

A machine learning approach for evaluation of battery state of health

Davide Aloisio^{1*}, Giuseppe Campobello², Salvatore Gianluca Leonardi¹, Francesco Sergi¹, Giovanni Brunaccini¹, Marco Ferraro¹, Vincenzo Antonucci¹, Antonino Segreto², Nicola Donato^{2*}

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

² *University of Messina, Department of Engineering, C.da di Dio, Vill. S.Agata, 98166 Messina, Italy*

* *E-mail: aloisio@itae.cnr.it; ndonato@unime.it*

Abstract – Ageing estimation of lithium ion (Li-Ion) batteries is a key point for their massive application in the market. In this work, different Machine Learning (ML) techniques were applied and compared to evaluate the State of Health (SoH) of a cobalt based Li-Ion battery, cycled under a stationary application profile. Experimental results show that ML can be profitably used for SoH estimation.

I. INTRODUCTION

As well known, Machine Learning (ML) is a subfield of computing, a type of Artificial Intelligence (AI), which provides machines with the ability to learn from the field without explicit programming [1]. In particular, it is well known that ML can be really useful in applications in which a lot of data are obtained from the field and you want to extract some information or unknown properties (‘features’) from the dataset, the so-called ‘training set’. The information extracted from this kind of data analysis can be used for prediction of the systems behavior when subject to certain operative conditions and under some constraints. In particular, the battery behavior is very complex to understand because many parameters act in determining the ageing evolution such as 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 above parameters makes the systems hard to model via analytical equations. This is particularly true in the case of Li-Ion batteries, where it is more difficult to translate electrochemical processes into analytical equations, due to the intrinsic and strong nonlinearities present in this type of batteries. The analytical models, in fact, 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 can provide a viable alternative and a useful tool for modelling the battery behavior. ML algorithms learn directly from experimental data, reducing the complexity of modelling, usually due to the high number of parameters and empirical corrections 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.

In this paper we exploit ML to identify possible methodologies for estimating the State of Health (SoH) of Li-Ion battery with the main aim of developing a feasible model which can be easily integrated in a battery management system (BMS). Implementation in BMSs of techniques able to lengthen batteries useful life, estimating the possible replacement time (estimation of Remaining Useful Life, or RUL), is considered a key research activity in the field [5].

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

In this section a few state-of-the-art of ML techniques applied to SoH and RUL estimation are reviewed.

Thanks to the remarkable computational capabilities of today systems, learning algorithms applied to large quantities of data have often become the preferred approach in the search and identification of complex system behaviors, 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, extracts non-linear relationships among the various parameters. The knowledge derived from this kind of information can allow a careful management of the battery, thus helping in prolonging 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 estimation is performed by determining battery capacity and internal resistance, parameters strictly related to SoH, from input variables behavior analysis (current, temperature, voltage, etc.). An application of Fuzzy logic with a potential use in portable device is shown in [6]. Here Electrochemical Impedance Spectroscopy (EIS) technique was used for the dataset creation. However, improper hypotheses in the Fuzzy rules [3] and reduced set of observation can lead to substantial errors on the result. The Support Vector Machine is a regression algorithm which converts nonlinearities in a lower dimension space to a linear model developed in a higher dimension one [7]. Examples of application of this technique applied to SoH are reported in several works [8-12]. In particular, in [12] 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 and other data manipulation techniques (particle filter, Bayesian technique) have to be used in conjunction with SVM to increase the robustness of the estimation [13], thus increasing the complexity of implementation. Relevance Vector Machine or RVM is suggested as a possible improvement of this approach in [7]. Artificial Neural Networks (ANN) are probably the most used approach for modeling of nonlinear systems which is inspired by the biological functioning of the human brain. SoH estimation using an independently recurrent neural network (in RNN) was realized in [3]. Here SoH is predicted accurately with root means square error (RMSE) of 1.33% and mean absolute error (MAE) of 1.14%. The main limitation underlined here is the need of a detailed analysis on experimental dataset. Different chemistries can require a precise identification and understanding of input parameters. In [14] an improved neural network method based on the combination of a LSTM (long-short-term memory) and (PSO) Particle Swarm Optimization was developed. The methodology proposed here uses some additional techniques in each part of the learning process, such as PSO for optimization 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 [14]. Another hybrid approach can be found in [15] where false nearest neighbor method is used in conjunction with a mixed LSTM and convolutional neural network (CNN) as 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 over different other common techniques showing its potential and advantages of data driven approaches is presented in [4]. The outcomes show the goodness of deep neural networks (DNN) which are suitable when high accuracy is needed. However, also this technique is not very easily implementable due to higher computational complexity and resources needed [4].

