Journal of The
Electrochemical Society

Frequency-Based Parameterization of Semi-Empirical Models for State-of-Health Estimation in Lithium-Ion Batteries

To cite this article before publication: Colin J. Chu *et al* 2025 *J. Electrochem. Soc.* in press https://doi.org/10.1149/1945-7111/ae135d

Frequency-Based Parameterization of Semi-Empirical Models for State-of-Health Estimation in Lithium-Ion Batteries

Journal:	*Journal of The Electrochemical Society*
Manuscript ID	JES-114911.R1
Manuscript Type:	Research Paper
Date Submitted by the Author:	15-Sep-2025
Complete List of Authors:	Chu, Colin; Stanford University, Energy Science and Engineering Thatipamula, Sai; Stanford University, Energy Science and Engineering Onori, Simona; Stanford University, Energy Resources Engineering
Keywords:	Lithium-ion battery, state-of-health, modeling and estimation, electrochemical impedance spectroscopy, equivalent circuit models

Frequency-Based Parameterization of Semi-Empirical Models for State-of-Health Estimation in Lithium-Ion Batteries

Colin J. Chu[a], Sai Thatipamula[b], and Simona Onori*[b]

[a]The Nueva School, San Mateo, CA 94403, USA
[b]Department of Energy Science & Engineering, Stanford University, Stanford, CA 94305, USA
*Corresponding author: sonori@stanford.edu

Abstract

The accurate estimation of state-of-health (SoH) in lithium-ion batteries is crucial for enabling safe and improved battery usage. Yet the complexity of electrochemical aging processes within the cell and their dependence on differing operating conditions make this task challenging. This paper proposes a novel SoH estimation method using equivalent circuit model (ECM) parameters identified based on electrochemical impedance spectroscopy and regression analysis. Six different ECM architectures are calibrated and compared based on quantitative metrics including root mean square error, Akaike information criterion, and Bayesian information criterion. Based on this analysis, the second-order constant phase element model is chosen and its parameters are used to perform three supervised tree-based regressions and one supervised linear regression. The random forest regression model is found to be the most accurate in estimating capacity-based SoH and a rigorous correlation analysis is conducted to examine the relationship between model parameters and SoH. This study utilizes data collected from an aging campaign of 22 5Ah nickel-manganese-cobalt lithium-ion cells and demonstrates the effectiveness of ECMs and regression models for SoH estimation.

Keywords: Lithium-ion battery, state-of-health, modeling and estimation, electrochemical impedance spectroscopy, equivalent circuit models.

1. Introduction

Lithium-ion (Li-ion) batteries deliver long-term efficiency and offer unique characteristics, including high gravimetric and volumetric energy density [1–3]. Li-ion batteries exhibit additional properties, including low self-discharge rates, extended cycle lives, and fast

charging capabilities [4–7]. Nonetheless, Li-ion cells degrade due to several mechanisms, including solid electrolyte interface growth, lithium plating, and particle cracking [8, 9]. In addition, further degradation can be triggered by transition metal dissolution and the decomposition of the electrolyte [10, 11]. Accurate modeling of capacity-based state-of-health (SoH) is necessary for the optimization of battery management system (BMS) performance [12–14]. Capacity-based SoH is defined by:

$$SoH(t) = \frac{Q(t)}{Q_0} \cdot 100\% \tag{1}$$

where $SoH(t)$ is expressed as a percentage of the measured capacity at time t, Q_t, and the nominal capacity when the battery is new, Q_0 [15]. Existing SoH estimation methods are constrained by the complexity of the electrochemical degradation mechanisms they can model, as well as limitations in robustness and accuracy when implemented in BMS applications [16, 17].

In this work, we use electrochemical impedance spectroscopy (EIS) data, which measures the impedance, or frequency-dependent opposition of the battery to current flow, to estimate SoH [18]. The EIS data used in this paper, as described in the Data section, provides key insights into the electrochemical processes in Li-ion batteries such as charge transfer, ion diffusion, and double-layer capacitance [19–21].

While previous literature has focused on equivalent circuit model (ECM)-based methods for real-time SoH estimation, this work proposes a novel approach that leverages regression techniques to effectively develop a semi-empirical pipeline for SoH estimation [22, 23] We propose a reliable and comprehensive method to predict capacity-based SoH by extracting the parameters of the best-fitting ECM, which was determined via a comparative analysis. ECMs simulate the electrochemical response of the cell by modeling it using electrical components [25–28]. However, under certain dynamic conditions, ECMs may exhibit limitations due to their inability to account for the non-linear characteristics and non-idealities of the cells [29, 30].

Previous work has also included the use of single-particle model parameters, which can be uniquely determined and fit using frequency-domain data for reliable time-domain simulations [31]. Subsequent research has refined the parameterization of physics-based models across states of charge using grouped-variable approaches, streamlining interpretation while preserving electrochemical relevance. These refinements also incorporate double-layer capacitance into the models to replicate the semicircle observed in Nyquist plots [32]. Recently, hybrid manual-automatic feature extraction techniques and lightweight convolutional neural network (CNN) architectures have also been proposed to deliver fast, computationally efficient capacity estimation methods that remain robust across varying cell chemistries and measurement uncertainties [33, 34]. In addition, recent studies have advanced data-driven SoH estimation, including LSTM-based frameworks for degradation pattern recognition, physics-informed neural networks embedding incremental

capacity peak constraints, and hybrid CNN–TCN–LSTM fusion architectures optimized via Bayesian methods for multi-feature learning [35–37]. Finally, Giazitzis et al. introduce a TinyML framework for SoH estimation from EIS data, though it is limited to only five cells under fixed conditions without temperature variation and does not benchmark different ECM families [38]. In contrast, our work uses 22 cells aged under varied conditions, systematically compares six ECMs, and applies regression with correlation analysis to deliver model-specific interpretability of feature contributions to SoH.

In this work, three supervised regression models are individually trained and tested, using the identified parameters from the most accurate and efficient ECM as input features. Regression models consider complex processes and tolerate noisy data; in addition, many models exhibit high interpretability in providing key insights, such as feature importance [39–42]. Nonetheless, it is important to acknowledge that regression models alone may also be oversensitive to specific data, resulting in overfitting [43, 44]. Thus, the combination of ECMs and regression models provides an efficient and interpretable approach to SoH estimation.

The paper structure is as follows: the Experimental Data section discusses the EIS data used in this work; the Methods section details the parametrization and fitting of the different ECMs, the selection of the most parsimonious ECM (i.e., the model that best balances accuracy with simplicity), and the construction of the supervised regression models to estimate SoH; the Results section reports the best model architecture and fitted parameters; and the Conclusions section summarizes the novelty of the research and proposes potential applications and future directions of work.

2. Experimental Data

The experimental data used in this work is discussed rigorously in Khan et al [45]. EIS data was collected from an aging campaign performed on 5Ah NMC (nickel-manganese-cobalt 1-1-1)/graphite prismatic Li-ion batteries using a frequency response analyzer and potentiometer/galvanostat by KRI, Inc. The dataset consists of 22 cells labeled S1 through S24, with S2 and S18 excluded due to evaluation errors. Cells S1, S3, S4, S8, and S9 were calendar-aged and were neither charged nor discharged during the experimental campaign. The remaining 17 cells were cycle-aged. Calendar-aged versus cycle-aged cells did not exhibit statistically significantly different results.

3. Methods

In this work, we compare ECMs based on their ability to inform Li-ion cells' capacity-based SoH. As shown in Figure 1, the methods used to achieve this goal involve several steps. After the EIS data is collected and processed, various ECMs are fit to the data to extract electrochemical parameters. The most appropriate ECM is then selected based on fitting

accuracy and efficiency. Finally, data-driven SoH estimation models are built and compared using the extracted parameters, including feature attribution and error analysis.

The above pipeline leads to a framework to estimate SoH of Li-ion cells using frequency-domain data. This work makes three key contributions. First, this study conducts a direct and systematic evaluation of six ECMs across three distinct circuit families (Randles, ZARC, and Transmission Line) [22, 47]. Second, while prior work has leveraged deep learning architectures, such as CNN–BiLSTM (Bidirectional Long Short-Term Memory) or GPR (Gaussian Process Regression)–Elman networks, for SoH prediction using modified Randles or ZARC parameters, this study focuses on interpretable supervised regressors (decision tree, random forest, gradient boosting, and linear regression) to allow for feature attribution [22, 48]. Third, this work incorporates rigorous interpretability by applying SHapley Additive exPlanations (SHAP) analysis to quantify the contribution of each ECM-derived parameter to the SoH prediction, enabling feature-level insights often absent in prior studies. Together, these contributions establish a scalable and interpretable framework for SoH estimation.

3.1 Model Parameterization and Fitting

Six ECMs were fitted with the EIS data: two types of second-order Randles circuits (e.g. 1W Randles, 2W Randles), three ZARC circuits of different orders (e.g. 1 ZARC, 2 ZARC, 3 ZARC), and one second-order Transmission Line model. All three families of models have been explored independently in Li-ion batteries in past literature [50–52], but a direct comparative analysis between the three ECMs is missing. The circuit diagrams of the ECMs used in this study are shown in Figure 2.

