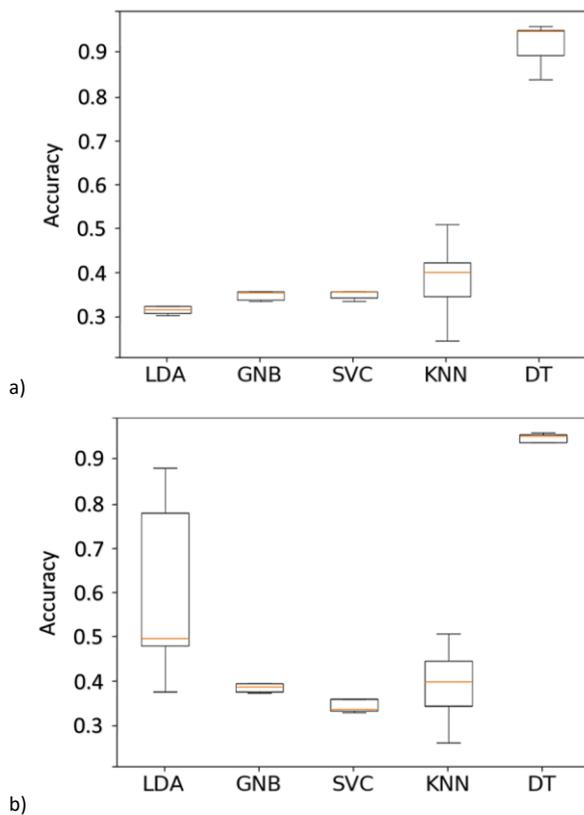


Figure 3. Accuracy of machine learning algorithms on the SoC estimation a) unfiltered, b) filtered data.

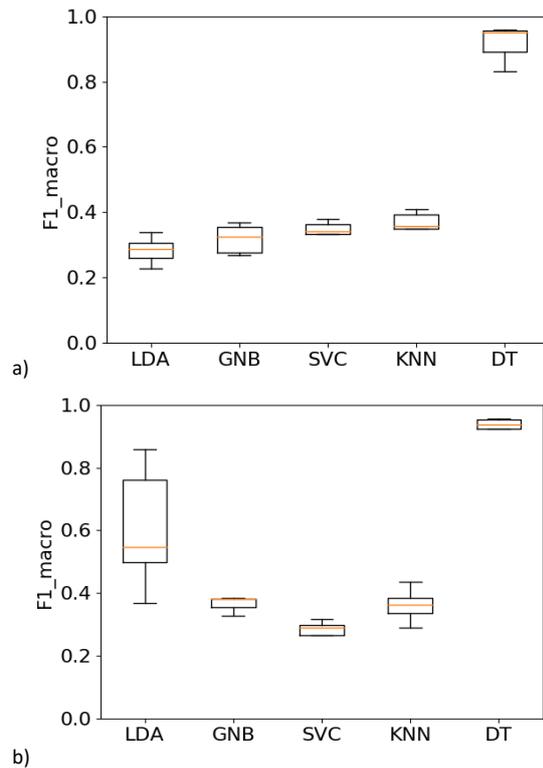


Figure 4. F1 metric results for a) unfiltered and b) filtered data.

increasing one or both of such parameters improve performance at the cost of greater complexity and estimation time.

Table 7 shows the  $R^2$  and MAE metrics obtained with Random Forest for some combinations of  $n\_estimators$  and  $max\_depth$  considering the original set, i.e. unfiltered data. It is possible to observe that  $R^2$  and MAE metrics are most affected by the  $max\_depth$  parameter. In particular, a maximum value of  $R^2$  equal to 0.97 is achieved by setting  $max\_depth = 30$ . The use of higher values increases computational complexity without significant performance advantages.

As regards the other parameter investigated (i.e.,  $n\_estimators$ ), there is no substantial difference in the values of  $R^2$  and MAE obtained by fixing  $max\_depth = 30$  and using  $n\_estimators$  values higher than 100. This analysis leads to conclude that, in the case of unfiltered data, the optimal values of the Random Forest parameters that maximise performance are  $max\_depth = 30$  and  $n\_estimators = 100$ , which are the parameters used in [37].

