35th International Electric Vehicle Symposium and Exhibition (EVS35)
Oslo, Norway, June 11-15, 2022

Li-ion Battery State-of-Charge estimation algorithm with CNN-LSTM and Transfer Learning using synthetic training data

M. Azkue[1,2], M. Lucu[1], L. Oca[2], E. Martinez-Laserna[1], U. Iraola[2]

[1] *Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA). P° J.M. Arizmendiarrieta, 2, 20500 Arrasate-Mondragón, Spain*

[2] *Mondragon Unibertsitatea, Department of Electronics and Computer Science, 20120 Hernani, Spain*

Summary

The development of State-of-Charge (SoC) algorithms for Li-ion batteries involves carrying out different laboratory tests with the money and time that this entails. Furthermore, such laboratory labours must typically be repeated for each new Li-ion cell reference. In order to minimise this issue, this work proposes a new approach for developing SoC algorithms, using an Artificial Neural Network in combination with a Transfer Learning method. The latter technique will make possible to take advantage of the data generated for previously tested cell references and use it for the development of a SoC estimation algorithm for a new cell reference. This work provides a proof-of-concept for the proposed approach, using synthetic data generated from electrochemical models, which describes the behaviour of different Li-ion cell references.

Keywords: battery SoC (state of charge), energy storage, training, lithium battery, smart

1 Introduction

In order to ensure a safe and optimal use of Li-ion batteries, it is necessary to estimate their State-of-Charge (SoC) accurately [1]. The development of SoC estimation algorithms requires performing several tests on Li-ion cells at different operating conditions (temperature, current or State-of-Health, SoH), typically in laboratory environment, which is time and cost intensive. An important issue in this field of research is that a SoC estimation model developed for one particular cell reference is not necessarily adequate for a different cell reference (different manufacturer, size, chemistry, etc.). This implies that every time a SoC estimation algorithm has to be developed for new cells, all the laboratory testing labours have to be started from scratch, inducing an important and periodical waste of time and economical resources [2].

The main objective of this research is to propose a new approach for developing SoC estimation algorithms, taking advantage of the data and prior knowledge generated from previously tested or deployed cell references. In this work, this is tackled using Artificial Neural Networks (ANN) algorithms in combination with Transfer Learning (TL) methods. Such methods consist of: i) first, training the ANN with data obtained from a previously tested and/or deployed cell reference, and ii) then retrain the ANN using a reduced amount of data from the new cell reference, deriving the corresponding SoC estimation algorithm. In this way, the obtained SoC estimation algorithms are expected to provide accurate estimations, while reducing the laboratory labours and/or the field data gathering period required for the new cell references [3]–[5]. This process is depicted in Figure 1.

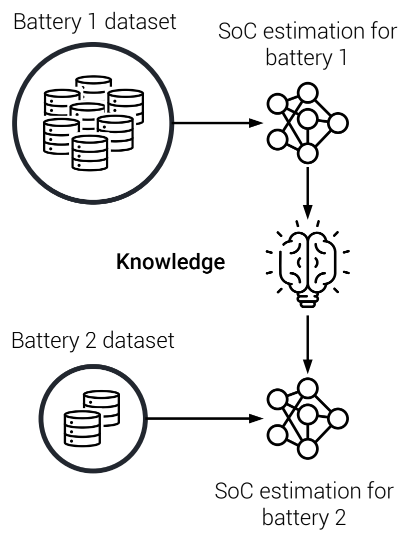

Figure 1. Conceptual flowchart summarising the SOC estimation algorithm development method, based on TL.

In this work, a proof-of-concept of this new SoC estimation algorithm approach is provided, using for that two different datasets which represent the cycling behaviour of two different cell references. The first dataset consists on synthetic data generated from a widely accepted electrochemical model. The second dataset, used to apply the TL method, was generated in laboratory environment by cycling real battery cells.

2 SoC estimation algorithm

The SoC estimation algorithms developed in this work are based on ANN. For this application, Convolutional Neural Networks (CNN) in combination with Recurrent Neural Networks (RNN) are of especial interest, as CNN are good for feature extraction between different inputs and RNN are particularly adequate for time-series data [6], [7]. Moreover, CNNs are a type of neural network specialised in reducing the number of parameters in the processing data due to their capacity to automatically detect the important features in the input data [8], this is of particular interest as it will extract the most important features and look for relationships between the different input parameters. Figure 2 illustrates an example of a 1D CNN architecture for time series processing.