Model	Equation
1W Randles	$Z(\omega)_{\text{1W Randles}} = j\omega \boldsymbol{L_0} + \boldsymbol{R_0} + \left(\frac{1}{\boldsymbol{R_1}} + j\omega \boldsymbol{C_1}\right)^{-1} + \left(\frac{1}{\boldsymbol{R_2}} + j\omega \boldsymbol{C_2}\right)^{-1} + \frac{\boldsymbol{W_0}}{\sqrt{\omega}}(1-j)$
2W Randles	$Z(\omega)_{\text{2W Randles}} = j\omega \boldsymbol{L_0} + \boldsymbol{R_0} + \left(\frac{1}{\boldsymbol{R_1}} + j\omega \boldsymbol{C_1}\right)^{-1} + \left(\frac{1}{\boldsymbol{R_2}+\boldsymbol{W_1}} + j\omega \boldsymbol{C_2}\right)^{-1} + \frac{\boldsymbol{W_0}}{\sqrt{\omega}}(1-j)$
1 ZARC	$Z(\omega)_{\text{1 ZARC}} = j\omega \boldsymbol{L_0} + \mathbf{R_0} + \frac{\boldsymbol{R_1}}{1+(j\omega \boldsymbol{\tau_1})^{\alpha_1}} + \frac{\boldsymbol{W_0}}{\sqrt{\omega}}(1-j)$
2 ZARC	$Z(\omega)_{\text{2 ZARC}} = j\omega \boldsymbol{L_0} + \boldsymbol{R_0} + \frac{\mathbf{R_1}}{1+(j\omega \boldsymbol{\tau_1})^{\alpha_1}} + \frac{\mathbf{R_2}}{1+(j\omega \boldsymbol{\tau_2})^{\alpha_2}} + \frac{\boldsymbol{W_0}}{\sqrt{\omega}}(1-j)$
3 ZARC	$Z(\omega)_{\text{3 ZARC}} = j\omega \boldsymbol{L_0} + \boldsymbol{R_0} + \frac{\mathbf{R_1}}{1+(j\omega \boldsymbol{\tau_1})^{\alpha_1}} + \frac{\mathbf{R_2}}{1+(j\omega \boldsymbol{\tau_2})^{\alpha_2}} + \frac{\mathbf{R_3}}{1+(j\omega \boldsymbol{\tau_3})^{\alpha_3}} + \frac{\boldsymbol{W_0}}{\sqrt{\omega}}(1-j)$
2 TL	$Z(\omega)_{\text{2 TL}} = j\omega \boldsymbol{L_0} + \boldsymbol{R_0} + \sqrt{\frac{\boldsymbol{R_{ion1}}}{\boldsymbol{Q_{s1}}(i\omega)^{\alpha_1}}} \coth(\sqrt{\boldsymbol{Q_{s1}}(i\omega)^{\alpha_1}\boldsymbol{R_{ion1}}}) + \sqrt{\frac{\boldsymbol{R_{ion2}}}{\boldsymbol{Q_{s2}}(i\omega)^{\alpha_2}}} \coth(\sqrt{\boldsymbol{Q_{s2}}(i\omega)^{\alpha_2}\boldsymbol{R_{ion2}}}) + \frac{\boldsymbol{W_0}}{\sqrt{\omega}}(1-j)$

Table 1: The equations of the six ECMs used in this study. The bolded variables denote ECM parameters that were optimized during fitting.

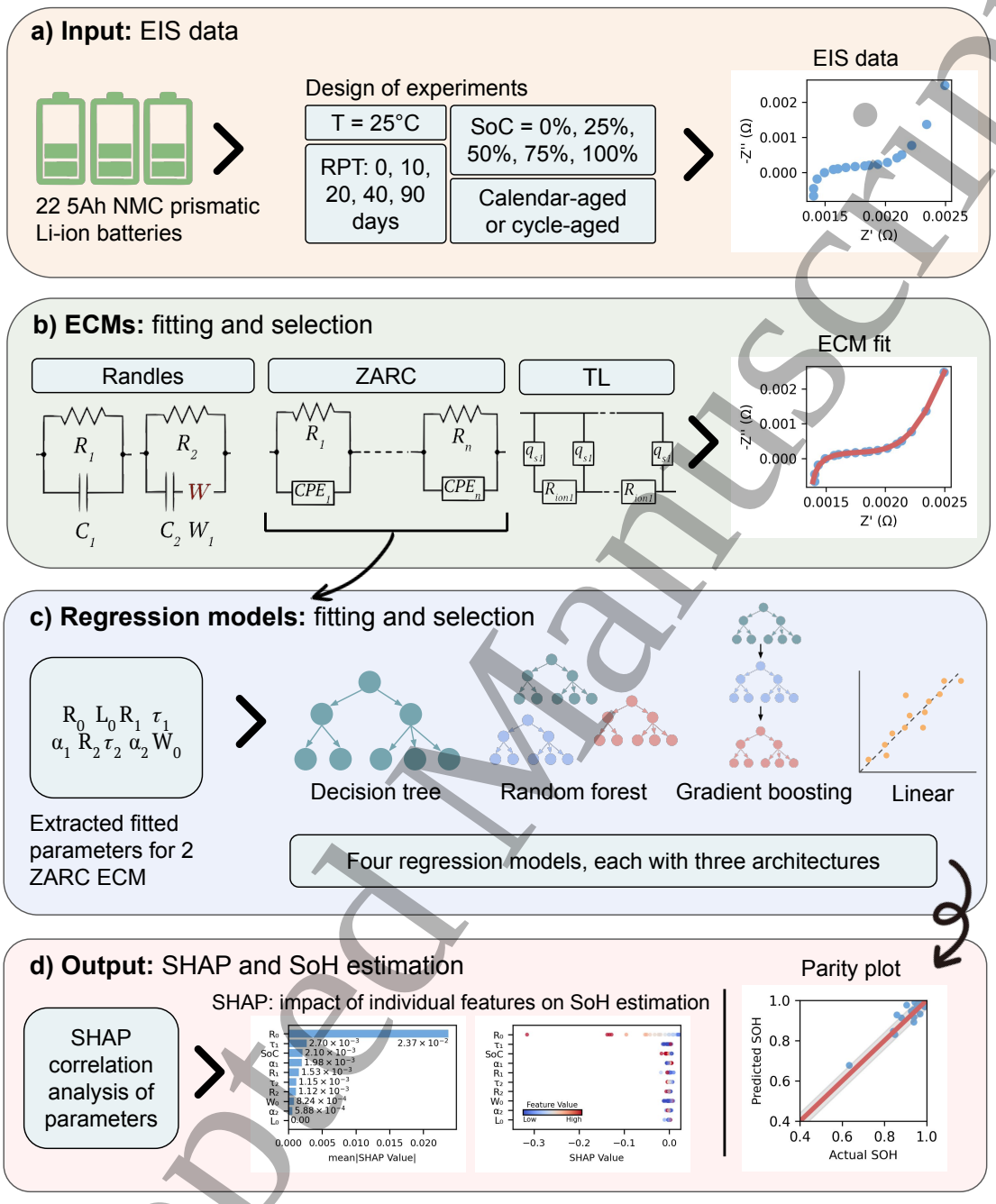

Figure 1: SoH estimation framework. (a) EIS data is collected from 22 NMC Li-ion cells under varied SoC and aging conditions, (b) multiple ECMs are fitted to extract electrochemical parameters, (c) regression models are trained on selected parameters to predict battery SoH, and (d) SHAP analysis quantifies feature importance, and model performance is evaluated.

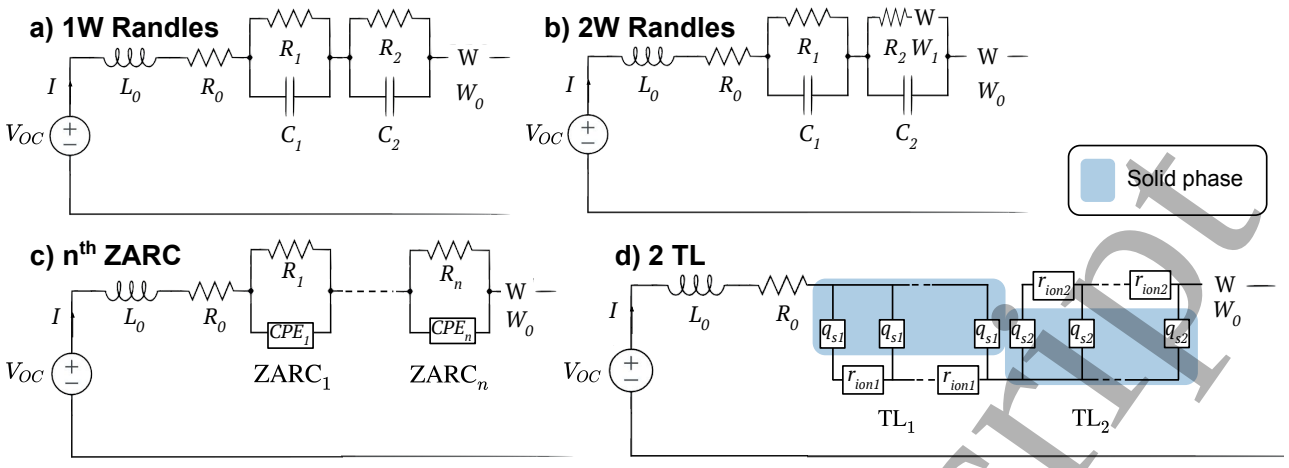

Figure 2: In this study, six different ECMs are used: 1W Randles, 2W Randles, 1 ZARC, 2 ZARC, 3 ZARC, and 2 TL. This figure includes the circuit diagrams of the a) 1W Randles, b) 2W Randles, c) n^{th} ZARC, and d) 2 TL ECMs.

Using ECM parameters fitted from EIS data is more advantageous than estimating SoH directly from raw EIS data. ECM fitting reduces the dimensionality of the problem by compressing the full impedance data into a set of physically interpretable features. This reduces the risk of overfitting. In addition, given the interpretability of ECM parameters, the resulting model is able to more clearly explain the underlying degradation mechanisms.

The Randles circuit is a widely recognized model that fits EIS data by simulating behaviors at the electrode interface [53,54]. This paper specifically evaluates a second-order Randles circuit with one Warburg component, referred to as the 1W Randles circuit, and a second-order Randles circuit with two Warburg components, referred to as the 2W Randles circuit. The 1W Randles circuit consists of the ohmic resistor, inductor, two resistor-capacitor (RC) components, and a Warburg element. Each part can be explained in terms of what electrochemical phenomenon it models and how it contributes to the EIS fit. The inductance element (L_0) models opposition to changes in current flow and appears as a vertical line in the positive imaginary region. The resistor (R_0) models resistance of the electrolyte, electrodes, and current collectors and appears as the EIS data's intercept with the real axis. The first resistor-capacitor element (R_1C_1) models the solid-electrolyte interface (SEI) while the second resistor-capacitor element (R_2C_2) models charge-transfer resistance and double-layer capacitance [55, 56]. Both RC elements appear as semi-circles. Finally, the Warburg element (W_0) models the diffusion process and mass-transfer limitations, appearing as a 45-degree tail at low frequencies. The 2W Randles circuit consists of an additional Warburg element in series with the resistor of the second RC element [57].

