Accelerated State of Health Estimation of Second life Lithium-ion Batteries via Electrochemical Impedance Spectroscopy Tests and Machine Learning Techniques

Mona Faraji-Niri [a, *], Muhammad Rashid [a], Jonathan Sansom [a], Muhammad Sheikh [a,b], Dhammika Widanage [a], James Marco [a]

[a] WMG, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK
[b] School of Digital, Technologies and Arts, Staffordshire University, College Road, University Quarter, Stoke-on-Trent, ST4 2DE, UK
*mona.faraji-niri@warwick.ac.uk

Abstract

Estimating the State of health (SoH) of Lithium-ion (Li-ion) batteries is a challenging task due to cross coupling and dependency between ageing mechanisms. An accurate estimation is particularly essential for second-life batteries to facilitate their successful implementation in the new application. By adopting the Electrochemical Impedance Spectroscopy (EIS) test and a machine learning (ML) approach, the accelerated SoH estimation problem is studied here. For this purpose, 325 experiments for 30 Li-ion cells were conducted at various SoH, temperature, and state of charge. First an optimised Gaussian Process Regression model is developed and validated for SoH estimation. Then the sensitivity of the model is evaluated relative to measurement noise. Finally, the model's robustness is quantified through a case study involving cells that have been characterised with different physical test equipment. The results demonstrate that the model can predict the SoH of Li-ion cells with an error about 1.1% and is reasonably robust to the various testing conditions of the battery. The methodology for handling the EIS data within a machine learning framework, the sensitivity analysis and the robustness quantification techniques are the main novelties of this study in the context of grading Li-ion batteries for second-life applications

Keywords: Electrochemical Impedance spectroscopy, Second life Lithium-ion batteries, Machine Learning, State of health, Prediction

I. Introduction

With the drastic increase in the number of hybrid and electric vehicles (EVs) in recent years, the global demand for lithium-ion batteries as their main power source has gone up significantly. This demand is expected to grow by 26% from 2021 to 2031, with a market value of about $70 billion by 2026 [1]. This record-breaking demand and the high production cost of lithium-ion batteries have attracted attentions toward the reuse, recycle and disposal management of those soon after they are technically at their end of life (EoL) in the EVs [2].

Reusing the batteries has significant economic benefits and reduced environmental impacts. In the case of lithium-ion batteries for EVs, the remaining energy capacity of cells has a direct impact on the vehicles remaining range and its safety. Generally, when the cell's state of health (SoH), which is the ratio of the present capacity to the initial capacity [3, 4], falls below 80%, it is considered retired due to reduced ability of supplying the energy and power for the EVs [5]. The retired batteries while not of any use for EVs can still be repurposed and utilised in many other applications, in fact they play an important role in energy storage systems that require relatively lower energy or power. Examples of such applications include load-shifting and frequency regulation in grid-scale [6, 7]. A comprehensive review of the applications of second-life batteries can be found in [5].

An accurate estimation of SoH is necessary for its safe and reliable implementation in the next life application. There exist several techniques to test retired batteries. In [8], constant current pulses are utilised to evaluate the cycle life of cells. Electrochemical models are developed by [9, 10] to capture the ageing mechanism of batteries by quantifying the change in the volume of the cyclable lithium-ions as well as the solid electrolyte interphase film resistance [11]. These electrochemical models are hard to be derived for second-life cells as their parametrisation is much more challenging, due to complex and nonlinear grading mechanisms, compared to the new cells. In search for methods with affordable cost and complexity for SoH estimation, a few studies have focused on the identification of certain parameters of cells as health indicators. Admitting that the cell energy capacity is a measure of its health, techniques based on incremental capacity and differential voltage are very common in this category [12, 13]. Beside energy capacity, the battery impedance can also serve as an indicator for its SoH, several researchers have focused on impedance measurement via an electrochemical impedance spectroscopy (EIS) test [14, 15]. EIS test determines the battery impedance at different frequencies and is considered as one of the off-line, non-intrusive and in-situ techniques for SoH characterisation [15]. It is favourable as only requires voltage and current measurements in the frequency domain.

EIS measurements have been used for SoH estimation of cells in a few studies of different applications. In [16] 3-electrode pouch cells were tested via EIS and modelled via equivalent circuits with a SoH estimation accuracy of 7%. An estimation of the Li-ion battery cell SoH during its operation is proposed in [17] by designing an EIS test for measuring the cell impedances at different SoC, and temperatures. Quasi-EIS tests are conducted in [18], and an empirical method with an error up to 1% is utilised in [19] to estimate the battery SoH considering its relaxation time. An identification of different ageing mechanisms addressing active material loss and loss of lithium-ions is performed in [20]. An online or so called in-operando version of the EIS test for the SoH estimation of cells is proposed in [21].

Most of the abovementioned studies have focused on just one or only on a handful number of cells, which makes generalisation of the algorithms to real-world problems quite challenging. Considering the high volume of data points and the information generated via energy capacity or impedance determination tests, data driven approaches such as machine learning (ML) are very suitable to perform SoH estimation on a large stack of cells. SoH estimation based on ML models has been recently addressed in a few studies. For example, an ensemble model based on usual capacity test data is developed in [22] for predicting the remaining useful life of the cells. The model is an infusion of an empirical and a polynomial regression model tested on six individual cells at room temperature. Extreme learning models are developed in [23, 24] for degradation estimation of cells tested under specific loading profiles and real-world cycling conditions [25]. Deep learning models based on capacity tests are proposed in [26] for SoH prognostics of connected cells in a series configuration. A combination of linear and Gaussian support vector machines is developed in [27] to improve the cycle life prediction. Long short-term memory models based on the capacity tests are developed in [28, 29] and tested on experimental and open-source data set with a SoH error estimation error below 1.46%.

Focusing on EIS test, while they can be very useful for SoH characterisation of cells in larger volumes, they have not yet been fully investigated with this aim. Example researches include EIS measurements combined with fuzzy logic algorithms for SoH characterisation of cells as addressed in [30, 31] and regression models based on twelve Li-ion cells EIS spectrum developed in [32]. A comprehensive study based on the EIS field data for the online estimation of SoH is performed in [33] where the degradation is monitored at the electrode level. A neural network based on EIS tests and equivalent circuit analysis is developed in [34] with the predictions of equivalent series resistance, charge-transfer resistance, and solid-electrolyte interphase resistance as representation of cells state of health by errors below

5%, 1.5%, and 1%, respectively. The zero-crossing frequency of the li-ion cells is considered as a feature to train a neural network for SoH estimation in [35]. A study based on EIS measurements at constant temperatures is performed via Gaussian process regression models for the estimation of remaining useful life of commercial cells for up to 300 cycles in [36].

While there are relatively large number of methods developed for SoC estimation via EIS tests and machine learning approaches [37, 38], not enough attention has been given to the SoH estimation by EIS experiments and there is an obvious research gap specially for second life battery applications. One of the main differences between the new and re-purposed cells is the inconsistency in between their characteristics [2, 39]. This inconsistency and difference between the cells usually lead to different ageing patterns and SoH for the cells. Therefore, tailoring the SoH estimation algorithms and methods for larger number of cells at their second life is still required. To address this issue and take the advantage of EIS measurements, this study focuses on the SoH estimation of NMC cells via EIS tests and a machine learning modelling approach. This research answers the following questions:

1. If the EIS test data can be directly used for the SoH estimation of cells or feature engineering is necessary?
2. What is the configuration of a ML model for SoH estimation?
3. How robust the method is to the measurement noise?
4. How generalisable are the ML-based models to real-world uncertainties arising from variations in test equipment?
5. What is the minimum amount of conditioning information required for an accurate SoH estimation?

To answer the above questions, first a series of EIS characterisation tests are conducted to provide training and validation data for the study. To factor in the impact of the cell-to-cell variation and facilitating an analysis based on the similarities and differences between the cells during the SoH estimation, each test is conducted on 4 batteries of the same type and from the same manufacturer. To uncover the impact of the testing temperature and the state of charge on the cells SoH predictability, the tests are run at three different temperatures of 15, 25, and 35 ^{o}C as well as 5 different SoC breakpoints between 5% to 95% (5, 20, 50, 70, 95). The correlation in between cell impedance, SoC, temperature and its SoH is built via a machine learning model. In this context, a Gaussian Process Regression model is developed and optimised via Bayesian methods. The model is created for two scenarios to identify the minimum amount of information for an accurate estimation and the necessity of feature engineering. In the first scenario, the full EIS data and in the second one, the selected features are utilised for building the ML model. For the first time in this study, a methodology is proposed to evaluate the model's sensitivity to the measurement noise by generating synthetic noisy inputs for the model. Additionally, the model's robustness is evaluated under a testing condition with different experimental equipment and cooling mechanism as well as the different set of model inputs for studying its generalisability. A brief diagram of this paper's proposal is as Figure 1.