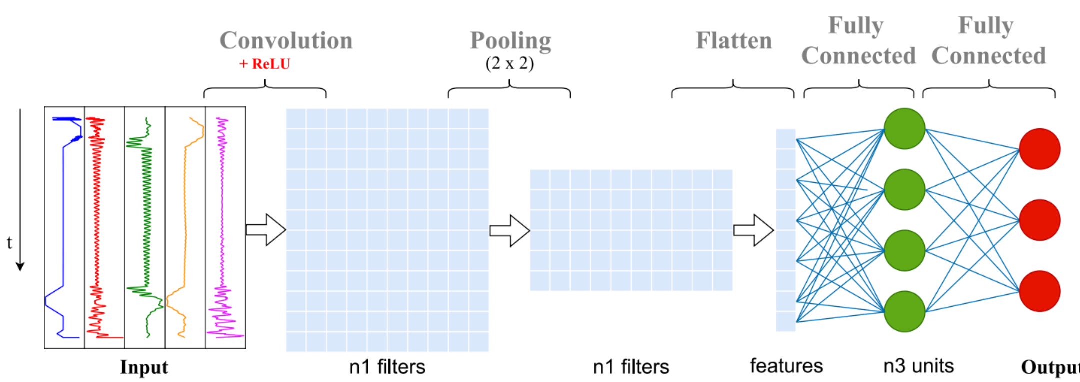

Figure 2. CNN architecture [9]

On the other hand, RNNs contain memory units in which they store the states estimated in the previous timesteps and take them into account to improve future estimations, as illustrated in Figure 3 [10], this type of neural network is of especial interest for the application as it will extract the temporal relationships of the timeseries that make up both the input and output data. More specifically, in this work, Long Short-Term Memory (LSTM) neurons will be used, as their memory capabilities are the solidest ones between different recurrent neurons [11].

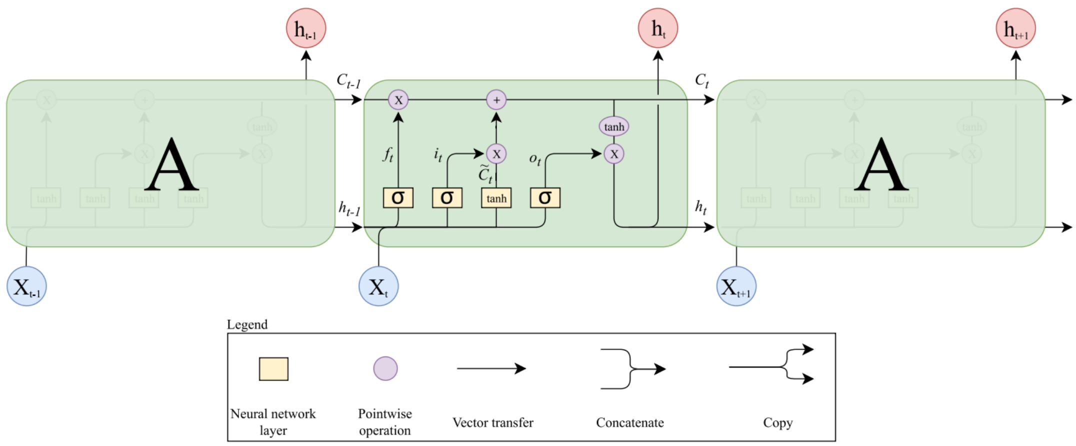

Figure 3. LSTM architecture [9]

To develop the cell SoC estimation algorithm, four different input variables were considered, namely the cell voltage, current, temperature and previous battery SoC estimation. The output of the algorithm was the SoC of the cell. In Figure 4, a scheme of the proposed algorithm is depicted.

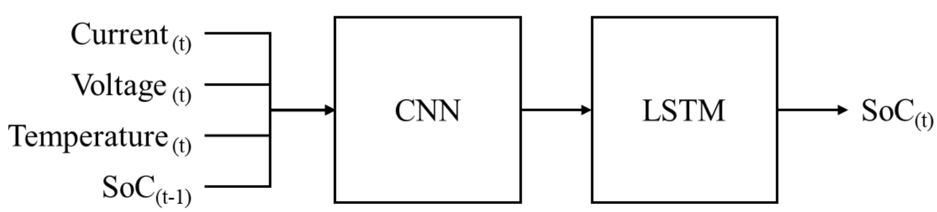

Figure 4. Proposed algorithm scheme.

As previously mentioned, a TL method was applied to develop a SoC estimation algorithm able to take advantage of the data and prior knowledge generated from previously tested or deployed cell references [12]. In fact, TL is a successful method which allows to apply in a new task the knowledge and skills acquired in previous tasks [13]. To do this, the neural network is first trained with one database. Afterwards, this neural network is retrained (that is, the internal weights of the neural network are recalculated) with a second database. In this work, the TL was carried out by completely retraining the neural network, without blocking any of the layers, in other words, retraining all the weights that compose the different layers of the algorithm.

3 Datasets

Two different datasets were used in this work, each one corresponding to a specific cell reference. Both datasets include different variables for each input data sample (temperature, voltage, current and previous SoC estimate) to allow the ANN to quantify their impact on the underlying SoC of the cell.

The first dataset was generated using a widely accepted electrochemical model to simulate the behaviour of an LCO-based cell, known as the Doyle cell [14]. A specific cycling profile was designed, which consists of different types of charging and discharging phases (constant current, constant voltage, or different charging/discharging levels), intercalated with pauses of different durations. This cycling profile was especially designed in order to facilitate the evaluation of the performances of the developed SOC estimation algorithm, in such a way that the weaknesses of the algorithm could be easily identified. The cycling profile was applied at several temperature values from 0 ºC up to 45 ºC. For each one of these temperatures, charge and discharge cycles were simulated at different C-rates, ranging from 0.5C to 4C in discharge and 0.5C to 1C in charge, as 1C is the maximum charging current. Table 1 summarises the cycling conditions and the cycling profile is represented in Figure 5.

