# SocHAP: a new data driven explainable prediction of battery state of charge

Théo Heitzmann, Ahmed Samet, Tedjani Mesbahi, Cyrine Soufi, Inès Jorge, and Romuald Boné

Université de Strasbourg, 67000, France Institut National des Sciences Appliquées (INSA Strasbourg), 67000, France CNRS, ICube Laboratory UMR 7357, Strasbourg, 67000, France firstname.lastname@insa-strasbourg.fr

Abstract. The performance and driving range of electric vehicles are largely determined by the capabilities of their battery systems. To ensure optimal operation and protection of these systems, Battery Management Systems rely on key information such as State of Charge, State of Health, and sensor readings. These critical factors directly impact the range of electric vehicles and are essential for ensuring safe and efficient operation over the long term. This paper presents the development of a battery State of Charge estimation model based on a 1-D convolutional neural network. The data used to train this model are theoretical operating data as well as driving cycles of lithium-ion batteries. An Explainable Artificial Intelligence method is then applied to this model to verify the physical behavior of the black box model. Finally, a testing platform is currently under development to assess the effectiveness of the State of Charge estimation model. Our explainable model, called SocHAP, is compared to other contemporary methods to evaluate its predictive accuracy.

**Keywords:** Explainable AI · Battery degradation · Deep learning · Electric vehicle · Lithium-ion battery cell · State of Charge.

# 1 Introduction

The European Commission has approved the goal to reach a "net-zero" Greenhouse Gas (GHG) emissions level by 2050. Currently, more than 20% of the EU's GHG emissions are related to transport and almost 50% of those are caused by passenger vehicles. As opposed to other energy-intensive sectors, such as electricity generation and industry, emissions from transportation activities have been growing in the past years. Therefore, effective measures to reduce these emissions are urgently needed. The market for electric vehicles has been growing over the last few years and is becoming more and more interesting thanks to significantly reduced CO2 emissions compared to a thermal vehicle. However, if the performance of electric vehicles has been greatly improved thanks to the late research advances, there are still progress to be made concerning the range and the life span of their batteries.

The range of an electric vehicle is directly linked to its energy source, which is mainly a battery composed of cells. Lithium-ion batteries are becoming overstretched on the electric vehicle market. Indeed, they have a very high energy density, a long life span and a wide range of operating temperatures [8]. The 18650 lithium-ion cell has reached a nominal capacity of 2.4 Ah which corresponds to an energy density of 200 Wh/kg [22]. In order to improve the driving range of electric vehicles, it is necessary to slow down the aging process of batteries. Indeed, as the battery ages, its storage capacity decreases, its internal resistance increases and therefore, its global performances and driving range degrade. The Battery Management System (BMS) manages the charging and discharging of the cells according to the State of Charge (SoC) and the State of Health (SoH). The advancement of battery health assessment is critical for building a clean and sustainable society, but limited access to sufficient battery aging data poses a significant challenge. This work presents an alternative solution for producing large-volume, high-quality aging datasets, which avoids the need for carrying out large-scale aging experiments in the laboratory. Although several public datasets [5] such as MIT [20] or NASA exist [19], every dataset is intended to investigate the impact of several factors on the ageing process. In this paper, we introduce a new dataset as one of our contributions. The data used in this study was obtained from a test bench installed in a research platform at INSA Strasbourg, where the cells were cycled using the WLTC (Worldwide harmonized Light vehicles Test Cycles) cycles. The data is in the form of time series measured by sensors during the cycling of the cells.

In automotive applications, SoC is considered one of the most important parameters for maintenance. Predicting an accurate value of the SoC avoids incidents such as overcharging and deep discharge of the battery. It takes values between 0 and 1 that directly indicates the amount of energy left in a battery to power an electrical device. While some studies have focused on predicting SoC with a machine learning approach, few have attempted to explain the results with an explainable model.

