Energy 223 (2021) 120116

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

State of charge prediction of EV Li-ion batteries using EIS: A machine learning approach

Iman Babaeiyazdi*, Afshin Rezaei-Zare, Shahab Shokrzadeh

Department of Electrical Engineering and Computer Science, York University, Toronto, ON, M3J 1P3, Canada

ARTICLE INFO

Article history:
Received 6 October 2020
Received in revised form
29 December 2020
Accepted 12 February 2021
Available online 16 February 2021

Index Terms:
Electric vehicle
Electrochemical impedance spectroscopy
Li-ion batteries
Machine learning

ABSTRACT

Due to the significantly complex and nonlinear behavior of li-ion batteries, forecasting the state of charge (SOC) of the batteries is still a great challenge. Therefore, accurate SOC estimation is essential for the proper operation of batteries while the battery is monitored by the battery management system (BMS). To this end, this paper employs informative measurements of electrochemical impedance spectroscopy (EIS) in machine learning models (ML), i.e., linear regression model and Gaussian process regression (GPR), to accurately predict the SOC of li-ion batteries. First, a feature sensitivity analysis of the data is conducted to extract the most reliable features, i.e., the EIS impedances which are highly correlated with SOC, from EIS measurements. Then, the models are fed by the chosen features. The models are designed to train the input features and establish the mapping relationship between the selected features and the SOC. The results demonstrate that the error of the GPR model was found to be less than 3.8%. Considering onboard EIS measurements, this method can be practically embedded in the battery management system for accurate measurements of SOC of li-ion batteries and ensure the proper and efficient operation of battery-powered electric vehicles.

1. Introduction

Lithium-ion batteries have been widely utilized in Plug-in Electric Vehicles (PEVs), by virtue of their high energy and long life duration [1]. However, the performance of PEVs may be impacted by low performance of the battery management system (BMS) due to the unpredictability of battery's chemical reactions. Estimation of the battery state of charge (SOC) as one of the main functions of the BMS contributes greatly to proper operation, charging/discharging cycling, and lifespan of the EV batteries [2]. Also, Reliable and accurate SOC estimation can have other important applications such as baseload power generation of intermittent sources in the electric grid [3], and the safe operation of EV fast-charging stations integrated with battery storage system [4]. SOC is defined as the capacity of the battery at the current state compared to the battery's capacity at fully charged state [2]. Moreover, SOC cannot be measured directly from inner quantities such as internal resistance and capacitance of the battery, and as a result, the external quantitative indices are utilized for SOC estimation [5].

Myriads of literature propose various SOC estimation methods such as coulomb counting method, data-driven, and model-based estimation methods [5]. Coulomb counting or ampere-hour counting is one of the widely used methods in the laboratory, but due to cumulative current measurement errors, this method is not considered a highly accurate method for SOC estimation [5]. In the model-based methods, equivalent circuit models (ECMs) and electrochemical impedance models (EIMs) are the major models derived from empirical data to predict the SOCs. ECMs and EIMs are combined with various adaptive algorithms such as Kalman filter [6,7], extended and unscented Kalman filter [8,9], and Particle filter [10] to calculate the SOC of the battery based on the charge and discharge voltage and current. Moreover, electrochemical impedance spectroscopy (EIS) measurements are utilized in ECMs to estimate SOC. In Refs. [2,11], and [12], the authors identify the parameters of ECMs based on the EIS measurements. In Ref. [2], the EIS data is derived only at one SOC, which prevents the model from being an inclusive model, but on the other hand, a wide range of temperature is considered for modeling the battery based on ECM. The EIS data at above-zero temperature and SOCs between 10% to 90% and 10%−100% has been derived in Refs. [11,12], respectively, which, however, decrease the accuracy and reliability of the

* Corresponding author.
E-mail addresses: eiman@eecs.yorku.ca (I. Babaeiyazdi), rezaei@yorku.ca (A. Rezaei-Zare), shokrzadeh@ieee.org (S. Shokrzadeh).

https://doi.org/10.1016/j.energy.2021.120116
0360-5442/

estimation for SOCs at sub-zero temperatures and SOCs below 10%. Therefore, due to the internal complex chemical reaction process and uncertain external operating conditions of batteries, modeling the batteries based on the ECM methods is challenging for estimating the battery characteristics in real-life operation [13,14]. Physic-based models (PBMs) demonstrate insights through chemical and electrochemical dynamics, such as li-ion diffusion and Ohmic effects [13]. However, to estimate SOC using PBMs, partial derivatives equations should be solved by the BMS controller, which is a highly intensive computational burden [15]. Data-driven models are only dependent on historical data, and they do not need complicated equivalent or mathematical models. However, the challenge of data-driven models is the acquirement of informative inputs to construct a robust model for predicting the battery characteristics. Additionally, effective extraction of the features from historical data still remains a challenging task [16]. In Ref. [17], the SOC of the battery is predicted by a neural network (NN) which utilizes voltage, current, temperature, and power of the battery as the input features. Prediction of SOC also has been conducted in Ref. [18] employing NN and random forest/tree. Voltage, current, and cycling number contribute as the inputs of the machine learning (ML) black box in the mentioned paper. In another study [19], support vector machine along with Gaussian methods estimate the SOC of the battery and extracts feature variables based on the charging curve. However, all of these data-driven models that use terminal voltage as input feature may lose the accuracy as the terminal voltage of battery suddenly drops at the end of discharge which accordingly does not provide reliable data for low SOCs [20]. Thus, identifying and extraction of reliable features become the main bottleneck of the adoption of the data driven approaches and thus, more research is required in this regard.

On the other hand, EIS measurements over a wide range of frequency provide rich information about the dynamic characteristics of the battery and pave the way for precise estimation of the battery status. Nevertheless, none of the reviewed papers have adopted the EIS measurements directly as input data for machine learning models to predict SOC except for [21], in which the EIS data obtained for SOCs above 30% and at room temperature have been utilized in a deep NN. The model does not employ the EIS data in a wide range of temperature and at different SOC points [21], while such a data exclusion decreases the accuracy and reliability of the model. Also, the reported error of the model of [21] is less than 5%.

