Auto-encoder LSTM for Li-ion SOH prediction : a comparative study on various benchmark datasets

Paul AUDIN
ICube, CNRS (UMR 7357)
INSA Strasbourg
University of Strasbourg
F-67000 Strasbourg, France
paul.audin@insa-strasbourg.fr

Inès JORGE
ICube, CNRS (UMR 7357)
INSA Strasbourg
University of Strasbourg
F-67000 Strasbourg, France
ines.jorge@insa-strasbourg.fr

Tedjani MESBAHI
ICube, CNRS (UMR 7357)
INSA Strasbourg
University of Strasbourg
F-67000 Strasbourg, France
tedjani.mesbahi@insa-strasbourg.fr

Ahmed SAMET
ICube, CNRS (UMR 7357)
INSA Strasbourg
University of Strasbourg
F-67000 Strasbourg, France
ahmed.samet@insa-strasbourg.fr

François DE BERTRAND DE BEUVRON
ICube, CNRS (UMR 7357)
INSA Strasbourg
University of Strasbourg
F-67000 Strasbourg, France
francois.debertranddebeuvron@insa-strasbourg.fr

Romuald BONÉ
ICube, CNRS (UMR 7357)
INSA Strasbourg
University of Strasbourg
F-67000 Strasbourg, France
romuald.bone@insa-strasbourg.fr

***Abstract*—Lithium-ion batteries are used in most battery powered devices. Today's research on Lithium-ion batteries mainly focuses on better energy management strategies and predictive maintenance. In this paper, a new approach based on auto-encoders and long short-term memory neural networks applied to usage data (voltage, current, temperature) is used to make a State of Health prediction. Encouraging results are obtained when conducting tests on various battery ageing datasets published by Sandia National Laboratories, the Massachusetts Institute of Technology and NASA's Prognostics Center of Excellence.**

***Index Terms*—Lithium Ion batteries, Electric Vehicles, Predictive prognostics, Machine learning, Deep Learning, Feature extraction, State of Health, Remaining Useful Life**

I. Introduction

The need for reliable energy sources is getting stronger everyday. The search for clean sources of energy and the development of electrical mobility and isolated electrical systems has inevitably led to the pursuit of an economical, safe and reliable energy storage system. Lithium-ion (Li-ion) batteries are probably the most prominent technology, due to their high energy density and output power, fast charging capabilities and their long lifespan. The supply and demand of lithium grows each year, as does the research for better recycling of Li-ion batteries [1].

An important aspect in pursuing the most reliable energy storage is an optimised maintenance, which would allow for the least down times of the systems. The development of predictive maintenance has led researchers to work on several approaches to predict the approximate time of the end of life of batteries. The degradation of a battery can be tracked using indicators such as its State of Health (SOH) or Remaining Useful Life (RUL). The RUL of a battery corresponds to the number of charging and discharging cycles a battery can still withstand before being considered worn out for mobility applications (20% loss of the initial capacity, or a 80% SOH).

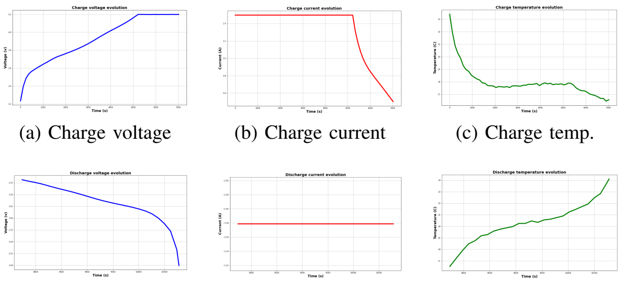
Two main classes of predictions have emerged : model-based methods, aiming to model the physical aspects of the battery to predict its behaviour, and data-driven approaches, which use historical ageing data in order to make a prediction. Data-driven models are currently thriving, as more and more data become available. The main advantage of these approaches are their simplicity, since no deep understanding of the electro-chemical processes of the battery is needed.

Most data-driven models are based on windows of historical SOH values [2], [3]. Only a few papers [4]–[6] take advantage of battery usage data (voltage, current, temperature) reflecting the use conditions of the battery. These approaches respectively use the I/V curves of charge cycles, charging voltage curves and time features. Battery degradation is directly linked to its usage (charge and discharge protocols, environment) so it makes sense to use this information to predict the capacity degradation. Examples of this data are presented in Figure 1.

Unfortunately, to the best of our knowledge, only a few approaches have investigated the use of all available battery usage data. You *et al.* [7] use Voltage (V), Current (I) and Temperature (T) data from a custom dataset to make a SOH prediction while in [6], a RUL prediction on the NASA dataset using time features extracted from V, I and T time series is made. To make use of all available usage data, and make a robust and generalizable model, we combine an auto-encoder to a long short-term memory neural network (LSTM) and apply it to each of the individual time series (charging and discharging voltage, current and temperature) to predict the

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

978-1-6654-4337-1/21/$31.00 ©2021 IEEE
DOI 10.1109/ICMLA52953.2021.00246



(a) Charge voltage (b) Charge current (c) Charge temp.