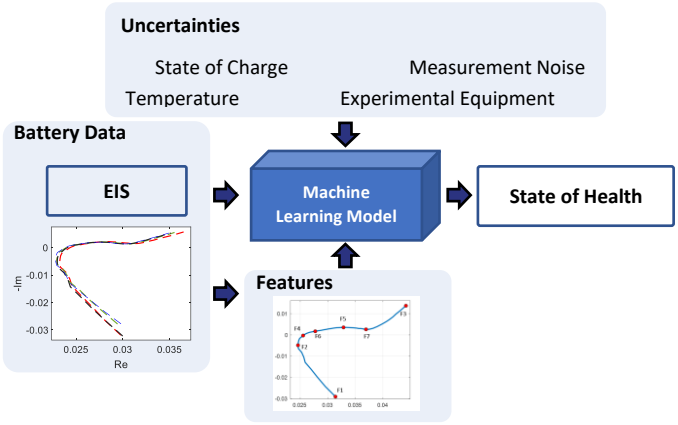

Figure 1: The diagram of the machine learning model for the prediction of the cell SoH via EIS test

The remainder of this paper is as follows, Section II, give the experimental details, the test protocols, and the plans. Section III articulates the machine learning modelling approach. In this section, the list of inputs and outputs, the definition of features and their extraction process is also given. Section VI includes the main results related to the prediction of the SoH. Section V includes discussions regarding the sensitivity and robustness of models. The final section summarises the methodology, results and findings and concludes the paper.

II. Experiments

To conduct this study, 30 Lithium-ion commercial LGM50 cells with NMC chemistry have been utilised. The rated capacity of cells was 5 Ah defined by the manufacturer. The initial capacity of all the cells has been measured as the reference for their state of health at C/3 current charge-discharge between 4.2 V to 2.5 V with a C/20 cutoff current at constant voltage mode during charge at 25 °C. Based on the reference capacity, the cell's state of health has been defined via Eq (1).

$$SoH(\%) = \frac{Current\,Capacaity(Ah)}{Reference\,Capacaity(Ah)} \times 100 \quad (1)$$

From the initial 30 cells, subsets of cells have been aged to produce 4 additional case breaks of 80, 85, 90 and 95% SoH where it has been assumed that the reference case is related to the 100% SoH. For each SoH subset in addition to the 1st cell tested 4 further cells have been aged similarly and stored under similar conditions, where their performance has been examined subsequently. This was required to evaluate cell-to-cell variation during analysis. The cycling test for aging was performed using a 10 A Digatron battery cycler with 192 channels where the cells were maintained at 25 °C within a thermal chamber (Binder™). During the cycling process the cells have been charged at C/2 constant-current mode until 4.2 V followed by a constant voltage charge mode until the current drops to C/20 and finished with a discharge at 1C constant current until the cell voltage reaches to 2.5 V. the charge and discharge process has been repeated till the cell reaches the required SoH. For different SoHs, the number of cycles has been different and varying between 60 cycles to reach a 95% SoH and 250 cycles to reach the 80% SoH.

The current capacity of the cells has been quantified for Eq. (1) via a C/3 charge discharge test profile, similar to the reference capacity test case for consistency. After the ageing process, the cells have been subject to a performance test. This comprises the impedance measurement via EIS test at different SoCs (5, 20, 50, 70 and 95%) and temperatures (15, 25 and 35 °C). Before starting the tests, all the cells have been immersed in an oil bath (Kryo-51 dielectric oil) for about 4 hours to guarantee

a thermal homogeneity within the cells and across their surface, a Lauda cooling system has been used to keep the cell's temperature at the defined condition. Figure 2 (a) and (b) show the test and the experimental setup configuration.

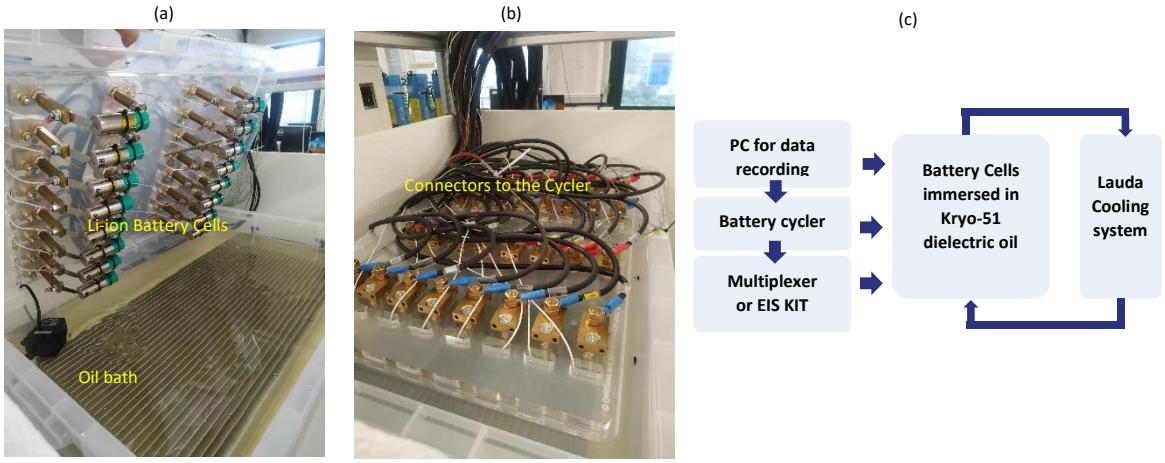

Figure 2: (a) Battery test setup configuration, (b) Test rig after immersion in oil, (c) Block diagram of the test setup

In performance tests, the cells have been discharged to the desired SoC (using a Coulomb counting method) at C/3 current and relaxed for 4 hours to minimize the potential/concentration gradients across the electrode/cell domain caused by the applied current [40, 41]. Once the cell voltage has reached equilibrium, the EIS measurement is conducted via the Multiplexer equipment in Figure 2 (c) configuration, between frequencies of 10kHz to 10mHz with a sinusoidal current superimposed on a DC current at 250 mA. An example EIS Nyquist spectrum is given in Figure 3 (a), where the horizontal axis refers to the real part of the impedance and the vertical one is for the imaginary part. It is noteworthy that the frequency values defined in Figure 3 (a) are approximate and the values for a given cell are known to vary depending on the cell manufacturers, electrochemistry, and form-factor.

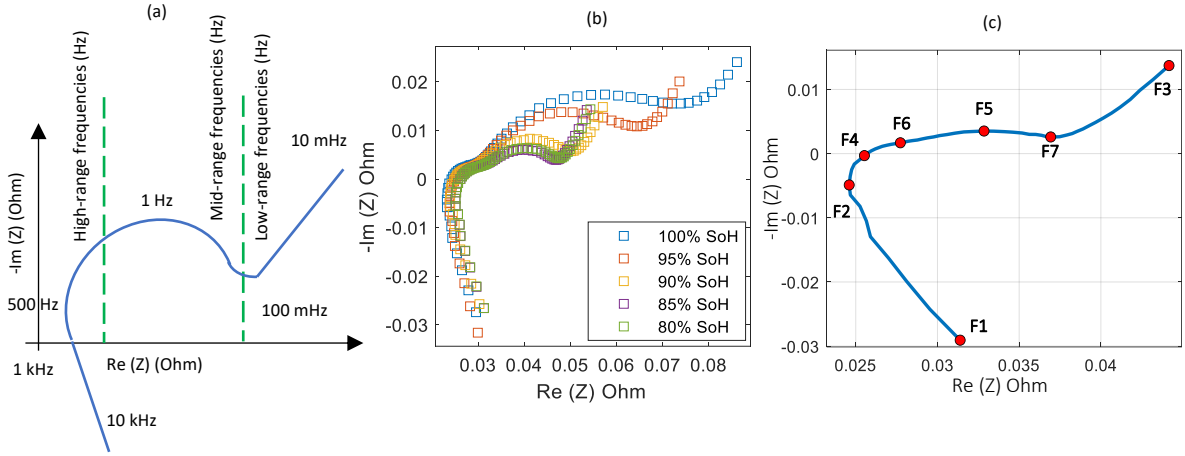

Figure 3: (a) An example EIS spectrum of the Lithium-ion battery, (b) visualisation of some EIS graphs at various SoH, (c) the feature definition for EIS graphs

For each cell, the impedance (Ohm), $Z(f)$ is quantified via Eq. (2), where I and V refer to the sinusoidal current and voltage at a particular frequency f (Hz). Here, ϕ refers to the phase angle of the voltage.

$$dI = I_{\max}\sin(2\pi ft)$$
$$dV = V_{\max}\sin(2\pi ft + \phi) \quad (2)$$
$$Z(f) = \frac{V_{\max}}{I_{\max}} e^{j\phi}$$

The EIS curve is a good representative of the cell's characteristics, the low frequency section (10 - 100 mHz) is associated with the diffusion processes that take place inside the cell, while the medium frequency range (circa 1 Hz) is indicating its double-layer capacitance effect. In the high frequency range (greater than 500 Hz), the interception between the curve and the real axis indicates the cell's Ohmic resistance [37].