This study investigates the effectiveness of the EIS measurement data for estimating the SOC of the li-ion batteries using machine learning techniques. In opposition to Ref. [21], which uses the whole EIS impedances from the EIS spectrum to estimate SOC, only highly correlated EIS impedances with SOC are used in this paper. The proposed method's advantages are higher accuracy of the models and lower computational burden by eliminating irrelevant input features, i.e., EIS impedances with low correlations. Therefore, highly correlated impedances are first identified and then extracted from EIS spectrum measurements obtained at SOCs from 0% to 100%. The chosen impedances are utilized as input features for the linear regression model and Gaussian process regression (GPR). The models are designed to train the input features and establish the mapping relationship between the selected frequencies and the SOC. Finally, the trained models are employed to achieve SOC prediction.

Moreover, since the machine learning algorithm is neither dependent on the model of the battery nor the method that the battery is charged/discharged, and only the input and output of the dataset matter here, the model can predict the SOC by interpolating or extrapolating the dataset, regardless of charging or discharging mode of the battery. The SOC can be precisely estimated for aged batteries if the EIS measurements dataset is available for degraded batteries with the state of health (SOH) between 60% and 100%. The reason that the battery's degradation was not considered in this paper is due to the unavailability of EIS measurements for different SOHs for the dataset utilized in the paper. In contrast to many other studies that only take into account the EIS data obtained at above-zero temperatures, this study considers the EIS data for both above-zero and sub-zero temperatures, i.e., as low as $-20\ ^{\circ}C$. The results demonstrate an error of less than 3.8% for the GPR model. Considering the online and on-board EIS measurement [22–24], this method can be practically embedded in the BMS for accurate measurements of SOC.

The paper is organized as follows: in section 2, the electrochemical impedance spectroscopy measurement is fully explained. In the next section, the methodology for extracting reliable features and building the prediction models based on linear regression and GPR algorithm is discussed. Section 4 introduces the result of the built-up models for predicting the SOC, and the last section is dedicated to the conclusions.

2. Electrochemical impedance spectroscopy

EIS is a non-destructive and information-rich test which is conducted by galvanostatic or potentiostatic excitation signal over a wide range of frequency to obtain the impedance of the battery during charging and discharging [25]. The excitation signals in galvanostatic and potentiostatic methods are commonly sinusoidal current and voltage and the corresponding response will be voltage and current, respectively. Based on these waveforms, the electrochemical impedance of the battery can be calculated. The impedance of the battery is obtained based on the following equations in galvanostatic mode [26]:

$$\Delta I = I_{max}\sin(2\pi ft), \tag{1}$$

$$\Delta V = V_{max}\sin(2\pi ft + \varnothing), \tag{2}$$

$$Z(f) = \frac{V_{max}}{I_{max}}e^{j\varnothing}, \tag{3}$$

where ΔI is a sinusoidal current at frequency f, which is superimposed on the dc charging or discharging current and results in ΔV and phase angle $\varnothing$. Accordingly, Eq. (3) shows that the battery's impedance is frequency-dependent and characterized by its magnitude and phase angle. Fig. 1 indicates a typical EIS spectrum. The horizontal axis indicates the real part of the impedance, and the vertical axis shows the negative of the imaginary part of the impedance. The EIS spectrum is drawn over a wide range of

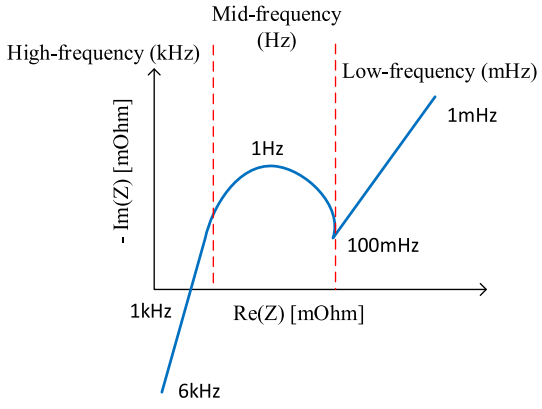

Fig. 1. Typical EIS spectrum of li-ion battery.

Table 1
Panasonic 18650 PF cell parameters.

Nominal Open Circuit Voltage	3.6 V
Min/Max Voltage	2.5 V/4.2 V
Mass/Energy storage	48 g/9.9 Wh
Capacity	2.75 Ah
Cycles to 80% Capacity	500 (100% DOD, 25 °C)
Minimum Charging Temperature	10 °C

frequency. The low-frequency tail indicates the diffusion processes inside the active material of the battery, the mid-frequency semicircle indicates the double-layer capacitance effect, and in the high-frequency region, the intercept of the EIS curve with the real axis is the indicator of Ohmic resistance of the battery.

2.1. EIS experimental data

In this study, the experimental data from Ref. [27] have been utilized, where a Panasonic NCR18650PF lithium-ion battery, an NCA chemistry cell similar to the cells used in Tesla's electric cars [28], was tested. The battery specifications have been presented in Table 1. In the test, EIS measurements were conducted over SOCs from 0% to 100% and temperature range of −20 °C − 25 °C for frequency sweep of 1 mHz−6 kHz. Fig. 2 shows the battery's impedance spectroscopy at different temperatures for the given SOC of 50%. It can be observed from Fig. 2(a) that the semicircle enlarges as the temperature decreases. It is due to the fact that the charge transfer rate decreases in the solid electrolyte interface (SEI) layer and electrode-electrolyte interface at lower temperatures [29]. Fig. 2(b) shows that as the temperature decreases, the polarization resistance of the battery (where $-\mathrm{Im}\,Z = 0$) increases due to a decrease in ion transfer in the electrolyte at lower temperatures [29]. Fig. 3(a), Fig. 3(b) and 3(c) plot the EIS measurements of the battery at different SOCs and temperatures of +25 °C, 0 °C, and −20 °C, respectively. This figure manifests that at above zero temperatures, the SOC is less effective on impedance spectra, while at zero and sub-zero temperatures, SOC significantly affects the impedance curve, especially in the mid-frequency regions. Thus, the effects of SOC on the impedance spectra at different temperatures are different.