In contrast to the 1W Randles circuit, the 2W Randles configuration not only captures a semi-infinite diffusion process but also a limited or localized diffusion process as well [58]. The equations for the impedance of the two Randles circuits are shown in Table 1, where j is the imaginary unit $\sqrt{-1}$; ω is the angular frequency [rad/s]; L_0 is the inductance

$[H]$; R_0, R_1, and R_2 are resistors $[\Omega]$; C_1 and C_2 are capacitors $[F]$; and W_0 and W_1 are Warburg coefficients $[\Omega/s^{1/2}]$ [59, 60].

The second and more nuanced ECM examined is the ZARC circuit, which is widely used to represent the non-ideal behavior of the double-layer capacitance and charge transfer in Li-ion batteries [61]. The ZARC model is a physically informed extension of the classic parallel resistor-capacitor (RC) circuit, designed to capture real-world imperfections in interfacial processes at the electrode–electrolyte boundary. This circuit introduces a constant phase element (CPE), which is defined by two parameters. First, τ is the characteristic time constant, physically interpreted as the product of the charge-transfer resistance and a generalized capacitance. It determines the characteristic frequency range over which the interfacial process occurs. Second, α is the CPE exponent, bounded between 0 and 1. It quantifies the degree of deviation from ideal capacitive behavior: $\alpha = 1$ corresponds to an ideal capacitor with uniform surface and perfect time response while $\alpha = 0$ behaves like a pure resistor.

A ZARC element is a resistor in parallel with a CPE, and has the effect of depressing the semicircular shape that an RC element would produce in a Nyquist plot, as shown in Figure 3. In this paper, the first-order, second-order, and third-order ZARC circuits (referred to as 1 ZARC, 2 ZARC, and 3 ZARC, respectively) are evaluated. The equations for all three ZARC circuits are shown in Table 1.

The third ECM family is the Transmission Line (TL), which models the electrochemical behavior of electrolyte-filled porous electrodes [62–64]. In this work, we focus on the second-order TL model (2 TL), which consists of a repeating unit of two physically derived elements: q_s and R_{ion} [65, 66]. First, q_s models the charge transfer processes at the interface between the electrode's solid matrix and the liquid electrolyte. It typically behaves like a constant phase element (CPE), reflecting the distributed nature of electrochemical reactions across the porous surface. The CPE is characterized by Q_s, the charge defined as the sum of all q_s elements, and α, the exponent of the CPE, or the degree of non-ideality in the interfacial charge transfer. R_{ion}, which is defined as the sum of all r_{ion} elements, represents the resistance to movement within the liquid electrolyte occupying the pore volume. This is a purely resistive component and is directly related to the ionic conductivity and tortuosity of the electrolyte pathways. Higher R_{ion} indicates either longer or more convoluted ionic paths, or lower electrolyte conductivity [67, 68].

In the 2 TL structure, multiple R_{ion} and q_s units may be connected in series, mimicking a finite section of a porous electrode where each pair represents a small slice of the pore depth. This ladder-like structure captures the distributed nature of ion transport and interfacial reaction across the electrode depth. The resulting impedance equation for the 2 TL model, provided in Table 1, is derived from solving the equivalent circuit based on Ohm's law and the current-voltage relationships of CPEs, reflecting a physics-based approximation of ion diffusion and charge transfer in porous media [65].

The Python-based package `impedance.py` was utilized for the fitting of all six ECMs [69].

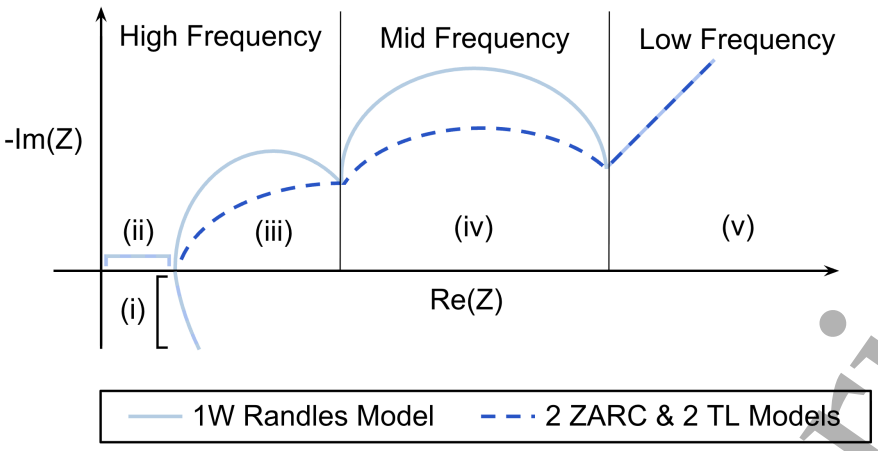

Figure 3: Nyquist plot demonstrating the theoretical impact of individual ECM components: (i) inductance element L_0; (ii) ohmic resistance element R_0; (iii) R_1-C_1 component for 1W Randles Model, $ZARC_1$ component for 2 ZARC and 2 TL Models; (iv) R_2-C_2 component for 1W Randles Model, $ZARC_2$ component for 2 ZARC and 2 TL Models; and (v) Warburg element W_0.

`impedance.py` is capable of fitting EIS data to custom circuits composed of up to 14 unique pre-defined elements. The package utilizes `curve_fit` from `scipy.optimize` function, which is a non-linear least squares regression fitting method. As no bounds were defined in these methods, the `curve_fit` function defaulted to using the Levenberg-Marquardt optimization algorithm to minimize the objective function, defined as:

$$\chi^2 = \sum_{n=1}^{N} [Z'_{EIS}(\omega_n) - Z'_{ECM}(\omega_n)]^2 + [Z''_{EIS}(\omega_n) - Z''_{ECM}(\omega_n)]^2, \quad (2)$$

where Z'_i is the real component of impedance and Z''_i is the imaginary component of impedance, where the i subscript distinguishes between the experimental EIS data and the ECM fit, respectively [70]. N is the number of frequencies at which the EIS measurements were taken; in the case of the data set used in this work, $N = 22$. For each parameter of the selected ECM, an initial guess was required. To ensure that the goodness-of-fit was as high as possible, for each cell, the EIS sample collected at Days $= 0$ and SoC $= 0\%$ was fit to a set of default initial guesses. For each subsequent EIS sample, the fitted parameter values obtained by the model for the initial EIS sample were used as initial guesses.

To effectively fit the EIS data to each ECM, all data in the positive real region (above the real axis) and the three highest points in the negative real region (below the real axis) were used.

3.2 Model Selection

The most parsimonious ECM was identified by calculating the root mean square error (RMSE) and information criteria values for each model. The RMSE was determined as

follows:

$$\mathrm{RMSE_{ECM}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left([Z'_{\mathrm{EIS}}(\omega_n) - Z'_{\mathrm{ECM}}(\omega_n;\boldsymbol{\Theta})]^2 + [Z''_{\mathrm{EIS}}(\omega_n) - Z''_{\mathrm{ECM}}(\omega_n;\boldsymbol{\Theta})]^2\right)} \quad (3)$$

where Θ is the set of parameters for each ECM (with k as the number of parameters):

$$\begin{bmatrix} \Theta_{\text{1W Randles}} &= [L_0\ R_0\ R_1\ C_1\ R_2\ C_2\ W_0] & k=7 \\ \Theta_{\text{2W Randles}} &= [L_0\ R_0\ R_1\ C_1\ R_2\ C_2\ W_0\ W_1] & k=8 \\ \Theta_{\text{1 ZARC}} &= [L_0\ R_0\ R_1\ \tau_1\ \alpha_1\ W_0] & k=6 \\ \Theta_{\text{2 ZARC}} &= [L_0\ R_0\ R_1\ \tau_1\ \alpha_1\ R_2\ \tau_2\ \alpha_2\ W_0] & k=9 \\ \Theta_{\text{3 ZARC}} &= [L_0\ R_0\ R_1\ \tau_1\ \alpha_1\ R_2\ \tau_2\ \alpha_2\ R_3\ \tau_3\ \alpha_3\ W_0] & k=12 \\ \Theta_{\text{2 TL}} &= [L_0\ R_0\ R_{\text{ion1}}\ Q_{s1}\ \alpha_1\ R_{\text{ion2}}\ Q_{s2}\ \alpha_2\ W_0] & k=9 \end{bmatrix}$$

.

Information criteria were also calculated to estimate the efficiency of each model by maximizing model fit and minimizing model complexity. Two types of information criteria were used: Akaike information criterion (AIC) and Bayesian information criterion (BIC). AIC uses a fixed penalty while BIC uses a penalty that increases logarithmically with sample size N. As a result, both information criteria favor the most parsimonious model; however, BIC favors simpler models more strongly than AIC. For both information criteria, a lower value indicates a more parsimonious model. The methods of calculation for each information criterion [71, 72] are defined as:

$$AIC_{\mathrm{ECM}} = 2k + N\ln(\mathrm{RMSE}^2_{\mathrm{ECM}}) \quad (4)$$

$$BIC_{\mathrm{ECM}} = k\ln(2k+N) + 2k + N\ln(\mathrm{RMSE}^2_{\mathrm{ECM}}). \quad (5)$$

3.3 SoH Estimation

While purely data-driven SoH estimation methods, such as GPR, support vector machines (SVMs), and CNNs, can achieve high accuracy, they often lack physical interpretability and may overfit limited datasets. On the other hand, integrating physics-based models with machine learning may constrain predictions to electrochemically explainable relationships. Our hybrid approach combines the predictive strength of machine learning with the interpretability of physics-informed features. Therefore, in our work, the estimation of SoH is carried out using the parameters of the most parsimonious model as input features for four supervised models: decision tree, random forest, gradient boosting, and linear regression. For each supervised regression model, three different model architectures were tested, as defined in Figure 4. Architecture 1 attempts to use all parameters of the most parsimonious model and SoC values as the input features. Architecture 2 attempts to use the single parameter that trends most with SoH (as determined by the SHAP analysis, explained in detail below) as the only input feature to determine if it is sufficient for