(d) Discharge voltage (e) Discharge current (f) Discharge temp.

Fig. 1: Example of usage time series - Sandia National Laboratories dataset

SOH. We also make a comparative study on various publicly available battery ageing datasets. In this paper, we propose a three fold scientific contribution: *(i)* an AE-LSTM approach based on an auto-encoder to extract knowledge from current, voltage and temperature curves and a LSTM to predict the evolution of SOH; *(ii)* the design of a model that makes long term predictions of SOH from a limited window of battery usage data; and *(iii)* an exhaustive performance study on the Sandia national laboratories (SNL) battery ageing dataset complemented by the MIT and NASA PCoE datasets, since few works have been conducted on the MIT and SNL datasets.

The rest of this paper is structured as follows : **Section II** addresses a quick state of the art on current modelling techniques. **Section III** defines several concepts linked to battery ageing and neural networks. **Section IV** describes the presented approach. **Section V** details the datasets used for this paper, reports the results and compares our approach to state of the art models. A conclusion is drawn in **Section VI** followed by future prospects of this work.

II. Related work

RUL and SOH estimation of Li-ion batteries is currently of big interest for the scientific community. The interests linked to the development of a reliable model that would provide precise estimations have led researchers to work on several approaches and models. These approaches can be separated into two groups: the *model-based* approaches and the *data-driven* approaches. The model-based approaches aim to develop models of the physical events happening inside of the battery. On the other hand, the data-driven approaches make use of the multiple databases related to Li-ion batteries to develop models based on historical data.

A. Model based

Model-based methods aim to capture the physical processes happening inside the battery. Chemical, thermal and electrical processes can be used, as well as more abstract processes such as active material losses. More details can be found in references [8]–[12].

B. Data driven

Li *et al*. [2] make use of frequency decomposition using Empirical Mode Decomposition (EMD). The different frequency components are then treated individually through Elman neural networks and a LSTM. The different outputs are used to predict the RUL. Using I/V charging data, You *et al*. [4] developed two approaches based on LSTM to predict SOH. One of them uses density distributions of the I/V data and the other uses the raw data. The team achieved good results using their own data and combinations of their two approaches. Wu *et al*. [5] use charging voltage curves and importance sampling to feed a neural network and predict the RUL of the battery. The RUL is estimated in [13] using an auto-encoder feeding a four-layer deep neural network. Wavelet denoising is used in [14] before using a relevance vector machine (RVM) to get a probabilistic interpretation of the RUL prediction. The method used in [15] is based on particle filter, or sequential Monte-Carlo, coupled with an exponential model. Good RUL predictions are achieved when using a major part of the entire cell life data. A combination of auto-encoder, LSTM and CNN is used in [6] to make a RUL prediction. The auto-encoder feeds the LSTM and the CNN in parallel , and their outputs are then used to make the prediction. In references [3] and [16], empirical mode decomposition (EMD) is applied to historical capacity data, and the result of EMD is given as input to predictive models based on LSTM to forecast the future trend of SOH degradation. Feature extraction from charging voltage curves is implemented in reference [17], and used as input to an ANN to estimate future SOH values.

As previously stated, several models can be used to study battery degradation. We chose to adopt a data-driven approach because of the increasing availability of extensive data. Two main parameters are predicted, the SOH and the RUL, which both represent a battery's health. We chose to focus on SOH predictions.

The majority of papers focuses on using historical capacity values as inputs. In most approaches, the conclusion is made the the larger the input window, the better the prediction. We propose a predictive approach based on a sliding window of limited size that performs well on long term predictions of SOH values.

Our approach bases its prediction on usage data, namely the voltage, current and temperature data, as it can be used to translate what a battery is used for. Tests on multiple datasets are conducted to confirm the adaptability of the approach. The following section sets the context for the research and provides necessary definitions for its understanding.

III. Background

A. Battery ageing definitions

Definition 1 (Capacity): A battery is an electro-chemical device used to store electrical energy under a chemical form. Its capacity depends on its chemical composition and size, and is measured in Ampere-Hours (Ah). For example, a 1Ah battery can output 1A for an hour before being depleted.

Definition 2 (State of Health (SOH)): The State of Health of a battery directly reflects its ability to function properly. It gives an indication on how much capacity was lost by doing the ratio between the actual capacity and the nominal capacity of the battery :

$$SOH = \frac{Q_{actual}}{Q_{nominal}} \quad and \quad SOH\% = \frac{Q_{actual}}{Q_{nominal}} * 100 \quad (1)$$

The SOH is expressed as a percentage, and a battery is considered worn out for electric vehicle (EV) applications when reaching a SOH of 80%.

While capacity values are easy to estimate by integrating the battery's current over time, developing a model that could translate the usage conditions of the battery into its capacity degradation would be really useful. In a real-world scenario, using the estimated capacity values as well as the current, voltage and temperature time series would be ideal since the capacity measurement could be used to predict the long-term trend of degradation of the battery, and the usage-related time series would allow to take into account any change in the battery's environment and operation.

B. Neural Networks background

This section aims to describe the different types of neural networks used in this study. A description of auto-encoders and their ability for dimensional reduction is followed by a definition of LSTM.