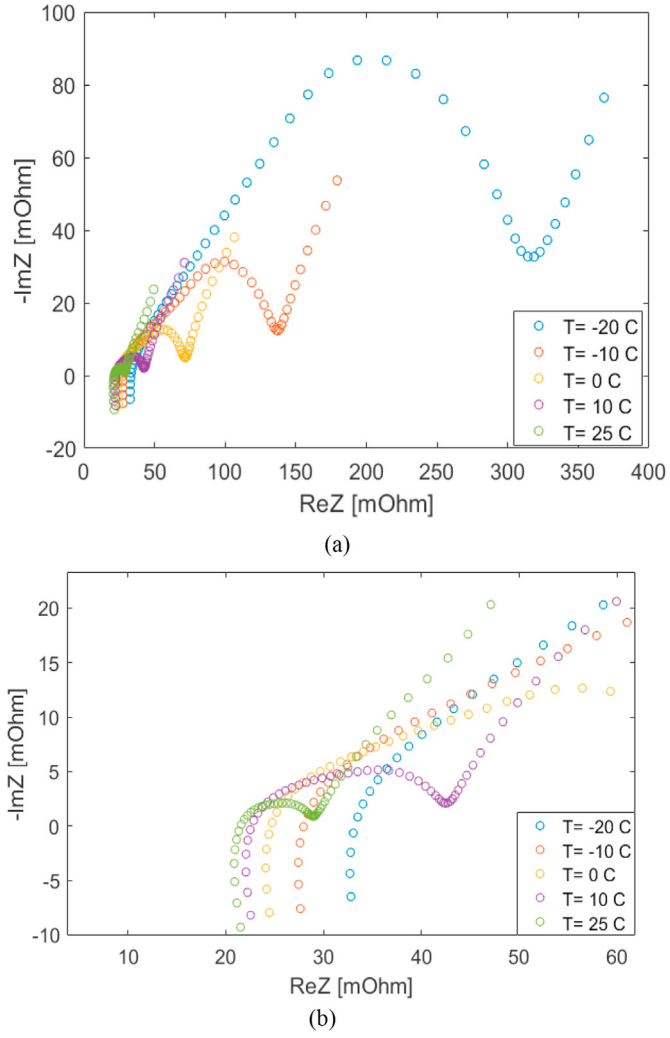

Fig. 2. EIS spectrum of the battery at (a) SOC of 50% and different ambient temperature, (b) zoomed-in version of (a).

3. Methodology

This section is dedicated to the feature sensitivity analysis to capture the highly correlated EIS features, i.e., highly correlated EIS impedances with SOC of the battery, and then the selected reliable features are utilized for training and testing of the machine learning models.

3.1. Feature sensitivity analysis

Extracting highly relevant features to the machine learning models' output, i.e., SOC, is essential for accurately predicting the output. To this end, the correlation matrix, which indicates the dependency of two or more variables on each other, is calculated and then shown on a color-coded image plot. The calculation of the correlation is performed by the Pearson correlation coefficient as follows:

$$\rho_{X,Y} = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{E\left(X^2\right) - E(X)^2}\cdot\sqrt{E\left(Y^2\right) - E(Y)^2}}, \tag{4}$$

where E is the expected value operator, and X and Y are two random variables. Fig. 4 shows the heatmap of the dataset [27] that we used for this study at different temperatures. Fig. 4 is a 2-D graphical representation that indicates the dependency of the features of the dataset. The features in the used dataset are electrochemical impedances at the corresponding frequencies that they were measured. In this case, the number of the features is 54 since the impedances were measured at 54 frequencies, sweeping from 1 mHz to 6 kHz. In Fig. 4, the correlation between the features varies from −1 to 1. The positive correlation has been shown in the spectrum of light to dark red, and the negative correlation has been shown in the spectrum of light to dark blue. The positive and negative correlations mean that the output varies in the same or opposite direction of the input variables' variations. The heatmap is a symmetric figure; thus, the last row or the last column represents the relation of the input variables, i.e., impedances at different frequencies with the output, i.e., SOC of the battery. Fig. 4(a)–(e) show the heatmap of the dataset at temperatures of (25 °C) − (−20 °C), respectively. Fig. 4 (a) shows that the first few features are highly and negatively correlated with the SOC and these features lie in the high and mid-frequency regions of the EIS spectrum. As also depicted in Fig. 4 (a), it is apparent from Fig. 4(b)–(e) that as the temperature decreases, some other features from mid-frequency appear to be positively correlated with the SOC. Another remarkable result deduced from Fig. 4 is that the low frequencies are significantly less correlated with the SOC. The reason is that

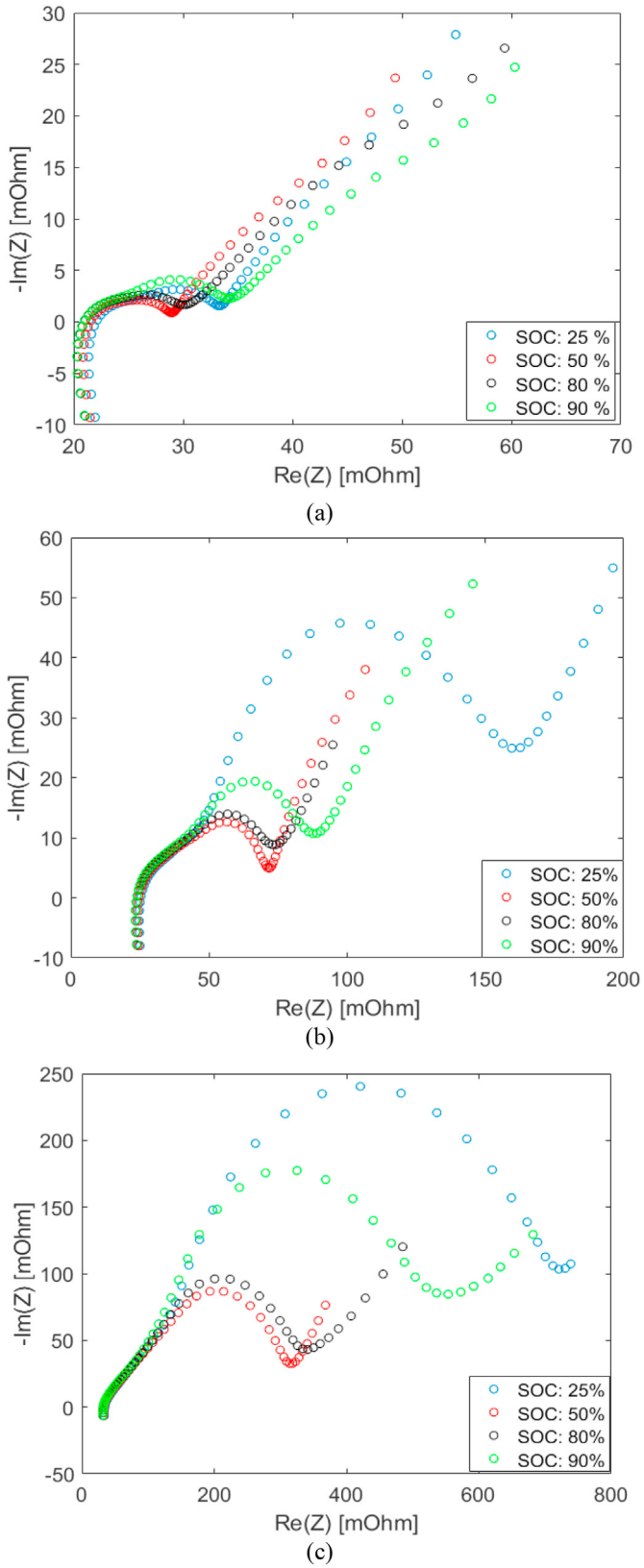

Fig. 3. EIS spectrum of the battery at (a) 25 °C, (b) 0 °C, and (c) −20 °C and at different SOC levels.

