Energy 283 (2023) 128461

Contents lists available at ScienceDirect

Energy

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

Battery SOC estimation from EIS data based on machine learning and equivalent circuit model

Emanuele Buchicchio [a,*], Alessio De Angelis [a], Francesco Santoni [a], Paolo Carbone [a], Francesco Bianconi [a], Fabrizio Smeraldi [b]

[a] University of Perugia, Department of Engineering, Via Goffredo Duranti 93, Perugia, 06125, PG, Italy
[b] Queen Mary University of London, School of Electronic Engineering and Computer Science, Mile End Road, London, E1 4NS, United Kingdom

ARTICLE INFO

Dataset link: https://data.mendeley.com/datasets/zdsgxwksn5, https://data.mendeley.com/datasets/mbv3bx847g

MSC:
68T99

Keywords:
Battery
State of charge
SOC
Electrochemical impedance spectroscopy
EIS

ABSTRACT

Estimating the state of charge (SOC) of batteries is fundamental for the proper management and safe operation of numerous systems, including electric vehicles, smart energy grids, and portable electronics. While there is no practical method for direct measurement of SOC, several estimation approaches have been developed, including a growing number of machine-learning-based techniques. Machine learning methods are intrinsically data-driven but can also benefit from a-priori knowledge embedded in a model. In this work, we first demonstrate, through exploratory data analysis, that it is possible to discriminate between different SOC from electrochemical impedance spectroscopy (EIS) measurements. Then we propose a SOC estimation approach based on EIS and an equivalent circuit model to provide a compact way to describe the frequency domain and time-domain behavior of the impedance of a battery. We experimentally validated this approach by applying it to a dataset consisting of EIS measurements performed on four lithium-ion cylindrical cells at different SOC values. The proposed approach allows for very efficient model training and produces a low-dimensional SOC classification model that achieves above 93% accuracy. The resulting low-dimensional classification model is suitable for embedding into battery-powered systems and for online SOC estimation.

1. Introduction

Battery-powered applications such as smartphones, tablets, laptops, unmanned aerial vehicles, and electric vehicles are now part of our daily life. Furthermore, energy storage systems using battery packs are also widely used in renewable energy generation to ensure stable and smooth electricity transmission from renewable resources to the primary grid. In this context, information about the remaining charge within the battery is essential for battery management systems (BMS) and for the end-users of most battery-operated systems. Given that lithium rechargeable batteries are the most common choice for many applications, estimating the remaining battery capacity is fundamental for their management because extremely high or extremely low state-of-charge (SOC) conditions can irreversibly damage the battery and pose safety hazards [1].

In a controlled laboratory setting, the most common method for direct SOC estimation is Coulomb counting, also known as Ampere-hour counting [2]. However, this method suffers from accumulation of error, due to its integrative nature, which makes it inaccurate in many practical scenarios. Moreover, the relationship between the battery's directly-measurable signals and the estimated SOC is highly non-linear, varying with temperature and discharge/charge currents [3]. For this reason, it is widely recognized that there is no practical method for SOC direct measurement outside laboratory settings [4]. As a consequence, much research has focused on the development of secure, practical, and reliable methods for SOC estimation in recent years [5,6].

With the continuous development of artificial intelligence, machine learning (ML) approaches gained popularity [7], especially deep learning (DL) methods based on neural networks [8–10]. Some recent works, such as [11,12], extend the machine learning-based SOC estimation to cells in battery packs.

Features derived from the current–voltage measurement, performed by BMS during charging and discharging curve are the most commonly used inputs, but electrochemical impedance spectroscopy (EIS), which obtains the impedance over a wide range of frequencies by measuring the current response to a voltage perturbation or vice versa is known to contain rich information on all materials properties, interfacial phenomena and electrochemical reactions [13]. This directly relates to

* Corresponding author.
E-mail addresses: emanuele.buchicchio@ieee.org (E. Buchicchio), alessio.deangelis@unipg.it (A. De Angelis), francesco.santoni@unipg.it (F. Santoni), paolo.carbone@unipg.it (P. Carbone), francesco.bianconi@unipg.it (F. Bianconi), f.smeraldi@qmul.ac.uk (F. Smeraldi).

https://doi.org/10.1016/j.energy.2023.128461
Received 22 December 2022; Received in revised form 10 May 2023; Accepted 15 July 2023
Available online 20 July 2023

changes and possible degradation of different internal parts of the battery and is able to track the status of the battery [14].

However, several limitations exist in ML methods for SOC estimation that should be tackled for real-world applications outside laboratory settings. In particular: (i) many proposed methods require a massive amount of data; (ii) the training of the DL models may require several hours or days to complete; (iii) deep neural networks tend to overfit when trained on a limited amount of data [10]; (iv) many results in literature do not take into account the variability across different instances of the same battery model. (v) EIS data are often fitted into an equivalent electrical circuit model (ECM), but the effectiveness of this approach, compared to the use of raw impedance data, has not been clearly demonstrated; (vi) purely model-based parametric methods rely on the assumption that the battery model is accurately established. This last condition is hard to realize in real-world applications due to measurement noise, model parameter drift with aging, and temperature [4].

Most SOC estimation studies rely on data obtained with laboratory battery analysis instrumentation with measurement accuracy not available to BMS in field applications. A realistic dataset obtained from an onboard measurement system should be used to investigate the effectiveness of SOC estimation methods. These limitations are the primary motivations for this work. In this paper, we demonstrate the feasibility of SOC estimation based on EIS data and low-complexity machine learning algorithms. To this end, we present the results of an exploratory data analysis applied to a dataset of EIS measurements performed on lithium-ion batteries at different SOC values. Our results show that it is possible to discriminate between different SOC based on EIS measurements. Then, we propose a low-complexity ML approach for SOC classification, which exploits EIS data and an equivalent circuit model of the battery.

Many existing works report accurate SOC estimation using data-driven or model-based approaches. Direct comparison of performance metrics such as RMSE, MAE or accuracy it is not feasible because results are strongly related to experimental conditions, variability across different batteries, and the dataset used for method evaluation. Our experimental results demonstrate that the proposed method extracts almost all available information from the input dataset. The comparison of results with and without an ECM shows that using an ECM to incorporate the available a-priori information increases the estimation accuracy with respect to the pure data-driven estimation on the same dataset.

Finally, we experimentally validate this approach by applying it to the EIS dataset. The proposed approach is different from DL-based approaches, since it does not require time-consuming training. Moreover, this approach is validated extensively by EIS measurements performed multiple times on the same battery and on different batteries of the same type, therefore providing insight on its generalization ability.

Therefore, the main contributions of this work can be summarized as follows:

1. we propose a lightweight SOC estimation method based on EIS and an equivalent circuit model;
2. we present a pipeline for exploratory data analysis on EIS datasets based on PCA, LDA and MCA techniques;
3. the proposed methods are validated on a real-world public available EIS dataset;
4. the experiments demonstrate the feasibility of SOC estimation based on sparse sampling of the EIS spectra (14 frequencies);
5. furthermore, we show that bespoke dimensionality reduction based on ECM parameter fitting achieves better accuracy than obtained with raw impedance data.