Auto-encoders

Auto-Encoders Neural Networks (AENN) are used in [6], [18] for feature extraction. These NN are mainly used for their dimensional reduction [13] and augmentation [6] capabilities, but are also used for feature extraction and data fusion [18]. As can be seen in Figure 2, they have a symmetrical structure.

This unique architecture consists in its most basic form of an input layer and an output layer of the same size separated by a hidden layer of a lesser size. The first stage is called an encoder, and aims to reduce the input data into a code, which represents the most fundamental characteristics of the data. This code is then used by the decoder to reconstruct the input as accurately as possible, or used directly in the case of feature extraction.

Long Short-Term Memory (LSTM)

LSTM belong to the category of recurrent neural networks (RNN). RNNs are neural networks which use the previous output as well as the current input to make their predictions. In response to RNNs being subjects to the vanishing gradient problem, which could cause an ineffective or faulty training, LSTM RNNs were developed. Making use of three gates (input, forget and output) to regulate the data flow, LSTM RNNs are widely used for RUL and SOH estimation [2]–[4], [6], [15], [16], [19], [20] as a standalone model or in combination with other techniques. Due to its ability to get the long-term trend of the data, this architecture is well adapted for battery ageing data and showed excellent results with proper training methodology [21].

As can be seen in Figure 3, a LSTM block comprises three gates :

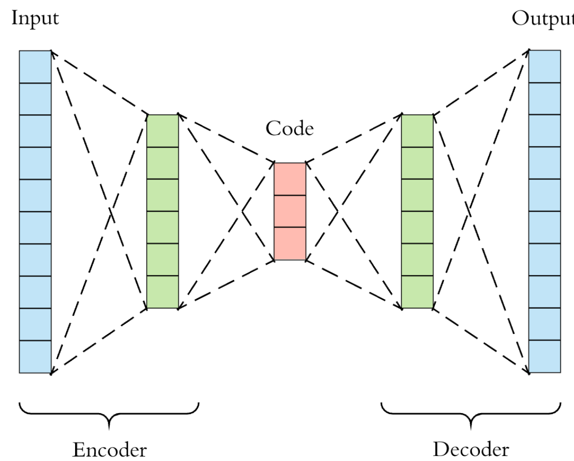


Fig. 2: Basic structure of an Auto-Encoder

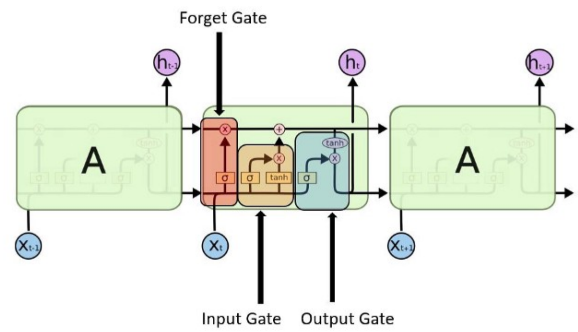


Fig. 3: Long Short-Term Memory blocks

1) *Input gate*: Allows to determine which input value is used to modify the memory. The *sigmoid* function determines which value to let through in a [0,1] range and the *tanh* function applies weights to the values that got through, and range their importance in a [-1,1] range.
2) *Forget gate* : Determines what should be discarded using a *sigmoid* function. Using the last output and the actual input, a number ranging from 0 (to be discarded) to 1 (to keep) is computed.
3) *Output gate* : Determines the output using the input and the memory. The *sigmoid* and *tanh* functions are also used here.

IV. AE-LSTM based model for SOH prediction

The proposed architecture is based on auto-encoders as a mean of feature extraction and dimensional reduction. Voltage, current, and temperature time series vary with the usage conditions of the battery and its state of degradation. As a battery degrades, its internal resistance rises and leads to more power loss which can be noticed in the evolution of its SOH and in usage time series. Usage time series are real time indicators of how the battery is being used by a driver, and we believe that they are key factors to observe in our prognostics and health management strategy. The following six time series are used to make a prediction: charge voltage (Uc), discharge voltage (Ud), charge current (Ic), discharge current (Id), charge temperature (Tc) and discharge temperature (Td).

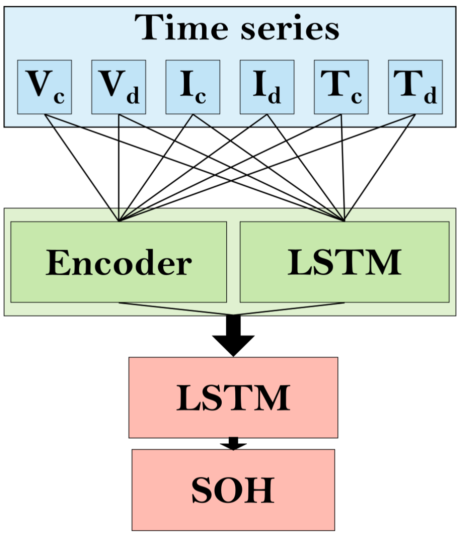


Fig. 4: Architecture summary

Figure 4 presents the architecture of our AE-LSTM approach. Windows of raw time series are used as inputs to make a SOH prediction. Each of these time series are then distributed to two branches. One of them is made of the encoding layers of an auto-encoder trained to reconstruct this particular time series. This encoder outputs a code, corresponding to a vector of data representative of the input. The second branch is composed of LSTM layers.