The same analysis was conducted for filtered data, and the related results are summarised in Table 8. In this case, better

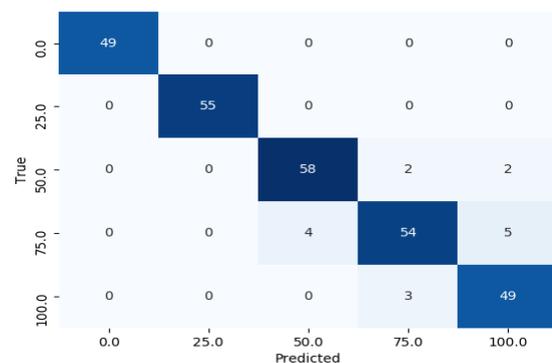


Figure 5. Confusion matrix of DT classifier.

Table 7.  $R^2$  and MAE values obtained by Random Forest technique varying parameters  $n\_estimators$  and  $max\_depth$  (on unfiltered data).

$n\_estimators$	$max\_depth$	$R^2$	MAE
10	5	0.79 (0.01)	10.47 (0.29)
10	10	0.93 (0.01)	4.49 (0.38)
10	30	0.96 (0.01)	3.32 (0.50)
10	50	0.96 (0.01)	3.34 (0.26)
25	30	0.97 (0.01)	3.11 (0.24)
50	50	0.97 (0.01)	3.02 (0.34)
100	5	0.80 (0.01)	10.43 (0.24)
100	10	0.93 (0.01)	4.43 (0.29)
100	30	0.97 (0.01)	3.02 (0.36)
100	50	0.97 (0.01)	3.03 (0.34)
1000	30	0.97 (0.01)	2.99 (0.30)

Table 8.  $R^2$  and MAE values obtained by Random Forest technique varying parameters  $n\_estimators$  and  $max\_depth$  (on filtered data).

$n\_estimators$	$max\_depth$	$R^2$	MAE
10	5	0.95 (0.01)	4.62 (0.30)
10	10	0.98 (0.00)	1.94 (0.22)
10	30	0.98 (0.00)	1.94 (0.22)
10	50	0.98 (0.00)	1.84 (0.29)
25	30	0.98 (0.00)	1.89 (0.15)
50	50	0.98 (0.00)	1.91 (0.18)
100	5	0.95 (0.01)	4.50 (0.20)
100	10	0.98 (0.00)	1.89 (0.13)
100	30	0.98 (0.00)	1.8 (0.21)
100	50	0.98 (0.00)	1.87 (0.22)
1000	30	0.98 (0.00)	1.83 (0.18)

results are achieved even with lower values of the parameters. For instance, the performance obtained using filtered data with  $max\_depth=10$  and  $n\_estimators=10$  is better than when using unfiltered data for  $max\_depth=30$  and  $n\_estimators=100$ .

Thus, by training the algorithm with filtered data we obtained models with better performance and lower complexity.

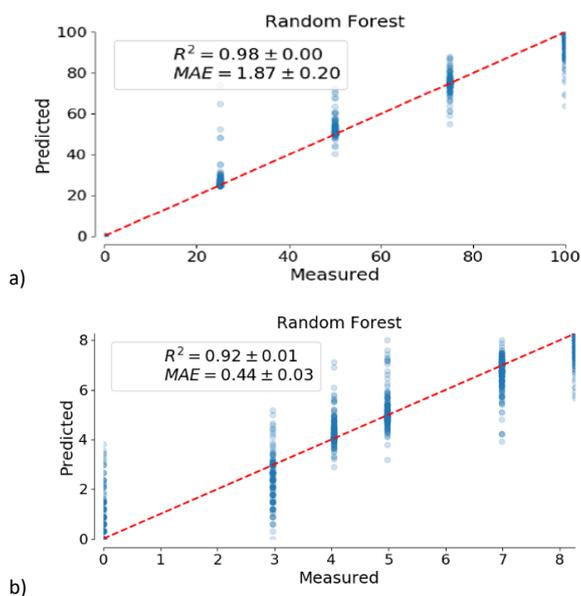


Figure 6. Random Forest distribution on filtered data for a) SoC and b) capacity loss.