The remainder of the paper is organized as follows: Section 2 reviews related work in SOC estimation; Section 3 illustrates the EIS dataset employed for characterizing the proposed approach; Section 4 describes the main structure of the proposed approach; Section 5 describes the ECM used; Section 6 details the exploratory data analysis results; Section 7 summarizes SOC classification results; Section 7 draws conclusions, and the final section provides additional information on data availability and result reproducibility.

2. Related works

In the literature SOC is usually defined as the ratio of the current available capacity $Q_{current}$ to the maximum available capacity Q_{max} and is expressed as a percentage [9]:

$$SOC = Q_{current}/Q_{max} \tag{1}$$

A comprehensive overview of different SOC estimation techniques in recent literature is given in [5]. The existing SOC estimation methods can be summarized into two categories: direct and indirect methods [4]. The most common direct methods are look-up approaches and ampere-hour counting (AHC). Indirect methods include model-based, data-driven, and hybrid techniques [15].

Direct look-up methods require tabulations of an external measurable characteristic such as open circuit voltage (OCV) and the battery's impedance at a different SOC. This approach is only applicable when the battery is in a static state, not connected to any load, and allowed to rest to achieve an equilibrium stage — hence is not feasible for online SOC estimation. In general, battery impedance and OCV testing are very time-consuming and require disconnecting the equipment during the offline measurement procedure. The multi-sine EIS implemented in [16] allows for a shorter measurement time with respect to the conventional stepped-sine technique and potentially enables in-situ and in-operando monitoring.

In the AHC method (also known as Coulomb Counting), the SOC is estimated by integrating the measured discharging current of a battery over time. The SOC at time t_1 is calculated by the following equation:

$$SOC\left(t_1\right) = SOC_0 - \frac{\eta}{C_n}\int_{t_0}^{t_1} I\left(t\right) dt \tag{2}$$

where $SOC_0 \triangleq SOC(t_0)$ denotes the initial state of charge (at time t_0), η the Coulombic efficiency, C_n the rated capacity, and $I(t)$ the instantaneous discharge current of the battery. To evaluate (2), the initial SOC of the battery must be known. Given the open-loop nature of the algorithm, errors in the SOC estimation accumulate over time, and even a small measurement error results in a significant integration error.

Model-based indirect SOC estimation methods leverage on a model to simulate the relationship between the battery's external measurable physical properties and the SOC. The latter is eventually estimated from the modeled relations. Model-based methods include electrochemical models and equivalent circuit models (ECMs). Electrochemical models require an in-depth understanding of the battery's internal chemical and electrical characteristics and describe its internal processes in terms of partial differential equations [7] or fractional order differential equations [17,18]. Basic ECMs include the Rint model, first-order RC model, second-order RC model, and Thevenin model [19]. The SOC estimation performance of the model-based method requires the assumption that the battery model is accurately established. The estimation accuracy can be improved by advanced filter algorithms [20] such as Sliding Mode Observer, Luenberger Observer, Kalman filter [21], extended and unscented Kalman filter [4,20] and Particle filter [1]. While a model-based approach can result in reliable and accurate models, it requires extensive domain knowledge, rigorous feature engineering, and a relatively long development time. An extensive review and comparison of battery modeling can be found in [19,22–25].

By contrast, data-driven SOC estimation approaches, such as ML methods, require only limited knowledge of the battery internals and can result in RMS lower than 1% [20,21]. Such approaches are becoming more popular for estimating the SOC and battery state-of-health (SOH) due to the increasing availability of battery data and

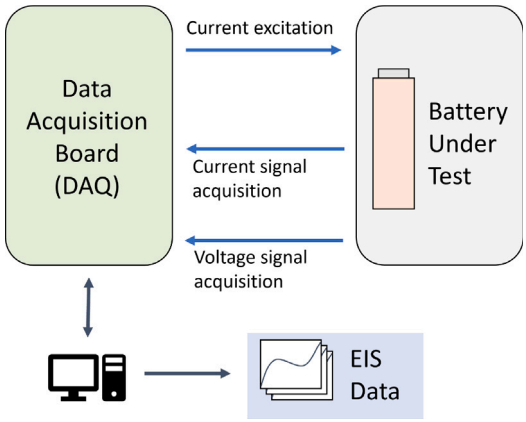

Fig. 1. Custom-built system for acquiring EIS data from a battery under test.

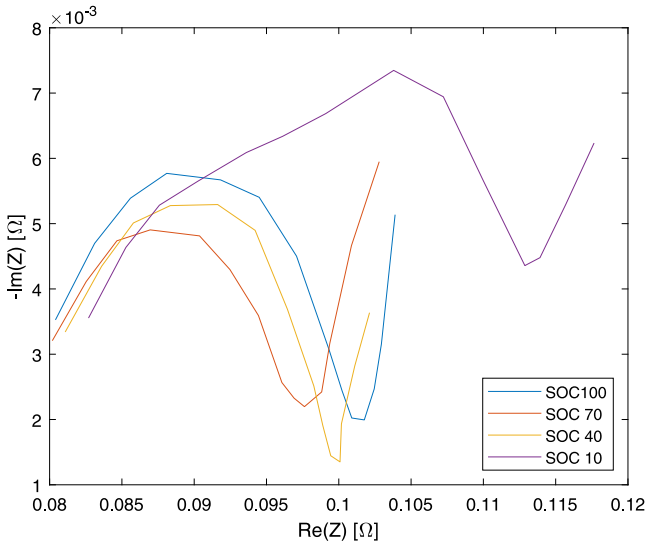

Fig. 2. Cole-Cole plot representations of a set of EIS measurement results, consisting of the complex values of the impedance Z, extracted from the dataset available in [31]. The EIS curves are measured at different values of the SOC on the same lithium-ion cell.

advances in computational power. A comprehensive review of SOH estimation method can be found in [26,27]. Some of the methods such as [11], allow for SOC and SOH co-estimation. ML methods are intrinsically data-driven but can also benefit from a-priori knowledge, which is embedded in a model. Therefore, the hybrid strategies that combine the model-driven and data-driven approaches are recognized as a promising research direction.

Although the on-board usage of EIS in battery management systems surely poses some implementation issues, different solutions have already been proposed. Kondratiev and Kuznietsov [28] recently described an EIS measurement equipment with unipolar signal based on relatively simple schematics; other on-board EIS systems for battery impedance estimation have been described by Koseoglou et al. [29] and, earlier on, by Wang et al. [30].

3. EIS dataset

For characterizing the proposed SOC estimation approach, we use an EIS dataset made publicly available in [31]. The proposed approach involves processing both the raw impedance data from the dataset and the lower-dimensional parametric representation obtained by fitting the ECM presented in Section 5 to such data. A description of the dataset and the measurement system used for its creation is provided here below.

3.1. EIS measurement system architecture

The basic architecture of the custom-built EIS measurement system is shown in Fig. 1. The data acquisition (DAQ) board generates an arbitrary voltage excitation signal. This signal is converted into a current signal by a Howland current pump, which makes use of a power operational amplifier. The current signal is then fed to the battery under test. A readout electronics section samples the current signal by means of a shunt resistor and an instrumentation amplifier. The readout section also samples the voltage at the battery terminals.