A lot of other techniques and approaches can be found in literature on this topic. Although they are not described here for sake of space, 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. 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 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 [1,16].

III. DATASET COLLECTION AND CREATION

The work was aimed for 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 [17].

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 ₂ -LiNiCoMnO ₂ /Graphite

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 were accelerated by operating at an ambient temperature of 45 °C. In fact, it is well known that the degradation processes of Li-Ion batteries are speeded up by temperature increase [18].

The ageing tests were performed by a dual-channel Bitrode FTV1 battery cycler. 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.

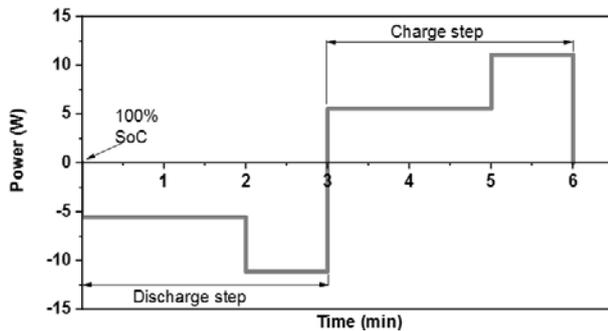


Fig. 1. FR ageing profile.

The full ageing protocol consisted of a first charge of the cell up to 100 % SoC and then an execution of the FR profile. Whenever 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 has been 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 by a high reliability Autolab 302N potentiostat/galvanostat, which exhibits applied potential accuracy of $\pm 0.2\%$ and current accuracy of $\pm 0.2\%$ of the used range. It is worth nothing that, due to instruments calibration and performances, measures were considered reliable and no impact of uncertainty on the model was evaluated. Only 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 state of health of the cell (SoH). The recorded discharge curves at 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. 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 limits.

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

E_d was obtained by integrating the actual power (i_d) between begin of discharge (t_0) and end of discharge (t_f), within the upper and lower voltage cut off limits.

$$E_d = \int_{t_0}^{t_f} P_d(t) dt \quad (2)$$

Where " $P_d(t) = V(t) \cdot I_d(t)$ ", with $V(t)$ and $I_d(t)$ representing the instantaneous values of voltage and current, respectively.

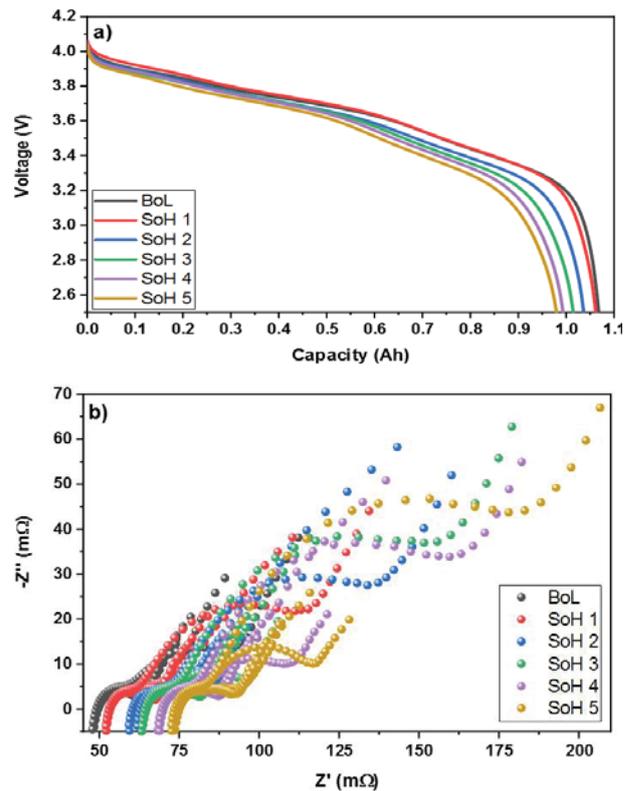


Fig. 2. a) Discharge curves for extrapolation of output variables; b) Nyquist plot of the impedance used as input variables of the database.