The encoder's weights are frozen during the training of the final model. This allows the first branch to act as a static feature extraction method while the second branch gives more context as it learns from time series of varying lengths.

The outputs of the two sub-systems are then concatenated and fed through LSTM layers to make the SOH prediction. A detailed description of this architecture is available in Sections IV-B and IV-C.

A. Data pre-processing

Data pre-processing plays a key role in data-driven strategies. Data that is not properly processed and shaped can't possibly lead to satisfying results. This is why a step of de-noising, outlier excluding and normalization is needed. During data acquisition, many things can impact measurements and add unwanted noise (electromagnetic noise, vibrations, ...).

As explained earlier, the model was developed on the Sandia National Laboratories [1] [22] dataset and tested on it as well as on the MIT Life Cycle dataset [2] [23] and the NASA PCoE battery dataset [3] [24].

The aim of the Sandia National Laboratories dataset is to study the effect of depth of discharge (DoD), load current and temperature on battery degradation. 86 cells of three different chemistries (LFP, NMC and NCA) were considered in this study. Three different DoD were used, 0%-100%, 20%-80% and 40%-60% at three different temperatures, 15°C, 25°C and 35°C. The discharge rates used are 0.5C, 1C, 2C and 3C. All tests are conducted beyond 80% SOH and batteries were charged at 0.5C. The NCA cells were not discharged at 3C since it would be destructive.

In this study, only the data measured from the NMC cells tested at a 0-100% depth of discharge is used. Further research on the impact of the depth of discharge on capacity degradation and model performance should be done, as the results presented in [22] show clear evidence that the depth of discharge has major repercussions on a battery's capacity.

The NASA PCoE dataset is the most widely used and contains 6 groups of cells for a total of 34. The cycling experiment was segmented into three parts : charging, discharging and impedance spectroscopy. Impedance measurements were carried out through an EIS frequency sweep from 0.1Hz to 5kHz. The charging protocol is the same for all tests, a CC-CV charging protocol with 1.5A and a 4.2V threshold. Various different discharging currents were used.

The MIT Cycle Life dataset contains data from 124 LFP cells cycled using various fast-charging protocols consisting of one-step or two-step constant current charging, switching to 1C CC-CV charging at 80% state of charge (SOC). The cells were divided into three batches, and tested under a constant temperature of 30°C. Internal resistance is measured during charging at 80% SOC, using an average of ten 30ms or 33ms pulses of $\pm 3.6C$.

The datasets contain data about cycle indexes, voltage, current, temperature, charge and discharge capacity, charge and discharge energy and battery impedance.

A first look at the SNL data showed some spikes in most capacity evolution curves, which had to be dealt with using a simple function which detects these spikes and computes their new value using the neighboring ones. These spikes are artifacts indicated on the dataset's download page, and are due to the transitions between normal cycling and capacity checks. After excluding these outliers, the data could be normalized before being used.

The NASA PCoE battery dataset, while being the most used dataset, needed several steps of pre-processing in order to be usable. The data is gathered during three distinct steps : charge, discharge and electrochemical impedance spectroscopy (EIS). In this study, we do not use the impedance data and therefore it is excluded. What remains of the data structure makes a clear distinction between charge and discharge data, which had to be grouped in order to get both information for each cycle. A

[1]Accessed here or at https://www.batteryarchive.org in March 2021

[2]Accessed here in March 2021

[3]Accessed here in March 2021

few irregularities such as two subsequent charging processes should be excluded. Finally, due to several gaps in the battery capacity measurements, several batteries are excluded from this work. We use batteries #5-7, #18, #32-36, #45-48 and #53-56.

The MIT Life Cycle dataset is the largest and the most recent battery ageing dataset. This dataset does not need a lot of pre-processing as the data is exempt of noise. The data is divided into three batches which corresponds to three series of tests. Several batteries from each of the batches were selected to conduct performance tests.

The process of normalization is an important step to avoid imbalance between different input data samples. Here, Min-Max normalization is applied to the data :

$$MinMax = \frac{data - min(data)}{max(data) - min(data)} \tag{2}$$

The output data ranges from 0 to 1. Note that this normalization process is applied to the input of the models, as the target data is the SOH which is obtained by dividing the capacity values by the nominal capacity of the battery. The batteries are split as follows : 60% for training, 20% for testing and 20% for validation.

B. Encoding branch

As described in III-B, auto-encoders are models used to reduce their input to a code representative enough to be reconstructed. In this work, the inputs are windows of padded time series since each and every charge and discharge cycle is of different length. The sequence-to-sequence auto-encoders are made of LSTM layers of the following sizes : [256,16,RL,256,IS] where RL is a RepeatVector layer with a repetition factor equal to the size of the input window and IS is the size of the input time series. The encoding layers of these models are then used in the main AE-LSTM model.