according to Figs. 2 and 3, the number of features, i.e., EIS impedances at lower frequencies, is significantly less than the number of features at higher frequencies. In other words, most of the EIS impedances which have lower magnitudes exist in all of the conditions in Figs. 2 and 3, and they lie in high and mid-frequency regions; however, the EIS impedances with larger magnitudes only exist in extreme conditions, i.e., at the temperature of −20 °C in Fig. 2 and SOC of 20% in Fig. 3.

Therefore, heatmaps are useful for extracting reliable features to build an accurate and fast model for prediction purposes. Moreover, pairgrid plot is also utilized to show the relationships of highly correlated features obtained from heatmap with the SOC and their distribution in the dataset. In Fig. 5, only four high-correlation features f1, f2, f4, and f5 have been shown as examples at 25 °C. The diagonal plots represent the distribution of single-variable and off-diagonal plots, which are the mirror image of each other and show the relations between every two variables of the dataset along with the regression plots. It is worthy to note that Fig. 5 shows only the negatively correlated features with the SOC. It is observed that with the increase of features f1, f2, f4, f5 in magnitude, the SOC decreases. Therefore, feature analysis helps identify the most influential features that contribute to the variation of the output variable.

3.2. Linear regression algorithm

A linear regression algorithm is used for identifying the relationship between a dependent variable and one or more independent variables. In this case, the impedances at different frequencies are the independent variables, and the SOC is the dependent variable. The basic multiple regression model of a dependent variable Y on a set of k independent variables (x_k) can be expressed as [30]:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{1k} + e_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \ldots + \beta_k x_{2k} + e_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \ldots + \beta_k x_{nk} + e_n \end{cases} \tag{5}$$

Therefore:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + e_i \tag{6}$$

where y_i is the i-th case of the dependent variable Y, x_{ij} is the value of the j-th independent variable (X_j) for the i-th case of the dependent variable, β_0 is the Y-intercept of the regression surface, each β_j is the slope of the regression surface with respect to variable X_j, and finally e_i is the random error component for the i-th case. In each equation in Eq. (5) the error is distributed with zero mean and standard deviation, and it is independent of the errors in the other equations. Since the variables are fixed quantities, the randomness of Y results from the randomness of error terms in each equation; although, in terms of correlation, the input variable are taken into account random variables, and the input variables are independent of the error terms. In matrix notation, Eq. (5) can be written as [30]:

$$Y = X\beta + e \tag{7}$$

where:

$$\begin{aligned} Y &= [y_1\ y_2 \cdots y_n]^T \\ \beta &= [\beta_1\ \beta_2 \cdots \beta_{k+1}]^T \\ X &= \begin{bmatrix} x_{11} & \cdots & x_{1(k+1)} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{n(k+1)} \end{bmatrix} \end{aligned} \tag{8}$$