The resulting current and voltage time series are transferred to a PC, where the impedance of the battery under test is estimated by performing the Discrete Fourier Transform (DFT) of the current and voltage signals and then dividing the DFT of the voltage by that of the current.

The choice of the excitation signal plays an important role in determining the accuracy and speed of the EIS measurement [32]. In this context, a broadband random-phase multi-sine excitation signal allows for simultaneous excitation at a wide range of frequencies, thus providing a shorter measurement time compared with the conventional single-sine approach. Random-phase excitation was chosen to avoid the increase of the crest factor (see [33] on this), thus obtaining an energy-efficient way to excite the battery under test. Spectral leakage was avoided using coherent sampling; i.e., by ensuring that the acquired time series contained an integer number of periods of all excited sinusoidal components of the multi-sine. A more detailed description of the EIS measurement system, together with a characterization of its measurement uncertainty, is available in [16].

3.2. Dataset description

The dataset consists of the results of EIS measurements on a set of Samsung ICR18650-26J cylindrical rechargeable Lithium-Ion cells [31]. It contains the complex impedance of the batteries measured at fourteen logarithmically spaced frequencies (0.05, 0.1, 0.2, 0.4, 1, 2, 4, 10, 20, 40, 100, 200, 400 and 1000 Hz) using a random-phase multi-sine excitation signal. For each excited frequency the current amplitude was 50 mA, resulting in a measurement uncertainty of approximately 0.1 mΩ [16]. Repeated EIS measurements on four different brand-new batteries and six separate discharge cycles at ten different states of charge were obtained. All measurements were performed in a temperature-controlled environment at 25 ± 1 °C. For illustrative purposes a set of EIS curves corresponding to different SOCs is reported in Fig. 2. As can be seen, a simple relationship between the EIS shape and the SOC is hard to infer: this motivates the use of ML techniques.

4. Proposed approach

Our solution combines data-driven ML methods and model-based processing with an equivalent circuit model. The goal is to estimate the SOC of a battery starting from EIS measurement data. We treat the problem as a supervised classification task, and we solve it using different feature sets and classification algorithms. In particular, we show that transforming the original EIS data through domain-specific previous knowledge in the form of a circuit-equivalent model (see below) improves accuracy by a large margin.

Our approach includes two main phases as described in Fig. 3: an exploratory data analysis followed by a SOC estimation that exploits the insights gained with data analysis. In the following sections we first detail the equivalent circuit model, which enables injecting previous

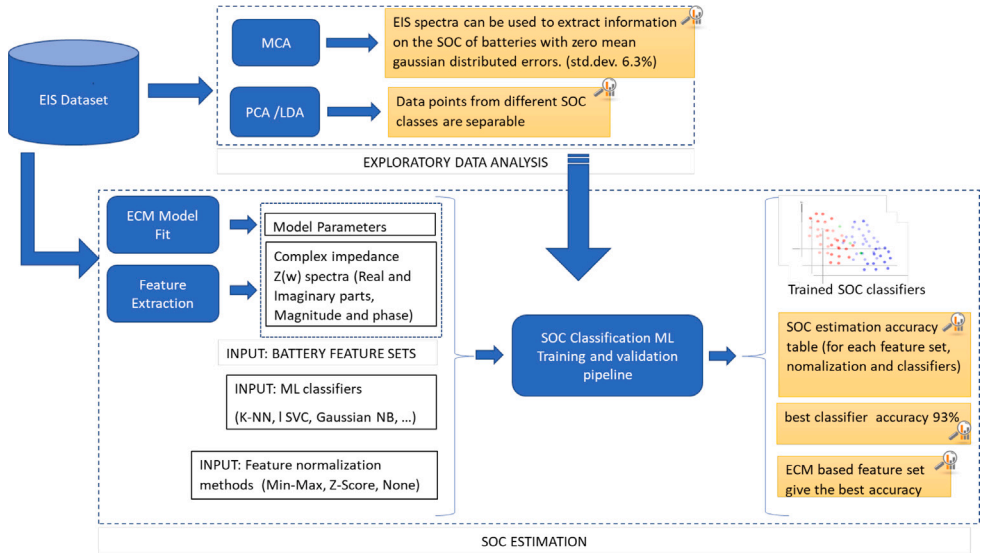

Fig. 3. Overview of the proposed two phases approach to battery SOC estimation. (1) Exploratory data analysis [Top]. (2) SOC estimation [Bottom] designed on.

knowledge into the EIS data (Section 5). Then we describe EDA (Section 6), the aim of which is to assess, qualitatively, the ability of EIS data (with and without ECM fitting) to discriminate between the different SOCs. Finally, we present the ML approaches for SOC estimation along with experimental validation and benchmarking (Section 7).

5. Equivalent circuit model

Equivalent circuit models provide a compact way to describe the frequency-domain and time-domain behavior of the impedance of a battery. Therefore, they can allow for a transformation from the raw EIS data to a smaller number of parameters. This can be seen as a bespoke dimensionality reduction stage, resulting in a lower-dimensional data representation that is then used in the SoC estimation process. Crucially, this allows to incorporate prior knowledge on the structure of EIS data.

In this work, we use the 7-parameter equivalent circuit model shown in Fig. 4. This representation has proven useful to investigate the EIS profile of lithium-ion batteries [18]. As can be seen in Fig. 4, the circuit model contains two resistors (R_0 and R_1), one inductor (L) and two constant-phase elements (CPE_1 and CPE_2). The latter are 2-parameter fractional-order components widely used for modeling electrochemical processes [34]. The complex impedance of the *CPE* is $Z_{\mathrm{CPE}} = \frac{1}{Qs^{\alpha}}$, where $s = i\omega$ (ω being the angular frequency in radians), α is a dimension-less parameter accounting for non-ideality (in the ideal case $\alpha = 1$, the CPE is a capacitor), and Q is a generalized capacitance measured in $\mathrm{s}^{\alpha}\Omega^{-1}$ ($\mathrm{s}\Omega^{-1} \equiv \mathrm{F}$ in the ideal case). When $\alpha_2 = 0.5$, CPE_2 is the so-called Warburg element, accounting for ion transport across an ideal semi-infinite electrolyte layer [35,36]. On the complex plane, the impedance of the Warburg element is a straight line with a slope of 45° with respect to the real axis. When CPE_2 is a Warburg element, the circuit loop in Fig. 4 is commonly known as the Randles–Ershler circuit. For a real finite layer, the slope of Z_{CPE_2} may change, and α_2 must be fitted to experimental data [18]. The total impedance of the circuit model in Fig. 4 is:

$$\begin{aligned} Z(s) &= R_0 + sL + \left[\left(Z_{\mathrm{CPE}_1} + R_1\right) \parallel Z_{\mathrm{CPE}_2}\right] = \\ &= R_0 + sL + \frac{R_1 Q_1 s^{\alpha_1} + 1}{Q_1 s^{\alpha_1} + R_1 Q_1 Q_2 s^{\alpha_1+\alpha_2} + Q_2 s^{\alpha_2}}. \end{aligned} \tag{3}$$

This model is fitted to the EIS data by numerically solving a nonlinear least squares problem with a multistart algorithm as in [37].