Table 1: Cycling conditions involved in the dataset generated using the electrochemical model of the Doyle cell.

Temperature	C-rate DCH	C-rate CHA
45 ºC	0.5 – 1 – 2 – 4	0.5 - 1
25 ºC		
10 ºC		
0 ºC		

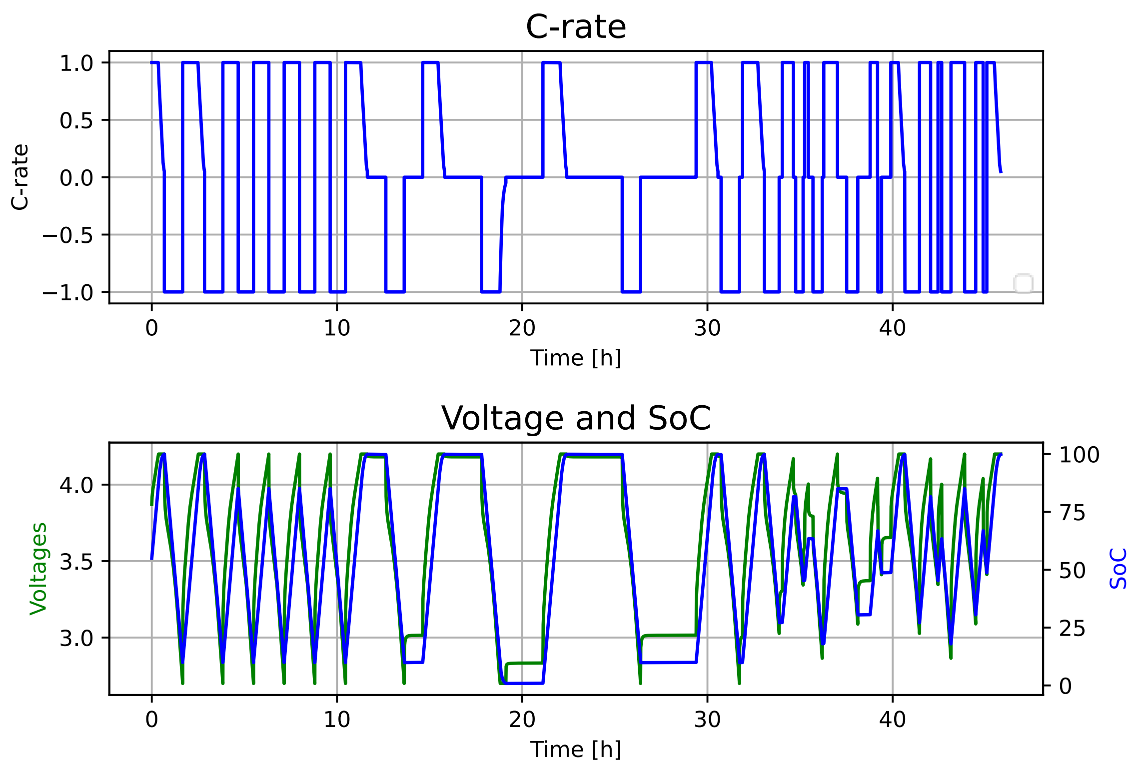

Figure 5. Designed profile: C-rate, SoC and Voltage vs. Time @ 25 ºC and 1C CHA/DCH

The second dataset was generated in laboratory environment, carrying out different cycling tests on commercial NMC-based 58 Ah cells. These tests included capacity, Hybrid Pulse Power Characterization (HPPC) and quasi-OCV tests. The capacity test consisted on three charge/discharge cycles at 1C. The HPPC was carried out using current pulses at different currents, varying from 0.5C to 3C at different SoC levels. The quasi-OCV test was done discharging and charging the cell at 0.05C.

In addition, the cells were tested using the above-mentioned designed profile to which WLTC profiles were applied at different SoCs, at the end of the profile. Three different cells were tested in each condition to ensure the repeatability of the results.

4 Results and Discussion

As previously mentioned, in this work a baseline ANN model was developed using the data obtained synthetically using the electrochemical model of the Doyle cell (hereinafter called “baseline model”), and then the TL method was used to retrain such model with the data generated in the laboratory using real commercial cells.

In addition, for comparative purposes and in order to evaluate the benefits of applying the TL method, a new ANN model with identical configuration was trained using only NMC cells data (hereinafter called “NMC model”). This section depicts the results obtained with each of the models.

4.1 Baseline model

As mentioned in Section 2, CNN and LSTM architectures will be combined to develop the SoC estimation algorithm. More specifically, a CNN layer using 16 filters and a bidirectional LSTM layer with 5 neurons will be used.

To train, validate and test this model, the data corresponding to the Doyle cell, previously introduced in section 3, was used. This data was divided as shown Table 2.

Table 2. Baseline model train/validation/test configuration.