Therefore, our contributions in this paper are as follows: (i) a newly built dataset that contains time series data of battery cell meant for SoC prediction;(ii) a SHAP-based explainable on a top of a convolutional neural network approach, called SocHAP, to predict the SoC using multiple features; (iii) the development of a test bench implementing our estimation model for real-time battery cell SoC estimates. The paper is organized as follows, we begin by introducing the study context, then present the results of the SoC estimation training of the model. Next, we use Explainable Artificial Intelligence (XAI) to recover the physical reality of the functioning of the SoC of a battery learned by the model. Finally, we present a test bench for cell SoC estimation.

# 2 Battery basics and related works

A lithium-ion battery is composed of several cells connected in series or in parallel. Each cell is composed of three main elements: a positive electrode (the cathode), a negative electrode (the anode) which are separated by an electrolyte [6,8].

In order to control the cells in their operation, the BMS requires several pieces of information related to the battery. While voltage, current and temperature are easily accessible via a series of sensors, SoC and SoH are cell state quantities that cannot be directly measured.

The SoC is a metric used to describe the amount of energy left in the energy storage system [21]. SoC is not a physical quantity that can be directly measured; instead, it can only be estimated by measuring strongly correlated proxy quantities such as voltage, current, and temperature [13]. Typically, SoC is expressed as a value in the range of 0 to 1, and it is defined in the literature as the ratio of the available amount of charge to the maximum amount of charge of the battery. It is computed as follows:

$$SoC = \frac{C_{rest}}{C_{nom}} \tag{1}$$

$$C_{rest} = \int i.dt \tag{2}$$

where  $C_{rest}$  is the remaining releasable capacity of the battery at a certain level of charge and  $C_{nom}$  is the nominal capacity of the battery.  $C_{rest}$  is also equal to the integral of the current *i* over time *t*. Accurate measurement of SoC is crucial for determining the appropriate charging and discharging strategies of batteries and thus avoiding any permanent damage to their internal structure [16]. The SoH of a battery is defined by the following relation :

$$SoH = \frac{C_{act}}{C_{nom}} \tag{3}$$

where  $C_{nom}$  represents the nominal capacity of the battery and  $C_{act}$  the current total capacity of the battery. This quantity gives direct information about the aging and degradation of a battery [4, 12]. In the context of electric mobility, a battery is considered no longer usable and must be replaced when its SoH value reaches 80% [17].

In the literature, a few data-driven approaches and techniques have already proven to be effective in terms of accurately estimating the SoC of a lithium-ion battery cell. One of the first approaches is to use a Feed-forward Neural Network (FNN) with regression to estimate the SoC [9]. This model consists of two hidden layers with 5 neurons each, and takes voltage, current, and temperature measurements of a battery cell in the form of time series as input. Chemali et al. introduced a Long Short-Term Memory - Recurrent Neural Network (LSTM-RNN) to predict the SoC value of a cell from time series of voltage, current and temperature [3]. The model is composed of 500 stacked LSTM cells. Machine learning and deep learning models have demonstrated good performance

in classification and regression tasks across various domains. Typically, accuracy is used as an index of model quality. However, these models are often considered "black boxes" [1] because they take input features and produce output values without providing any information about the reasoning process that underlies their predictions. It is essential to provide users with explanations, and therefore, Explainable Artificial Intelligence (XAI) is employed.

Gu et al. [7] proposed an explainable model based on the SHapley Additive exPlanation (SHAP) method to determine the importance of the input parameters in a battery cell SoC estimation model. The prediction model, called SW-SHAP-LSTM, is based on a recurrent network of LSTM cells.

In this paper, we propose an explainable approach based on a Convolutional Neural Network (CNN) to estimate the SoC of a lithium-ion battery cell.

The interest in applying CNNs to time series data is that they would be able to learn filters that represent repeated patterns in the series [2, 11]. The estimation model would get a better understanding of the relationships hidden in the input time-series.

We also introduce a dataset of lithium-ion cell usage cycles reproducing driving cycles of electric vehicles in urban environment. Finally, an implementation of the estimation model has been performed on an embedded system.