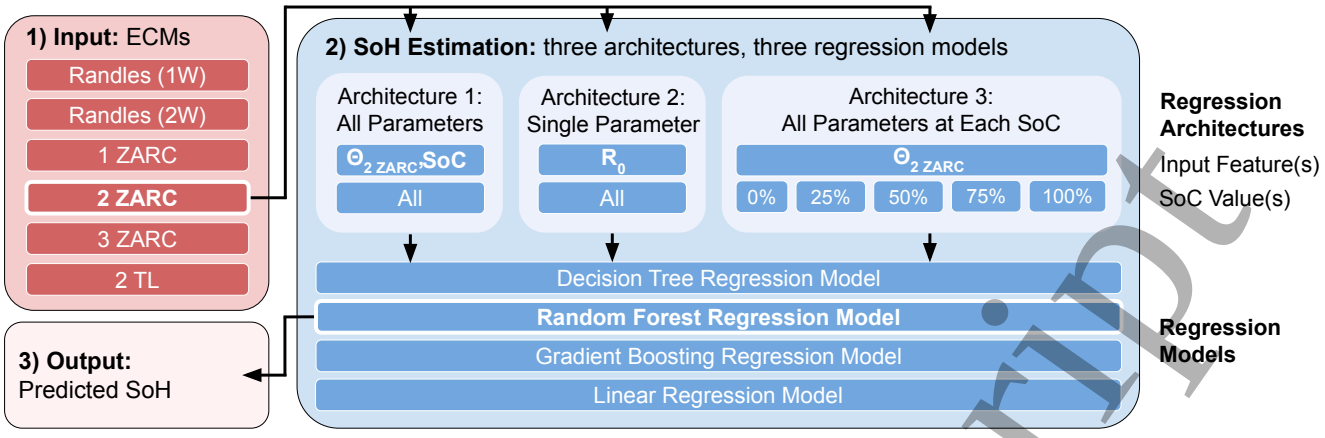

Figure 4: The three model architectures tested in the context of the entire model.

accurate SoH estimation. Finally, Architecture 3 attempts six different models, each using all parameters of the most parsimonious ECM as input features, at each SoC to determine if SoH estimation is more or less accurate at any single SoC.

The decision tree model was performed using `DecisionTreeRegressor` from the `sklearn.tree` module while the random forest and gradient boosting models used `RandomForestRegressor` and `GradientBoostingRegressor`, respectively, from the `sklearn.ensemble` module. The linear regression model was performed using `LinearRegression` from the `sklearn.linear_model` module. Each of the four regression models were evaluated using three different train/test splits: 70%/30%, 80%/20%, and 90%/10% of the entire dataset. The regression models were also evaluated at 21 random states, from 30 to 50, meaning that the models were trained and tested using initial random seeds within this range to assess performance consistency across varying random initializations. A random state controls the random subsampling of data for each estimator, or tree, as well as the random selection of features for splitting. In addition, the random forest and gradient boosting regression models were evaluated with 100 estimators.

For each regression model, the average RMSE was calculated across all train/test splits and random states. The RMSE of each regression model is defined as:

$$\mathrm{RMSE}_{\mathrm{Regression}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(SoH_{\mathrm{EIS}} - SoH_{2\ \mathrm{ZARC}})^2}. \tag{6}$$

After the regression model with the lowest RMSE was selected, a follow-up SHapley Additive exPlanations (SHAP) multicollinearity correlation analysis was performed. SHAP is used to determine which features — including SoC and the parameters of the selected ECM — are most important in SoH estimation. The SHAP-explained impact of a feature is defined as:

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!}[f_x(z') - f_x(z' \setminus i)], \tag{7}$$

where $\phi_i(f, x)$ is the SHAP value for feature i, f is the prediction model (e.g. decision tree, random forest, gradient boosting, or linear regression model), x is the input instance being explained (vector of feature values), $x' \in \{0, 1\}^M$ is a binary vector indicating which features are included, $z' \subseteq x'$ is a subset of included features, M is the total number of input features, $|z'|$ is the number of features in the subset, $f_x(z')$ is the model output when only the features in z' are known, and $f_x(z' \setminus i)$ is the model output when feature i is removed from the subset [73]. A higher mean absolute SHAP value indicates a stronger correlation with SoH.

As an example of the SHAP analysis, suppose our SoH regressor uses just two features: R_0 and SoC. For one EIS sample, the model's outputs for each subset of features are: with no features SoH = 0.82, with R_0 only SoH = 0.78, with SoC only SoH = 0.88, and with both SoH = 0.84. Since $M = 2$, the SHAP weight on each marginal contribution is $\frac{1}{2}$. Thus,

$$\phi_{R_0} = \frac{1}{2}(0.78 - 0.82) + \frac{1}{2}(0.84 - 0.88) \tag{8}$$

and

$$\phi_{SoC} = \frac{1}{2}(0.88 - 0.82) + \frac{1}{2}(0.84 - 0.78). \tag{9}$$

For each of the models used in this section, SHAP values are calculated for each of the respective input features with respect to SoH to determine the most impactful of these features. Architecture 2 in the following section uses only the most impactful feature, that is, the feature with the highest SHAP value, as the sole input.

4. Results

4.1 Assessment of ECMs

The EIS data from all 22 cells at different ages and SoCs were fit to the six ECMs. The RMSEs and individual parameter values from each fit were then extracted and averaged for each ECM. Table 2 shows the average value and standard deviation for each parameter. The best and worst fits for each ECM are shown in Figure 5. This work finds that the 2 ZARC model was the most accurate in fitting the data, as demonstrated by both the RMSE and information criterion evaluation results.

Parameter	Average Value	Average Standard Deviation
L_0	5.7988×10^{-8} $[H]$	2.03×10^{-8}
R_0	1.3525×10^{-3} $[\Omega]$	2.37×10^{-4}
R_1	3.9600×10^{-4} $[\Omega]$	5.75×10^{-4}
τ_1	1.3900×10^{-2} $[s]$	2.05×10^{-1}
α_1	7.9419×10^{-1}	2.14×10^{-1}
R_2	6.8001×10^{-1} $[\Omega]$	1.29×10^{1}
τ_2	4.1100×10^{4} $[s]$	6.92×10^{5}
α_2	8.7071×10^{-1}	1.27×10^{-1}
W_0	1.6603×10^{-4} $[\Omega/s^{1/2}]$	1.95×10^{-4}

Table 2: Average values

of extracted parameter and standard deviation values from 2 ZARC ECM fits of all EIS samples.

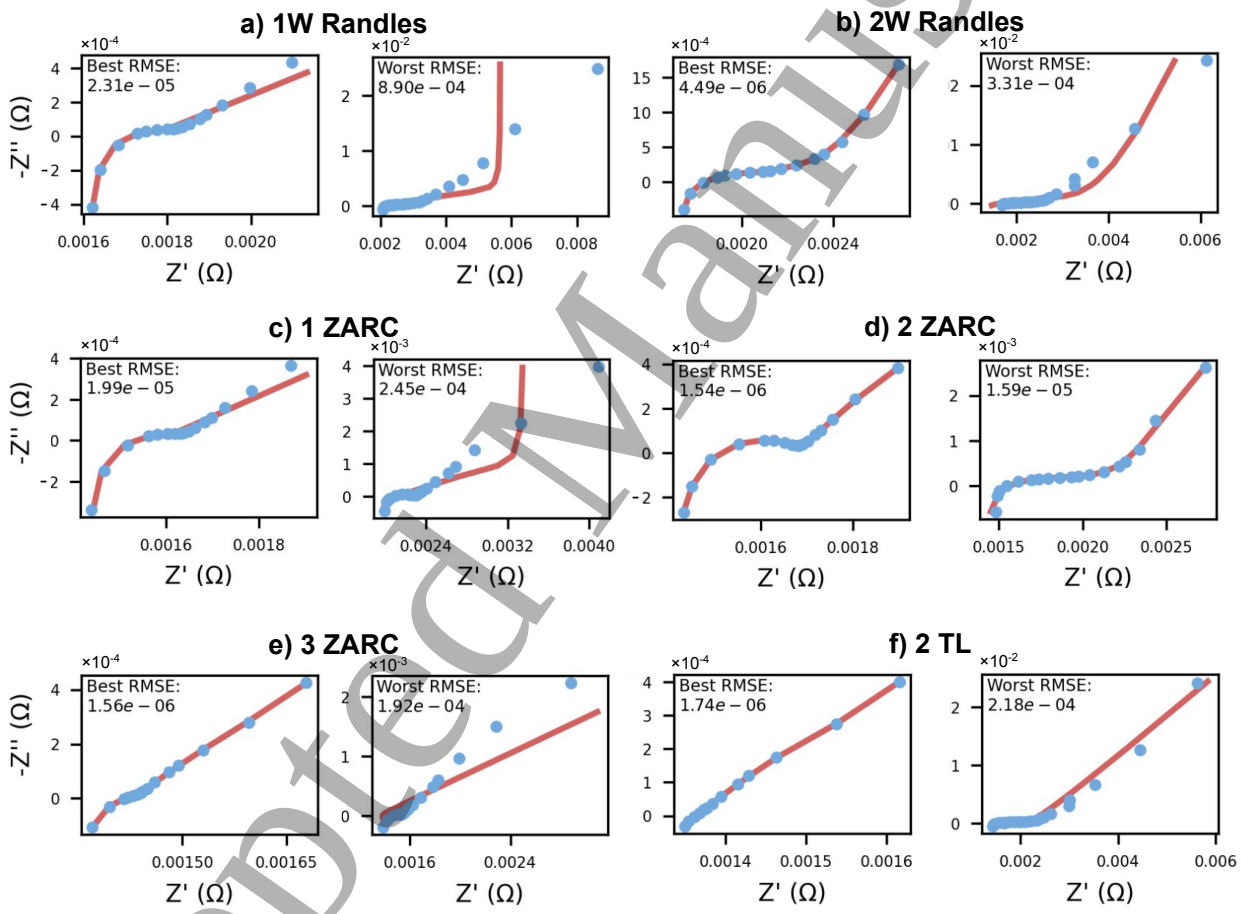

Figure 5: The best fit, corresponding to the lowest RMSE, and worst fit, corresponding to the highest RMSE, across all cells for each ECM: a) 1W Randles, b) 2W Randles, c) 1 ZARC, d) 2 ZARC, e) 3 ZARC, and f) 2 TL.