Following these experiments, a combination of the 30 cells at 5 different SoH, 5 different SoC and 3 temperatures have led to 300 cases for the modelling and analysis purposes. Example records of data are visualised on Figure 3 (b) at various SoH, 15 °C and 5% SoC.

III. Machine Learning Model

To uncover the relation between the EIS measurements of the cells and the associated SoH, the GPR model is created. For this purpose, first the fundamentals of the GPR are given and then the accuracy metrics are provided to evaluate the performance of the model for completeness.

GPR is a Bayesian method based on Gaussian processes for non-parametric regression. GPR is deemed to be suitable for dealing with small data set and is capable of providing uncertainty measurements on the predictions [42]. As a Bayesian method, it helps inferring the probability distribution of the parameters in a function that is meant to be learnt through the training process. For a given input indicated by x, the GPR creates a probability distribution function of $f(x)$ with the mean and covariance function of $m(x)$ and $C(x,x')$ as Eq. (3).

$$f(x) \sim GPR\big(m(x), C(x,x')\big)$$
$$m(x) = E(f(x)) \quad (3)$$
$$C(x,x') = E\Big[\big(m(x) - f(x)\big)\big(m(x') - f(x')\big)\Big]$$

Here, E is the expected value. $C(x,x')$ is alternatively called a kernel function to capture the relevance between the real responses and predicted values by the similarities of the associated inputs.

In the regression process, the prior distribution of the output is specified as Eq. (4), where D refers to the distribution function and its mean value is set to zero for simplicity during the computations.

$$y \sim D(0, C(x,x')) \quad (4)$$

Considering y' as the predicted output, x as the training and x' as the test dataset, the joint prior distribution of y' and y is specified by Eq. (5) assuming that the training and test data have the same Gaussian distribution.

$$\begin{pmatrix} y \\ y' \end{pmatrix} \sim D\left(0, \begin{bmatrix} C(x,x) & C(x,x') \\ C(x,x')^T & C(x',x') \end{bmatrix}\right) \quad (5)$$

These kernel functions of the GPR need to be carefully selected and their hyperparameters, which are denoted by θ_G, are required to be optimised to guarantee a desirable performance of the model. A review of various kernel functions, their details and hyperparameters is given in [43].

The hyperparameter optimisation is usually performed through minimising the negative logarithmic likelihood function $L(\theta_G)$ given as Eq. (6) using the gradient descent or heuristic approaches [44, 45].

$$L(\theta_G) = \frac{1}{2}\log\left[\det C(\theta_G)\right] + \frac{1}{2} y^T C^{-1}(\theta_G) y + \frac{n}{2}\log(2\pi) \tag{6}$$

Through this optimisation the predicted output of the test dataset can be obtained through its conditional distribution by Eq. (7),

$$p(y' \mid x, y, x') \sim N(y' \mid \bar{y}', \mathrm{cov}(y')) \tag{7}$$

where $\bar{y}'$ is the mean value of the predicted outputs and the $\mathrm{cov}(y')$ is their variance defined by $\bar{y}' = C(x,x')^T [k(x,x)]^{-1} y$ and $\mathrm{cov}(y') = C(x',x') - C(x,x')^T \left[k(x,x)\right]^{-1} k(x,x')$ respectively.

To validate the model, the cross-validation (CV) approach is taken [46]. Cross validation is a technique based on splitting data randomly into K folds and using those folds for training and validation iteratively. The folds are mutually exclusive portions of the whole dataset, and the idea is to use each portion at least once for validation. The number of folds is defined by the researcher and the folds usually have the same size. The diagram of Figure 4 shows the split of the data into K = 5 folds of training and validation. As the figure shows, at each iteration, K-1 folds are used for training the model and the remaining fold is used for validation. CV is very suitable for small to medium size datasets where there is limited data for validation. The performance of the model is only evaluated based on the accuracy metrics applied to the validation portion of the data.

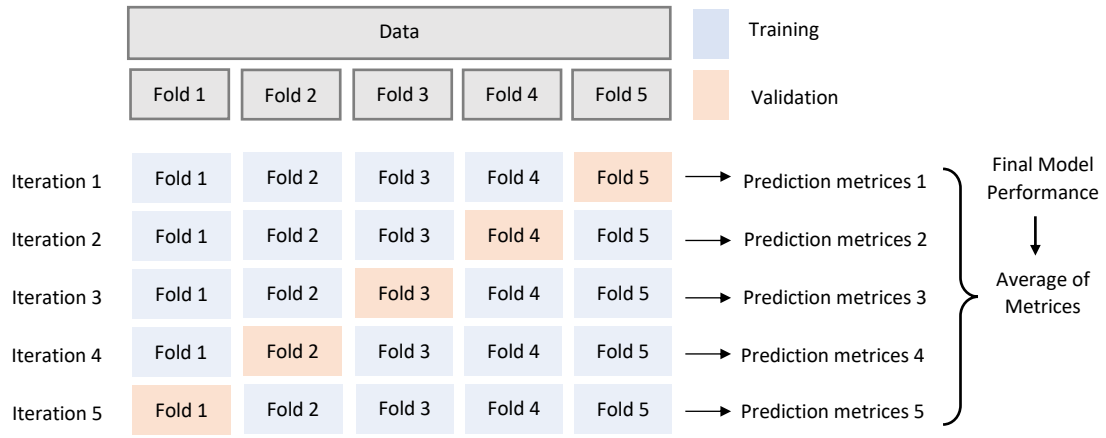

Figure 4: Cross validation approach for training and evaluating the model performance during validation

To evaluate the performance of the models for the prediction of the SoH values, four complimentary performance metrics are utilised.

Root mean squared error (RMSE), is the first measure which is calculated by Eq. (8) and indicates the difference between the predicted and real SoH values. Here N is referring to the number of data points. RMSE has the same unit of SoH so it is in %. Mean squared error (MSE) is the second metric which is the second power of the RMSE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(y_j - \hat{y}_j\right)} \tag{8}$$

The third metric is R-Squared (R^2), obtained by Eq. (9)

$$R^2 = 1 - \frac{SSE}{TSE}$$
$$SSE = \sum_{j=1}^{N} \left(y_j - \hat{y}_j \right)^2 \quad (9)$$
$$TSE = \sum_{j=1}^{N} \left(y_j - \bar{y} \right)^2$$

where SSE is the sum of the squared error and TSE is the total sum of squares obtained for the average of all output values $\bar{y} = \frac{1}{N} \sum_{j=1}^{N} y_j$ for $j = 1, \ldots, N$. R^2 is a value between 0 and 1, it does not have a unit and so reported here in percent. An R^2 of 100% means a perfect prediction where the modelled outputs and the real values are matched exactly. The final metric is the mean absolute error (MAE) which is given by Eq. (10).

$$MAE = \frac{1}{N} \sum_{j=1}^{N} \left| y_j - y_j ' \right| \quad (10)$$

The four metrics are only applied to the validation portion of the data during each iteration of cross-validation.

IV. Main Results

To analyse the performance of the GPR model for the prediction of SoH, two case studies are reported here. One with all the data points of the EIS graphs as inputs to the model beside the SoC and temperature, and the other one, with specific features extracted from the EIS graphs as inputs. Figure 3 (c) shows the approximate features on a sample EIS graph. The list of the inputs for each of the case studies has been given in the Table 1.

Table 1: The input and outputs for the two case study models based on EIS and conditioning data

Model	Case Study 1	Type	Case Study 2	Type
Inputs	EIS at 61 frequencies	Real Imaginary	F1: Highest Frequency point F2: Minimum Real Value point	Real Imaginary
	Temperature SoC		F3: lowest frequency point F4: Zero crossing point F5: Peak point F6: local minima between F4 and F5 F7: local minima between F5 and F3 Temperature SoC	Frequency
Response	SoH (100% to 80%)		SoH (100% to 80%)	

It is worth mentioning that one means to exploit the EIS data to train the ML algorithm is to extract key features from the spectrum that can be used to characterise the general shape of the EIS response and the ageing state of the battery. This has the potential advantage of reducing the volume of data that must be stored to underpin model training and validation. Here, the features are selected considering the variations in between the cells at different SoHs as well as the expert's view considering the fact that each of those is a representative of various characteristics of a cell's health as:

- F1 is related to the information of cell at the highest frequency
- F2 is a numerical choice
- F3 is related to the information of cell at lowest frequency, the slope of the line between F7 and F3 is related to the gradient of the Warburg impedance

- F4 is zero-crossing point and related to the DC resistance of cell
- F6 is related to the possibility of the build-up of formation of the solid electrode interphase layer
- F5 is related to the value of change transfer impedance
- F7 is the “turning-point” that represents the transition to diffusion

A detailed discussion on each of these factors is beyond the scope of this paper and can be found in a number of educational texts [47] and research articles highlighting how to interpret the EIS profile of the lithium-ion battery [48], build an equivalent circuit model of the battery for voltage prediction and to facilitate control algorithm design [49] and finally to infer battery ageing and the different degradation modes that impact battery life and performance [48].