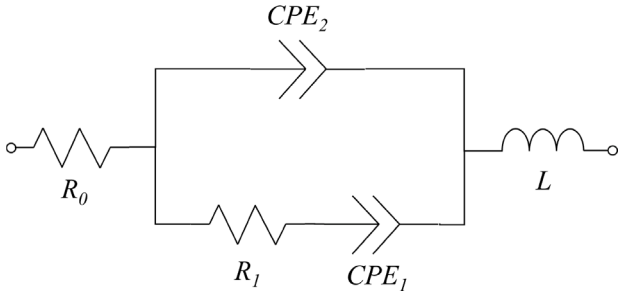

Fig. 4. Equivalent circuit used to model the impedance of the battery.

6. Exploratory data analysis

Exploratory data analysis (EDA) was carried out with the purpose of visualizing the data and assessing, qualitatively, their ability to separate the different states of charge. We carried out EDA on both the impedance values as described in Section 3.2 and the parameters of the equivalent circuit model as detailed in Section 5. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were used for both sets of features. Maximum Covariance Analysis (MCA) was used on the first set only.

6.1. PCA and LDA

Altogether we took into account seven sets of features — see also Table 1 for a recap: (a) real part of the impedance, (b) imaginary part of the impedance, (c) concatenation of (a) and (b), (d) module of the impedance, (e) phase of the impedance, (f) concatenation of (d) and (e), and (g) parameters of the equivalent circuit model.

Data normalization was carried out on each feature separately using three approaches: (1) no normalization, denoted as 'none' in the remainder, indicating that we used the raw features with no preprocessing altogether; (2) min–max linear scaling to $[0, 1]$ ('MinMax' in the remainder) and (3) zero-mean, unit-variance normalization ('Z-score'). For each of the $7 \times 3 = 21$ combined feature set/normalization methods we generated planar scatter plots of the data via PCA and LDA. As a result each point in the plots represents the projection of the corresponding data point on the first two PCA or LDA axes. Note that the projection maximizes the overall data variance in the case of PCA

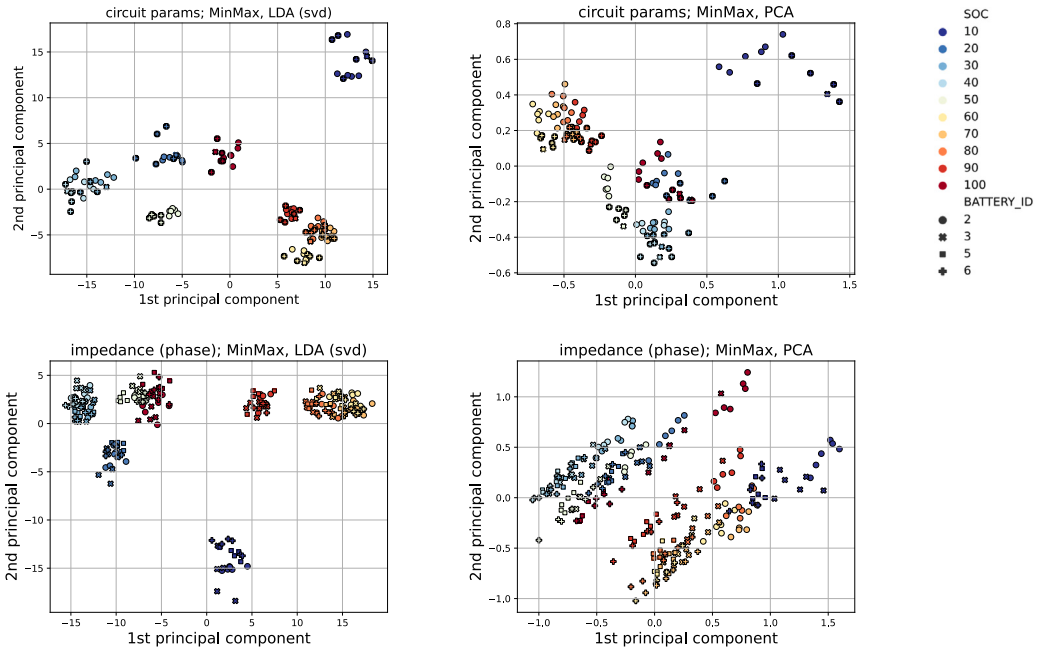

Fig. 5. Exploratory data analysis. The figure reports PCA and LDA projections of the two combinations feature set/normalization method that achieved the best prediction accuracy for the state of charge. In each plot markers denote battery id, and colors indicate state of charge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Sets of features considered in the EDA and for the estimation of the state of charge. Symbol '+' indicates feature concatenation.

Abbrev.	Description	Num. of features
real	Real part of the impedance	14
imag	Imaginary part of the impedance	14
real+imag	Real and imaginary part of the impedance	28
abs	Module of the impedance	14
phase	Phase of the impedance	14
abs+phase	Module and phase of the impedance	28
circparams	Parameters of the equiv. circuit model	7

and the ratio of between-class to intra-class variance in the case of LDA. Fig. 5 reports four sample plots — specifically PCA and LDA projections of the two combinations feature set/normalization method that were the best predictors for the state of charge (more on this in Section 7). The figure shows how data-points corresponding to different SOCs tend to group into separate clusters, which suggests a good prediction power of the features used. This is particularly evident with the parameter of the ECM (Fig. 5, first row). The complete scatter plots for all the feature sets are provided as supplementary material in a separate file (`scatter-plots.svg`).

6.2. MCA

Maximum covariance analysis aims at finding components that are maximally correlated. Specifically, we looked for linear combinations of the frequency components of the impedance spectrum that are maximally correlated with the SOC. The general technique can be described as follows. Let us consider two observation matrices $\mathbf{X} \in \mathbb{R}^{m\times n}$ and $\mathbf{Y} \in \mathbb{R}^{q\times n}$, where m and q are the number of observables, and n is the number of successive measurements. Observables are assumed to be standardized, i.e. their mean over the n measurements is subtracted, and they are divided by the standard deviation over the n measurements. Let $\mathbf{u}$ and $\mathbf{v}$ be two unit-vectors, and let us define:

$$\begin{aligned}\mathbf{a}^T &= \mathbf{u}^T\mathbf{X},\\ \mathbf{b}^T &= \mathbf{v}^T\mathbf{Y}.\end{aligned} \tag{4}$$

We seek the optimal $\mathbf{u}$ and $\mathbf{v}$ maximizing the covariance between $\mathbf{a}$ and $\mathbf{b}$:

$$\begin{aligned}\operatorname{cov}\left[\mathbf{a}^T,\mathbf{b}^T\right] &= \operatorname{cov}\left[\mathbf{u}^T\mathbf{X},\mathbf{v}^T\mathbf{Y}\right] =\\ &= \frac{1}{n-1}\left[\mathbf{u}^T\mathbf{X}\left(\mathbf{v}^T\mathbf{Y}\right)^T\right] = \mathbf{u}^T\mathbf{C}_{xy}\mathbf{v},\end{aligned} \tag{5}$$

where $\mathbf{C}_{xy} = \frac{1}{n-1}\mathbf{X}\mathbf{Y}^T \in \mathbb{R}^{m\times q}$ is the covariance matrix between $\mathbf{X}$ and $\mathbf{Y}$. In [38], it is proved that the maximum covariance is obtained from the leading modes of the singular value decomposition (SVD) of $\mathbf{C}_{xy}$, $\mathbf{u} = \mathbf{u}_1$ (the first left singular vector), and $\mathbf{v} = \mathbf{v}_1$ (the first right singular vector). The successive and mutually orthogonal singular vectors $\mathbf{u}_1,\dots,\mathbf{u}_m$ and $\mathbf{v}_1,\dots,\mathbf{v}_q$ return the successive maxima of the covariance. The components related to $\mathbf{u}_i$ and $\mathbf{v}_j$ for $i \neq j$ are uncorrelated, i.e., $\mathbf{u}_i^T\mathbf{C}_{xy}\mathbf{v}_j = 0$.