First, the average RMSE of the 2 ZARC circuit was found to be significantly lower than that of the 1W Randles, 2W Randles, 1 ZARC, and 2 TL circuits, and slightly lower than that of the 3 ZARC circuit, as shown in Figure 6a. The median RMSE of each ECM also presents similar trends, demonstrating that the performance of the 2 ZARC

circuit is consistent across most cells and not driven by outliers. This indicates that the ZARC element successfully captured the non-ideal behavior of the capacitor in both RC components. This improvement stems from the inclusion of a CPE in the ZARC, which generalizes ideal capacitive behavior by introducing a fractional-order impedance (α); unlike a standard RC element, the CPE can model the distributed and non-ideal electrochemical processes that arise from surface roughness, porosity, and inhomogeneities within the electrode [74]. In addition, because the ZARC component aimed to specifically model the behavior of the double-layer capacitance within a cell, which is represented by the second ZARC element, the 2 ZARC circuit further performed better than the 1 ZARC circuit.

Second, the 2 ZARC circuit was found to have the most optimal balance of accuracy and efficiency. Both the AIC and BIC tests found that the 2 ZARC had the lowest value among all six ECMs, indicating that it was the most parsimonious, as shown in Figure 6a. Similar to the RMSE results, the 1W Randles, 2W Randles, 1 ZARC, and 2 TL circuits had significantly higher AIC and BIC values than both the 2 ZARC and 3 ZARC circuits. For real-world conditions, this means that the 2 ZARC ECM would likely require comparatively less computational time and energy to simulate data in general.

Across the five different SoCs, the average RMSE of each ECM's fit was relatively equal, except for SoC = 0%, which had a significantly greater average RMSE. This can be attributed to the Warburg tail component at SOC = 0%, which is much harder to fit because it is not precisely 45 degrees from the x-axis. An example of this is the best fit in Figure 5.

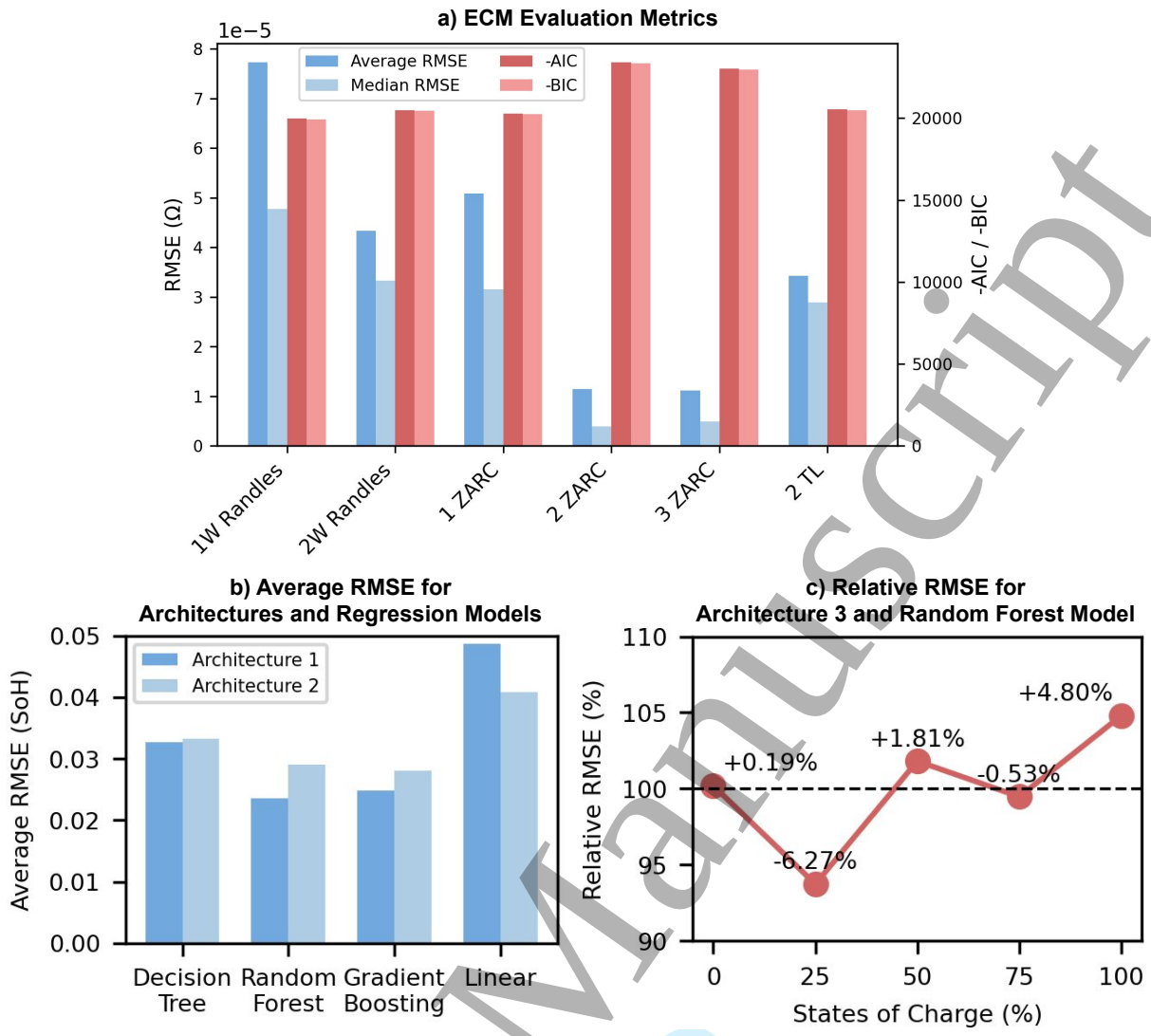

Figure 6: Across all cells, (a) average RMSE, median RMSE, -AIC, and -BIC comparison across ECMs, (where lower AIC and BIC values, shown as taller columns, indicate better model fit penalized by complexity), (b) the average RMSEs of Architectures 1 and 2 with decision tree, random forest, gradient boosting, and linear regression models, and (c) relative RMSEs of Architecture 3 with random forest regression model, where 100% relative RMSE indicates the average RMSE across all SoCs.

4.2 Evaluation of Regression Models

Next, the decision tree, random forest, and gradient boosting, and linear regression models were each executed under three distinct model architectures.

The regression models were first performed under Architecture 1, which used all of the parameters extracted from the 2 ZARC model as well as SoH, as input features. The results from Architecture 1, as shown in Figure 6b, suggested that the random forest model estimated SoH best. This specific architecture-model pipeline achieved an average RMSE of 0.0236 (where SoH is normalized to range from 0 to 1), a 96.4% accuracy within a 95% confidence interval, and a 0.8762 R^2 value (indicating that 87.62% of the variance in the actual SoH values is explained by the model's estimation) for the best train/test/random

state split of 90% train/0.05% test data and random state 42 (Figure 7a). This is consistent with the fact that the random forest model is generally more robust to noise and less prone to overfitting, as it averages estimations across several trees. The density plot (Figure 7b) of RMSE across different train/test splits indicates that the greater training proportions tend to result in relatively lower RMSE distributions than smaller training proportions.

Work utilizing other approaches on the same data has been performed [46]. While our work achieves an average RMSE of 2.36% for the hybrid ECM-Random Forest model and 4.86% for the hybrid ECM-Linear Regression model, an alternative approach using distribution of relaxation times (DRT) achieves an average RMSE of 1.69% for the hybrid DRT-Long Short-Term Memory model and 4.89% for the hybrid DRT-Linear Regression model.

While the random forest model is typically less interpretable than a single decision tree, the SHAP feature importance analysis provided a critical method for understanding the impact that individual features may have on the accuracy of SoH estimation. The SHAP feature importance analysis found that the ohmic resistance feature (R_0) is most significant in SoH estimation (Figure 7c) and indeed had a negative correlation with SoH (Figure 7d). A higher R_0 may indicate greater degradation, such as electrode cracking. These aging effects reduce ionic and electronic conductivity within the cell, resulting in a lower available capacity [75]. Following R_0, the τ_1 component showed a weaker, yet still valuable, feature importance with SoH. This is consistent with the fact that slower speeds of electrochemical reactions may represent slower charge transfer kinetics and increased diffusion resistance. A thicker insulating layer induces rapid performance degradation and longer ion diffusion times, resulting in a lower SoH [76].

To evaluate whether R_0 alone could predict SoH, Architecture 2 tested models using only this feature. The random forest model performed worse, with an average RMSE of 0.0290 (Figure 6b). Similar results were observed across the other three models, where Architecture 2 generally underperformed relative to Architecture 1. The only exception was linear regression, which improved slightly, likely because excluding less relevant features reduced sensitivity to multicollinearity. Overall, these results indicate that R_0 by itself is insufficient for accurate SoH estimation, and additional features are necessary.

Finally, Architecture 3 was performed to evaluate the accuracy of the random forest model on SoC-specific data. The data was separated by SoC into five testing sets. All parameters of the 2 ZARC model were used as features. The results of Architecture 3 suggest that the random forest model performs relatively similarly across all SoCs (Figure 6c). While the model at SoC = 25% did exhibit the lowest RMSE, its RMSE was only 6.27% less than that of the average RMSE. The model at SoC = 100% exhibited the highest RMSE, which was 4.80% greater than the average RMSE.