In order to show the importance and relevance of the selected features of the second case study, a correlation analysis is performed here. For this purpose the correlation coefficeints and their significane is calculated via Pearson product-moment correlation analysis method [51]. considering the features as variable F, and the SoH for each case, the coefficients of $r_{F,SoH}$ are obtained via equation (11). Where, μ is the mean value of each feature, σ is the standard deviation and $\mathbb{E}$ refers to the expectation.

$$r_{F,SoH} = \frac{\mathbb{E}\left[\left(F-\mu_F\right)\left(SoH-\mu_{SoH}\right)\right]}{\sigma_F \sigma_{SoH}} \tag{11}$$

The coefficients are values between zero and one and could be positive or negative which refer to direct or invers correlation. The closer the value to the one, the stronger the correlation. To be able to investigate if the correlation between each feature and SoH is demonstrating a significant relation and the analysis are generalisable from the samples to the population [50] a p-value is also calculated which denotes the probability of a ‘Null Hypothesis’. The null hypothesis is that there exists no relationship between the features and the SoH [51]. P-value is a number between zero and one and if are smaller than a threshold, here 0.1, then the correlation is considered significant. This means the results are valid in 90% of cases. It is worth to mention that this threshold is a common choice for most of the data analysis problem. Figure 5 shows the correlation coefficient and p-values of each feature in relation the cells’ SoH.

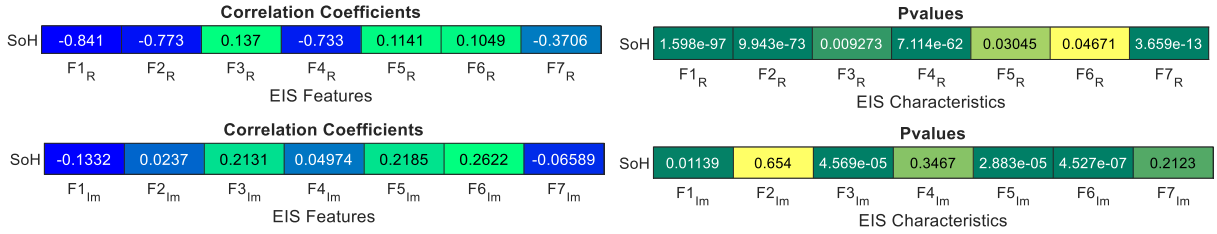

Figure 5: Correlation between EIS features and SoH of cells

As Figure 5 shows, the real part of all features is highly correlated with the SoH of cells, this is concluded by the correlation coefficients and the p values smaller than the mentioned threshold. It is worth to highlight that some correlations are direct, and some are inverse and the correlation of the real part of feature 6 is slightly weaker than others as its lower correlation coefficient shows. Similarly, the imaginary part of the features is significantly correlated with the SoH value, the weakest correlation is for the imaginary part of feature 2. Considering this correlation analysis in total all the 7 features could be valuable for the SoH estimation of cell, therefore all of them are considered for the modelling purposes in what follows.

For case study 1, the whole EIS data at various conditions are used for the model and as the frequency of the measurements has been the same for all cases, it is not a feature for the model. Figure 6 shows the regression results. Figure 6 (a) is the distribution of predicted and observed data with the record number, while Figure 6 (b) is showing the positioning of predictions and observations compared to a perfect prediction.

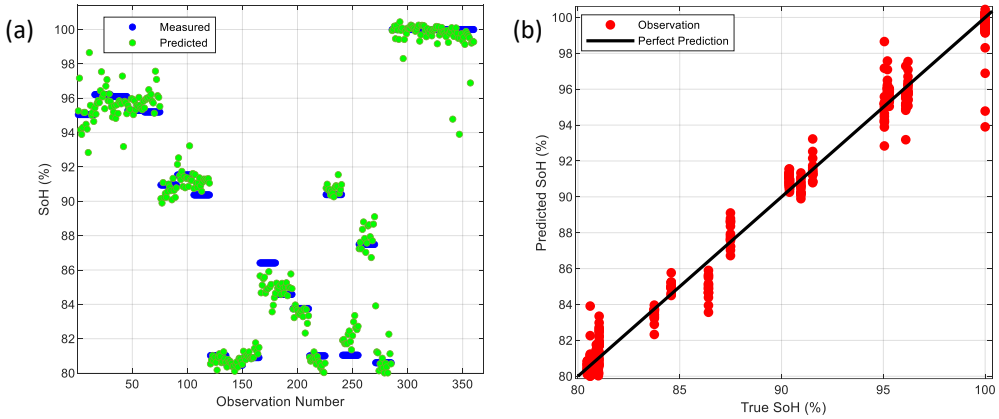

Figure 6: Regression results for case study 1, (a) predicted and observations SoH value distribution, (b) Predicted and true SoH values against perfect prediction, case study 1

As Figure 6 shows that most of the observations agree with the predicted values from the GPR model. For case study 2, the regression results are given in Figure 7.

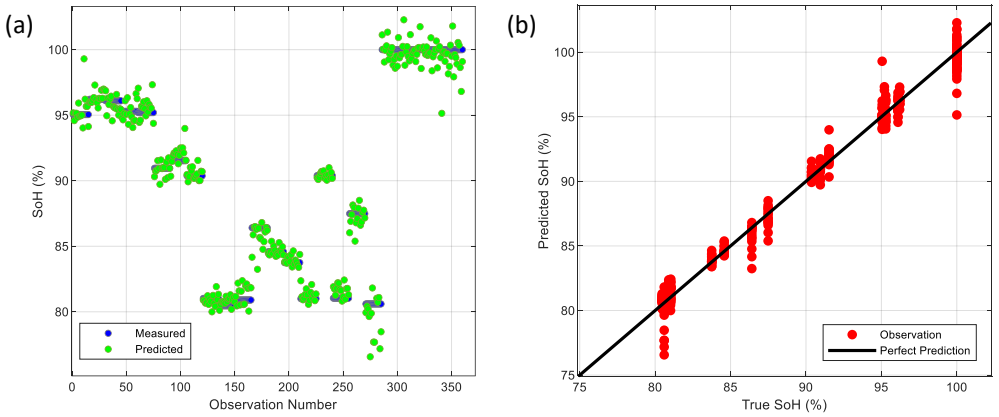

Figure 7: Regression results for case study 2, (a) predicted and observations SoH value distribution, (b) Predicted and true SoH values against perfect prediction, case study 2

Accuracy metrics of both models are listed in Table 2, the results are the mean and standard deviation (std) of multiple runs to facilitate a more comprehensive conclusion. Here the model has been iterated for 40 times as a compromise between run time and the stability of the results.

Table 2. Accuracy metrics for predicting the cells SoH via the ML model

Index Unit	RMSE (Mean) (%)	RMSE (Std) (%)	R^2 (Mean) (-)	R^2 (Std) (-)	MSE (Mean) (%)	MAE (Mean) (%)	Time (Mean) (sec)
Case Study 1	1.1096	0.0766	0.9759	0.0034	1.0374	0.6555	4011
Case Study 2	1.1397	0.0906	0.9750	0.0042	1.6407	0.8573	270

According to the table, the model with full data points, case study 1, has a higher accuracy, about 0.03% better than the feature-based model, case study 2. This is also observable for the R^2 value which is about 0.01 better than the feature-based model. The difference between the two models become more visible in the MSE and the MAE and the time indices. The full data point model has a MSE of

0.6% and a MAE of 0.2% less than the feature-based model. However, the run time of the full point model is much more than that of the feature-based model which confirm the complexity of its training, according to the last column of the table the run time in case 1 is about 50 times higher than case 2. The run time has been calculated via a personal computer with 1.90 GHz and 2.11 GHz CPU and 16.0 GB RAM. It is also worth noting that the standard deviation of the full point model is lower which implies that it is more stable that feature-based model.

To further highlight the performance of the model in different conditions, the distribution of the prediction error with respect to the conditioning features, (SoC and Temperature) is given in Figure 8. Here the y axis refers the prediction versus real SoH value, as it is a subtraction of the predicted and real SoH in percent.