# 3 SHAP Convolutional Neural Network for SOC prediction (SocHAP)

In this section, we introduce our SocHAP model that is described in Figure 1. First, we develop a Convolution Neural Network that estimates the SoC in realtime from a window of size W = 100 time points, consisting of voltage, current and temperature measurements of the cell. Then, the SHAP method is applied to compute the importance of each feature for a particular prediction made by the model.

# 3.1 Data used in the approach and introduction of a new dataset

The data used for the training of the degradation models and experimentation phase are two sets of lithium-ion cell degradation data. The first set is part of our contributions, which consists of battery cell usage data in the form of driving cycles. We used lithium ferrophosphate (LFP) battery cells APR18650M1A from A123 Systems to perform our tests. These are the same cells used in the MIT dataset. The tests were generated using the Worldwide Harmonized Light Vehicle Test Procedure (WLTP). A battery cycler is programmed to wear out cells through cycles consisting of a charge phase and a discharge phase. The charging phases are carried out using the classic method of charging cells CC-CV: a constant current charge followed by constant voltage charge. The discharge phase is constituted by using the driving cycles described previously by linking several of these cycles. This phase also integrates the phenomenon of regenerative braking



SocHAP: a new data driven explainable prediction of battery state of charge

Fig. 1. Explainable Convolutional Neural Network to estimate the SOC.

which recharges the cell during the driving cycle by converting the energy dissipated during the deceleration and braking of the vehicle. This process is repeated until the cell reaches a degradation of 80%. The test system used to generate this data comes from Basytec (Basytec XCTS) and allows twelve simultaneous battery tests. This data set is a useful resource for training battery degradation models and also for data processing in the context of electric mobility. A few samples of our dataset are available in the provided link <sup>1</sup>. Time series of cell voltage, current, and temperature measurements for one cycle of the dataset can be seen in Figures 2, 3, 4. The SoC time series (Figure 5) was computed using the integral of the current.





Fig. 3. INSA Cycle : Current

<sup>1</sup> https://github.com/thtzmn/INSA\_LFP\_DATASET

ICCS Camera Ready Version 2023 To cite this paper please use the final published version: DOI: 10.1007/978-3-031-36030-5\_37 5



The second dataset used is the one released by the Massachusetts Institute of Technology (MIT) [20], which contains data on the theoretical use of the battery with constant current discharges. To date, this is the largest public dataset containing lithium-ion cell cycling information.

# 3.2 State of Charge explainable prediction

The model used to estimate the SoC of battery cells is a convolutional neural network (CNN).

Our SocHAP model takes as input a sliding window of information of size W = 100 related to the battery usage, see Table 1. These data include cell voltage V, current I, and temperature T.

Features		Labels
Voltage (V)	$U_0, U_1, \ldots U_{98}, U_{99}$	
Current (A)	$I_0, I_1, \dots I_{98}, I_{99}$	SoC value at $t=99$
Temperature (°C)	$T_0, T_1, \ldots T_{98}, T_{99}$	

Table 1. Sliding window of 100 time points to perform a SoC estimation

A 1-D convolution layer with f=32 filters of size l=5 and a ReLU (Rectified Linear Unit) activation followed by a 1-D Average Pooling layer of size l = 3 allows to perform a feature selection and data reduction operation. A flattening layer is then applied with a layer of Dropout with a coefficient of 0.05 to avoid hyper-parameter overfitting in the training process. Three fully connected layers of sizes n = 512, n = 256 and n = 128 respectively with ReLU activation, and a final layer of size 1 are used to estimate the SoC value.

The values of the hyperparameters have been selected after several comparisons of different network configurations, see Table 2.