C. LSTM branch

In parallel to the encoding layers, the input time series are distributed towards LSTM layers. These two LSTM layers are of size 256 and 32. As the encoding layers extracted from the auto-encoders have their weight frozen, this structure allows the final model to learn while training the final model. After concatenating the outputs of the different branches, the SOH prediction is made using LSTM layers of sizes [512,256,256,32] and a final dense layer with 1 unit and a linear activation function giving the final output.

D. Smoothing

Certain predictions of the AE-LSTM model present some irregularities. In an attempt to improve these predictions, a post-processing method is investigated. One thing to consider when post-processing SOH predictions is that each value can only be smoothed using the previous ones. Our smoothing method uses a window of predicted capacity values to adjust the next one. These values are assigned a weight in order to give more importance to the most recent predictions. The weights are computed using a simple polynomial function applied to a vector of progressive integers. For example, if using a window of size 5 the vector is initialized as [1,2,3,4,5]. A function $f(x) = x^n$ is then applied to this vector to adjust the weights, where n is a power factor that can be adjusted. With $n = 3$, we get a weight vector equal to [1,8,29,64,125] which we can apply to the window of capacity values preceding the considered prediction.

The smoothing method shows promising results when applied to the predictions but is not consistent enough. The error values presented in the following sections are the one computed using the predictions without the post-processing stage.

V. Experiments

This section contains the results of the different experiments carried out as part of this work. An introduction to the data used is followed by the results obtained with this approach.

A. Error metrics

In order to compare different models, several indicators can be computed to express their performance. The most commonly used are the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Normalized Mean Standard Error (NMSE). A measure of the standard deviation of the MAE (STD) is also computed. These metrics are expressed as follows :

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{pred,i} - y_i)^2} \tag{3}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{pred,i} - y_i| \tag{4}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_{pred,i} - y_i|}{y_i} * 100 \tag{5}$$

$$NMSE = \frac{\sum_{i=1}^{N}(y_{pred,i} - y_i)^2}{N * V} \tag{6}$$

Where $y_{pred,i}$ is the model's prediction, y_i is the true SOH value, V is the variance of y and N is the number of samples from which the error is computed.

B. Auto-encoders for feature extraction

The training process of the model has two steps. First, each auto-encoder is trained to reconstruct a particular time series. Then, the encoding layers of the auto-encoders are used inside the general model with their weights frozen so that their output remains a representative code during training. The loss function used in all training sessions is the Mean Squared Error. The auto-encoders were trained 500 epochs and the general model 1000 epochs, both using a learning rate of $5 * 10^{-5}$ and the Adam optimizer.

For the six time series of charge and discharge voltage, current and temperature, six auto-encoders were trained to reconstruct them. A window of 25 time series with a length of around 750 timesteps each is used as input. The encoder reduces this input and condenses the information contained in each of the time series into a vector of 16 values. Table I presents the mean reconstruction errors of the models on the SNL dataset.

Errors	MAE	STD	RMSE
Uc	$7,1973.10^{-3}$	$3,7382.10^{-2}$	$3,8068.10^{-2}$
Ic	$5,0665.10^{-3}$	$2,5466.10^{-2}$	$2,5965.10^{-2}$
Tc	$4,9787.10^{-3}$	$1,8251.10^{-2}$	$1,8918.10^{-2}$
Ud	$4,7034.10^{-3}$	$1,3779.10^{-2}$	$1,4560.10^{-2}$
Id	$4,3670.10^{-3}$	$2,6193.10^{-2}$	$2,6555.10^{-2}$
Td	$6,6154.10^{-3}$	$2,1169.10^{-2}$	$2,2178.10^{-2}$

TABLE I: AE reconstruction performance

As can be seen in Table I, the MAE of the reconstruction of the different time series is low. The data is normalized between 0 and 1, and the MAE values are below 0,01 which indicates that the coded representation is reliable.

C. Training dataset structure

In order to select which of the six studied inputs (charge and discharge voltage, current and temperature) are the most impactful, a comparison was made between different combinations of input to determine which is the best. The tested combinations are shown in Table II:

Name	Uc	Ud	Ic	Id	Tc	Td
All	X	X	X	X	X	X
Charge	X		X		X	
Discharge		X		X		X
Voltage	X	X				
Current			X	X		
Temperature					X	X
Current / temp.			X	X	X	X

TABLE II: Tested combinations

Note that tests are not conducted on every possible combinations to limit computing time and that tested combinations are those we think are the most relevant. The results are shown in Table III. The error values were obtained by averaging the results of three distinct training and testing sessions, on different batteries from the SNL dataset.