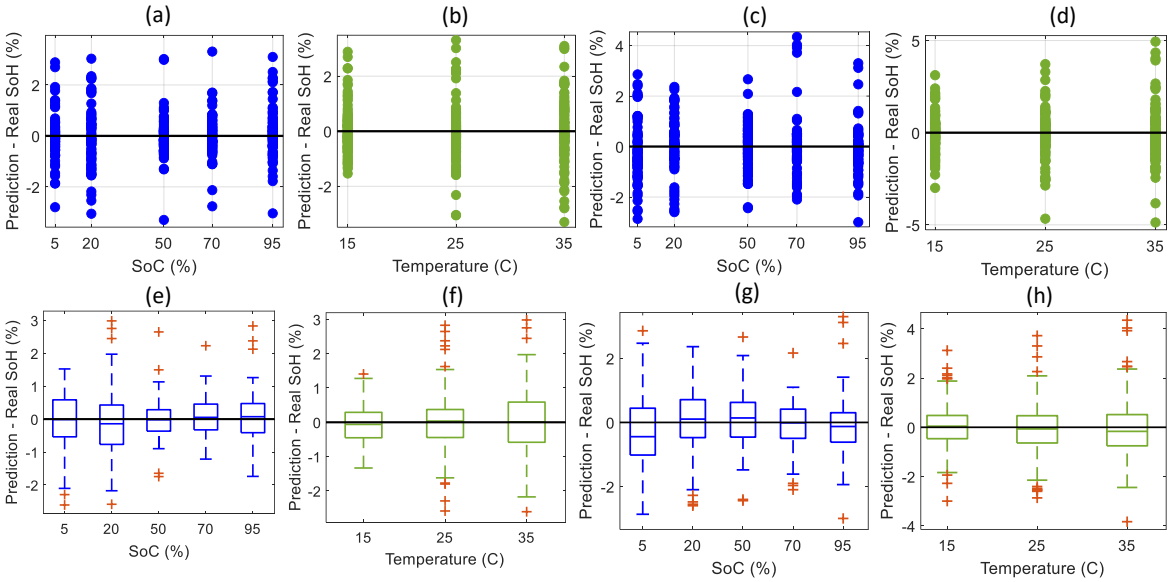

Figure 8: The distribution of prediction error at various conditions, (a) SoC case study 1, (b) temperature case study 1, (c) SoC case study 2, (d) temperature case study 2, (e) - (f) box plots case study 1, (g) - (h) box plots case study 2

According to the Figure 8, the distribution of the prediction error is quite different from one SoC point to another and from one temperature to another. The quantified errors are listed in Table 3.

Table 3. Accuracy metrics for predicting the cells SoH via the ML model at various SoCs

	SoC (%)					Temperature (C)		
	5	20	50	70	95	15	25	35
Index	**Case Study 1**							
Median	-0.0141	-0.1415	-0.0206	0.0533	0.0733	-0.0627	0.0358	0.0009
Lower Quartile	-0.5380	-0.7682	-0.3652	-0.3263	-0.4125	-0.4515	-0.4444	-0.5832
Higher Quartile	0.5852	0.4283	0.2854	0.4556	0.4726	0.2895	0.3725	0.5888
Index	**Case Study 2**							
Median	-0.4446	0.1041	0.1395	-0.0122	-0.1269	0.0436	-0.0669	-0.1658
Lower Quartile	-1.0177	-0.4737	-0.4578	-0.4968	-0.6158	-0.4629	-0.6372	-0.7526
Higher Quartile	0.4439	0.7113	0.6251	0.4143	0.3031	0.4800	0.4699	0.5152

For case study 1, the largest prediction error median is 0.1415% at 20% SoC, Figure 8 (e), the median of error is largest at 15 °C, Figure 8 (f). For case study 2, the graphs of Figure 8 and the indices on Table 3 show that the largest prediction error is related to the lowest SoC, (5%) with a median of 0.4446%. This error is minimum at mid-range SoC which is 70% with a median of 0.0122%. For temperature, the minimum prediction error is related to the room temperature (25 °C) with a median of 0.0669%, while the highest error is for highest temperature of 35 °C with a median of 0.1658 (%). It is also evident

that the prediction error at a fixed SoC across all temperatures, Figure 8 (g), is having larger errors compared to the prediction at a particular temperature across all SoCs, Figure 8 (h). Summarising both case studies, it can be concluded that model performance is generally better at middle range SoC and temperatures. The implications of this for the future deployment of the model within a practical use case is explored further in the following section.

V. Discussions

V.I Sensitivity and robustness to noise

Understanding the source of measurement noise associated with making EIS measurements is particularly important during the development of the SoH estimation models by ML. The noise in the data is usually caused by the integrated components of the EIS measurement setups such as the connection cables and can be very misguiding because of the slight excitation signals required for the EIS measurements [22]. Here the main source of noise is assumed to be originated from the SoC and Temperature measurements.

To evaluate the sensitivity of the developed SoH prediction model to noise, a noise analysis methodology is proposed here. The methodology is described in the Figure 9.

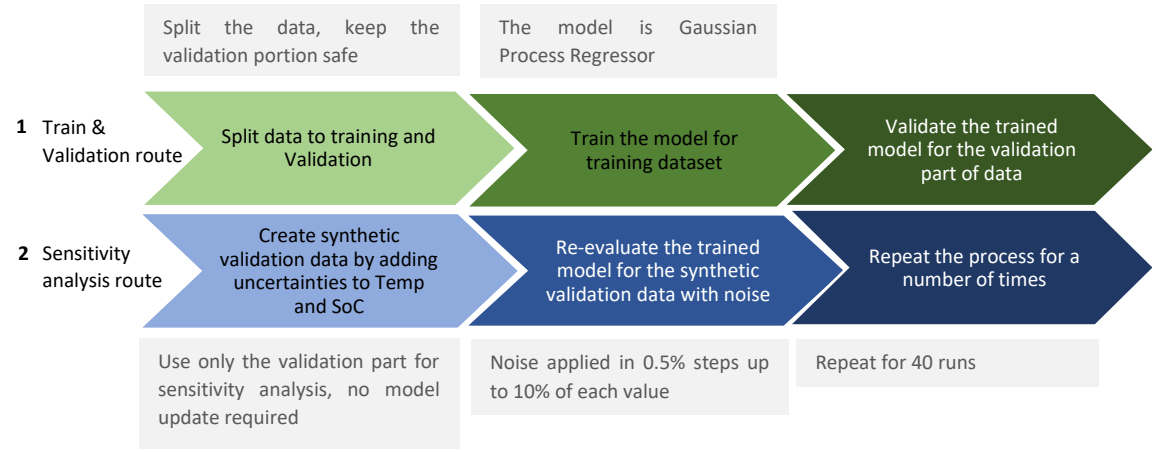

Figure 9: The methodology to analyse the impact of the measurement noise on SoH prediction model's performance

For this purpose, first the data is split into training and validation portions. The ratio of 80% for training and 20% for validation is preferred here. Then the model is trained on the training portion and validated on the remaining part. The validation portion is then altered by adding an uncertainty with the Gaussian distribution of mean zero and the standard deviation up to 10% of the temperature and SoC value at each record of data. This uncertainty is to represent the measurement noise for the two inputs synthetically. Then the model is validated against the new data without a change in its optimised hyperparameters. The whole process is repeated for multiple times to evaluate the repeatability of the method and increase the confidence in the results.

To capture the results, the models are iterated for 40 times to quantify the variation in the performance metrics and the impact of the random uncertainties. This number of repetitions is deemed suitable because of a balance between the stability of the results and the run time. The models start with a noise level of 0% and get evaluated by a stepwise increment in noise, 0.5% at each step, till 10% of the nominal value of the same record. The results at location 0 of x-axis contribute to the original data and the rest are based on synthetical noisy data. Figure 10 implies that the model

can tolerate up to 8.2% of noise for an R^2 value higher than 90% and 7.8% of noise for a RMSE lower than 1.5%. Considering both metrics, an overlapping region is visible for up to 7.8% of measurement noise. The overlapping regions are visible in red and blue, respectively for the R^2, Figure 10 (a), and RMSE, Figure 10 (b). The analysis confirms that the 7.8% of noise is still guaranteeing the model's performance to a good extent.

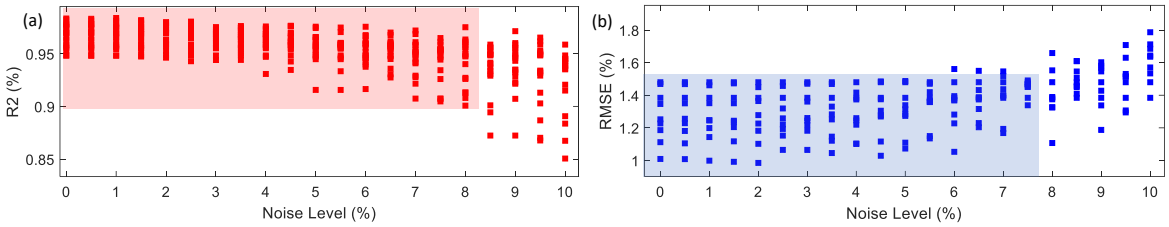

Figure 10: The SoH prediction model sensitivity to measurement noise in temperature and SoC

V.II Sensitivity and robustness to EIS characterisation equipment

Although the EIS test profiles of the cell are mostly affected by the characterisations of the cell in terms of its SoH, SoC and temperature, the testing equipment and the experimental apparatus have also an undeniable impact on the EIS data features. In this section the trained model for which the data have been collected using the Multiplexer equipment, is validated against a new dataset that has been collected via an alternative equipment, Gamry which is a DC and impedance analyser. In the validation experiments, 5 cells were put under EIS test at 25 °C and a range of SoC conditions. Beside the difference in the testing equipment, their cooling mechanism was selected to be via turbulent air inside a thermal chamber rather than the oil cooling which was used during the collection of the first data set. Representing different levels of experimental fidelity and the possible impact this may have on model prediction performance. In total 25 experiments were conducted.