Dataset	**Conditions**
Training data	0 ºC - 0.5C CHA - 0.5C DCH 10 ºC - 0.5C CHA - 0.5C DCH 10 ºC - 1C CHA - 4C DCH 25 ºC - 0.5C CHA - 0.5C DCH 25 ºC - 1C CHA - 1C DCH 25 ºC - 1C CHA - 4C DCH 45 ºC - 1C CHA - 4C DCH
Validation data	0 ºC - 1C CHA - 2C DCH 10 ºC - 1C CHA - 1C DCH
Test data	0 ºC - 1C CHA - 1C DCH 10 ºC - 1C CHA - 2C DCH 25 ºC - 1C CHA - 2C DCH 45 ºC - 1C CHA - 2C DCH

The algorithm was trained using the “training data” for 500 epochs, in order to determine the optimal weights of the developed ANN model. The model with the minimum Mean Absolute Error (MAE) in the validation data during training phase was saved. The algorithm was then evaluated using the “test data” to verify the accuracy of the algorithm on never observed data. Table 3 shows the MAE as well as the maximum error achieved for the training, validation and test datasets.

Table 3. MAE and maximum errors achieved with the baseline model, for the training, validation and test datasets (Doyle dataset).

Dataset	**Mean Absolut Error (MAE)**	**Maximum Error**
Training data	0.113%	1.435%
Validation data	0.097%	1.512%
Test data	0.077%	1.459%

Figure 6 depicts the SOC estimation obtained at 25 ºC, 1C charge and 2C discharge. In this case the MAE is 0.225% and the maximum error is 1.418%. It could be observed that the SOC estimation algorithm follows the real SoC curve with high accuracy.

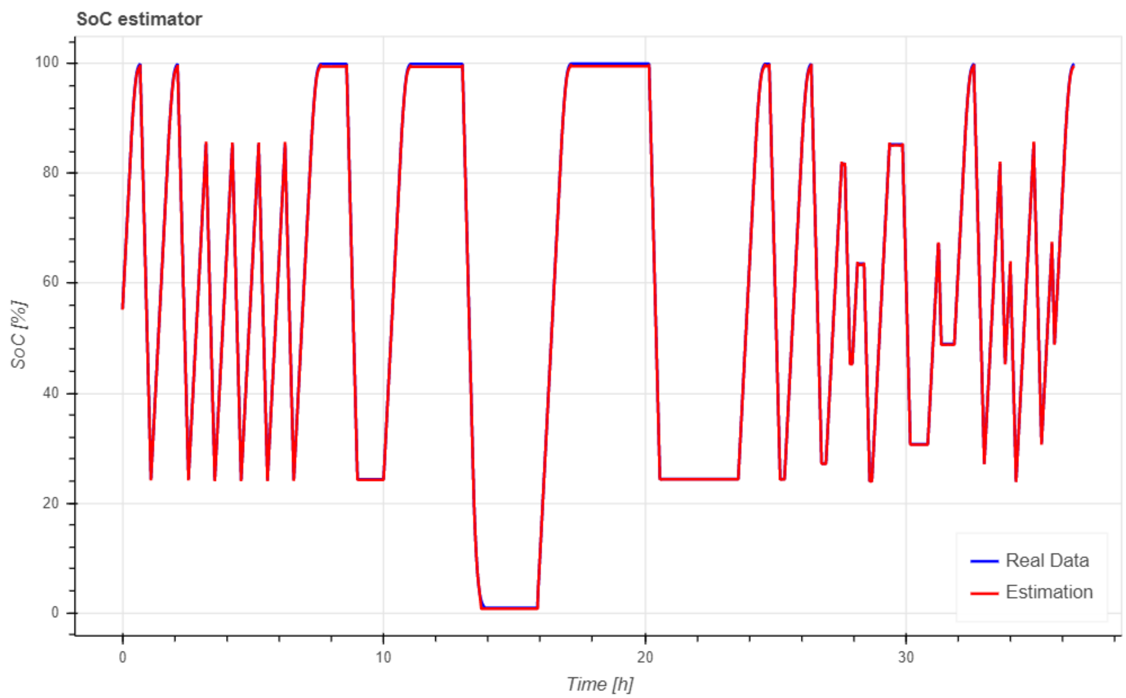

Figure 6. Estimation of the baseline model for de designed profile at 25 ºC, 1C CHA and 2C DCH.

4.2 NMC model

The NMC model was developed for comparative purposes, in order to evaluate later on the benefits of using TL method. Therefore, the ANN model used to train this model was identical to the baseline, i.e. composed of a CNN layer with 16 filters and a bidirectional LSTM layer with 5 neurons.

4.2.1 NMC model #1: model trained with reduced training data

To train the ANN, the data from the capacity test, qOCV and HPPC of one of the cells tested at 25 ºC was used. To validate this model, data corresponding to three other cells was used, which involved the same profiles at 10 ºC, 25 ºC and 45 ºC. Finally, to test the SoC estimation algorithm, the data from the additional cells was used. In addition, the designed profile explained in section 3 was also included in the test phase.

Table 4. Training, validation and test data split for the NMC model #1.