Name	MAE	STD	RMSE
All	$\mathbf{0,7331.10^{-2}}$	$\mathbf{0,7331.10^{-2}}$	$\mathbf{1,0861.10^{-2}}$
Charge	$1,7830.10^{-2}$	$1,7830.10^{-2}$	$2,0362.10^{-2}$
Discharge	$2,7485.10^{-2}$	$2,7485.10^{-2}$	$3,2076.10^{-2}$
Voltage	$0,9894.10^{-2}$	$0,9894.10^{-2}$	$1,3239.10^{-2}$
Current	$1,4353.10^{-2}$	$1,4353.10^{-2}$	$1,8131.10^{-2}$
Temperature	$1,1590.10^{-2}$	$1,1601.10^{-2}$	$1,4292.10^{-2}$
Current / temp.	$0,9923.10^{-2}$	$0,9996.10^{-2}$	$1,2944.10^{-2}$

TABLE III: Tested combinations performances

As can be seen in Table III, the best performing combinations are the one using all available inputs and the one using only current and temperature time series. This could have been expected, as temperature reflects both the battery's environment and use (even in a climate controlled chamber), and the current output is directly linked to the strain put on the battery.

We know that a battery's use conditions directly affects its capacity degradation. The models described here predict a SOH value in the future at a fixed step (for example, the prediction of the SOH value of the battery 200 cycles in the future). A study comparing different steps of prediction was carried out to determine the delay between the usage and the degradation.

Here, a comparison is made using different steps of prediction : 1, 25, 50, 100 and 200 steps ahead.

Step	MAE	STD	RMSE
1	$1,057.10^{-2}$	$2,988.10^{-2}$	$1,262.10^{-2}$
25	$1,120.10^{-2}$	$2,795.10^{-2}$	$1,468.10^{-2}$
50	$0,759.10^{-2}$	$2,458.10^{-2}$	$1,111.10^{-2}$
100	$0,905.10^{-2}$	$2,141.10^{-2}$	$1,270.10^{-2}$
200	$1,419.10^{-2}$	$1,907.10^{-2}$	$1,861.10^{-2}$

TABLE IV: Errors of different prediction steps

Table IV shows that prediction error is steady when the prediction step gets bigger, until reaching a value leading to a bigger error. This result could be expected, but the aim is to determine at which point the predictions become unreliable. We still have to remember that the data used for this comparison is laboratory data using CC charging and discharging protocols. These protocols do not vary throughout the battery life, which means that the capacity degradation trend is relatively steady. This study could be conducted on data coming from experiments where the protocols change during testing.

We choose to build a model making a prediction 50 cycles ahead into the future.

D. Prediction performance and comparison

To evaluate the prediction performance of the proposed architecture, a dozen training were executed on the SNL data to evaluate the prediction performance of the model. Each training was done using a different distribution of training, validation and testing data. Tests on the MIT Life cycle dataset and on the NASA PCoE battery dataset are also conducted subsequently in order to study the effect of different charging protocols. Following the previous comparisons made in section V-C, the model uses a 25 series input window of all available time series to make a prediction of the SOH value 50 cycles ahead into the future.

Table V shows the mean of the error metrics computed during twelve different training sessions. The MAE value of around 0,01 on the SOH prediction reflects an average error of 1% on the capacity loss prediction, which is a good result knowing that the model does not use historical SOH values.

As a benchmark, we study a well performing model based on LSTM layers and historical capacity values only. A window of previously measured SOH values is used as an input to predict a future value of SOH. This approach works well on laboratory data as the test conditions are not representative of a real world scenario and therefore the battery degradation trend is steady. LSTM layers being inherently good at capturing the temporal trend of data, they are ideal for this task. One of these models is studied along the main approach described in this paper. All models used in this comparison use an input window of 25 samples and make a prediction 50 cycles ahead in the future. Note that when a dataset is indicated, it means that the model has been trained and tested exclusively on it.

SNL	Usage	History
	AE-LSTM	LSTM
MAE	$\mathbf{1,030.10^{-2}}$	$\mathbf{0,962.10^{-2}}$
STD	$\mathbf{0,934.10^{-2}}$	$2,995.10^{-2}$
RMSE	$\mathbf{1,320.10^{-2}}$	$\mathbf{1,235.10^{-2}}$
MAPE	1,3107187	1,2858368
NMSE	0,1805153	**0,1406527**

TABLE V: Prediction errors comparison - SNL

	Usage - AE-LSTM	
	PCoE	MIT
MAE	$6,847.10^{-2}$	$2,433.10^{-2}$
STD	$3,700.10^{-2}$	$1,218.10^{-2}$
RMSE	$9,009.10^{-2}$	$2,777.10^{-2}$
MAPE	11,000188	2,6664528
NMSE	9,2791106	0,8071475

TABLE VI: Prediction errors - Other datasets

Table V shows the prediction performances of the models on the SNL dataset. The SNL data leads to the best prediction performances, our AE-LSTM approach shows comparable performance with the history-based LSTM model, with a MAE and a standard deviation of about 1%. The model structure and hyperparameters have been developed on this data so it makes sense that we get the best results on it. The tests on the other datasets are conducted using the exact same model structure developed using the SNL data.