Convolution layer	layers of neurons	MAE	RMSE
f = 16	2 fully connected : $n_1 = 256, n_2 = 128$	0.045	0.0246
f = 16	3 fully connected : $n_1 = 512, n_2 = 256, n_3 = 128$	0.0269	0.0156
f = 32	2 fully connected : $n_1 = 256, n_2 = 128$	0.0424	0.0213
f = 32	3 fully connected : $n_1 = 512, n_2 = 256, n_3 = 128$	0.0095	0.0139

Table 2. Comparison of the different configurations of the CNN estimation model

This model was trained using the Mean Absolute Error (MAE) as a loss function :

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}$$
(4)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}$$
(5)

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} |\frac{y_i - \hat{y}_i}{y_i}|$$
(6)

with N being the number of samples,  $\hat{y}_i$  and  $y_i$  the estimated value of sample *i* and the actual value of sample *i* respectively. We also computed the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) to compare the estimation performances of our model with those of the state-of-the-art over the test samples of our training set. The AdaMax [10] gradient descent algorithm was used to train the model. It is more suitable than Adam for learning time-variant processes since it adapts the learning rate for each parameter during training.

An Explainability method is applied to determine the influence of the current as an input parameter on the SoC estimates during driving cycles. The objective is to recover the physical reality of the battery's SoC functioning learned by the model. SHAP was introduced by Lundberg and Lee in 2017 [14]. This explanation method belongs to the family of additive feature contribution methods. It assigns to each feature a coefficient or an importance value for a particular prediction f(x), with f a prediction model and x a particular input of this model.

The final explanation take the form of a linear combination of the individual contributions of the input features x.

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{7}$$

where g(z') represents a local approximation function of the original model f and  $\phi_i$  represents the contribution of feature i to the prediction f(x).  $z' \in \{0, 1\}^M$  equals 1 when a feature is observed or 0 otherwise and M is the number of

simplified input features. The average of  $\phi_i$  is often used as a bias value for  $\phi_0$ . Shapley values are used to estimate the contribution  $\phi_i$  of each feature.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup i) - f(S)]$$
(8)

where f is the prediction model, F the set of all input features, and S the feature subsets.

# 4 Experiments and results

# 4.1 Evaluation of the SocHAP model

Before applying the XAI method to explain the estimates of the SoC estimation model, we first compared its performance with that of the state-of-the-art models. All models were trained with the same samples. Slightly more than 2 500 000 samples were used for the training phase of the models. 75% as training set, 15% as validation set and 15% as test set to evaluate the performance of the models. The training was performed with Tensorflow 2 on Python 3. Their estimation performances were then compared using the MAE, the RMSE and MAPE as indicators. The training set is composed of 70% samples of our dataset with a sampling time of 0.5 seconds and 30% of MIT data with a sampling time of 5 seconds. We compare SocHAP to two other architectures. The first is an LSTM based estimation model and the second is a feed-forward neural network.

The performance of the three SoC estimation model architectures can be observed in the following Table 3:

Model	MAE	RMSE	MAPE
SocHAP	0.0096	0.0172	17.26%
He et al. FNN [9]	0.0652	0.0963	56.68%
Chemali et al. LSTM-RNN [3]	0.0106	0.0203	9.10%

 Table 3. Comparison of performances between the different architectures.

Considering only the MAE and RMSE, our SocHAP approach obtains the best estimation performances. The LSTM-based model obtains the best score for the MAPE. This means that the LSTM-based model produces more accurate estimates when the values to be estimated of SoC are close to 0 if we consider the fact that our model has a lower MAE.

Estimates of the model over the entire cycles of the MIT and INSA datasets are observable in Figures 6, 7.



Fig. 6. SoC estimation of the CNN over a Fig. 7. SoC estimation of the CNN over a MIT cycle driving cycle

#### 4.2 Explainability of the model's estimations

We evaluate the explanation given by SHAP over a driving cycle to understand why the model makes a certain estimate of SoC from the window of battery cell voltage, current, and temperature given as input.

By applying the SHAP and signal processing algorithms, we observed a relation between the increase of the current and the decrease of the SoC in the battery time series of a driving cycle (see Figure 8). Based on our findings, we can conclude that the model learned the physical behavior of the battery's SoC as the SoC can be defined as the integral of the current, the amount of energy stored compared to the battery's maximum capacity during its use.



Fig. 8. Superposition of current contributions to SoC estimates during part of a driving cycle.