Let $s_1,\dots,s_S$ be the values of the SOC, $f_1,\dots,f_F$ the values of EIS frequencies, and let us define $Z_i' = \Re\mathrm{e}Z_i$ and $Z_i'' = \Im\mathrm{m}Z_i$, where i is the index identifying N repeated measurements. Let us define the first observation matrix:

$$\mathbf{X} = \begin{pmatrix} Z_1'(f_1,s_1) & \cdots & Z_1'(f_1,s_S) & \cdots & Z_N'(f_1,s_1) & \cdots & Z_N'(f_1,s_S)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ Z_1'(f_F,s_1) & \cdots & Z_1'(f_F,s_S) & \cdots & Z_N'(f_F,s_1) & \cdots & Z_N'(f_F,s_S)\\ Z_1''(f_1,s_1) & \cdots & Z_1''(f_1,s_S) & \cdots & Z_N''(f_1,s_1) & \cdots & Z_N''(f_1,s_S)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ Z_1''(f_F,s_1) & \cdots & Z_1''(f_F,s_S) & \cdots & Z_N''(f_F,s_1) & \cdots & Z_N''(f_F,s_S)\end{pmatrix} \tag{6}$$

where $m = 2F$ and $n = NS$. Let us define the second observation matrix as:

$$\mathbf{Y} = \begin{pmatrix} s_{1,1} & \cdots & s_{S,1} & \cdots & s_{1,N} & \cdots & s_{S,N}\end{pmatrix}, \tag{7}$$

i.e. all the SOC values repeated N times in a row. The matrices are then standardized over the n columns. Since $q = 1$, $\mathbf{Y}$ is not changed by the SVD, and $\mathbf{v} = \mathbf{v}_1 = 1$ is the only possible choice. All the components obtained with $\mathbf{u}_2,\dots,\mathbf{u}_m$ are uncorrelated to the SOC.

Thus, the MCA processes the spectrum components $Z(f,s)$ (represented by $\mathbf{X}$) to produce the new set of components $\mathbf{a}_1,\dots,\mathbf{a}_{2F}$ maximally correlated with the SOC. These components are plotted versus

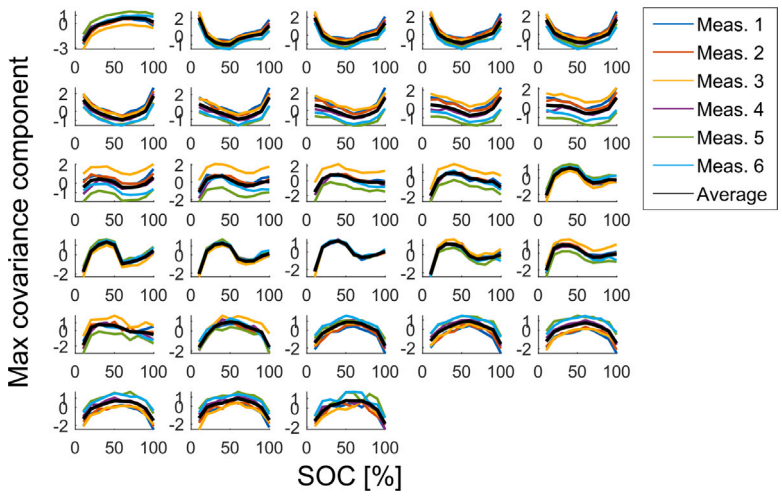

Fig. 6. MCA components $\mathbf{a}_1, \dots, \mathbf{a}_{2F}$ of the observation matrix (6) vs SOC for a single battery. Color: six repeated measurements. Black: average curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

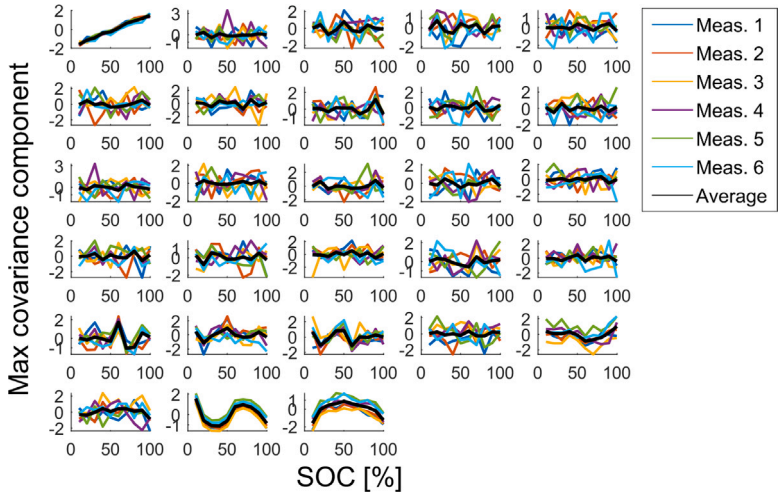

Fig. 7. MCA components $\mathbf{a}^{pc}_1, \dots, \mathbf{a}^{pc}_{2F}$ of the observation matrix $\mathbf{X}_{pc}$ vs SOC for a single battery. Color: six repeated measurements. Black: average curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SOC values in Fig. 6, for $\mathbf{v} = 1$ and $\mathbf{u} = \mathbf{u}_1, \dots, \mathbf{u}_m$ From a simple qualitative observation, it is clear that different components encode the same information. In fact, there are several curves that are very similar to each other. In order to eliminate this redundancy, the matrix $\mathbf{X}$ is replaced with its principal components $\mathbf{X}_{pc}$, i.e., given the covariance matrix $\mathbf{C}_X = \mathrm{cov}[\mathbf{X}] \in \mathbb{R}^{m\times m} \equiv \mathbb{R}^{2F\times 2F}$, it is diagonalized as $\mathbf{D}_X = \mathbf{U}^T\mathbf{C}_X\mathbf{U}$, the columns of $\mathbf{U}$ being a base of eigenvectors of $\mathbf{C}_X$; hence $\mathbf{X}_{pc} = \mathbf{U}^T\mathbf{X}$. The corresponding new set of MCA components $\mathbf{a}^{pc}_1, \dots, \mathbf{a}^{pc}_{2F}$ is plotted in Fig. 7. As exptected, only the first component is correlated to the SOC. Components $\mathbf{a}^{pc}_2, \dots, \mathbf{a}^{pc}_{26}$ are not correlated to the SOC, and they also seem independent of the SOC. A clear non-linear dependence can instead be observed for the last two components $\mathbf{a}^{pc}_{27}$ and $\mathbf{a}^{pc}_{28}$, indeed non-correlation does not imply independence. A qualitative observation suggests that the first MCA component $\mathbf{a}^{pc}_1$ encodes all the relevant information about the SOC. Indeed a good linear correlation between $\mathbf{a}^{pc}_1$ and SOC appears, as confirmed by the following quantitative analysis.