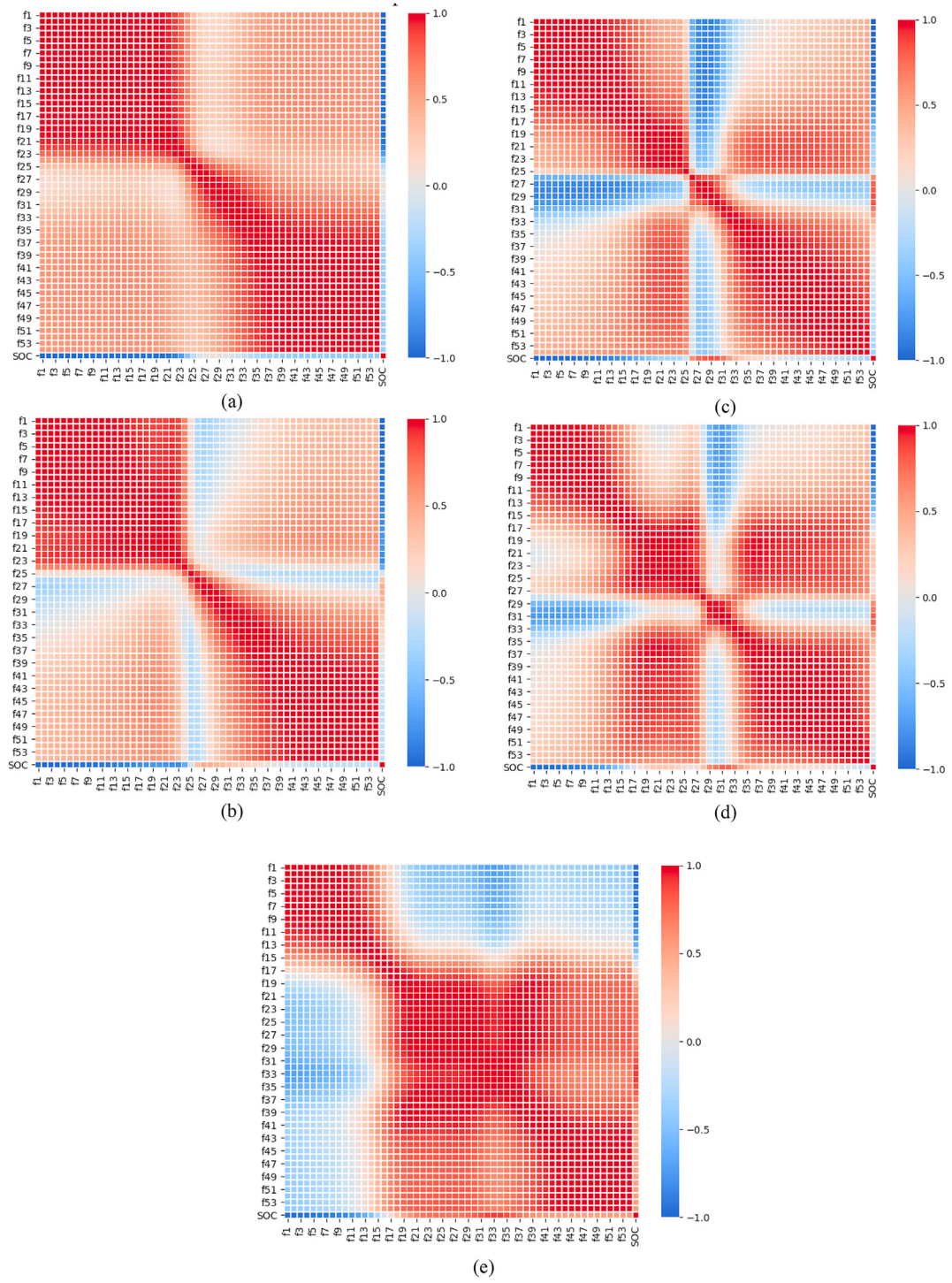

Fig. 4. Heatmap for feature sensitivity analysis of EIS spectrum at (a) 25 °C. (b) 10 °C, (c) 0 °C, (d) −10 °C, (e) −20 °C.

and Y is the target vector, e is the error vector which is a column vector of length n, and β is the vector of parameters, which is a column vector of length $k+1$. Matrix X is the input matrix, which is n by $k+1$ matrix. To do prediction, β and e should be calculated. The structure of the regression model has been shown in Fig. 6.

3.3. Gaussian process regression (GPR)

For a given training dataset of $T = \{(x_i, y_i), i = 1, 2, \ldots, n\}$ with n pairs of inputs x_i, which may have one or more than one features, and output y_i, the GPR model computes the predictive distribution of unobserved test datasets with y^* as output and x^* as input [31]. In this study, X and Y are defined as $X = [x_1, \ldots, x_n]^T$ and $Y = [y_1, \ldots, y_n]^T$, respectively. In this case $x_i = [EIS\ impedances]$ is the EIS impedances and the output y_i is the SOC of the cells. It is also assumed that $y_i = f(x_i + \varepsilon_i)$ where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is an independent and identically distributed Gaussian noise. The outputs $\mathbf{F} = (f(x_1) + \ldots + f(x_n))$ are modeled as Gaussian random field $\mathbf{F}$

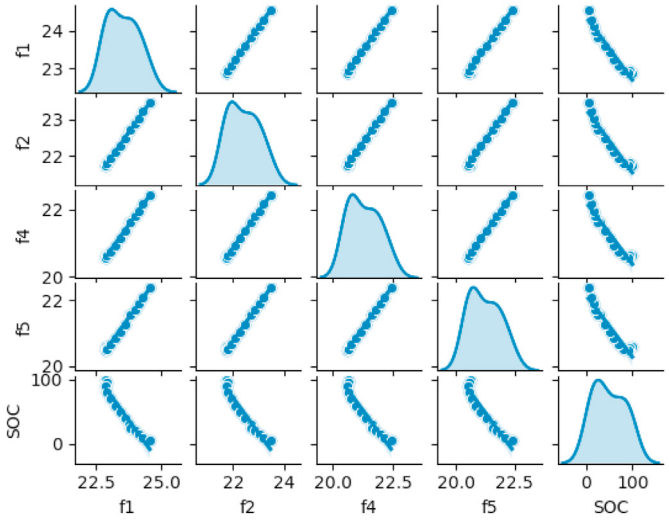

Fig. 5. PairGrid for reliable features of EIS spectrum at 25 °C.

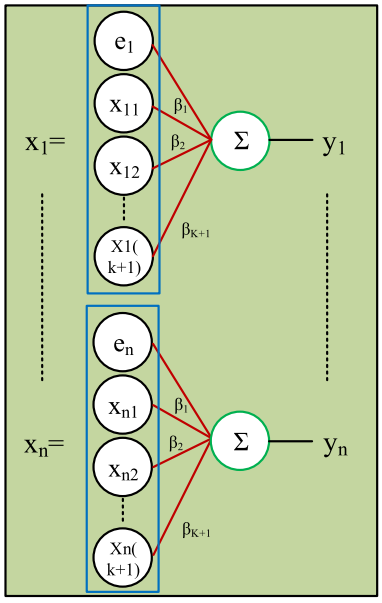

Fig. 6. Linear regression structure.

$\sim \mathcal{N}(0, K)$ where $K_{ij} = k(x_i, x_j)$ is the covariance kernel. In this case, the radial basis function is utilized for covariance kernel as well. The joint distribution of the training dataset $\{(x_i, y_i), i = 1, 2, \ldots, n\}$ and the predicted test output (x^*, y^*) is [31]:

$$\begin{bmatrix} Y \\ y^* \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma^2 I & K(X,x^*) \\ K(x^*,X) & K(x^*,x^*) \end{bmatrix}\right) \tag{9}$$

Conditioning on the training set yields the predicted mean on x^*:

$$\overline{y}^* = K(x^*,X)\left(K(X,X) + \sigma^2 I\right)^{-1} Y \tag{10}$$

And its predicted variance is:

$$\Delta^2 = K(x^*,x^*) - K(x^*,X)\left(K(X,X) + \sigma^2 I\right)^{-1} K(X,x^*) \tag{11}$$

3.4. Data partitioning

To avoid the malfunction of the ML models over a new dataset, the dataset should be split into two partitions, i) training set and ii) test set. The training set is used to construct the models and contains known output. The ML models are fitted to the training data and calculate the regression coefficients. Then, the test set is utilized to observe the performance of the ML models over unknown data. This implies that the test set serves as a criterion for the evaluation of the model predictions. In contrast to many other studies, the effect of different test-size (TS) is observed in this study for the dataset with the selected features and different correlation values.

3.5. Accuracy evaluation

The indices used for the evaluation of the performance of the proposed models are:

3.5.1. R-squared

Goodness-of-fit R-squared (R^2) is defined as [32]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{12}$$

where y_i is the actual value and $\widehat{y}_i$ is the predicted value for the *i-th* case. R^2 ranges between 0 and 1. The closer to 1, the better the prediction.

3.5.2. Mean absolute error (MAE)

MAE is used to compare the precision of the prediction and is defined as follows [32]:

$$MAE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{\widehat{y}_i}\right| \tag{13}$$

The smaller the MAE, the better the prediction.