The prediction performances of the AE-LSTM model on the MIT and NASA datasets is presented in VI. The NASA PCoE dataset has a very limited amount of exploitable battery data compared to the MIT and SNL datasets. The prediction error of our AE-LSTM on this dataset is a little higher than with the SNL data with a mean absolute error of about 0,068 but is acceptable knowing that both the training and testing is done using only this data, and that the model structure is developed on the SNL data. The MIT dataset is the largest battery ageing dataset and the one that needs the least pre-processing. The performance of the AE-LSTM model stands between the results obtained from the SNL and the NASA datasets. We observe a MAE of about 0,024 which is encouraging considering the test conditions.

Table VI proves that the AE-LSTM approach is adaptable to different datasets while giving satisfying results.

Table VII shows a comparison between our AE-LSTM approach and other papers. The AE-LSTM model used here still takes a window of 25 cycles as an input to make a 50 cycles ahead prediction.

PCoE	Usage		History	
	AE-LSTM	[17]	[3]	[16]
MAE	$6,847.10^{-2}$	$1,366.10^{-2}$	N/A	N/A
RMSE	$9,009.10^{-2}$	N/A	$0,54.10^{-2}$	$7,986.10^{-2}$

TABLE VII: Prediction errors comparison with other papers

Two of the considered models are based on historical capacity measurements [3], [16] and the third [17] on time features such as charging times, extracted from the CV charging phase of the batteries. These three papers only provide one error metric and are all based on the NASA PCoE dataset. No tests were done on other datasets. These papers were selected for comparison because some of their characteristics are close to those of our AE-LSTM approach while using a publicly available dataset. Liu *et al.* [3] make a 24-step ahead prediction using historical capacity values and a LSTM+GPR model. [16] use 30% or more of the total battery life to predict a capacity degradation.

Although Table VII may make it seem that the performance of our AE-LSTM approach is only comparable to the one of [16], there are multiple reasons that can explain it. First, models based on historical capacity values are expected to perform better since the capacity degradation trend of laboratory data is relatively steady. Our model structure was also developed on data gathered from NMC cells and the comparison made in Table VII on NCA cells data shows that the AE-LSTM approach can adapt to different battery chemistries while maintaining an acceptable prediction performance.

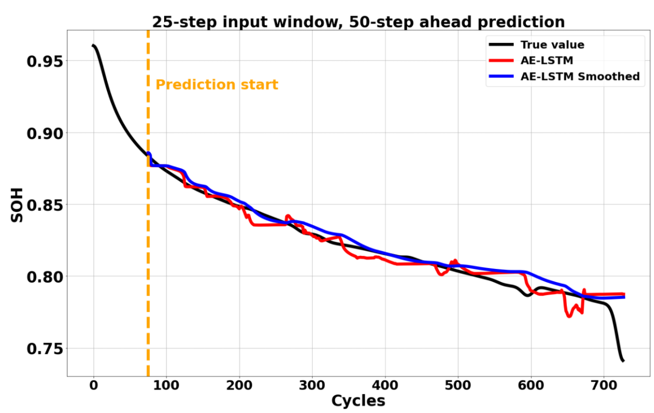


Fig. 5: Example of SOH prediction

Figure 5 shows a visual example of SOH prediction using the AE-LSTM model. It can be seen that the model performs well, and that the smoothing method presented in Section IV-D can lead to a better visualization.

VI. Conclusion and Prospects

A. Conclusion

In this paper, we established that a battery degradation model using usage data would make more sense in a real world scenario. Data such as voltage, current and temperature are easy to gather as current BMS circuits are able to measure them. The goal of the study was to develop a model capable of making a Li-ion battery SOH prediction exploiting usage data, and achieve a prediction performance comparable to a history-based model. For this, an approach making use of auto-encoders for feature extraction and LSTM layers to capture the long-term capacity degradation has been developed. Results are promising, with a MAE of about 0,01 when considering the SNL dataset. To prove the adaptability of the model, tests were also conducted on two other battery ageing datasets, namely the NASA PCoE and MIT datasets, and showed promising performances. The model structure was also developed on NMC cells data and showed acceptable performance on data gathered on NCA cells (NASA dataset) and LFP cells (MIT dataset).

B. Prospects

Using time series that are representative of the usage conditions of a battery becomes really interesting when considering data translating a real world usage scenario. With this in mind, future research work could focus on acquiring data measured on commercially available electric vehicles. An EV can be operated in radically different conditions in a small time frame (holidays, steep ascent, weather, secondhand use...) : the battery capacity degradation trend is more complex than those of batteries aged in laboratories. Therefore, in order to be able to make accurate predictions, future models should be developed based on real-world data. In the meantime, a dataset made of parts of all public battery ageing datasets could be developed to reflect different kind of battery usage conditions.

We have seen that models based on usage data and models based on historical capacity values both have advantages. Therefore, usage dependant time series could be used in combination with a LSTM model based on historical SOH values. This could lead to a hybrid model good at capturing the degradation trend using historical values, and that could adapt to different use cases through the exploitation of time series.

In order to obtain better results, our AE-LSTM model structure could be adapted to the type of data that is used. In this study, the structure used on all datasets was the one developed on the SNL dataset. Moreover, fine-tuning an already existing model to some new data should be considered.