The testing equipment and the difference in the connections, cabling and cooling mechanism all had an impact on the cells EIS curve. This difference was quite obvious specially at the inductive tail of the EIS curves as shown in Figure 11.

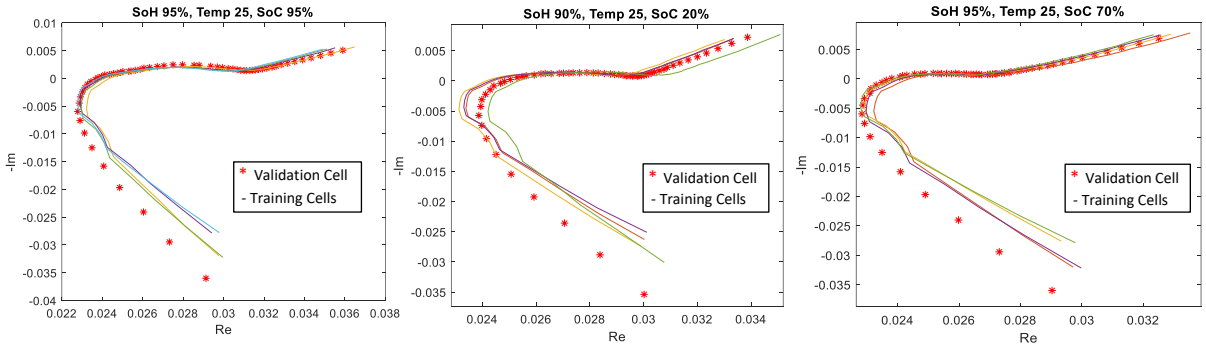

Figure 11: The impact of the testing equipment on the EIS measurements at different conditions (-) Multiplexer equipment (training data), (*) Gamry equipment (validation data)

Table 4. The performance of the model against the new EIS data with a different equipment

	Test Conditions			Model of Case Study 1		Model of Case Study 2	
	SoC	Temp	SoH (%)	Average Error (%)	RMSE (%)	Average Error (%)	RMSE (%)
Validation set 1	5-95%	25	89.4	2.86	1.38	3.61	1.62
Validation set 2	5-95%	25	95.0	-2.54	1.33	4.09	1.84
Validation set 3	5-95%	25	99.6	1.96	1.76	2.58	1.18
Validation set 4	5-95%	25	81.5	5.32	2.42	7.65	3.42
Validation set 5	5-95%	25	92.0	1.86	1.06	3.34	1.50

Considering the new data, Table 4 summarises the performance of the model in predicting the SoH of this new dataset while it is trained on the original data set. It is worth highlighting that the model has been completely blind to the new data during the training. Here, the Gamry equipment data are called the validation cell data for generality.

The results shown in Table 4, indicate that the methodology is reproducible. The model performs responds well to the change in the test equipment. Considering the model is trained by all the data points of the EIS curve, the average error of prediction for the 5 cases is 2.9%, while the RMSE is equal to 1.59%. Considering that the RMSE of the trained model via original dataset was actually 1.1096% these values are very much desirable. For the second model trained based on the features described in Table 1, the average prediction error for the 5 conditions is 4.25%, while the RMSE by average is 1.91%. Comparing the feature-based and full point models reveals that the full-point model is more robust to the change in the testing equipment, which is believed to be due to its ability in generalisation because of the larger volume of information that it receives during training (61 data points) compared to the 7 data points of the feature-based configuration.

V.III SoC and Temperature contributions

The characterisation of second-life Lithium-ion batteries is a complex process which is affected by various factors such as the cells temperature, their SoC, their ageing history, the first life conditions as well as storage conditions. Measuring and quantifying all these factors is costly and non-practical specially when it comes to high volumes of cells and systems received from the provider. To realise the significance of each factor and its impact on the accuracy of the estimated SoH, an in-depth feature contribution analysis is crucial. In this section, the contribution of SoC and temperature on the predicted SoH accuracy is quantified. For this purpose, first the model is trained with all the predictors, which is SoC, temperature and all EIS datapoints, then various attempts have been made to train the model with a different set of predictors. A case involving only SoC and EIS data (without temperature), a case with temperature and EIS data (without SoC), and finally a case with only the EIS data (omitting SoC and Temperature) is created. A summary of the accuracy metrics of the models is provided at Figure 12. According to the comparative results, the model with full predictors has the best performance (lowest errors and highest R^2). Considering the model with all predictors as the baseline, it can easily be seen that removing the temperature from the predictor pool increases the prediction RMSE by about 12.84% and decreases the R^2 value by 0.77. A model without SoC increases the RMSE by 19.96% while reducing the R^2 by 1.09.

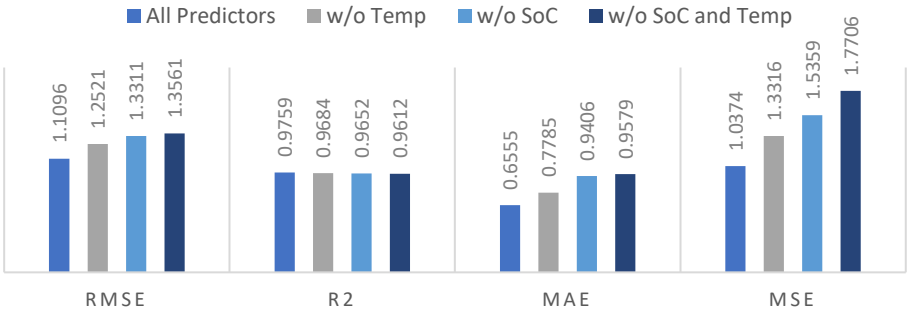

Figure 12. The impact of SoC and Temperature on the SoH prediction model performance

A model without SoC and temperature leads to predictions with about 22.21% higher RMSE and circa 1.50 lower R^2 which is the poorest performance. Considering the accuracy indices reported above, the temperature and SoC play an important role in the predictability of the SoH values. Comparing the

two cases that work without SoC or temperature as predictors, it is clear that SoC is a more contributing factor to the accuracy of the model than temperature. This contribution analysis is crucial when dealing with limitations regarding the measurements of SoC or temperature of cells for their SoH estimation. In summary the SoC and temperature measurements are not necessary to be obtained if the error levels of SoH prediction are considered tolerable or acceptable by the application engineers during the characterisation process.

V.VI ML conflicts and limitations

Generally, when it comes to the application of machine learning techniques for discoveries regarding engineering problems, including those within the context of energy storage systems and Li-ion batteries, three major challenges arise that need to be addressed reasonably to gain the confidence of users. According to the literature and expert's view [52], these contradictions are about the three following items. In this section is it explained how each item has been addressed in this study.

- *Data and Sample Size*

The second-life battery state of health is affected by a wide variety of factors such as its first life conditions, storage conditions, its state of health, ageing mechanism and many more; however, it is very costly and resource consuming to measurement and quantify all the mentioned factors. In a pilot or laboratory scale characterisation line such as what is available in WMG, University of Warwick, it has taken about 10 months to gather data for about 325 cases (for 35 cells). This time has been invested to make sure that a reasonable volume of data is available to train the machine learning models and make conclusions as general as possible. Furthermore, a second dataset is created for validation purposes in a different testing condition than the original one used for model creation, and a feature contribution analysis is performed to ease the decision making regarding the quantification of less important variables for larger-scale experiments.

- *Complexity, Accuracy, and Ease of Use*

When it comes to the state of health estimation of Li-ion batteries, selecting the model configuration and specially its hyperparameters are a major concern due to the nonlinear interconnections between the factors and the response of cells. The strongest model if not tuned and optimised properly might not be able to generalise the decisions from the training to the validation and test scenarios. In general, linear models such as support vector machine with linear kernels, or simpler models such as decision trees, although easy to train and use, usually have low accuracy and limited capability for interpretability. Therefore, the application of more complex models is inevitable in this case. In this study the GPR models are utilised to achieve a reasonable accuracy in balance with model training complexities. For the maximum performance, Bayesian optimisation is utilised for the automatic tuning of hyperparameters. The Bayesian optimisation technique while might be resource demanding during the training process, is not raising a particular concern for this problem as the training of such models usually take place in an offline configuration and only the model deployment which does not have a specific resource requirement takes place in real-time.