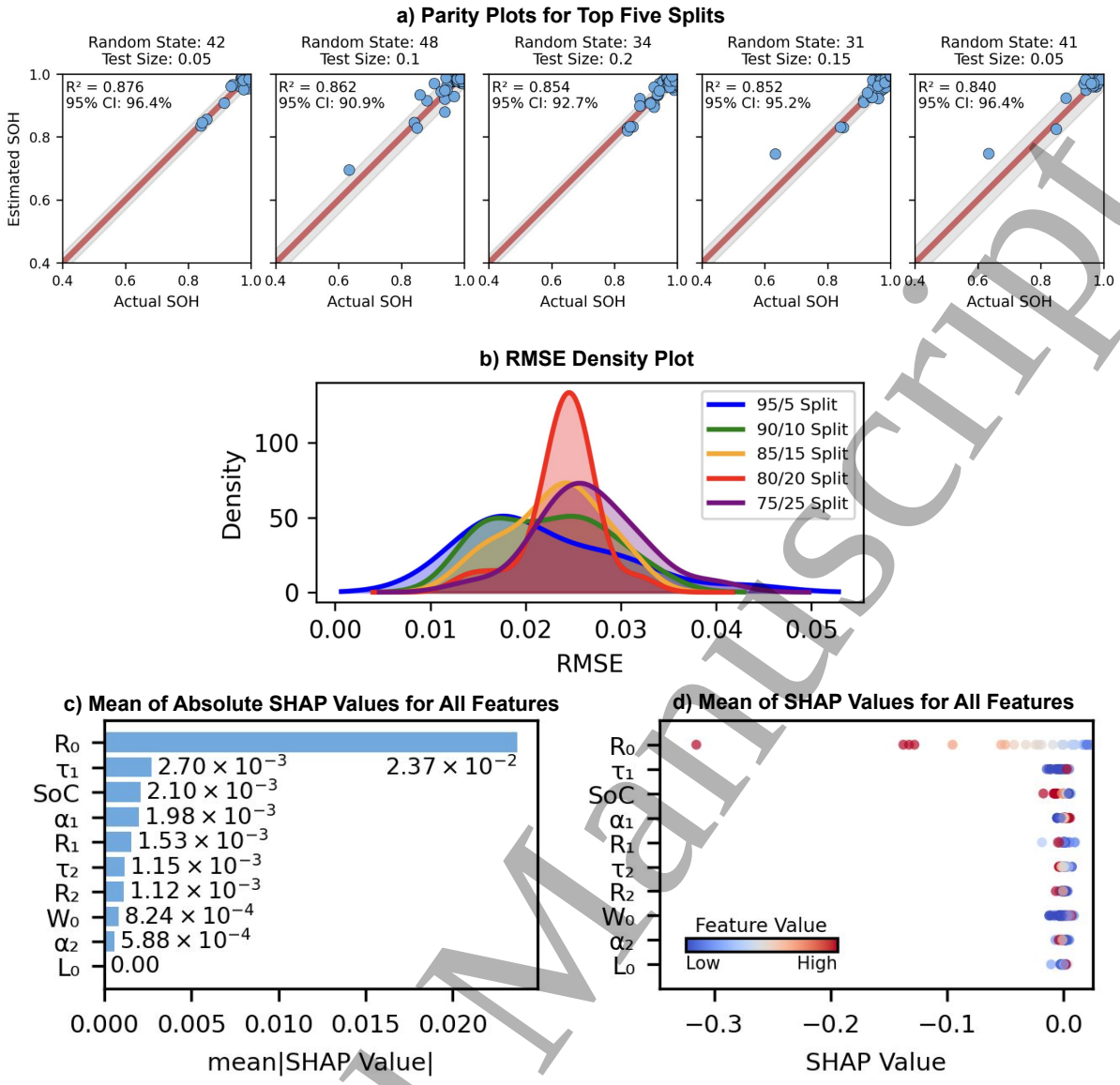

Figure 7: For the random forest regression model under Architecture 1: a) parity plots for the top five splits (highest R^2 across all combinations of random states 30–50 and select test sizes 0.05, 0.10, 0.15, 0.20, and 0.25), where "95% CI" indicates the percentage of predictions falling within 95% confidence interval (gray area), b) density plot of RMSE for the selected test sizes, each evaluated 100 times, c) absolute mean SHAP values for all features, and d) mean SHAP values colored by feature value.

5. Conclusions

This study provides the groundwork for a comprehensive and explainable estimation tool that fits EIS data to multiple ECMs, uses comparative analysis to determine the most accurate model, and ultimately uses the parameters from this model to estimate SoH. The results notably show that the ohmic resistance of a cell's electrolyte, electrodes, and current collectors is a critical factor in estimating its health. However, the inclusion of additional parameters is also vital for accurate estimation.

Furthermore, two regression model architectures, one that includes all parameters of

the most accurate and efficient ECM and one that includes the ohmic resistance element, prove that the latter is not sufficient for accurate battery health estimation. A third model architecture finds that SoH estimation is relatively consistent across all states of charge of the battery.

Although our work is validated on 5 Ah NMC/graphite cells, the proposed framework, which includes SHAP-based feature importance analysis, can be applied to other Li-ion chemistries. Because cells of different chemistries are all characterized by degradation mechanisms such as SEI growth, loss of active material, and electrolyte decomposition, the application of the model framework will likely remain similar, even though the specific values and types of extracted ECM parameters may differ due to material properties.

Finally, limitations include that EIS is not widely feasible for continuous monitoring in commercial BMSs due to hardware limitations and measurement constraints. Future work could focus on current-voltage pulse responses or drive-cycle data, which are more readily available onboard. The deployment of on-board EIS technologies for diagnostics would also benefit from the pipeline presented in this work [77].

6. Acknowledgements

Data was provided by Nuvoton Technology Corporation [45]. In addition, financial support was provided from Nuvoton Technology Corporation through Stanford SystemX Alliance. The research was conducted through the Stanford Young Investigators internship program.

References

[1] Sotomayor, M. E., de La Torre-Gamarra, C., Levenfeld, B., Sanchez, J. Y., Varez, A., Kim, G. T., ... & Passerini, S. (2019). Ultra-thick battery electrodes for high gravimetric and volumetric energy density Li-ion batteries. *Journal of Power Sources*, *437*, 226923.

[2] Son, Y., Cha, H., Jo, C., Groombridge, A. S., Lee, T., Boies, A., ... & De Volder, M. (2021). Reliable protocols for calculating the specific energy and energy density of Li-Ion batteries. *Materials Today Energy*, *21*, 100838.

[3] McCloskey, B. D. (2015). Attainable gravimetric and volumetric energy density of Li–S and Li ion battery cells with solid separator-protected Li metal anodes. *The Journal of Physical Chemistry Letters*, *6*(22), 4581-4588.

[4] Miao, Y., Hynan, P., Von Jouanne, A., & Yokochi, A. (2019). Current Li-ion battery technologies in electric vehicles and opportunities for advancements. *Energies*, *12*(6), 1074.

[5] Jaguemont, J., Boulon, L., & Dubé, Y. (2016). A comprehensive review of lithium-ion batteries used in hybrid and electric vehicles at cold temperatures. *Applied Energy*, *164*, 99-114.

[6] Duru, K. K., Karra, C., Venkatachalam, P., Betha, S. A., Madhavan, A. A., & Kalluri, S. (2021). Critical insights into fast charging techniques for lithium-ion batteries in electric vehicles. *IEEE Transactions on Device and Materials Reliability*, *21*(1), 137-152.

[7] Al-Saadi, M., Olmos, J., Saez-de-Ibarra, A., Van Mierlo, J., & Berecibar, M. (2022). Fast charging impact on the lithium-ion batteries' lifetime and cost-effective battery sizing in heavy-duty electric vehicles applications. *Energies*, *15*(4), 1278.

[8] Li, Y., Guo, W., Stroe, D. I., Zhao, H., Kristensen, P. K., Jensen, L. R., ... & Gurevich, L. (2024). Evolution of aging mechanisms and performance degradation of lithium-ion battery from moderate to severe capacity loss scenarios. *Chemical Engineering Journal*, *498*, 155588.

[9] Takahashi, K., & Srinivasan, V. (2015). Examination of graphite particle cracking as a failure mode in lithium-ion batteries: a model-experimental study. *Journal of The Electrochemical Society*, *162*(4), A635.

[10] Lin, X., Khosravinia, K., Hu, X., Li, J., & Lu, W. (2021). Lithium plating mechanism, detection, and mitigation in lithium-ion batteries. *Progress in Energy and Combustion Science*, *87*, 100953.

[11] Huang, D., Engtrakul, C., Nanayakkara, S., Mulder, D. W., Han, S. D., Zhou, M., ... & Tenent, R. C. (2021). Understanding degradation at the lithium-ion battery cathode/electrolyte interface: connecting transition-metal dissolution mechanisms to electrolyte composition. *ACS Applied Materials & Interfaces*, *13*(10), 11930-11939.

[12] Wang, Y., Tian, J., Sun, Z., Wang, L., Xu, R., Li, M., & Chen, Z. (2020). A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. *Renewable and Sustainable Energy Reviews*, *131*, 110015.

[13] Park, S., Ahn, J., Kang, T., Park, S., Kim, Y., Cho, I., & Kim, J. (2020). Review of state-of-the-art battery state estimation technologies for battery management systems of stationary energy storage systems. *Journal of Power Electronics*, *20*, 1526-1540.

[14] Demirci, O., Taskin, S., Schaltz, E., & Demirci, B. A. (2024). Review of battery state estimation methods for electric vehicles-Part II: SOH estimation. *Journal of Energy Storage*, *96*, 112703.

[15] Ha, S., Pozzato, G., & Onori, S. (2024). Electrochemical characterization tools for lithium-ion batteries. *Journal of Solid State Electrochemistry*, *28*(3), 1131-1157.

[16] Xiong, R., Li, L., & Tian, J. (2018). Towards a smarter battery management system: A critical review on battery state of health monitoring methods. *Journal of Power Sources*, *405*, 18-29.

[17] Yang, B., Qian, Y., Li, Q., Chen, Q., Wu, J., Luo, E., ... & Wang, J. (2024). Critical summary and perspectives on state-of-health of lithium-ion battery. *Renewable and Sustainable Energy Reviews*, *190*, 114077.

[18] Zhai, N. S., Li, M. W., Wang, W. L., Zhang, D. L., & Xu, D. G. (2006, October). The application of the EIS in Li-ion batteries measurement. In *Journal of Physics: Conference Series* (Vol. 48, No. *1*, p. 1157). IOP Publishing.

[19] Westerhoff, U., Kroker, T., Kurbach, K., & Kurrat, M. (2016). Electrochemical impedance spectroscopy based estimation of the state of charge of lithium-ion batteries. *Journal of Energy Storage*, *8*, 244-256.