## 6. CONCLUSIONS

Starting from impedance measurements, different machine learning techniques were analysed as predictors of the state of charge and the loss of capacity of a lithium battery, subjected to a frequency regulation profile for grid applications. According to the results, the following conclusions can be drawn:

- for the training of machine learning techniques, the use of impedance values expressed in polar form is to be preferred;
- Decision Trees and Random Forest provided superior performance compared to the other machine learning techniques analysed;
- using low frequency data for training Random Forest regressor improved performance in terms of  $R^2$  and MAE for both state of charge and capacity loss prediction and largely reduced overall complexity.

## ACKNOWLEDGEMENT

Special thanks to the Italian Ministry of Economic Development for funding this activity.

## REFERENCES

- [1] G. Hackeling, *Mastering Machine Learning With scikit-learn*: Packt Publishing, 2014.
- [2] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: A deep learning approach, *IEEE Access* 6 (2018), pp. 50587-50598. DOI: [10.1109/ACCESS.2018.2858856](https://doi.org/10.1109/ACCESS.2018.2858856)
- [3] P. Venugopal, State-of-health estimation of Li-ion batteries in electric vehicle using IndRNN under variable load condition, *Energies* 12(22) (2019), art. 4338. DOI: [10.3390/en12224338](https://doi.org/10.3390/en12224338)
- [4] P. Khumprom, N. Yodo, A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm, *Energies* 12(4) (2019), art. 660. DOI: [10.3390/en12040660](https://doi.org/10.3390/en12040660)
- [5] J. Meng, G. Luo, M. Ricco, M. Swierczynski, D.I. Stroe, R. Teodorescu, Overview of lithium-ion battery modeling methods for state-of-charge estimation in electrical vehicles, *Applied Sciences* 8(5) (2018), art. 659. DOI: [10.3390/app8050659](https://doi.org/10.3390/app8050659)
- [6] C. Lin, A. Tang, W. Wang, A review of SOH estimation methods in Lithium-ion batteries for electric vehicle applications, *Energy Procedia* 75 (2015), pp. 1920-1925. DOI: [10.1016/j.egypro.2015.07.199](https://doi.org/10.1016/j.egypro.2015.07.199)
- [7] C. Weng, Y. Cui, J. Sun, H. Peng, On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression, *Journal of Power Sources*, 235 (2013), pp. 36-44. DOI: [10.1016/j.jpowsour.2013.02.012](https://doi.org/10.1016/j.jpowsour.2013.02.012)
- [8] R. R. Richardson, C. R. Birkl, M. A. Osborne, D. A. Howey, Gaussian process regression for in situ capacity estimation of lithium-ion batteries, *IEEE Transactions on Industrial Informatics* 15(1) (2019), pp. 127-138. DOI: [10.1109/TII.2018.2794997](https://doi.org/10.1109/TII.2018.2794997)
- [9] X. Xu, N. Chen, A state-space-based prognostics model for lithium-ion battery degradation, *Reliability Engineering and System Safety* 159 (2017), pp. 47-57. DOI: [10.1016/j.res.2016.10.026](https://doi.org/10.1016/j.res.2016.10.026)
- [10] M. A. Patil, P. Tagade, K. S. Hariharan, S. M. Kolake, T. Song, T. Yeo, S. Doo, A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation, *Applied Energy* 159 (2015), pp. 285-297. DOI: [10.1016/j.apenergy.2015.08.119](https://doi.org/10.1016/j.apenergy.2015.08.119)