Fig. 9. Representation of the studied driving cycle.

Now that the model is validated, we can try to find the causes of the battery degradation in a driving cycle by analyzing the values of the input parameters contributions to the SoC. Figure 10 represents the evolution of SHAP values of temperature for the same driving cycle as in Figure 9.



Fig. 10. Representation of the temperature evolution during the cycle with the associated SHAP values

Temperature plays a very important role in the life cycle of a lithium-ion battery. We consider as a general rule that the acceptable temperature range for the use of a lithium-ion battery is from -20 °C to 60 °C [15]. However, it has been shown that it is rather between 15 °C and 35 °C [18]. Any use of the battery outside this temperature interval can lead to an accelerated degradation of the battery. In addition, high operating temperatures will cause capacity and power losses.

The cycler used to perform battery ageing tests using driving cycles does not allow temperature control of the test chambers. The temperatures of the test chambers are therefore directly related to the temperature of the room where the battery cycling device is installed.

Based on the experiment carried out to study the temperature of the cell during a driving cycle, we initially observed that the temperature strongly increases during the discharge phase. This is due to the high current demands set by the cycler to simulate the driving cycle.

Moreover, the SHAP values related to the battery cell temperature are maximum during this strong temperature increase. By studying the SHAP values of the temperature inputs, we can isolate critical moments of battery use (e.g. high current demands, fast charging of batteries etc.). These moments of extreme battery use could then be identified and recorded. It would then be possible to compare their influence in a model predicting the SoH of the battery to quantify the degradation associated with these patterns.

### 4.3 Real-time cell SoC test bench under development

One of the perspectives of our work is to implement our SoC estimation model in a test bench. The objective is to be able to compare the estimates of our model with the real SoC of a battery cell in real conditions. Currently, we are still in the testing and adjustment phase.

This embedded estimation bench integrates a Jetson Nano which is a lowconsumption microcomputer powerful enough to deploy our model. Moreover, the Jetson Nano communicates via I2C with sensors in order to retrieve measurements of voltage, current and temperature during the cycling of the cell to estimate its SoC. These data are also stored for further experimentation, such as computing the real SoC value of the battery cell using the integral of the current method.

The bench displays through an LED screen the SoC value of a lithium-ion battery cell in real time with the measurements of the voltage, current and temperature of the battery cell (Figure 11).

The cell can be charged or discharged by connecting its power circuit either to a stabilized power supply or rheostat type resistor using banana plugs.

The operation of the bench is described by the diagram Figure 12.

# 5 Conclusion

This paper introduces a publicly available dataset containing measurements of LFP battery cell usage during driving cycles. The dataset was used to train our proposed explainable State of Charge (SoC) estimation approach, SocHAP. Additionally, an experiment was conducted to demonstrate the explainability of the SoC estimation for a driving cycle using SHAP values. The results of this experiment suggest that patterns of battery aging acceleration can be identified by analyzing temperature data using SHAP values.

The future challenges of this work include first completing the development of the embedded SoC estimator. The second objective is to further enhance the explainability aspect of the lithium-ion battery degradation models. The aim is to make the explanations provided by the explainability methods more interpretable and to provide advice to the electric vehicle users on how to use the battery efficiently. These recommendations could be applied to both the battery charging phase and the driving phase with the goal of extending the life cycle of batteries as well as increasing the autonomy of the vehicle.

**Acknowledgements** This work was carried out as part of HALFBACK and VEHICLE project, sponsored by INTERREG V A Upper Rhine Programme, FEDER and Franco-German regional funds (Bade-Wurtemberg, Rhénanie-Palatinat and Grand-Est).

This paper has received funding from the European Union under the program Horizon Europe and the innovation program under GAP-101103667.