- *Learning results and domain expert knowledge*

While ML models have been in wide use for the prediction of characteristics for lithium-ion related problems, it is still a challenge to effectively include the prior knowledge of domain experts in models. In this study, the authors have tried to incorporate this knowledge during the data collection, and

preparation stage of modelling and via expert-led feature engineering. It is worth mentioning that the findings of this research were mostly in one of the two categories, 1) results that our experts had an empirical prior expectation and understanding of those, and this research helped to quantify those, such as the contribution of temperate and SoC to the SoH and their relative importance. 2) results that were new to the experts such as the impact of measurement noise, testing conditions (EIS testing equipment and cooling mechanism) on the response, which can initiate new lines of research for future investigations. In general, it is acknowledged that creating a systematic approach for a weighted and scored inclusion of expert knowledge is imperative, however, this requires extensive research and it currently out of the scopes of our study.

VI. Conclusions

In this research, the prediction of the li-ion battery SoH using the EIS measurements and ML approaches is addressed. This study is built upon a large number of experimental cases, 325 cases, by taking into account the cell impedance across a desired range of frequencies between 10 mHz to 10 KHz, the effect of cell-to-cell variation, the cell state of charge, temperature, as well as characterisation uncertainties. It links the SoH of cells to their EIS features at various experimental equipment uncertainties and measurements noises. The methodology for creating the model beside the approaches taken for uncertainty and robustness analysis are the main novelties of this study in the context of Li-ion battery SoH estimation. Linking back to the research questions listed in the introduction section, the following observations and findings can be summarised after this study.

- The EIS data beside the conditioning information can be directly used to predict the SoH of cells with an accuracy up to 98.89%. Feature extraction and engineering can also be utilised for building a ML model, in such case the accuracy drops to 98.86% which is still very desirable. This accuracy reduction is due to the lost information when the EIS data are compressed from 61 frequency points into only 7 features. If the training time is factored in, then the full EIS model is more complex to be optimised with a training time of about 50 times higher than the feature-based model. But and if the robustness of model is factored in, the full point model is more efficient than the feature-based model (accuracy of 98.41% for new data set compared to 98.09%). Considering the trade-off between accuracy, robustness, and the training effort, one model can be preferred over the other one in a particular application. The accuracy obtained by either model implies that the ML-based approaches can facilitate an accelerated SoH estimation method with a reasonable effort.
- It is important that the model for SoH prediction be able to generalise the predictions from the training data set to the new data sets. To provide this capability, considering a cross validation approach during the training is necessary as it helps optimising the model's hyperparameters by virtually creating a larger data set and considering each data point for validation at least once. Testing the model's performance against a new dataset that was collected through a different experimental equipment and under a different cooling mechanism showed that the ML model is performing well with an error below 1.56% (case study 1) and 1.91% (case study 2). Further analysis showed that model is also robust to the measurement noise of up to 7.8% of the nominal EIS datapoints. This quantifies the level of noise that is tolerable when transferring the modelling technology from one application to another where the quality of measuring equipment is not necessarily the same.
- It is obvious that the SoH prediction accuracy depends on the information that is collected from the cells for their impedance, temperature, and SoH. Considering the effort and resource required for collecting this information, it is very valuable to be able to quantify each factor's impact on the

model accuracy. Through separating the impact of SoC and temperature on the model performance, it was noted that the model is has its best performance with both features, and the contribution of SoC to the model accuracy is higher than that of temperature. This means that in case of resource limitation, the SoC characterisation should be prioritised to the temperature measurements. It was also witnessed the model's performance is not the same across various SoC and temperatures. According to the analysis, the ML model has the largest median of the prediction error is at 20% SoC and 15 ^{o}C, which is believed to be due to the impact of other factors, such as the type of again mechanisms, that have not taken into account or quantifies at low SoC or temperatures.

The main future work planned in this research direction is about including further factors in cell characterisation for its SoH estimation. Factoring in the cell's first life information, such as the charge and discharge scenarios, the storage conditions and calender ageing is believed to enhance the model accuracy. However, this requires an agreement between the cell providers after their first life and the researchers in terms of data sharing and management. Another interesting direction for future studies in the application of ML models and EIS tests for online estimation of SoH. In such case the combination of the in-operando EIS test [21], and the proposed methodology for ML model creation could be a fair solution.

Supplementary Information

The experimental data of this study will be available through a data in Brief Submission. Contact Authors for further information.

Acknowledgement

This research was undertaken as part of the RECOVAS project, project funded by the UK Advanced Propulsion Centre (APC), Grant Number: TS/V012789/1.

Declaration of competing interest

The authors declare that there is no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT

Mona Faraji Niri: Conceptualization; Data curation; Formal analysis; Methodology; Software; Visualization; Validation; Investigation; Writing - original draft. Muhammad Rashid: Methodology; Data curation; Software; Investigation; Writing - original draft. Jonathan Sansom: Methodology; Investigation; Writing - review & editing. Dhammika Widanage: Software, James Marco: Funding acquisition; Supervision; Resources; Writing - review & editing.

References

[1] A. Holland, J. Edmondson and L. Gear , "Lithium-ion Batteries for Electric Vehicles 2021-2031," IDTechEX, 2021.

[2] Y. Jiang, J. Jiang, C. Zhang, W. Zhang and Y. Gao, “State of health estimation of second-life LiFePO4 batteries for energy storage applications,” *Journal of cleaner production,* vol. 205, pp. 754-762, 2018.

[3] L. Ungurean, G. Cârstoiu, M. Micea and V. Groza, “Battery state of health estimation: a structured review of models, methods and commercial devices,” *nternational Journal of Energy Research,* vol. 41, no. 2, pp. 151-181, 2017.

[4] L. Yao, S. Xu, A. Tang, F. Zhou, J. Hou, Y. Xiao and Z. Fu, “A Review of Lithium-Ion Battery State of Health Estimation and Prediction Methods,” *World Electric Vehicle Journal,* vol. 12, no. 3, p. 113, 2021.

[5] M. Haram, J. Lee, G. Ramasamy, E. Ngu, S. Thiagarajah and Y. Lee, “Feasibility of utilising second life EV batteries: Applications, lifespan, economics, environmental impact, assessment, and challenges,” *Alexandria Engineering Journal,* vol. 60, no. 5, pp. 4517-4536, 2021.

[6] M. Abdel-Monem, O. Hegazy, N. Omar, K. Trad, P. Van den Bossche and J. Van Mierlo, “Lithium-ion batteries: Comprehensive technical analysis of second-life batteries for smart grid applications,” in *19th European Conference on Power Electronics and Applications*, Warsaw, Poland , 2017.

[7] E. Martinez-Laserna, E. Sarasketa-Zabala, I. Sarria, D. Stroe, M. Swierczynski, A. Warnecke, J. Timmermans, S. Goutam, N. Omar and P. Rodriguez, “Technical viability of battery second life: A study from the ageing perspective,” *IEEE Transactions on Industry Applications,* vol. 54, no. 3, pp. 2703-2713, 2018.

[8] A. Burke and . M. Miller, “Life cycle testing of lithium batteries for fast charging and second-use applications,” in *2013 World Electric Vehicle Symposium and Exhibition*, 2013.

[9] S. Moura, N. Chaturvedi and M. Krstić, “Adaptive partial differential equation observer for battery state-of-charge/state-of-health estimation via an electrochemical model,” *Journal of Dynamic Systems, Measurement, and Control,* vol. 136, no. 1, p. 011015, 2014.

[10] Y. Gao, K. Liu, C. Zhu, X. Zhang and D. Zhang, “Co-estimation of state-of-charge and state-of-health for lithium-ion batteries using an enhanced electrochemical model,” *IEEE Transactions on Industrial Electronics,* vol. 69, no. 3, pp. 2684-2696, 2021.

[11] J. Li, K. Adewuyi, N. Lotfi, R. Landers and J. Park, “A single particle model with chemical/mechanical degradation physics for lithium ion battery State of Health (SOH) estimation,” *Applied energy,* vol. 212, pp. 1178-1190, 2018.

[12] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche and J. Van Mierlo, “A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter,” *Journal of Power Sources,,* vol. 373, pp. 40-53, 2018.

[13] M. Berecibar, M. Garmendia, I. Gandiaga, J. Crego and I. Villarreal, “State of health estimation algorithm of LiFePO4 battery packs based on differential voltage curves for battery management system application,” *Energy,* vol. 103, pp. 784-796, 2016.

[14] C. Pastor-Fernández, K. Uddin and G. Chouchelamane, “A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems,” *Journal of Power Sources,* vol. 360, pp. 301-318, 2017.

[15] D. Stroe, M. Swierczynski, A. Stan and V. Knap, “Diagnosis of lithium-ion batteries state-of-health based on electrochemical impedance spectroscopy technique,” in *IEEE Energy Conversion Congress and Exposition*, 2014.