3.5.3. Root mean squared error (RMSE)

The root square of the above equation provides the standard deviation of the random error term. Root mean square error is an estimate of the standard deviation of the random component in the data and is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{n}} \tag{14}$$

4. Prediction results and discussion

In this study, Python and MATLAB have been harnessed to perform the statistical analysis and prediction of the SOC. Fig. 7 shows the flowchart of the proposed methodology for predicting the SOC of the battery using EIS measurements. In this section, the models' performance at different temperatures is discussed,

considering the effect of reliable features extraction based on their correlation value and portioning of the dataset. The statistical evaluation indices discussed in the previous section have been tabulated in Table 2 in different conditions.

As mentioned earlier, some datasets have been introduced to the ML models with different portioning. The default portioning is that 80% of the dataset is dedicated to the training set, and 20% (test_size (TS) = 0.2) is dedicated to the test set. However, to see the effect of TS, other values for this variable have been also taken into account. For corr_value of 0.5, different evaluation indices have been obtained at different temperatures as presented in Table 2. The linear regression model can perfectly predict the battery's SOC based on the values of R_squared, MAE, and RMSE at the mentioned temperatures except for 10 °C and 0 °C. The MAE for 25 °C, −10 °C, and −20 °C temperatures is less than 4.9%, but for 10 °C and 0 °C, the MAEs are 8.9% and 17.5%, respectively. Moreover, it is clear that for TS of 0.4 the highest accuracy is achieved for temperatures 10 °C and 0 °C. As for the corr_value of 0.7, we can observe that for the temperatures of 25 °C, −10 °C, and −20 °C, the evaluation indices values have not changed, significantly. However, in the cases of 10 °C and 0 °C temperatures, the improvement of evaluation criteria is noticeable such that the MAEs have reduced to 5.5% and 9.7%, respectively. Moreover, one may observe the influence of TS on the mentioned temperatures, as the TS increases, an increase in R_squared, and a reduction in MAE and RMSE are observed. Considering the corr_value of 0.9, The MAE for all the cases is achieved with a value of less than 7%. Since the extracted features are reliable, it is expected that the MAE and RMSE decrease, but on the contrary, they increase. This is because when highly correlated features are selected, most of the other features are lost, and the machine learning model may lose accuracy if the dataset is not big enough. Thus, the performance of the model over a dataset is of importance. Although the linear regression model functions properly for corr_values of 0.9, with a maximum error of 7%, a more accurate and reliable model, i.e., GPR, is used for this corr_value. The GPR model results for corr_value of 0.9, and the best TS have been presented in Table 2. The MAE for 25 °C, −10 °C, and −20 °C temperatures is less than 2.8%, but for 10 °C and 0 °C, the MAEs are 3.8% and 8.7%, respectively. As an example, the training and test data and their predicted values at different temperatures have been shown in Fig. 8.

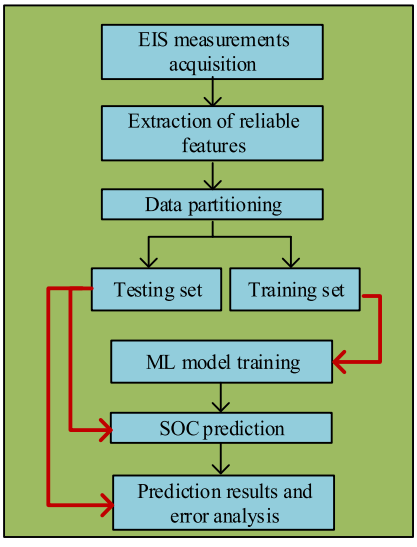

Fig. 7. Flowchart of the proposed model for predicting the SOC of the battery using EIS measurements.

Table 2
Evaluation indices under different conditions.

ML model	Corr-value	Temp(C)	R_squared	MAE	RMSE
Linear Regression	0.5	25 (TS = 0.2)	0.978	4.8456	5.9924
		10 (TS = 0.4)	0.74245	8.9317	20.098
		0 (TS = 0.4)	0.62	17.444	20.098
		−10 (TS = 0.2)	0.9894	2.9056	3.3387
		−20 (TS = 0.2)	0.9825	4.230	4.3
Linear Regression	0.7	25 (TS = 0.2)	0.975	4.8456	5.9924
		10 (TS = 0.4)	0.6983	9.5468	17.913
		10 (TS = 0.6)	0.9429	5.5629	8.4434
		0 (TS = 0.4)	0.7348	13.664	16.797
		0 (TS = 0.5)	0.8764	9.7318	11.481
		−10 (TS = 0.2)	0.9877	3.0525	3.602
		−20 (TS = 0.2)	0.977	4.9398	4.8428
Linear Regression	0.9	25 (TS = 0.2)	0.9524	6.2756	8.275
		10 (TS = 0.2)	0.78241	15.237	17.694
		10 (TS = 0.6)	0.899	6.9951	11.216
		0 (TS = 0.4)	0.4345	15.094	24.528
		0 (TS = 0.5)	0.8785	9.0877	11.382
		0 (TS = 0.6)	0.9259	6.9159	8.877
		−10 (TS = 0.3)	0.9828	3.2369	3.5628
		−20 (TS = 0.5)	0.875	6.5511	9.9834
		−20 (TS = 0.6)	0.93	6.11	6.8883
GPR	0.9	25 (TS = 0.2)	0.998	1.3409	1.4602
		10 (TS = 0.4)	0.9817	3.8178	4.4068
		0 (TS = 0.3)	0.9036	8.6819	9.9630
		−10 (TS = 0.3)	0.9838	2.7223	3.4537
		−20 (TS = 0.2)	0.9880	2.7493	3.5530

Fig. 8 showcases SOC's predicted values versus f1 and f2, among the highly correlated features.

The study results demonstrate that in addition to identifying and extracting reliable features, the learning ability of the model and partitioning of the data for training are highly crucial for precise prediction. Considering the above-mentioned elements' effects, we also observed that the GPR model outperforms the linear regression model. The proposed method will be implemented in BMS for online measurement of EIS and SOC prediction utilizing the potential approaches from Refs. [22–24] such as fractional-order equivalent circuit model (FOECM) and pseudo-random sequences (PRS), which are fast and easily implementable for measuring EIS at low measurement time and low complexity.