References

[1] Carlos M. Costa, João C. Barbosa, Renato Gonçalves, Helder Castro, F J Del Campo, and Senentxu Lanceros-méndez. Recycling and environmental issues of lithium-ion batteries : Advances , challenges and opportunities. *Energy Storage Materials*, 37(February):433–465, 2021.

[2] Xiaoyu Li, Lei Zhang, Zhenpo Wang, and Peng Dong. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. *Journal of Energy Storage*, 21(November 2018):510–518, 2019.

[3] Kailong Liu, Yunlong Shang, Quan Ouyang, and Widanalage Dhammika Widanage. A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery. *IEEE Transactions on Industrial Electronics*, 68(4):3170–3180, 2021.

[4] Gae Won You, Sangdo Park, and Dukjin Oh. Diagnosis of Electric Vehicle Batteries Using Recurrent Neural Networks. *IEEE Transactions on Industrial Electronics*, 64(6):4885–4893, 2017.

[5] Ji Wu, Chenbin Zhang, and Zonghai Chen. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Applied Energy*, 173:134–140, 2016.

[6] Lei Ren, Jiabao Dong, Xiaokang Wang, Zihao Meng, Li Zhao, and Jamal Deen. A Data-driven Auto-CNN-LSTM Prediction Model for Lithium-ion Battery Remaining Useful Life. *IEEE Transactions on Industrial Informatics*, 17(5):1–1, 2020.

[7] Gae won You, Sangdo Park, and Dukjin Oh. Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach. *Applied Energy*, 176:92–103, 2016.

[8] Wei He, Nicholas Williard, Michael Osterman, and Michael Pecht. Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method. *Journal of Power Sources*, 196(23):10314–10321, 2011.

[9] Austin Downey, Yu Hui Lui, Chao Hu, Simon Laflamme, and Shan Hu. Physics-based prognostics of lithium-ion battery using non-linear least squares with dynamic bounds. *Reliability Engineering and System Safety*, 182(August 2018):1–12, 2019.

[10] Ghoulam Yasser, Paul Theophile, Mesbahi Tedjani, and Durand Sylvain. Modeling, Identification and Simulation of Hybrid Battery/Supercapacitor Storage System Used in Vehicular Applications. *ICEVT 2019 - Proceeding: 6th International Conference on Electric Vehicular Technology 2019*, pages 156–162, 2019.

[11] James D. Kozlowski, Carl S. Byington, Amulya K. Garga, Matthew J. Watson, and Todd A. Hay. Model-based predictive diagnostics for electrochemical energy sources. *IEEE Aerospace Conference Proceedings*, 6:3149–3164, 2001.

[12] Yilu Zhang, Xinyu Du, and Mutasim Salman. Battery state estimation with a self-evolving electrochemical ageing model. *International Journal of Electrical Power and Energy Systems*, 85:178–189, 2017.

[13] Lei Ren, Li Zhao, Sheng Hong, Shiqiang Zhao, Hao Wang, and Lin Zhang. Remaining Useful Life Prediction for Lithium-Ion Battery: A Deep Learning Approach. *IEEE Access*, 6:50587–50598, 2018.

[14] Hong Li, Donghui Pan, and C. L.Philip Chen. Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(7):851–862, 2014.

[15] Yongzhi Zhang, Rui Xiong, Hongwen He, and Michael G. Pecht. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Transactions on Vehicular Technology*, 67(7):5695–5705, 2018.

[16] Jiantao Qu, Feng Liu, Yuxiang Ma, and Jiaming Fan. A Neural-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery. *IEEE Access*, 7:87178–87191, 2019.

[17] Haokai Ruan, Zhongbao Wei, and Hongwen He. State of health estimation of li-ion battery based on regional constant voltage charging. pages 950–952, 2021.

[18] Yu DIng, Chen Lu, and Jian Ma. Li-ion battery health estimation based on multi-layer characteristic fusion and deep learning. *2017 IEEE Vehicle Power and Propulsion Conference, VPPC 2017 - Proceedings*, 2018-Janua:1–5, 2018.

[19] Jialin Li, Xueyi Li, and David He. A Directed Acyclic Graph Network Combined With CNN and LSTM for Remaining Useful Life Prediction. *IEEE Access*, 7:75464–75475, 2019.

[20] Hao Yang, Penglei Wang, Yabin An, Changli Shi, Xianzhong Sun, Kai Wang, Xiong Zhang, Tongzhen Wei, and Yanwei Ma. Remaining useful life prediction based on denoising technique and deep neural network for lithium-ion capacitors. *eTransportation*, 5:100078, 2020.

[21] Benvolence Chinomona, Chunhui Chung, Lien-Kai Chang, Wei-Chih Su, and Mi-Ching Tsai. Long Short-Term Memory Approach to Estimate Battery Remaining Useful Life Using Partial Data. *IEEE Access*, 8:165419–165431, 2020.

[22] Preger et al. Degradation of Commercial Lithium-Ion Cells as a Function of Chemistry and Cycling Conditions. *Journal of The Electrochemical Society*, 2020.

[23] Severson et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4:383—-391, 2019.

[24] B. Saha and K. Goebel. Battery data set. *NASA Ames Prognostics Data Repository*, 2007.