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. This information came from EIS analysis carried out in correspondence of parametric check-ups. EIS output is often shown through the Nyquist diagram in which, inverse imaginary part of impedance is plotted as a function of the real one. In the case of Li-Ion batteries, the above diagram consists of four distinct regions typically belonging to the frequency range between 10 mHz to 10 kHz [19]. 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 are usually present, representative of the impedance of either charge transfer phenomena or passivation layers on the surface of the electrodes (solid electrolyte interphase-SEI). Another interesting point is represented by the intersection of the impedance spectrum with the real axis (pure ohmic impedance), which is related to the internal resistance of the cell. Finally, the high frequency region is representative of inductive phenomena. Since each one of these phenomena are strictly related to temperature, SoC and SoH of the cell, the analysis of the impedance data can be used to monitor instantaneous operating conditions and the degradation level induced by use.

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 of about 8% was reached. 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 range of frequency 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 1860 impedance measurements. Table 2 contains some statistical information on the dataset used.

Table 2. Statistical data of used dataset.

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

IV. METHODOLOGY

The data were partitioned completely randomly into two sets containing respectively 80% and 20% of the impedance measurements. This partitioning was preliminary used for a test with the aim of carrying out a first selection of the machine learning techniques. In particular, through the Scikit-learn tool [20], an open-source library for machine learning developed in Python, various classification and regression techniques were tested. 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.

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) and 5 testing rounds were carried out, using in each round a

different fold for tests and the remaining 4 folds for training. In the case of regressors, the value of the absolute mean error (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).

V. RESULTS

Table 3 shows the average values and the standard deviation of the accuracy obtained for some of the classifiers used for the prediction of the SoC, i.e. K-nearest Neighbors (KNN), Linear Discriminant Analyzes (LDA), Gaussian Naive Bayes (GNB), Support Vector Classification (SVC), Decision Tree (DT).

Table 3. Accuracy of some classifiers used for modeling the SoC starting from impedance values in both rectangular and polar form.

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

Each classifier was trained both with the original data, constituted by impedance expressed in rectangular coordinates (real part and imaginary part), and with pre-processed data, in order to use impedance values expressed in polar coordinates (module and phase).

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 has better performances, obtaining an average accuracy equal to 0.915 with the polar representation. It is worth noting that the default values preset by Scikit-learn have been used for all classifiers and no optimization has been made.

In addition, several regressors were tested for modeling the SoC, i.e. Linear Regressor, Lasso, Ridge and Random Forest. By the analysis carried out, the regressor with the best performance in terms of R^2 and MAE was the Random Forest. In particular, average value of R^2 and standard deviation are of 0.97 and 0.01, respectively (see Fig. 3a).

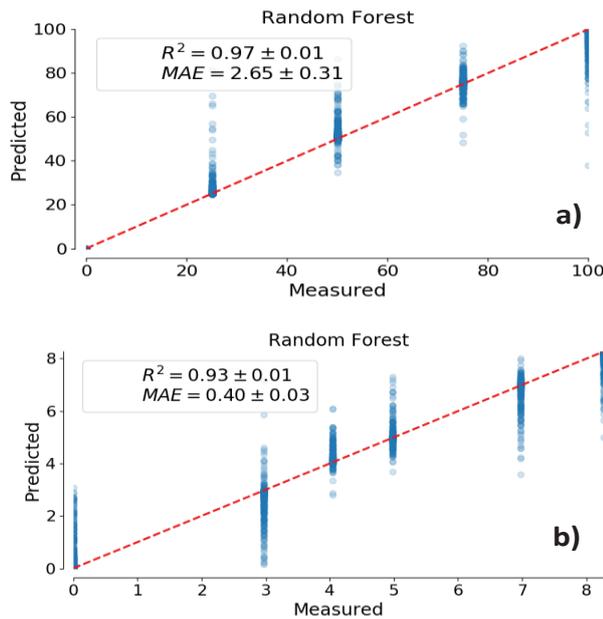


Fig. 3. Random forest distribution for a) SoC b) Capacity losses.