- [11] C. Lu, L. Tao, H. Fan, Li-ion battery capacity estimation: A geometrical approach, *Journal of Power Sources* 261 (2014), pp. 141-147.  
DOI: [10.1016/j.jpowsour.2014.03.058](https://doi.org/10.1016/j.jpowsour.2014.03.058)
- [12] D. I. Stroe, M. Swierczynski, A. I. Stan, V. Knap, R. Teodorescu, S. J. Andreasen, Diagnosis of lithium-ion batteries state-of-health based on electrochemical impedance spectroscopy technique, 2014 IEEE Energy Conversion Congress and Exposition (ECCE), Pittsburgh, PA, 14-18 September 2014, pp. 4576-4582.  
DOI: [10.1109/ECCE.2014.6954027](https://doi.org/10.1109/ECCE.2014.6954027)
- [13] D. Andre, M. Meiler, K. Steiner, C. Wimmer, T. Soczka-Guth, D. U. Sauer, Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation, *Journal of Power Sources* 196(12) (2011), pp. 5334-5341.  
DOI: [10.1016/j.jpowsour.2010.12.102](https://doi.org/10.1016/j.jpowsour.2010.12.102)
- [14] F. Huet, A review of impedance measurements for determination of the state-of-charge or state-of-health of secondary batteries, *Journal of Power Sources* 70(1) (1998), pp. 59-69.  
DOI: [10.1016/S0378-7753\(97\)02665-7](https://doi.org/10.1016/S0378-7753(97)02665-7)
- [15] I. Masmitjà Rusinyol, J. González, G. Masmitjà, S. Gomàriz, J. del Río-Fernández, Power system of the Guanay II AUV, *Acta IMEKO* 4(1) (2015), pp. 35-43.  
DOI: [10.21014/acta\\_imeko.v4i1.161](https://doi.org/10.21014/acta_imeko.v4i1.161)
- [16] S. Buteau, J. R. Dahn, Analysis of thousands of Electrochemical impedance spectra of lithium-ion cells through a machine learning inverse model, *Journal of the Electrochemical Society* 166(8) (2019), art. A1611.  
DOI: [10.1149/2.1051908jes](https://doi.org/10.1149/2.1051908jes)
- [17] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, A. A. Lee, Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning, *Nature Communications* 11 (2020), art. 1706.  
DOI: [10.1038/s41467-020-15235-7](https://doi.org/10.1038/s41467-020-15235-7)
- [18] F. Liu, X. Liu, W. Su, H. Lin, H. Chen, M. He, An online state of health estimation method based on battery management system monitoring data, *International Journal of Energy Research* 44(8) (2020), pp. 6338-6349.  
DOI: [10.1002/er.5351](https://doi.org/10.1002/er.5351)
- [19] P. Singha, R. Vinjamuria, X. Wang, D. Reisner, Design and implementation of a fuzzy logic-based state-of-charge meter for Li-ion batteries used in portable defibrillators, *Journal of Power Sources* 162(2) (2006), pp. 829-836.  
DOI: [10.1016/j.jpowsour.2005.04.039](https://doi.org/10.1016/j.jpowsour.2005.04.039)
- [20] S. B. Sarmah, P. Kalita, A. Garg, X.-d. Niu, X.-W. Zhang, X. Peng, D. Bhattacharjee, A review of state of health estimation of energy storage systems: Challenges and possible solutions for futuristic applications of Li-Ion battery packs in electric vehicles, *Journal of Electrochemical Energy Conversion and Storage* 16(4) (2019), art. 040801.  
DOI: [10.1115/1.4042987](https://doi.org/10.1115/1.4042987)
- [21] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods, *Journal of Power Sources* 239 (2013), pp. 680-688.  
DOI: [10.1016/j.jpowsour.2012.11.146](https://doi.org/10.1016/j.jpowsour.2012.11.146)
- [22] Z. Chen, M. Sun, X. Shu, R. Xiao, J. Shen, Online state of health estimation for lithium-ion batteries based on support vector machine, *Applied Sciences* 8(6) (2018), art. 925.  
DOI: [10.3390/app8060925](https://doi.org/10.3390/app8060925)
- [23] V. Klass, M. Behm, G. Lindbergh, A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation, *Journal of Power Sources*, vol. 270 (2015), pp. 262-272.  
DOI: [10.1016/j.jpowsour.2014.07.116](https://doi.org/10.1016/j.jpowsour.2014.07.116)
- [24] J. Meng, L. Cai, G. Luo, D.-I. Stroe, R. Teodorescu, Lithium-ion battery state of health estimation with short-term current pulse test and support vector machine, *Microelectronics Reliability* 88-90 (2018), pp. 1216-1220.  