[16] R. Mingant, J. Bernard, V. Moynot, A. Delaille, S. Mailley, J. Hognon and F. Huet, "EIS measurements for determining the SOC and SOH of Li-ion batteries," *ECS Transactions,* vol. 33, no. 39, p. 41, 2011.

[17] J. Kim, L. Krüger and J. Kowal, "On-line state-of-health estimation of Lithium-ion battery cells using frequency excitation," *Journal of Energy Storage,* vol. 32, p. 101841, 2020.

[18] R. Mingant, J. Bernard and V. Sauvant-Moynot, "Novel state-of-health diagnostic method for Li-ion battery in service," *Applied Energy,* vol. 183, pp. 390-398, 2016.

[19] M. Messing, T. Shoa and S. Habibi, "Estimating battery state of health using electrochemical impedance spectroscopy and the relaxation effect," *Journal of Energy Storage,* vol. 43, p. 103210, 2021.

[20] C. Pastor-Fernández, W. Widanage, J. Marco, M. Gama-Valdez and G. Chouchelamane, "Identification and quantification of ageing mechanisms in Lithium-ion batteries using the EIS technique," in *IEEE Transportation Electrification Conference and Expo*, Dearborn, MI , 2016.

[21] N. Hallemans, W. Widanage, X. Zhu, S. Moharana, M. Rashid, A. Hubin and J. Lataire, "Operando electrochemical impedance spectroscopy and its application to commercial Li-ion batteries," *Journal of Power Sources,* vol. 547, p. 232005, 2022.

[22] Y. Xing, E. Ma, K. Tsui and M. Pecht, "An ensemble model for predicting the remaining useful performance of lithium-ion batteries," *Microelectronics Reliability,* vol. 53, no. 6, pp. 811-820, 2013.

[23] L. Chen, H. Wang, B. Liu, Y. Wang, Y. Ding and H. Pan, "Battery state-of-health estimation based on a metabolic extreme learning machine combining degradation state model and error compensation," *Energy,* vol. 215, p. 119078, 2021.

[24] H. Pan, Z. Lü, H. Wang, H. Wei and L. Chen, "Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine," *Energy,* vol. 160, pp. 466-477, 2018.

[25] R. Li, J. Hong, H. Zhang and X. Chen, "Data-driven battery state of health estimation based on interval capacity for real-world electric vehicles," *Energy,* vol. In press, p. 124771, 2022.

[26] Y. Che, Z. Deng, P. Li, X. Tang, K. Khosravinia, X. Lin and X. Hu, "State of health prognostics for series battery packs: A universal deep learning method," *Energy,* vol. 238, p. 121857, 2022.

[27] M. Alipour, S. Tavallaey, A. Andersson and D. Brandell, "Improved Battery Cycle Life Prediction Using a Hybrid Data-Driven Model Incorporating Linear Support Vector Regression and Gaussian," *ChemPhysChem,* vol. 23, no. 7, p. e202100829, 2022.

[28] H. Goh, Z. Lan, D. Zhang, W. Dai, T. Kurniawan, and K. Goh, "Estimation of the state of health (SOH) of batteries using discrete curvature feature extraction," *Journal of Energy Storage,* vol. 50, p. 104646, 2022.

[29] Y. Gong, X. Zhang, D. Gao, H. Li, L. Yan, J. Peng and Z. Huang, "State-of-health estimation of lithium-ion batteries based on improved long short-term memory algorithm," *Journal of Energy Storage,* vol. 53, p. 105046, 2022.

[30] Z. Xia and J. Qahouq, "Evaluation of parameter variations of equivalent circuit model of lithium-ion battery under different SOH conditions," in *IEEE Energy Conversion Congress and Exposition*, Detroit, MI, USA , 2020.

[31] A. Zenati, P. Desprez and H. Razik, "Zenati, A., Desprez, P. and Razik, H., 2010, November. Estimation of the SOC and the SOH of li-ion batteries, by combining impedance measurements with the fuzzy logic inference," in *36th Annual Conference on IEEE Industrial Electronics Society*, 1773-1778, 2010.

[32] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming and A. Lee, "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning," *Nature communications,* vol. 11, no. 1, pp. 1-6, 2020.

[33] W. Li, J. Chen, K. Quade, D. Luder, J. Gong and D. U. Sauer, "Battery degradation diagnosis with field data, impedance-based modeling and artificial intelligence," *Energy Storage Materials,* vol. 53, pp. 391-403, 2022.

[34] Y. Li, B. Dong, T. Zerrin, E. Jauregui, X. Wang, X. Hua, D. Ravichandran, R. Shang, J. Xie, M. Ozkan and C. Ozkan, "State-of-health prediction for lithium-ion batteries via electrochemical impedance spectroscopy and artificial neural networks," *Energy Storage,* vol. 2, no. 5, p. 186, 2020.

[35] Z. Xia and J. Qahouq, "Adaptive and fast state of health estimation method for lithium-ion batteries using online complex impedance and artificial neural network," in *IEEE Applied Power Electronics Conference and Exposition*, Anaheim, CA, USA , 2019.

[36] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming and A. Lee, "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning," *Nature communications,* vol. 11, no. 1, pp. 1-6, 2020.

[37] I. Babaeiyazdi, A. Rezaei-Zare and S. Shokrzadeh, "State of charge prediction of EV Li-ion batteries using EIS: A machine learning approach," *Energy,* vol. 223, p. 120116, 2021.

[38] A. Densmore and M. Hanif, "Determining battery SoC using electrochemical impedance spectroscopy and the extreme learning machine," in *2nd International Future Energy Electronics Conference*, Taipei, Taiwan, 2015.

[39] Y. Jiang, J. Jiang, C. Zhang, W. Zhang and Y. Gao, "Recognition of battery aging variations for LiFePO4 batteries in 2nd use applications combining incremental capacity analysis and statistical approaches," *Journal of Power Sources,* vol. 360, pp. 180-188, 2017.

[40] M. Rashid and A. Gupta, "Effect of Relaxation Periods over Cycling Performance of a Li-Ion Battery," *Journal of The Electrochemical Society,* vol. 162, no. 2), pp. A3145-A3153, 2015.

[41] A. Barai, G. Chouchelamane, Y. Guo, A. McGordon and P. Jennings, "A study on the impact of lithium-ion cell relaxation on electrochemical impedance spectroscopy," *Journal of Power Sources,* vol. 280, pp. 74-80, 2015.

[42] C. Rasmussen and H. Nickisch, "Gaussian processes for machine learning (GPML) toolbox," *Journal of Machine Learning Research,* vol. 11 , pp. 3011-3015, 2010.

[43] K. Liu, X. Hu, Z. Wei, Y. Li and Y. Jiang, "Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries," *IEEE Transactions on Transportation Electrification,* vol. 5, no. 4, pp. 1225-1236, 2019.

[44] R. Richardson, C. Birkl, M. Osborne and D. Howey, "Gaussian process regression for in situ capacity estimation of lithium-ion batteries," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 1, pp. 127-138, 2018.

[45] D. Yang, X. Zhang, R. Pan, Y. Wang and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *Journal of Power Sources,* vol. 384, pp. 387-395, 2018.

[46] T. Fushiki, “Estimation of prediction error by using K-fold cross-validation,” *Statistics and Computing,* vol. 21, no. 2, pp. 137-146, 2011.

[47] B. Liaw and G. Pistoia, Behaviour of Lithium-Ion batteries in electric vehicles, Berlin, Germany: Springer International Publishing, 2018.

[48] C. Pastor-Fernandez, T. F. Yu, W. D. Widanage and J. Marco, “Critical review of non-invasive diagnosis techniques for quantification of degradation modes in lithium-ion batteries,” *Renewable and Sustainable Energy Reviews,* vol. 109, pp. 138-159, 2019.

[49] A. Bizeray, J. Kim, S. Duncan and D. Howey, “Identifiability and parameter estimation of the single particle lithium-ion battery model,” *IEEE Transactions on Control Systems Technology,* vol. 27, no. 5, pp. 1862-1877, 2018.

[50] E. Komaroff, “Relationships Between p-values and Pearson Correlation Coefficients, Type 1 Errors and Effect Size Errors, Under a True Null Hypothesis,” *Journal of Statistical Theory and Practice,* vol. 14, no. 3, pp. 1-13, 2020.

[51] N. Gogtay and U. Thatte, “Principles of correlation analysis,” *Journal of the Association of Physicians of India,* vol. 65, no. 3, pp. 78-81, 2017.

[52] Y. Liu, B. Guo, X. Zou, Y. Li and S. Shi, “Machine learning assisted materials design and discovery for rechargeable batteries,” *Energy Storage Materials,* vol. 31, pp. 434-450, 2020.

[53] J. Benesty, J. Chen, Y. Huang and I. Cohen, “Pearson Correlation Coefficient,” in *Noise Reduction in Speech Processing. Springer Topics in Signal Processing*, Berlin, Heidelberg, Springer, 2009.

546

547