In Fig. 8 six repeated measurements of first MCA component $\mathbf{a}^{pc}_1$ versus SOC of four different batteries are reported. Each bundle of curves has been fitted with a line by using the least square method. The slope, intercept and Pearson correlation coefficient R are reported in Table 2. Pearson coefficient values indicate a very good linear correlation. All the values of the slope and intercept are compatible within one sigma. Weighted means of slope and intercept have been used to estimate SOC from $\mathbf{a}^{pc}_1$. The error on the estimate is defined as the difference between the true SOC and its estimated value, and it is shown in the histogram of Fig. 9. The histogram has been fitted with a Gaussian distribution. The quantile–quantile plot shows that the distribution is compatible with a Gaussian. A Kolmogorov–Smirnov test strongly confirms the null hypothesis that the data are extracted from a Gaussian distribution with a p-value of 0.99. This proves that the information on the SOC is largely encoded in $\mathbf{a}^{pc}_1$; once this information is extracted, only zero-mean uncorrelated uncertainty is left. This analysis strongly suggests that impedance spectra can be used to extract information on the SOC of batteries.

Pearson correlation coefficients between SOC and all the other components $\mathbf{a}^{pc}_2, \dots, \mathbf{a}^{pc}_{2F}$ are practically zero, but the last two components $\mathbf{a}^{pc}_{27}$ and $\mathbf{a}^{pc}_{28}$ can be respectively fitted with a third and second order polynomial. We tried to extract information on the SOC from these non-linear functions, similarly to what we did for the linear fit, but the accuracy on SOC estimation was not improved.

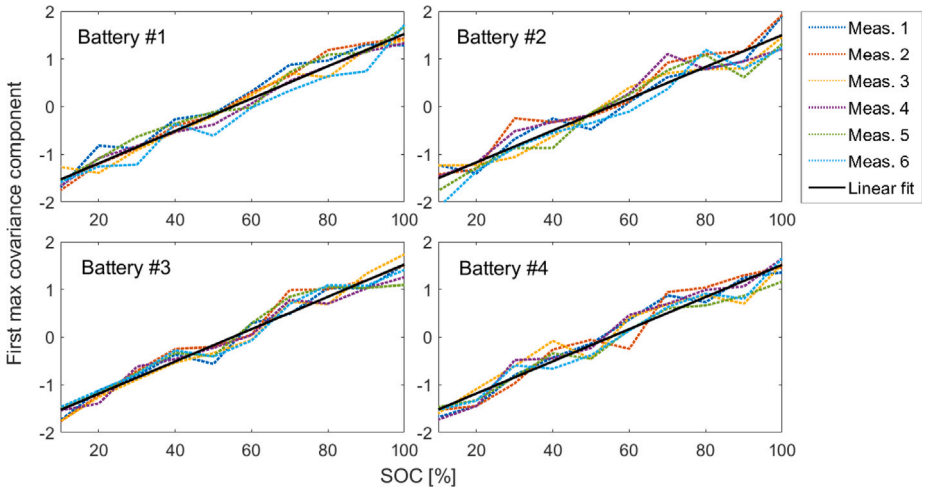

Fig. 8. Six repeated measurements of first MCA component $\mathbf{a}_1^{pc}$ versus SOC of four different batteries. The linear fit is shown in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

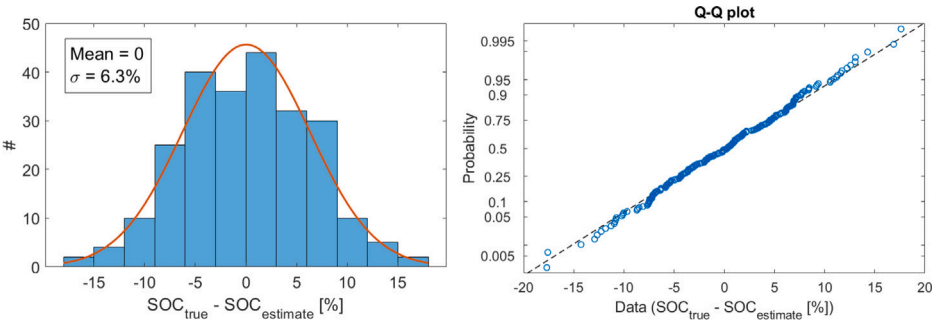

Fig. 9. Distribution of the error on the estimate of the SOC from $\mathbf{a}_1^{pc}$.

Table 2
Pearson coefficient R, slope and intercept of the linear fit of Fig. 8.

Battery #	R	Slope	Intercept
1	0.983	0.0339 ± 0.0008	−1.87 ± 0.05
2	0.966	0.0333 ± 0.0012	−1.83 ± 0.07
3	0.982	0.0339 ± 0.0009	−1.86 ± 0.05
4	0.977	0.0337 ± 0.0010	−1.85 ± 0.06
Weighted mean		0.0338 ± 0.0005	−1.85 ± 0.03

In conclusion, the standard deviation $\sigma = 6.3\%$ of the distribution in Fig. 9, obtained from the linear fit alone, represents the best accuracy on SOC estimation that can be attained. No additional information can be extracted from the present EIS data to improve the accuracy of SOC estimation.

7. Prediction of the state of charge

7.1. Accuracy estimation

The ability to predict the state of charge from the sets of features and data normalization methods described in Section 6 was assessed through a series of supervised classification experiments. We treated the problem as a multi-class classification one where each state of charge (10%, 20%, … , 100%) represents one class label. To predict the accuracy we adopted a leave-one-out cross-validation procedure protected for battery ID. Specifically, we picked one observation (features + battery ID + SOC label) at a time as the test set, then trained the classifier with all the remaining observations minus those with the same battery ID as the one in the test set. We iterated the process over all the data points and estimated the accuracy as the fraction of observations whose SOC label was identified correctly.

The procedure was implemented using three classification models: Gaussian Naïve Bayes ('Gaussian NB' in the remainder), k-Nearest Neighbors ('k-NN') and Linear Support Vector Classifier with one-vs-the-rest multiclass formulation ('lSVC'). For hyperparameter tuning we searched over $k = \{1, 2, 3\}$ nearest neighbors for the k-NN, and over penalty factor $C = \{0.01, 0.1, 1.0, 10.0\}$ with maximum number of iterations $= 10^4$ for the lSVC.

7.2. Results

Table 3 summarizes the results of the ten combinations feature set, normalization method and classifier that achieved the best prediction accuracy. The complete results for all the combinations are available as supplementary material (`classification-results.txt`).

We observe that in the best scenario it was possible to predict the state of charge with a promising 93.9% accuracy. Also note that, in general, the use of circuit parameters as features resulted in better accuracy than achieved with impedance values. Indeed the six best performing prediction models all use the parameters of the equivalent circuit model as features; by contrast, the best performing model based on impedance values (phase) only comes seventh in the table with an accuracy nearly 10% lower than obtained with the circuit parameters.