[20] Zhang, X., Zhang, X., Sun, X., An, Y., Song, S., Li, C., ... & Ma, Y. (2021). Electrochemical impedance spectroscopy study of lithium-ion capacitors: Modeling and capacity fading mechanism. *Journal of Power Sources*, *488*, 229454.

[21] Meddings, N., Heinrich, M., Overney, F., Lee, J. S., Ruiz, V., Napolitano, E., ... & Park, J. (2020). Application of electrochemical impedance spectroscopy to commercial Li-ion cells: A review. *Journal of Power Sources*, *480*, 228742.

[22] Li, C., Yang, L., Li, Q., Zhang, Q., Zhou, Z., Meng, Y., ... & Lv, F. (2024). SOH estimation method for lithium-ion batteries based on an improved equivalent circuit model via electrochemical impedance spectroscopy. *Journal of Energy Storage*, *86*, 111167.

[23] Rodriguez-Cea, A. I., Morinigo-Sotelo, D., & Tinaut, F. V. (2025). A procedure for evaluating the SOH of Li-ion batteries from data during the constant voltage charge phase and the use of an ECM with internal resistance. *Journal of Energy Storage*, *108*, 115074.

[24] Choi, W., Shin, H. C., Kim, J. M., Choi, J. Y., & Yoon, W. S. (2020). Modeling and applications of electrochemical impedance spectroscopy (EIS) for lithium-ion batteries. *Journal of Electrochemical Science and Technology*, *11*(1), 1-13.

[25] Hu, X., Li, S., & Peng, H. (2012). A comparative study of equivalent circuit models for Li-ion batteries. *Journal of Power Sources*, *198*, 359-367.

[26] Zhang, L., Peng, H., Ning, Z., Mu, Z., & Sun, C. (2017). Comparative research on RC equivalent circuit models for lithium-ion batteries of electric vehicles. *Applied Sciences*, *7*(10), 1002.

[27] Li, Y., Vilathgamuwa, M., Farrell, T., Tran, N. T., & Teague, J. (2019). A physics-based distributed-parameter equivalent circuit model for lithium-ion batteries. *Electrochimica Acta*, *299*, 451-469.

[28] Madani, S. S., Schaltz, E., & Kær, S. K. (2018). A review of different electric equivalent circuit models and parameter identification methods of lithium-ion batteries. *ECS Transactions*, *87*(1), 23.

[29] You, H. W., Bae, J. I., Cho, S. J., Lee, J. M., & Kim, S. H. (2018). Analysis of equivalent circuit models in lithium-ion batteries. *AIP Advances*, *8*(12).

[30] Tao, Z., Zhao, Z., Wang, C., Huang, L., Jie, H., Li, H., ... & See, K. Y. (2024). State of charge estimation of lithium Batteries: Review for equivalent circuit model methods. *Measurement*, 115148.

[31] Bizeray, A. M., Kim, J. H., Duncan, S. R., & Howey, D. A. (2018). Identifiability and parameter estimation of the single particle lithium-ion battery model. *IEEE Transactions on Control Systems Technology*, *27*(5), 1862-1877.

[32] Hallemans, N., Courtier, N. E., Please, C. P., Planden, B., Dhoot, R., Timms, R., ... & Duncan, S. (2025). Physics-based battery model parametrisation from impedance data. *Journal of The Electrochemical Society*.

[33] Ning, Z., Venugopal, P., Soeiro, T. B., & Rietveld, G. (2024). Computation-light AI models for Robust Battery Capacity Estimation based on Electrochemical Impedance Spectroscopy. *IEEE Transactions on Transportation Electrification*.

[34] Li, Y., Ye, M., Wang, Q., Lian, G., Kemény, M., Xia, B., & Zhang, B. (2025). Capacity estimation of lithium-ion battery through interpretation of electrochemical impedance spectroscopy combined with machine learning. *Measurement*, *243*, 116374.

[35] Li, J., Deng, Z., Che, Y., Xie, Y., Hu, X., & Teodorescu, R. (2023). Degradation pattern recognition and features extrapolation for battery capacity trajectory prediction. *IEEE Transactions on Transportation Electrification*, *10*(3), 7565-7579.

[36] Sun, G., Liu, Y., & Liu, X. (2025). A method for estimating lithium-ion battery state of health based on physics-informed machine learning. *Journal of Power Sources*, *627*, 235767.

[37] Sonthalia, A., Josephin, J. F., Varuvel, E. G., Chinnathambi, A., Subramanian, T., & Kiani, F. (2025). A deep learning multi-feature based fusion model for predicting the state of health of lithium-ion batteries. *Energy*, *317*, 134569.

[38] Giazitzis, S., Isse, A. A., Blasuttigh, N., Pavan, A. M., Raimondo, D. M., Badha, S., ... & Ogliari, E. (2025). TinyML models for SoH estimation of lithium-ion batteries based on Electrochemical Impedance Spectroscopy. *Journal of Power Sources*, *653*, 237568.

[39] Chen, Z., Sun, M., Shu, X., Shen, J., & Xiao, R. (2018, February). On-board state of health estimation for lithium-ion batteries based on random forest. In 2018 IEEE International Conference on Industrial Technology (ICIT) (pp. 1754-1759). *IEEE*.

[40] Yang, N., Song, Z., Hofmann, H., & Sun, J. (2022). Robust State of Health estimation of lithium-ion batteries using convolutional neural network and random forest. *Journal of Energy Storage*, *48*, 103857.

[41] Lu, J., Xiong, R., Tian, J., Wang, C., & Sun, F. (2023). Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nature Communications*, *14*(1), 2760.

[42] Wang, G., Lyu, Z., & Li, X. (2023). An optimized random forest regression model for li-ion battery prognostics and health management. *Batteries*, *9*(6), 332.

[43] Lu, D., Cui, N., & Li, C. (2024). A Novel Transfer Learning Framework Combining Attention Mechanisms and Random Forest Regression for State of Health Estimation of Lithium-Ion Battery With Different Formulations. *IEEE Transactions on Industry Applications*.

[44] Marri, I., Petkovski, E., Cristaldi, L., & Faifer, M. (2023). Comparing Machine Learning Strategies for SoH Estimation of Lithium-Ion Batteries Using a Feature-Based Approach. *Energies*, *16*(11), 4423.

[45] Khan, M. A., Chu, C., & Onori, S. (2025). Calendar and cycle aging dataset for lithium-ion batteries with multi-SOC electrochemical impedance spectroscopy measurements. *Data in Brief*, under review.

[46] Khan, M. A., Thatipamula, S., & Onori, S. (2024). Onboard Health Estimation using Distribution of Relaxation Times for Lithium-ion Batteries. *IFAC-PapersOnLine*, *58*(28), 917-922.

[47] Nunes, H., Martinho, J., Fermeiro, J., Pombo, J., Mariano, S., & do Rosário Calado, M. (2024). Impedance analysis and parameter estimation of lithium-ion batteries using the eis technique. *IEEE Transactions on Industry Applications*.

[48] Li, D., Yang, D., Li, L., Wang, L., & Wang, K. (2022). Electrochemical impedance spectroscopy based on the state of health estimation for lithium-ion batteries. *Energies*, *15*(18), 6665.

[49] Amuta, O., & Kowal, J. (2023). State of Health Assessment of Spent Lithium–Ion Batteries Based on Voltage Integral during the Constant Current Charge. *Batteries*, *9*(11), 537.

[50] Hasan, R., & Scott, J. (2020). Extending randles's battery model to predict impedance, charge–voltage, and runtime characteristics. *IEEE Access*, *8*, 85321-85328.

[51] Farmann, A., Waag, W., & Sauer, D. U. (2015). Adaptive approach for on-board impedance parameters and voltage estimation of lithium-ion batteries in electric vehicles. *Journal of Power Sources*, *299*, 176-188.

[52] Moškon, J., Žuntar, J., Talian, S. D., Dominko, R., & Gaberšček, M. (2020). A powerful transmission line model for analysis of impedance of insertion battery cells: a case study on the NMC-Li system. *Journal of The Electrochemical Society*, *167*(14), 140539.

[53] Keshavarzi, M., Derakhshan, M., Gilaki, M., L'Eplattenier, P., Caldichoury, I., Soudbakhsh, D., & Sahraei, E. (2021, December). Coupled electrochemical-mechanical modeling of lithium-ion batteries using distributed randle circuit model. In 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). *IEEE*.

[54] Nejad, S., Gladwin, D. T., & Stone, D. A. (2016). A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states. *Journal of Power Sources*, *316*, 183-196.

[55] Gomez, J., Nelson, R., Kalu, E. E., Weatherspoon, M. H., & Zheng, J. P. (2011). Equivalent circuit model parameters of a high-power Li-ion battery: Thermal and state of charge effects. *Journal of Power Sources*, *196*(10), 4826-4831.

[56] Soudbakhsh, D., Gilaki, M., Lynch, W., Zhang, P., Choi, T., & Sahraei, E. (2020). Electrical response of mechanically damaged lithium-ion batteries. *Energies*, *13*(17), 4284.

[57] Li, S. E., Wang, B., Peng, H., & Hu, X. (2014). An electrochemistry-based impedance model for lithium-ion batteries. *Journal of Power Sources*, *258*, 9-18.

[58] Huang, J. (2018). Diffusion impedance of electroactive materials, electrolytic solutions and porous electrodes: Warburg impedance and beyond. *Electrochimica Acta*, *281*, 170-188.

[59] Zhou, Z., Li, Y., Wang, Q. G., & Yu, J. (2023). Health indicators identification of lithium-ion battery from electrochemical impedance spectroscopy using geometric analysis. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-9.

[60] Choi, W., Shin, H. C., Kim, J. M., Choi, J. Y., & Yoon, W. S. (2020). Modeling and applications of electrochemical impedance spectroscopy (EIS) for lithium-ion batteries. *Journal of Electrochemical Science and Technology*, *11*(1), 1-13.

[61] Zheng, Y., Shi, Z., Guo, D., Dai, H., & Han, X. (2021). A simplification of the time-domain equivalent circuit model for lithium-ion batteries based on low-frequency electrochemical impedance spectra. *Journal of Power Sources*, *489*, 229505.