Dataset	**Conditions**	**Tests**
Training data	CELL1: 25 ºC	Capacity Quasi-OCV HPPC
Validation data	CELL2: 10 ºC CELL3: 25 ºC CELL4: 45 ºC	
Test data	CELL5: 10 ºC CELL6: 10 ºC CELL7: 25 ºC CELL8: 45 ºC CELL9: 45 ºC	
	CELL10: 25 ºC	Designed profile

Table 5 indicates the MAE and the maximum error achieved for the training, validation and test datasets.

Figure 7 depicts the SOC estimation of the algorithm on the designed profile. The MAE of the algorithm in this profile is 2.48% with a maximum error of 12.294%.

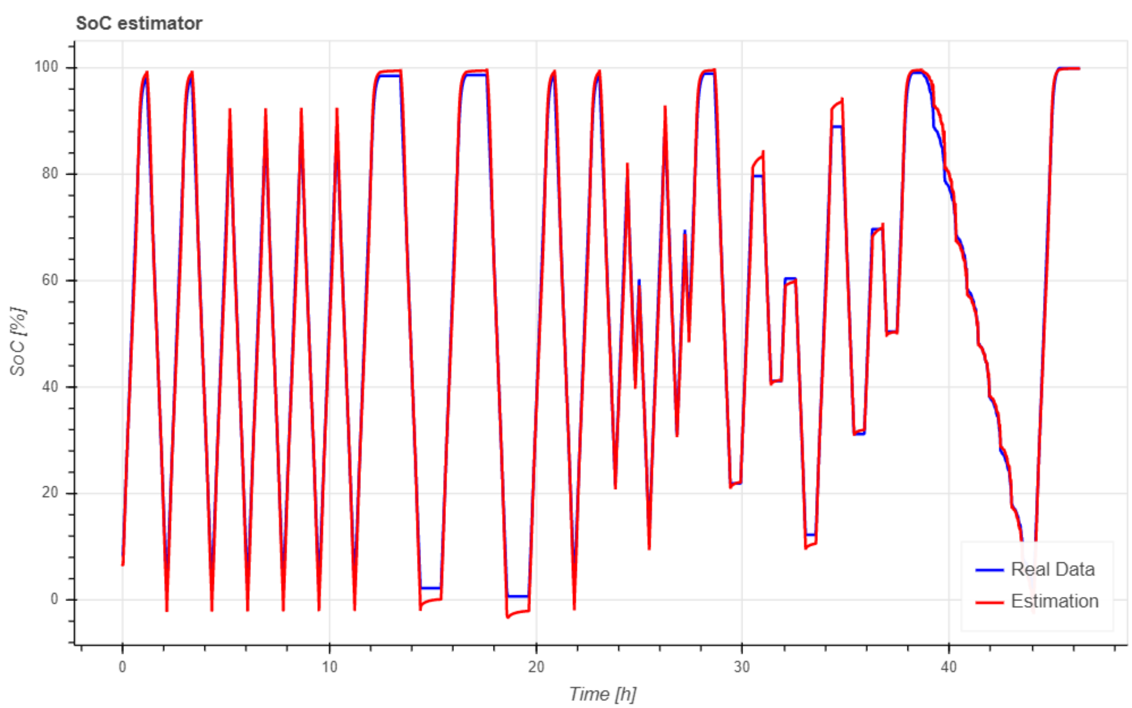

Figure 7. Estimation of the SoC made by the model trained with reduced training data for the designed profile at 25 ºC and 1C CHA/DCH.

4.2.2 NMC model #2: model trained with increased training data

In order to obtain more accurate results and complementary information for later comparison, a new training was carried out including more data in the training dataset. In this case, data from CELL5 and CELL8 (Table 4) was added to the algorithm training. Therefore, the validation dataset remain constant and test dataset was reduced as

CELL5 and CELL8 were used to train. In the the MAE and the maximum error achieved during the training, validation and testing is depicted.

It is noteworthy that the MAE decreased by more than 1% compared to the results achieved in Section 4.2.1, and that the maximum error was also reduced. Figure 8 shows the designed profile, and it could be observed that the estimations followed more accurately the real SOC value. The MAE of the algorithm in this profile was 0.941% and the maximum error was 2.895%.