DOI: [10.1016/j.microrel.2018.07.025](https://doi.org/10.1016/j.microrel.2018.07.025)
- [25] X. Feng, C. Weng, X. He, X. Han, L. Lu, D. Ren, M. Ouyang, Online state-of-health estimation for Li-Ion battery using partial charging segment based on support vector machine, *IEEE Transactions on Vehicular Technology* 68(9) (2019), pp. 8583-8592.  
DOI: [10.1109/TVT.2019.2927120](https://doi.org/10.1109/TVT.2019.2927120)
- [26] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, P. Van den Bossche, Critical review of state of health estimation methods of Li-ion batteries for real applications, *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 572-587.  
DOI: [10.1016/j.rser.2015.11.042](https://doi.org/10.1016/j.rser.2015.11.042)
- [27] J. Qu, F. Liu, Y. Ma, J. Fan, A neural-network-based method for RUL prediction and SOH monitoring of lithium-ion battery, *IEEE Access* 7 (2019), pp. 87178-87191.  
DOI: [10.1109/ACCESS.2019.2925468](https://doi.org/10.1109/ACCESS.2019.2925468)
- [28] G. Ma, Y. Zhang, C. Cheng, B. Zhou, P. Hu, Y. Yuan, Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network, *Applied Energy*, 253 (2019), art. 113626.  
DOI: [10.1016/j.apenergy.2019.113626](https://doi.org/10.1016/j.apenergy.2019.113626)
- [29] R. La Rosa, A. Y. S. Pandiyan, C. Trigona, B. Andò, S. Baglio, An integrated circuit to null standby using energy provided by MEMS sensors, *Acta IMEKO* 9(4) (2020), p. 144 -150.  
DOI: [10.21014/acta\\_imeko.v9i4.741](https://doi.org/10.21014/acta_imeko.v9i4.741)
- [30] G. Campobello, D. Dell'Aquila, M. Russo, A. Segreto, Neuro-genetic programming for multigenre classification of music content, *Applied Soft Computing* 94 (2020), art. 106488.  
DOI: [10.1016/j.asoc.2020.106488](https://doi.org/10.1016/j.asoc.2020.106488)
- [31] International Standard IEC 61427-2: Secondary cells and batteries for renewable energy storage- General requirements and methods of test - Part 2: on-grid applications, ed, 2015.
- [32] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, W. Shang, Temperature effect and thermal impact in lithium-ion batteries: A review, *Progress in Natural Science: Materials International* 28(6) (2018), pp. 653-666.  
DOI: [10.1016/j.pnsc.2018.11.002](https://doi.org/10.1016/j.pnsc.2018.11.002)
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, *The Journal of machine Learning research* 12 (2011), pp. 2825-2830. Online [Accessed 09 June 2021]  
<http://jmlr.org/papers/v12/pedregosa11a.html>
- [34] V. J. Ovejas, Impedance Characterization of an LCO-NMC/Graphite Cell: Ohmic Conduction, SEI Transport and Charge-Transfer Phenomenon, *Batteries* 4(3) (2018), art. 43.  
DOI: [10.3390/batteries4030043](https://doi.org/10.3390/batteries4030043)
- [35] T. F. Landinger, G. Schwarzberger, A. Jossen, A novel method for high frequency battery impedance measurements, *IEEE International Symposium on Electromagnetic Compatibility, Signal & Power Integrity (EMC+SIPI)*, New Orleans, LA, USA, 22-26 July 2019, pp. 106-110.  
DOI: [10.1109/ISEMC.2019.8825315](https://doi.org/10.1109/ISEMC.2019.8825315)
- [36] M. L. Zhang, Z. H. Zhou, A review on multi-label learning algorithms, *IEEE Transactions on Knowledge and Data Engineering* 26(8) (2014), pp. 1819-1837.  
DOI: [10.1109/TKDE.2013.39](https://doi.org/10.1109/TKDE.2013.39)
- [37] D. Aloisio, G. Campobello, S. G., Leonardi, A. Segreto, N. Donato, A machine learning approach for evaluation of battery state of health, 24th IMEKO TC4 International Symposium and 22nd International Workshop on ADC and DAC Modelling and Testing, Palermo, Italy, 14-16 September 2020, pp. 129-134. Online [Accessed 09 June 2021]  
<https://www.imeko.org/publications/tc4-2020/IMEKO-TC4-2020-25.pdf>