The same regressors have been also used for the modeling of capacity loss, which is an indication of SoH. Similarly, in the case of SoC, the regressor with the best performance in terms of R^2 and MAE was the Random Forest. In particular, average value of R^2 and standard deviation are 0.93 and 0.01, respectively (see Fig 3b). The distributions of the predicted values by Random forest regressor in relation to the measured data are reported in figures 3a and 3b for the modeling of SoC and Capacity losses, respectively.

VI. CONCLUSIONS

Different machine learning techniques have been analyzed 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, starting from impedance measurements. According to the results, the following conclusions were drawn:

- for the training of machine learning techniques, the use of impedance values expressed in module and phase is preferable;
- both for classification and regression models, techniques based on decision trees provided superior performance compared to the other machine learning techniques analyzed;
- Random forest regressor showed the best performance in terms of R^2 and MAE for both state of charge and capacity loss prediction.

ACKNOWLEDGMENTS

This research was funded by the Italian Ministry of Economic Development under the programme “Ricerca di Sistema”, project electrochemical storage.

REFERENCES

- [1] G. Hackeling, *Mastering Machine Learning With scikit-learn*: Packt Publishing, 2014.
- [2] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, and L. Zhang, "Remaining useful life prediction for lithium-ion battery: A deep learning approach," *IEEE Access*, vol. 6, pp. 50587-50598, 2018.
- [3] P. Venugopal, "State-of-Health Estimation of Li-ion Batteries in Electric Vehicle Using IndRNN under Variable Load Condition," *Energies*, vol. 12, p. 4338, 2019.
- [4] P. Khumprom and N. Yodo, "A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm," *Energies*, vol. 12, p. 660, 2019.
- [5] F. Liu, X. Liu, W. Su, H. Lin, H. Chen, and M. He, "An online state of health estimation method based on battery management system monitoring data," *International Journal of Energy Research*, 2020, in press.
- [6] P. Singha, R. Vinjamuria, X. Wangb, 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*, vol. 162, pp. 829-836, 2006.
- [7] S. B. Sarmah, P. Kalita, A. Garg, X.-d. Niu, X.-W. Zhang, X. Peng, and 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*, vol. 16, 2019.
- [8] 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*, vol. 239, pp. 680-688, 2013.
- [9] Z. Chen, M. Sun, X. Shu, R. Xiao, and J. Shen, "Online State of Health Estimation for Lithium-Ion Batteries Based on Support Vector Machine," *Applied Sciences*, vol. 8, p. 925, 2018.
- [10] V. Klass, M. Behm, and 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, pp. 262-272, 2014.
- [11] J. Meng, L. Cai, G. Luo, D.-I. Stroe, and R. Teodorescu, "Lithium-ion battery state of health estimation with short-term current pulse test and support vector machine," *Microelectronics Reliability*, vol. 88-90, pp. 1216-1220, 2018.

- [12] X. Feng, C. Weng, X. He, X. Han, L. Lu, D. Ren, and 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*, vol. 68, pp. 8583-8592, 2019.
- [13] M. Bercibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 572-587, 2016.
- [14] J. Qu, F. Liu, Y. Ma, and J. Fan, "A Neural-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery," *IEEE Access*, vol. 7, pp. 87178-87191, 2019.
- [15] G. Ma, Y. Zhang, C. Cheng, B. Zhou, P. Hu, and Y. Yuan, "Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network," *Applied Energy*, vol. 253, p. 113626, 2019.
- [16] G. Campobello, D. Dell'Aquila, M. Russo, A. Segreto, "Neuro-genetic programming for multigenre classification of music content", *Applied Soft Computing*, vol. 94, 2020.
- [17] "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.
- [18] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, and W. Shang, "Temperature effect and thermal impact in lithium-ion batteries: A review," *Progress in Natural Science: Materials International*, vol. 28, pp. 653-666, 2018.
- [19] D. Andre, M. Meiler, K. Steiner, C. Wimmer, T. Soczka-Guth, and D. U. Sauer, "Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation," *Journal of Power Sources*, vol. 196, pp. 5334-5341, 2011.
- [20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.