Power circuit for charging or discharging the cell Measurement of input Temperature sensor Data transfer with an Jetson Nano with the implemented SOC estimation model SOC display of the cell LCD screen

Fig. 12. Operating diagram of the SoC estimator

# References

- 1. Alicioglu, G., Sun, B.: A survey of visual analytics for explainable artificial intelligence methods. Computers & Graphics 102, 502–520 (2022)
- 2. Borovykh, A., Bohte, S., Oosterlee, C.W.: Conditional time series forecasting with convolutional neural networks (2018)
- 3. Chemali, E., Kollmeyer, P.J., Preindl, M., Ahmed, R., Emadi, A.: Long short-term memory networks for accurate state-of-charge estimation of li-ion batteries. IEEE Transactions on Industrial Electronics 65(8), 6730–6739 (2018)
- 4. Chen, Z., Mi, C.C., Fu, Y., Xu, J., Gong, X.: Online battery state of health estimation based on genetic algorithm for electric and hybrid vehicle applications. Journal of Power Sources 240, 184–192 (2013)
- 5. dos Reis, G., Strange, C., Yadav, M., Li, S.: Lithium-ion battery data and where to find it. Energy and AI 5, 100081 (2021)
- 6. Gomadam, P.M., Weidner, J.W., Dougal, R.A., White, R.E.: Mathematical modeling of lithium-ion and nickel battery systems. Journal of Power Sources 110(2), 267-284 (2002)
- 7. Gu, X., See, K., Wang, Y., Zhao, L., Pu, W.: The sliding window and shap theory—an improved system with a long short-term memory network model for state of charge prediction in electric vehicle application. Energies 14(12) (2021)

SocHAP: a new data driven explainable prediction of battery state of charge

- Hannan, M., Lipu, M., Hussain, A., Mohamed, A.: A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations. Renewable and Sustainable Energy Reviews 78, 834–854 (2017)
- He, W., Williard, N., Chen, C., Pecht, M.: State of charge estimation for li-ion batteries using neural network modeling and unscented kalman filter-based error cancellation. International Journal of Electrical Power & Energy Systems 62, 783– 791 (2014)
- Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014), https: //arxiv.org/abs/1412.6980
- Koprinska, I., Wu, D., Wang, Z.: Convolutional neural networks for energy time series forecasting. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2018). https://doi.org/10.1109/IJCNN.2018.8489399
- Lipu, M.H., Hannan, M., Hussain, A., Hoque, M., Ker, P.J., Saad, M., Ayob, A.: A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations. Journal of Cleaner Production 205, 115–133 (2018)
- Lipu, M.H., Hussain, A., Saad, M., Ayob, A., Hannan, M.: Improved recurrent narx neural network model for state of charge estimation of lithium-ion battery using pso algorithm. In: 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). pp. 354–359. IEEE (2018)
- Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions (2017)
- Ma, S., Jiang, M., Tao, P., Song, C., Wu, J., Wang, J., Deng, T., Shang, W.: Temperature effect and thermal impact in lithium-ion batteries: A review. Progress in Natural Science: Materials International 28(6), 653–666 (2018)
- Ng, K.S., Moo, C.S., Chen, Y.P., Hsieh, Y.C.: Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. Applied Energy 86(9), 1506–1511 (2009)
- Perner, A., Vetter, J.: 8 lithium-ion batteries for hybrid electric vehicles and battery electric vehicles. In: Scrosati, B., Garche, J., Tillmetz, W. (eds.) Advances in Battery Technologies for Electric Vehicles, pp. 173–190. Woodhead Publishing Series in Energy, Woodhead Publishing (2015)
- Pesaran, A., Santhanagopalan, S., Kim, G.H.: Addressing the impact of temperature extremes on large format li-ion batteries for vehicle applications (presentation) (5 2013)
- Saha, B., Goebel, K.: Battery data set. NASA Ames Prognostics Data Repository (2007)
- Severson, K., Attia, P., Jin, N., et al.: Data-driven prediction of battery cycle life before capacity degradation. Nature Energy 4 (2019)
- 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 6, 35957–35965 (2018)
- 22. Yoshio, M., Brodd, R.J., Noguchi, H.: Lithium-ion batteries: Science and Technologies. Springer (2009)