[62] Rodríguez-Iturriaga, P., Anseán, D., Rodríguez-Bolívar, S., García, V. M., González, M., & López-Villanueva, J. A. (2024). Modeling current-rate effects in lithium-ion batteries based on a distributed, multi-particle equivalent circuit model. *Applied Energy*, *353*, 122141.

[63] Geng, Z., Wang, S., Lacey, M. J., Brandell, D., & Thiringer, T. (2021). Bridging physics-based and equivalent circuit models for lithium-ion batteries. *Electrochimica Acta*, *372*, 137829.

[64] Graule, A., Oehler, F. F., Schmitt, J., Li, J., & Jossen, A. (2024). Development and evaluation of a physicochemical equivalent circuit model for lithium-ion batteries. *Journal of The Electrochemical Society*, *171*(2), 020503.

[65] Landesfeind, J., Hattendorff, J., Ehrl, A., Wall, W. A., & Gasteiger, H. A. (2016). Tortuosity determination of battery electrodes and separators by impedance spectroscopy. *Journal of The Electrochemical Society*, *163*(7), A1373.

[66] Moškon, J., & Gaberšček, M. (2021). Transmission line models for evaluation of impedance response of insertion battery electrodes and cells. *Journal of Power Sources Advances*, *7*, 100047.

[67] Nara, H., Mukoyama, D., Yokoshima, T., Momma, T., & Osaka, T. (2015). Impedance analysis with transmission line model for reaction distribution in a pouch type lithium-ion battery by using micro reference electrode. *Journal of The Electrochemical Society*, *163*(3), A434.

[68] Cruz-Manzo, S., & Greenwood, P. (2020). An impedance model based on a transmission line circuit and a frequency dispersion Warburg component for the study of EIS in Li-ion batteries. *Journal of Electroanalytical Chemistry*, *871*, 114305.

[69] Murbach, M. D., Gerwe, B., Dawson-Elli, N., & Tsui, L. K. (2020). impedance. py: A Python package for electrochemical impedance analysis. *Journal of Open Source Software*, *5*(52), 2349.

[70] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261-272.

[71] Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. Wiley *Interdisciplinary Reviews: Computational Statistics*, *11*(3), e1460.

[72] Emiliano, P. C., Vivanco, M., Menezes, F. S. D., & Avelar, F. G. (2009). Foundations and comparison of information criteria: Akaike and Bayesian. *Revista Brasileira de Biometria*, *27*(3), 394-411.

[73] Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint* arXiv:1705.07874.

[74] Locorotondc, E., Pugi, L., Berzi, L., Pierini, M., Scavuzzc, S., Ferraris, A., ... & Carello, M. (2019, September). Modeling and simulation of constant phase element for battery electrochemical impedance spectroscopy. In 2019 *IEEE* 5th International forum on Research and Technology for Society and Industry (RTSI) (pp. 225-230). IEEE.

[75] Su, J., Lin, M., Wang, S., Li, J., Coffie-Ken, J., & Xie, F. (2019). An equivalent circuit model analysis for the lithium-ion battery pack in pure electric vehicles. *Measurement and Control*, *52*(3-4), 193-201.

[76] Pinson, M. B., & Bazant, M. Z. (2012). Theory of SEI formation in rechargeable batteries: capacity fade, accelerated aging and lifetime prediction. *Journal of the Electrochemical Society*, *160*(2), A243.

[77] Geng, Z., & Thiringer, T. (2022). In situ key aging parameter determination of a vehicle battery using only CAN signals in commercial vehicles. *Applied Energy*, 314, *118932*.

7. Appendix

It is important to note that a fourth model architecture was tested, which had the same input features as Architecture 1, except for R_0. This model architecture performed worse that both Architectures 1 and 2, with an average RMSE of 0.0371 for the Random Forest model. Table 4 is a look-up table for the fitted parameters of the 2 ZARC model and their associated SoH.

Cycle Day	SoC	L_0	R_0	R_1	τ_1	α_1	R_2	τ_1	α_1	W_0	SoH
0	0	6.11E-08	1.32E-03	3.00E-04	9.17E-04	8.40E-01	8.82E-01	2.04E+05	8.92E-01	1.98E-04	1.000
10	0	5.51E-08	1.29E-03	3.94E-04	3.38E-02	7.80E-01	4.66E-02	9.03E+01	8.79E-01	1.55E-04	0.986
20	0	5.48E-08	1.32E-03	3.98E-04	2.87E-03	7.90E-01	1.09E-01	3.85E+04	8.51E-01	1.43E-04	0.978
40	0	5.23E-08	1.35E-03	4.25E-04	3.65E-02	7.91E-01	3.41E-02	2.55E+03	8.57E-01	1.55E-04	0.961
90	0	5.80E-08	1.46E-03	4.13E-04	7.81E-03	7.88E-01	1.63E-02	6.90E+01	8.69E-01	1.70E-04	0.918
0	25	6.11E-08	1.32E-03	3.00E-04	9.17E-04	8.40E-01	8.82E-01	2.04E+05	8.92E-01	1.98E-04	1.000
10	25	5.51E-08	1.29E-03	3.94E-04	3.38E-02	7.80E-01	4.66E-02	9.03E+01	8.79E-01	1.55E-04	0.986
20	25	5.48E-08	1.32E-03	3.98E-04	2.87E-03	7.90E-01	1.09E-01	3.85E+04	8.51E-01	1.43E-04	0.978
40	25	5.23E-08	1.35E-03	4.25E-04	3.65E-02	7.91E-01	3.41E-02	2.55E+03	8.57E-01	1.55E-04	0.961
90	25	5.80E-08	1.46E-03	4.13E-04	7.81E-03	7.88E-01	1.63E-02	6.90E+01	8.69E-01	1.70E-04	0.918
0	50	6.11E-08	1.32E-03	3.00E-04	9.17E-04	8.40E-01	8.82E-01	2.04E+05	8.92E-01	1.98E-04	1.000
10	50	5.51E-08	1.29E-03	3.94E-04	3.38E-02	7.80E-01	4.66E-02	9.03E+01	8.79E-01	1.55E-04	0.986
20	50	5.48E-08	1.32E-03	3.98E-04	2.87E-03	7.90E-01	1.09E-01	3.85E+04	8.51E-01	1.43E-04	0.978
40	50	5.23E-08	1.35E-03	4.25E-04	3.65E-02	7.91E-01	3.41E-02	2.55E+03	8.57E-01	1.55E-04	0.961
90	50	5.80E-08	1.46E-03	4.13E-04	7.81E-03	7.88E-01	1.63E-02	6.90E+01	8.69E-01	1.70E-04	0.918
0	75	6.11E-08	1.32E-03	3.00E-04	9.17E-04	8.40E-01	8.82E-01	2.04E+05	8.92E-01	1.98E-04	1.000
10	75	5.51E-08	1.29E-03	3.94E-04	3.38E-02	7.80E-01	4.66E-02	9.03E+01	8.79E-01	1.55E-04	0.986
20	75	5.48E-08	1.32E-03	3.98E-04	2.87E-03	7.90E-01	1.09E-01	3.85E+04	8.51E-01	1.43E-04	0.978
40	75	5.23E-08	1.35E-03	4.25E-04	3.65E-02	7.91E-01	3.41E-02	2.55E+03	8.57E-01	1.55E-04	0.961
90	75	5.80E-08	1.46E-03	4.13E-04	7.81E-03	7.88E-01	1.63E-02	6.90E+01	8.69E-01	1.70E-04	0.918
0	100	6.11E-08	1.32E-03	3.00E-04	9.17E-04	8.40E-01	8.82E-01	2.04E+05	8.92E-01	1.98E-04	1.000
10	100	5.51E-08	1.29E-03	3.94E-04	3.38E-02	7.80E-01	4.66E-02	9.03E+01	8.79E-01	1.55E-04	0.986
20	100	5.48E-08	1.32E-03	3.98E-04	2.87E-03	7.90E-01	1.09E-01	3.85E+04	8.51E-01	1.43E-04	0.978
40	100	5.23E-08	1.35E-03	4.25E-04	3.65E-02	7.91E-01	3.41E-02	2.55E+03	8.57E-01	1.55E-04	0.961
90	100	5.80E-08	1.46E-03	4.13E-04	7.81E-03	7.88E-01	1.63E-02	6.90E+01	8.69E-01	1.70E-04	0.918

Table 3: Look-up table for the fitted parameters of the 2 ZARC model and the associated SoH at five different SoCs and five different days of aging.

In addition, to better understand how feature importance impacts estimation accuracy at certain SoHs, Figure 8 visualizes the actual versus estimated SoH across top-performing splits.

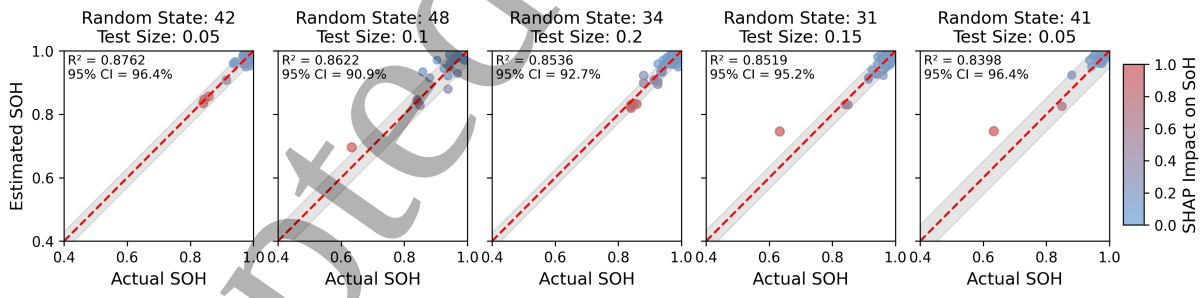

Figure 8: Parity plots for the top five splits (highest R^2 across all combinations of random states 30–50 and select test sizes 0.05, 0.10, 0.15, 0.20, and 0.25), colored by normalized SHAP impact on SoH to represent the total contribution of all features to the estimated SoH.