Table 3
Accuracy of SoC estimation for the top-10 combinations of feature set, normalization method and classifier.

Feature set	Normalization	Classifier	Classifier hyperparameters	Accuracy [%]
circparams	MinMax	k-NN	$k = 1$	93.9
circparams	MinMax	k-NN	$k = 2$	93.0
circparams	MinMax	k-NN	$k = 3$	92.2
circparams	Z-score	k-NN	$k = 1$	91.7
circparams	Z-score	k-NN	$k = 2$	90.9
circparams	Z-score	k-NN	$k = 3$	90.0
phase	MinMax	k-NN	$k = 1$	84.2
phase	MinMax	lSVC	$C = 10.0$	83.8
circparams	MinMax	Gaussian NB	–	83.0
circparams	Z-score	Gaussian NB	–	83.0

8. Conclusions

In this paper we have presented a hybrid approach for the critical task of estimating battery SOC from EIS data. Our method combines data-driven ML with equivalent-circuit models, and was validated experimentally on EIS data acquired from multiple lithium-ion cells.

The exploratory data analysis confirmed that EIS data can discriminate different states of charge with good reliability. Furthermore, we demonstrated that the use of circuit parameters instead of raw EIS impedance values resulted in better accuracy, lower dimensionality and altogether a more efficient model for SOC estimation. Compared with other data-driven approaches (e.g. DL) our solution also avoids the need for a costly and time-consuming training phase. We conclude that the approach presented here shows potential for translation into practical applications, such as online SOC estimation in battery-powered systems.

9. Computational reproducibility

We implement Reproducibility Enhancement Principles (REP) from [39] publishing on Code Ocean a working version of the code together with all data required to reproduce the finding of this work in a Code Capsule hosted on Code Ocean.

Code Ocean is a cloud-based executable repository and computational reproducibility platform allowing researchers to collaborate and execute code. The code capsule is available at https://codeocean.com/capsule/9473632.

A detailed description of the software is available in [40] and the complete source code and documentation are publicly accessible on GitHub.

CRediT authorship contribution statement

Emanuele Buchicchio: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **Alessio De Angelis:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **Francesco Santoni:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Validation. **Paolo Carbone:** Conceptualization, Methodology, Supervision. **Francesco Bianconi:** Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **Fabrizio Smeraldi:** Conceptualization, Methodology, Supervision, Writing – review & 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.

Data availability

The raw current and voltage data acquisition are published on Mendeley Data repository and available to download from https://data.mendeley.com/datasets/zdsgxwksn5.

The EIS dataset required to reproduce the above findings is published on Mendeley Data repository and available to download from https://data.mendeley.com/datasets/mbv3bx847g.

A detailed description of the EIS dataset is also available in [31].

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.energy.2023.128461.

References

[1] Chen Y, Kang Y, Zhao Y, Wang L, Liu J, Li Y, et al. A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards. J Energy Chem 2021;59:83–99. http://dx.doi.org/10.1016/j.jechem.2020.10.017, URL https://www.sciencedirect.com/science/article/pii/S2095495620307075.

[2] Waag W, Fleischer C, Sauer DU. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. J Power Sources 2014;258:321–39. http://dx.doi.org/10.1016/j.jpowsour.2014.02.064, URL https://www.sciencedirect.com/science/article/pii/S0378775314002572.

[3] Chemali E, Kollmeyer PJ, Preindl M, Emadi A. State-of-charge estimation of li-ion batteries using deep neural networks: A machine learning approach. J Power Sources 2018;400:242–55. http://dx.doi.org/10.1016/j.jpowsour.2018.06.104, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775318307080.

[4] Hu C, Ma L, Guo S, Guo G, Han Z. Deep learning enabled state-of-charge estimation of LiFePO4 batteries: A systematic validation on state-of-the-art charging protocols. Energy 2022;246:123404. http://dx.doi.org/10.1016/J.ENERGY.2022.123404, URL https://linkinghub.elsevier.com/retrieve/pii/S0360544222003073.

[5] Hossain M, Haque M, Arif M. Kalman filtering techniques for the online model parameters and state of charge estimation of the li-ion batteries: A comparative analysis. J Energy Storage 2022;51:104174. http://dx.doi.org/10.1016/j.est.2022.104174, URL https://www.sciencedirect.com/science/article/pii/S2352152X22002067.

[6] Babaeiyazdi I, Rezaei-Zare A, Shokrzadeh S. State of charge prediction of EV li-ion batteries using EIS: A machine learning approach. Energy 2021;223:120116. http://dx.doi.org/10.1016/j.energy.2021.120116, URL https://linkinghub.elsevier.com/retrieve/pii/S0360544221003650.

[7] Ng MF, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge and health of batteries using data-driven machine learning. Nat Mach Intell 2020;2(3):161–70. http://dx.doi.org/10.1038/s42256-020-0156-7, URL https://www.nature.com/articles/s42256-020-0156-7.

[8] Cui Z, Kang L, Li L, Wang L, Wang K. A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF. Energy 2022;259:124933. http://dx.doi.org/10.1016/J.ENERGY.2022.124933.

[9] Guo S, Ma L. A comparative study of different deep learning algorithms for lithium-ion batteries on state-of-charge estimation. Energy 2023;263:125872. http://dx.doi.org/10.1016/j.energy.2022.125872, URL https://www.sciencedirect.com/science/article/pii/S036054422202758X.

[10] Messing M, Shoa T, Ahmed R, Habibi S. Battery SoC estimation from EIS using neural nets. In: 2020 IEEE transportation electrification conference & expo. 2020, p. 588–93. http://dx.doi.org/10.1109/ITEC48692.2020.9161523.

[11] Tang X, Zhou Y, Gao F, Lai X. Joint estimation of state-of-charge and state-of-health for all cells in the battery pack using "leader-follower" strategy. eTransportation 2023;15:100213. http://dx.doi.org/10.1016/j.etran.2022.100213, URL https://www.sciencedirect.com/science/article/pii/S2590116822000583.

[12] Guo Y, Yang Z, Liu K, Zhang Y, Feng W. A compact and optimized neural network approach for battery state-of-charge estimation of energy storage system. Energy 2021;219:119529. http://dx.doi.org/10.1016/j.energy.2020.119529, URL https://linkinghub.elsevier.com/retrieve/pii/S0360544220326360.

[13] Zhang Y, Tang Q, Zhang Y, Wang J, Stimming U, Lee AA. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nature Commun 2020;11(1). http://dx.doi.org/10.1038/s41467-020-15235-7.

[14] Li D, Yang D, Li L, Wang L, Wang K. Electrochemical impedance spectroscopy based on the state of health estimation for lithium-ion batteries. Energies 2022;15(18). http://dx.doi.org/10.3390/en15186665, URL https://www.mdpi.com/1996-1073/15/18/6665.

[15] How DNT, Hannan MA, Lipu MSH, Ker PJ. State of charge estimation for lithium-ion batteries using model-based and data-driven methods: A review. IEEE Access 2019;7:136116–36. http://dx.doi.org/10.1109/ACCESS.2019.2942213, URL https://ieeexplore.ieee.org/document/8843918/.