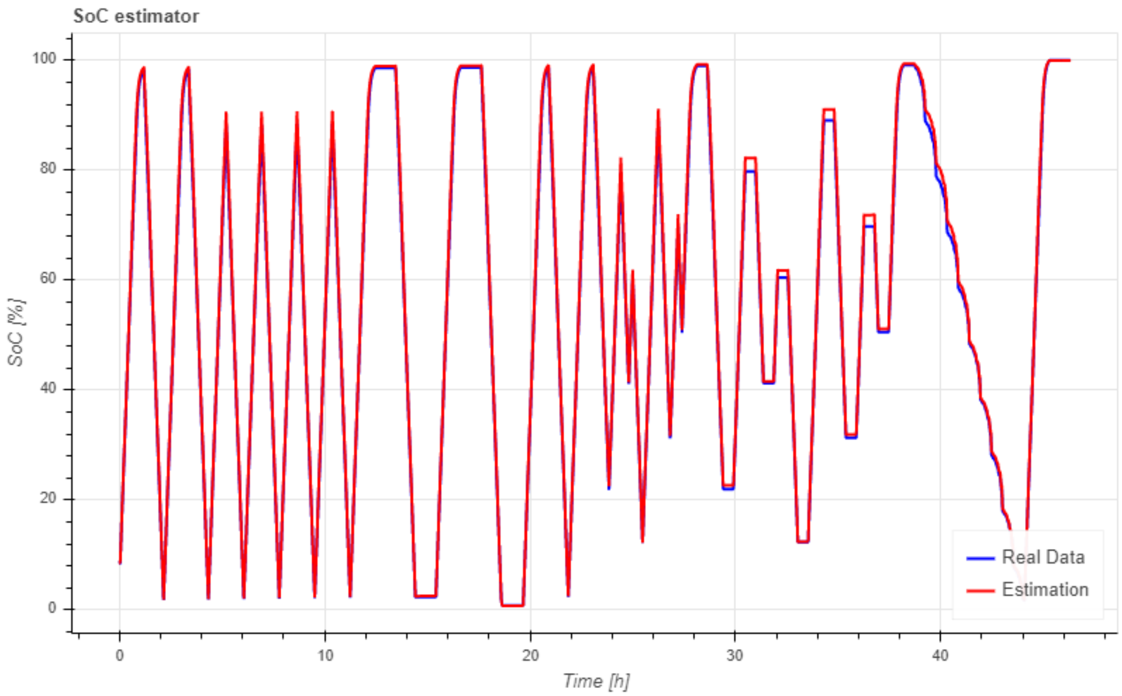

Figure 8. Estimation of the SoC made by the model trained with increased training data for the designed profile at 25 ºC and 1C CHA/DCH.

4.3 TL model

The last step in the development of the SoC algorithm was to apply the TL to the model trained with the synthetic Doyle cell data in section 4.1.

The training data defined in section 4.2.1 (reduced amount of training data, with just one single cell data employed for training) was used here to retrain the algorithm (Table 4). For the validation dataset, the data from three other cells cycled at 10 ºC, 25 ºC and 45 ºC was used. Data from the remaining cells, as well as the data corresponding to the designed profile, was included in the test dataset.

Table 5 shows the errors obtained after applying the TL method. It could be observed that the error is reduced by more than 1.2% compared to the model trained with reduced training data (NMC model #1) achieved in section 4.2.1 and 0.25% compared to the model trained with increased training data (NMC model #2) achieve in section 4.2.2.

Table 5. MAE and maximum errors achieved with the different models, for the training, validation and test datasets (NMC experimental dataset).

Model	Dataset	Mean Absolut Error (MAE)	Maximum Error
Model trained with reduced training data (NMC model #1)	Train	1.497%	16.787%
	Validation	2.653%	17.458%
	Test	3.168%	18.369%
Model trained with increased training data (NMC model #2)	Train	0.487%	**2.083%**
	Validation	0.486%	**2.102%**
	Test	0.494%	**2.504%**
TL model	Train	**0.251%**	4.045%
	Validation	**0.243%**	4.086%
	Test	**0.243%**	3.411%

Figure 9 depicts the SOC estimation obtained for the designed profile, using the model updated by TL. The SOC estimation significantly improved compared to the results depicted in Figure 7. For this profile, the MAE is 0.450% and the maximum error is 1.408%.