5. Conclusions

In this investigation, the prediction of li-ion battery SOC using EIS measurements was performed based on an ML approach. This study was conducted based on extracting reliable features according to their correlation value with the SOC of the battery. The features are the impedances of EIS measurements over the range of desired frequencies, i.e., from 1 mHz to 6 kHz. After selecting the reliable features of different datasets at various temperatures and different SOCs, the linear regression model and GPR were trained, and the prediction was performed by the trained models over the test set. Statistical indices such as R_squared, MAE, and RMSE were used to evaluate the accuracy and robustness of the models. The results indicated that the proposed models are able to precisely predict the SOC of the battery using the reliable features. The models trained by the features with corr_value of above 0.9 indicated the best performance among the others, such that the error of the GPR model was found to be less than 3.8%. Furthermore, the impact of test-size on the model precision was evaluated. It was observed that for some cases, the larger test-size results in higher accuracy. Therefore, considering the online and onboard EIS measurement, this method can be practically embedded in the BMS for accurate measurements of SOC of li-ion batteries and ensure the proper operation of PEVs.

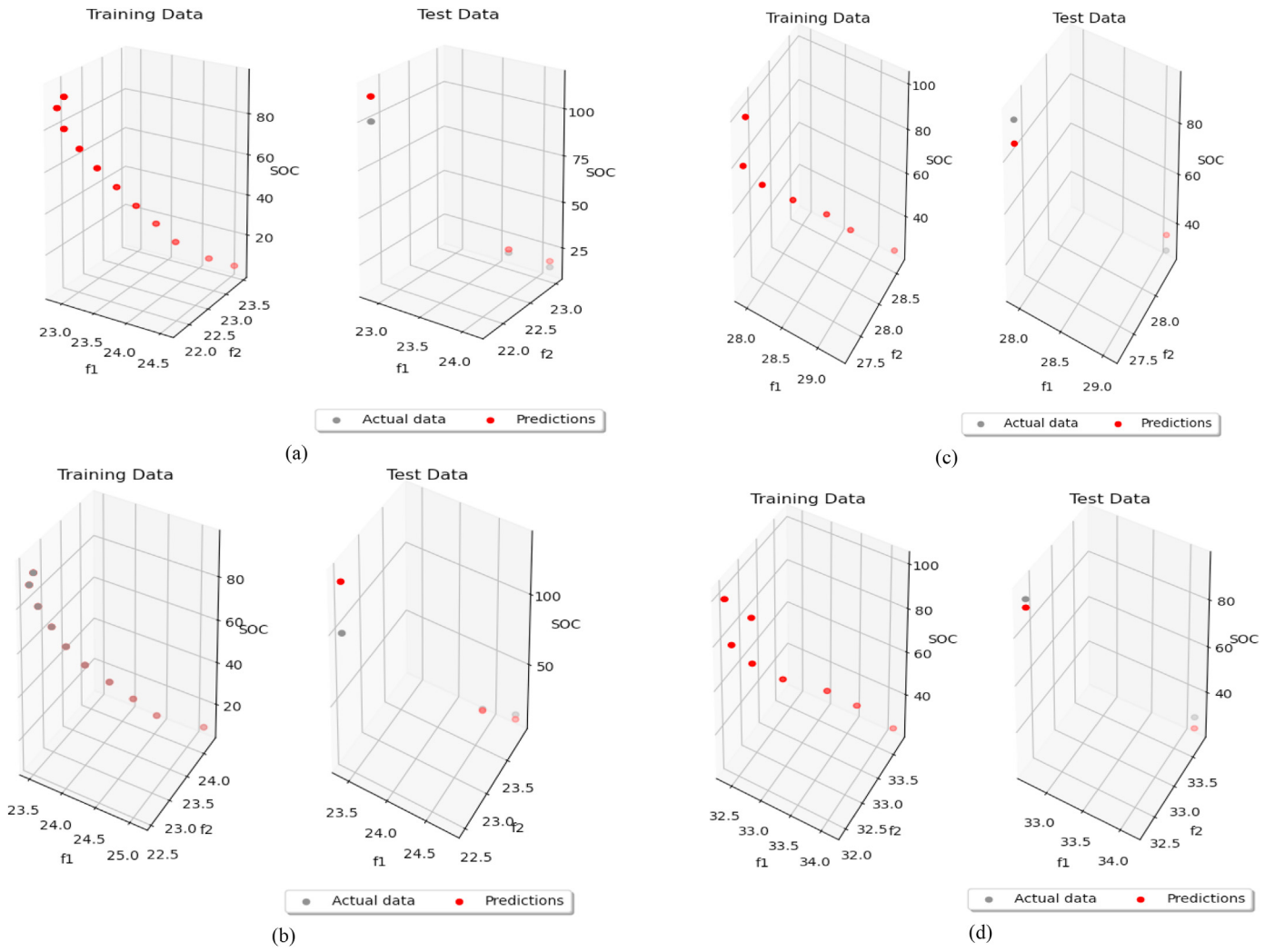

Fig. 8. SOC prediction of the proposed model at temperatures of (a) 25 °C, (b) 10 °C, (c) −10 °C, (d) −25 °C.

Credit author statement

Iman Babaeiyazdi: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization, Afshin Rezaei-Zare: Supervision, writing – Reviewing & editing. Shahab Shokrzadeh: Supervision, writing – Reviewing & editing.