[16] De Angelis A, Buchicchio E, Santoni F, Moschitta A, Carbone P. Uncertainty characterization of a practical system for broadband measurement of battery EIS. IEEE Trans Instrum Meas 2022;71. http://dx.doi.org/10.1109/TIM.2022.3156994.

[17] Hu X, Yuan H, Zou C, Li Z, Zhang L. Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus. IEEE Trans Veh Technol 2018;67(11):10319–29. http://dx.doi.org/10.1109/TVT.2018.2865664.

[18] Mirzaei H, Li Z, Parvini Y. Validation and sensitivity analysis of a fractional order model of a lithium ion battery via impedance spectra and temporal duty cycles. In: 2020 American control conference. 2020, p. 359–64. http://dx.doi.org/10.23919/ACC45564.2020.9147573.

[19] Hu X, Li S, Peng H. A comparative study of equivalent circuit models for li-ion batteries. J Power Sources 2012;198:359–67. http://dx.doi.org/10.1016/j.jpowsour.2011.10.013, URL https://www.sciencedirect.com/science/article/pii/S0378775311019628.

[20] Zhang S, Zhang C, Jiang S, Zhang X. A comparative study of different adaptive extended/unscented Kalman filters for lithium-ion battery state-of-charge estimation. Energy 2022;246:123423. http://dx.doi.org/10.1016/J.ENERGY.2022.123423, URL https://linkinghub.elsevier.com/retrieve/pii/S0360544222003267.

[21] Hannan MA, How DNT, Hossain Lipu MS, Mansor M, Ker PJ, Dong ZY, et al. Deep learning approach towards accurate state of charge estimation for lithium-ion batteries using self-supervised transformer model. Sci Rep 2021;11:19541. http://dx.doi.org/10.1038/s41598-021-98915-8, URL https://www.nature.com/articles/s41598-021-98915-8.

[22] Zhang M, Fan X. Review on the state of charge estimation methods for electric vehicle battery. World Electr Veh J 2020;11:23. http://dx.doi.org/10.3390/wevj11010023, URL https://www.mdpi.com/2032-6653/11/1/23.

[23] Wang X, Wei X, Zhu J, Dai H, Zheng Y, Xu X, et al. A review of modeling, acquisition, and application of lithium-ion battery impedance for onboard battery management. eTransportation 2021;7:100093. http://dx.doi.org/10.1016/j.etran.2020.100093, URL https://linkinghub.elsevier.com/retrieve/pii/S2590116820300515.

[24] Lipu MH, Hannan M, Hussain A, Ayob A, Saad MH, Karim TF, et al. Data-driven state of charge estimation of lithium-ion batteries: Algorithms, implementation factors, limitations and future trends. J Clean Prod 2020;277:124110. http://dx.doi.org/10.1016/j.jclepro.2020.124110, URL https://linkinghub.elsevier.com/retrieve/pii/S095965262034155X.

[25] Espedal IB, Jinasena A, Burheim OS, Lamb JJ. Current trends for state-of-charge (SoC) estimation in lithium-ion battery electric vehicles. Energies 2021;14:3284. http://dx.doi.org/10.3390/en14113284, URL https://www.mdpi.com/1996-1073/14/11/3284.

[26] Ge MF, Liu Y, Jiang X, Liu J. A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries. Measurement 2021;174:109057. http://dx.doi.org/10.1016/J.MEASUREMENT.2021.109057.

[27] Yang S, Zhang C, Jiang J, Zhang W, Zhang L, Wang Y. Review on state-of-health of lithium-ion batteries: Characterizations, estimations and applications. J Clean Prod 2021;314:128015. http://dx.doi.org/10.1016/j.jclepro.2021.128015, URL https://doi.org/10.1016/j.jclepro.2021.128015.

[28] Kondratiev S, Kuznietsov A. Application of embedded electrochemical impedance spectroscopy for on-board battery diagnostics. In: 2022 International symposium on power electronics, electrical drives, automation and motion. 2022, p. 41–5. http://dx.doi.org/10.1109/SPEEDAM53979.2022.9842275.

[29] Koseoglou M, Tsioumas E, Papagiannis D, Jabbour N, Mademlis C. A novel on-board electrochemical impedance spectroscopy system for real-time battery impedance estimation. IEEE Trans Power Electron 2021;36(9):10776–87. http://dx.doi.org/10.1109/TPEL.2021.3063506.

[30] Wang X, Wei X, Dai H, Wu Q. State estimation of lithium ion battery based on electrochemical impedance spectroscopy with on-board impedance measurement system. In: 2015 IEEE vehicle power and propulsion conference. 2015, p. 1–5. http://dx.doi.org/10.1109/VPPC.2015.7353021.

[31] Buchicchio E, De Angelis A, Santoni F, Carbone P, Bianconi F, Smeraldi F. Dataset on broadband electrochemical impedance spectroscopy of lithium-ion batteries for different values of the state-of-charge. Data in Brief 2022;45:108589. http://dx.doi.org/10.1016/j.dib.2022.108589, URL https://www.sciencedirect.com/science/article/pii/S235234092200796X.

[32] Relan R, Firouz Y, Timmermans J-M, Schoukens J. Data-driven nonlinear identification of li-ion battery based on a frequency domain nonparametric analysis. IEEE Trans Control Syst Technol 2017;25(5):1825–32. http://dx.doi.org/10.1109/TCST.2016.2616380.

[33] Pintelon R, Schoukens J. System identification: A frequency domain approach. John Wiley & Sons; 2012.

[34] Orazem ME, Tribollet B. Electrochemical impedance spectroscopy. 2008, p. 383–9,

[35] Jossen A. Fundamentals of battery dynamics. J Power Sources 2006;154(2):530–8. http://dx.doi.org/10.1016/j.jpowsour.2005.10.041.

[36] Dhillon S, Kant R. Theory for electrochemical impedance spectroscopy of heterogeneous electrode with distributed capacitance and charge transfer resistance. J Chem Sci 2017;129:1277–92.

[37] Crescentini M, De Angelis A, Ramilli R, De Angelis G, Tartagni M, Moschitta A, et al. Online EIS and diagnostics on lithium-ion batteries by means of low-power integrated sensing and parametric modeling. IEEE Trans Instrum Meas 2021;70. http://dx.doi.org/10.1109/TIM.2020.3031185.

[38] Bretherton CS, Smith C, Wallace JM. An intercomparison of methods for finding coupled patterns in climate data. J Clim 1992;5(6):541–60. http://dx.doi.org/10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2, URL https://journals.ametsoc.org/view/journals/clim/5/6/1520-0442_1992_005_0541_aiomff_2_0_co_2.xml.

[39] Stodden V, McNutt M, Bailey DH, Deelman E, Gil Y, Hanson B, et al. Enhancing reproducibility for computational methods. Science 2016;354(6317):1240–1. http://dx.doi.org/10.1126/science.aah6168, arXiv:https://www.science.org/doi/pdf/10.1126/science.aah6168.

[40] Buchicchio E, De Angelis A, Santoni F, Carbone P, Bianconi F, Smeraldi F. LiBEIS : A software tool for broadband electrochemical impedance spectroscopy of lithium-ion batteries. Softw Impacts 2022;100447. http://dx.doi.org/10.1016/j.simpa.2022.100447, URL https://www.sciencedirect.com/science/article/pii/S2665963822001312.