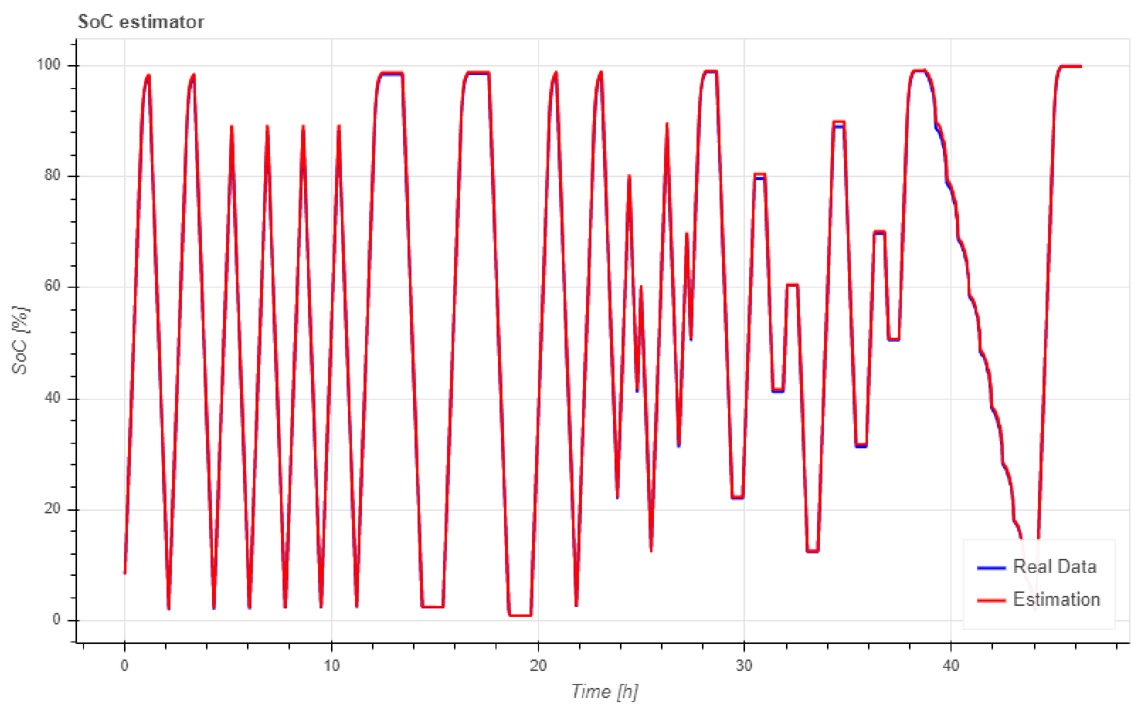

Figure 9. Estimation of the SoC made by the TL model for the designed profile at 25 ºC and 1C CHA/DCH.

5 Conclusions and Future works

This work introduced a proof-of-concept of a new approach for the development of SOC estimation algorithm, based on the TL method. The main objective of such approach was to reduce the need for exhaustive laboratory testing, typically required for SOC algorithm development each time a new cell reference is desired to be deployed. This was possible by leveraging the data previously available for different cell references, already tested or deployed, which could encode useful information usable for the new cell reference.

In this work, a SoC estimation algorithm was developed using a neural network composed of a CNN layer and a LSTM layer. This network was trained with four different training datasets, leading to four different models: i) the "baseline model" was trained only with LCO-based synthetic data, generated using a previously validated electrochemical model, ii) the "NMC model #1" was trained exclusively with a reduced number of NMC-based data, generated in laboratory by cycling commercial cells, iii) the "NMC model #2" was trained exclusively with NMC-based data (higher amount of training data), also generated in laboratory by cycling commercial cells, and iv) the "TL model" was trained combining LCO-based synthetic data and a reduced number of NMC-based data.

Comparing the performances of these models, the benefits of the proposed TL-based approach were highlighted: the number of data to be generated in laboratory and required to train the algorithm is significantly reduced, and at the same time, the accuracy of the SOC estimations is improved considerably. This could be explained by the fact that, during the first training phase based on LCO synthetic data, the neural network was able to learn about the cell behaviour at different conditions of C-rates or temperatures. Part of this knowledge was also relevant for NMC cell's SOC estimation, and was then exploited by the TL during the retraining phase, making possible to reduce the amount of NMC cell data required from laboratory testing. Compared to an analogue algorithm developed from scratch for NMC cells (NMC model #2), the error was reduced 50% and the training dataset required from laboratory was reduced 40%.

Results obtained lead the way towards ultra-fast SOC algorithm development, as baseline synthetic data generated by means of well-established electrochemical models serves to minimise the required experimental tests for any given new cell reference to which the algorithm needs to be tailored. However, it should be noted that, so far, the developed model does not take into account the degradation of the cell. This model can be extended, and an algorithm capable of estimating the SoC of the battery can be developed by including the actual SoH of the battery as an input variable.

Finally, the two cells used in this work have similar behaviour, so it would be interesting to analyse how the TL behaves when data from a different chemistry is applied during retraining, such as data from LFP-based cells. Further research will be carried out to identify the limitations of the proposed TL-based approach.

Acknowledgments

This investigation work was financially supported by the project LIBERTY, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 963522.

This document reflects only the author's view, the Agency is not responsible for any use that may be made of the information it contains.

References

[1] I. B. Espedal, A. Jinasena, O. S. Burheim, and J. J. Lamb, "Current trends for state-of-charge (SoC) estimation in lithium-ion battery electric vehicles," *Energies*, vol. 14, no. 11. MDPI AG, Jun. 01, 2021. doi: 10.3390/en14113284.

[2] J. P. Rivera-Barrera, N. Muñoz-Galeano, and H. O. Sarmiento-Maldonado, "Soc estimation for lithium-ion batteries: Review and future challenges," *Electronics (Switzerland)*, vol. 6, no. 4. MDPI AG, Dec. 01, 2017. doi: 10.3390/electronics6040102.

[3] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[4] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, “Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries,” *Applied Energy*, vol. 260, no. November 2019, p. 114296, 2020, doi: 10.1016/j.apenergy.2019.114296.

[5] M. Savargaonkar and A. Chehade, “An Adaptive Deep Neural Network with Transfer Learning for State-of-Charge Estimations of Battery Cells,” in *2020 IEEE Transportation Electrification Conference & Expo (ITEC)*, Jun. 2020, pp. 598–602. doi: 10.1109/ITEC48692.2020.9161464.

[6] C. Vidal, P. Kollmeyer, E. Chemali, and A. Emadi, “Li-ion Battery State of Charge Estimation Using Long Short-Term Memory Recurrent Neural Network with Transfer Learning,” in *2019 IEEE Transportation Electrification Conference and Expo (ITEC)*, Jun. 2019, pp. 1–6. doi: 10.1109/ITEC.2019.8790543.