Declaration of competing interest

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

References

[1] Cai L, Meng J, Stroe D, Luo G, Teodorescu R. An evolutionary framework for lithium-ion battery state of health estimation. J Power Sources 2019;412:615–22. July 2018.
[2] Chen N, Zhang P, Dai J, Gui W. Estimating the state-of-charge of lithium-ion battery using an H-infinity observer based on electrochemical impedance model. IEEE Access 2020;8:26872–84.
[3] Shokrzadeh S, Jafari M, Bibeau E, Molinski T. A statistical algorithm for predicting the energy storage capacity for baseload wind power generation in the future electric grids. Energy 2015;89:793–802.
[4] Babaeiyazdi I, Rezaei-zare A, Shokrzadeh S. Fast charging systems to enable electrification of Transportation : an operational constrained based analysis. In: 2019 IEEE transportation electrification conference and expo (ITEC); 2019. p. 1–6.
[5] Sidhu MS, Ronanki D, Williamson S. State of charge estimation of lithium-ion batteries using hybrid machine learning technique. In: Iecon 2019 - 45th annual conference of the IEEE industrial electronics society, vol. 1; 2019. p. 2732–7.
[6] Sepasi S, Ghorbani R, Yann B. A novel on-board state-of-charge estimation method for aged Li-ion batteries based on model adaptive extended Kalman fi lter. J Power Sources 2014;245:337–44.
[7] Baccouche I, Jemmali S, Manai B, Omar N, Essoukri N, Amara B. Improved OCV model of a Li-ion NMC battery for online SOC estimation using the extended kalman filter. Energies 2017;10(6):764.
[8] Lee J, Nam O, Cho BH. Li-ion battery SOC estimation method based on the reduced order extended Kalman filtering. J Power Sources 2007;174(1):9–15.
[9] Wang W, Wang X, Xiang C, Wei C, Zhao Y. Unscented kalman filter-based battery SOC estimation and peak power prediction method for power distribution of hybrid electric vehicles. IEEE Access 2018;6:35957–65.
[10] Xia B, et al. A comparative study of three improved algorithms based on Particle filter algorithms in SOC estimation of lithium ion batteries. Energies 2017;10(8):1149.
[11] Xu J, Mi CC, Cao B, Cao J. A new method to estimate the state of charge of lithium-ion batteries based on the battery impedance model. J Power Sources 2013;233:277–84.
[12] Waag W, Kabitz S, Sauer DU. Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application. Appl Energy 2013;102:885–97.
[13] Ng M, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge nad health of batteries using data-driven machine learning. Nat. Mach. Intell. 2020;2:161–70.
[14] Xiong R, Cao J, Yu Q, He H, Sun F. Critical review on the battery state of charge estimation methods for electric vehicles. IEEE Access 2017;6:1832–43.
[15] Lacroix M, Jokar A, Rajabloo B, Martin D. Review of simpli fi ed Pseudo-two-

Dimensional models of lithium-ion batteries. J Power Sources 2016;327:44–55.
[16] Wu Y, et al. State of health estimation for lithium-ion batteries based on healthy features and long short-term memory. IEEE Access 2020;8:28533–47.
[17] Zahid T, Xu K, Li W, Li C, Li H. State of charge estimation for electric vehicle power battery using advanced machine learning algorithm under diversi fi ed drive cycles. Energy 2018;162:871–82.
[18] Donato THR, Quiles MG. Machine learning systems based on xgBoost and MLP neural network applied in satellite lithium-ion battery sets impedance estimation. Adv. Comput. Intell 2018;5:1–20.
[19] Hu C, Jain G, Schmidt C, Strief C, Sullivan M. Online estimation of lithium-ion battery capacity using sparse Bayesian learning. J Power Sources 2015;289:105–13.
[20] Chang W-Y. The state of charge estimating methods for battery: a review. *Int. Sch. Res. Not*; 2013.
[21] Messing M, Shoa T, Ahmed R, Habibi S. Battery SoC estimation from EIS using neural nets. In: 2020 IEEE transportation electrification conference & expo (ITEC). IEEE; 2020. p. 588–93.
[22] Guha A, Patra A. Online estimation of the electrochemical impedance spectrum and remaining useful life of lithium-ion batteries. IEEE Trans. Instrum. Meas. 2018;67(8):1836–49.
[23] Crescentini M, et al. Online EIS and diagnostics on lithium-ion batteries by means of low-power integrated sensing and parametric modeling. IEEE Trans. Instrum. Meas. 2020;9456(1).
[24] Sihvo J, Stroe D, Messo T, Roinila T. Fast approach for battery impedance identification using pseudo-random. IEEE Trans Power Electron 2020;35(3):2548–57.
[25] Din E, Schaef C, Member S, Moffat K, Member S, Stauth JT. A scalable active battery management system with embedded real-time electrochemical impedance spectroscopy. IEEE Trans Power Electron 2017;32(7):5688–98.
[26] Huet F. A review of impedance measurements for determination of the state-of-charge or state-of-health of secondary batteries. J Power Sources 1998;70:59–69.
[27] Kollmeyer P. Panasonic 18650PF Li-ion battery data. 2018.
[28] Tesla model S 18650 cell test data (March 30, 2015), Tesla Motors Club Discussion Forum.".
[29] Wang Q, He Y-J, Hu X, Ma Z-F. State of charge-dependent polynomial equivalent circuit modeling for electrochemical impedance spectroscopy of lithium-ion batteries. IEEE Trans Power Electron 2018;33(10):8449–60.
[30] Brandt BS. Linear and polynomial regression. In: Data analysis. New York: SPRINGER; 1999.
[31] Zhang Y, Tang Q, Zhang Y, Lee AA, Wang J, Stimming U. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nat Commun 2020;11(6–11).
[32] Makridakis R, Wheelwright S, Hyndman S. Forecasting methods and applications. NewYork: Wiley; 1998.

Iman Babaeiyazdi was born in Yazd, Iran. He received B·Sc. from Amirkabir University of Technology, Tehran, Iran, and his M.Sc. from York University, Toronto, Canada, both in electrical engineering, in 2017 and 2019, respectively. Currently, he is perusing his Ph.D. in electrical engineering at York University.

Afshin Rezaei-Zare (M′08–SM′10) received the B.Sc., M.Sc., and Ph.D. degrees (Hons.) in electrical engineering from the University of Tehran, Tehran, Iran, in 1998, 2000, and 2007, respectively. He was a Postdoctoral Fellow from 2007 to 2009 at the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. From 2010 to 2017, he was with the Department of Special Studies, Hydro One Networks Inc., Toronto. In 2017, he joined the Department of Electrical Engineering and Computer Science, York University, Toronto, as associate professor. His research interests include power system resilience to geomagnetic disturbance, electromagnetic transients in power systems, and renewable and energy storage systems. Dr. Rezaei-Zare is a registered Professional Engineer in the Province of Ontario, Canada, and an Associate Editor of the IEEE Transactions on Power Delivery and IEEE Power Engineering Letters.

Shahab Shokrzadeh (M ′15) received the B·Sc. degree in mechanical engineering from Iran University of Science and Technology, Tehran, Iran, in 2004, the M.Sc. degree in production engineering and management from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2009, and his PhD in mechanical engineering from the University of Manitoba, Winnipeg, Canada in 2015. He was the co-founder and Associate Director of the International Renewable Energy Academy at York University in Toronto, Canada. He is currently a Project Manager for Alternative Energy in the Standards Division at CSA Group in Toronto, Canada. Dr. Shokrzadeh has several years of research and industry experience in the energy and automotive sectors, with a focus on electrified transportation, renewable energy, and advanced energy storage technologies.Dr. Shokrzadeh is a registered Professional Engineer in Ontario.