[7] K. Kaur, A. Garg, X. Cui, S. Singh, and B. K. Panigrahi, “Deep learning networks for capacity estimation for monitoring SOH of Li-ion batteries for electric vehicles,” *International Journal of Energy Research*, vol. 45, no. 2, pp. 3113–3128, Feb. 2021, doi: 10.1002/er.6005.

[8] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D convolutional neural networks and applications: A survey,” *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, Apr. 2021, doi: 10.1016/J.YMSSP.2020.107398.

[9] M. Cañizo Zubizarreta, “Large-scale anomaly detection and diagnosis on industrial heterogeneous multi-sensor systems using deep learning,” Universidad de Deusto, 2020.

[10] C. Li, F. Xiao, and Y. Fan, “An Approach to State of Charge Estimation of Lithium-Ion Batteries Based on Recurrent Neural Networks with Gated Recurrent Unit,” *Energies (Basel)*, vol. 12, no. 9, p. 1592, Apr. 2019, doi: 10.3390/en12091592.

[11] X. Song, F. Yang, D. Wang, and K.-L. Tsui, “Combined CNN-LSTM Network for State-of-Charge Estimation of Lithium-Ion Batteries,” *IEEE Access*, vol. 7, pp. 88894–88902, 2019, doi: 10.1109/ACCESS.2019.2926517.

[12] J. Tian, R. Xiong, W. Shen, and J. Lu, “State-of-charge estimation of LiFePO4 batteries in electric vehicles: A deep-learning enabled approach,” *Applied Energy*, vol. 291, Jun. 2021, doi: 10.1016/j.apenergy.2021.116812.

[13] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller, “Transfer learning for time series classification,” *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, pp. 1367–1376, Jan. 2019, doi: 10.1109/BIGDATA.2018.8621990.

[14] “Doyle et al. - 1996 - Comparison of Modeling Predictions with Experimental Data from Plastic Lithium Ion Cells”.

Authors

	**Markel Azkue** received his M.Sc in Industrial Engineering from the University of Mondragon, in 2020. He joined the Energy Storage and Management unit of Ikerlan Technology Research Centre (BRTA), Spain, in 2016. He is currently coursing there his Ph.D. in Applied Engineering in collaboration with Mondragon Unibertsitatea, Spain. His research interests include machine learning algorithms applied to electrochemical energy storage systems state estimation algorithms.
	**Dr. Mattin Lucu** received his M.Sc in Integration of Renewable Energy Sources into the Electricity Grid from the University of the Basque Country, UPV-EHU (Spain) in 2016. During his graduate studies, he worked as R&D intern successively at EneR-GEA research group (ESTIA Engineering School, France) in wind turbine emulation and control, and at EDP in the analysis of photovoltaic power insertion in low-voltage distribution networks. In 2016, he joined the Ikerlan Technology Research Centre, where he carried out his Ph.D. degree on data-driven ageing models for Li-ion batteries, in collaboration with the University of the Basque Country. His research interests include modelling and control of electrochemical energy storage systems, machine learning applied to battery diagnostic and prognostic algorithms, fleet battery data analytics and cloud-based battery management.
	**Dr. Laura Oca** received the degree in management engineering (End of Degree Awards for the best academic records) from Mondragon Unibertsitatea, Basque Country, Spain, in 2015 and the M.S. degrees in Industrial engineering from Mondragon Unibertsitatea in 2017. She received the Ph.D. degree in applied engineering (Predoctoral Grant from the Basque Country) from Mondragon Unibertsitatea in 2020. She is currently a researcher and lecturer at Mondragon Unibertsitatea. Her research includes physics-based modelling of different electrochemical energy storage devices (i.e. lithium-ion batteries, sodium-ion batteries, lithium-ion capacitors, solid-state batteries, fuel cells), parameter measurement and estimation of physics-based models and lithium-ion cell design fabrication and optimization through physics-based models.
	**Dr. Egoitz Martinez-Laserna [M]**: received his M.Sc in Electronic Engineering from the University of Mondragon, Spain, in 2013. He joined the Energy Business Unit of IKERLAN Technological Research Centre, Spain, in 2013. He obtained his Ph.D. (Cum Laude) in Engineering with international merit in 2017, focused on 2nd life battery performance, sizing and integration. His research interests include electrochemical energy storage systems, electric vehicles, state estimation algorithms, energy management and BMS design. He has been co-author of various scientific publications and has contributed to multiple international conferences. Since 2020 he is the coordinator of the H2020-LC-BAT-10 project *LIBERTY*.
	**Dr. Unai Iraola** obtained his Doctor degree from Mondragon University in 07/2014 in the field of the electro-thermal optimization of battery-based energy storage systems for power applications. After defending his thesis work in 2014, Dr. Iraola continues working at the Faculty of Engineering of Mondragon University as a Research Professor. He is working both in private and public projects and he has supervised different PhD works until now. Dr. Iraola´s current research area focuses on energy storage systems technologies, mainly, lead acid, lithium-ion and solid-state batteries. His research is mainly focused on physics-based battery modeling and implementation within BMS systems with the objective of minimizing degradation via modular battery packs.