CHEMICAL ENGINEERING TRANSACTIONS

VOL. 94, 2022

A publication of AIDIC

The Italian Association
of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Petar S. Varbanov, Yee Van Fan, Jiří J. Klemeš, Sandro Nižetić
Copyright © AIDIC Servizi S.r.l.
ISBN 978-88-95608-93-8 **ISSN** 2283-9216

DOI: 10.3303/CET2294131

Cyclic Degradation Prediction of Lithium-Ion Batteries using Data-Driven Machine Learning

Lerissah D. Lim[a], Andrei Felix J. Tan[a], Jan Goran T. Tomacruz[a], Michael T. Castro[a], Miguel Francisco M. Remolona[b], Joey D. Ocon[a,]*

[a]Laboratory of Electrochemical Engineering, Department of Chemical Engineering, University of the Philippines Diliman Quezon City 1101, Philippines
[b]Chemical Engineering Intelligence Learning Laboratory, Department of Chemical Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines
jdocon@up.edu.ph

Accurately estimating the capacity degradation of lithium-ion (Li-ion) batteries is vital in ensuring their safety and reliability in electric vehicles and portable electronics. Future capacity estimation using machine learning (ML) models allow battery lifetime predictions with minimal cycling data in the train set, well before capacity degradation occurs within the cell. The use of ML methods removes the need for prior knowledge of cell chemistry and the physical and chemical behaviors of batteries. In this paper, the data-driven ML models Gaussian process regression (GPR) and recurrent neural network – long short-term memory (RNN-LSTM) estimated the charge capacity of Li-ion batteries from the Oxford Battery Dataset, using only the battery's cycle index and capacity as input. With only 15 % of the battery’s lifetime as training data, the GPR model achieved a mean average percent error (MAPE) of 8.335 % and an R^2 of 0.9755, while the LSTM model achieved a MAPE of 9.984 % and an R^2 of 0.9898. These results indicate the goodness of fit and are comparable to results from similar models in the literature (MAPE = 9.1 to 15 %). The methodology may be applied to different features to help establish the relationship between health indicators and capacity fade and can be used in applications that require early capacity prediction such as in space technologies where lifetime and capacity are crucial in ensuring success and safety. This successful estimation highlights the promise and potential of accurately predicting Li-ion battery capacity degradation using a single-feature approach.

1. Introduction

In the shift towards the widespread application of renewable energy technologies, the role of battery energy storage systems (BESS) has expanded in terms of power generation, the progress of electric vehicles, and high-capacity mobile devices. Specifically, the technology of rechargeable Li-ion batteries positioned it on the frontline of mobile consumer electronics. Its features include high energy density in compact packaging, allowing its use in various mobile applications, such as in electric vehicles, aerospace, consumer electronics, and industrial applications (Severson et al., 2019). Apart from its expansion in the market, Li-ion batteries are continuously the subject of competitive research because of promising technology, with rapid developments in the areas of efficiency, reliability, safety, and management systems. However, Li-ion is still restricted by limited lifespan, high costs, and safety issues. These factors affect the overall performance and applicability of Li-ion batteries (Park et al., 2020). Understanding and predicting the capacity fade will aid in the mitigation of battery degradation. Furthermore, the study may also give an insight into improving battery design.

Though several approaches such as experimental and model-based estimation models have been explored to predict battery capacity, they are barred by limitations, such as the large amounts of calculation required, that make them unsuitable for real-time capacity estimation. Some of these difficulties and restraints are evident in acquiring the required parameters, the knowledge of aging mechanisms, and the high dependence of the accuracy on the model. Data-driven machine learning methods can be used to overcome these limitations and estimate battery capacity with no prior knowledge of the battery chemistry. The advanced capabilities of the ML algorithms allow it to build analytical models that will allow the prediction of the capacity fade of the battery.

Paper Received: 15 April 2022; Revised: 01 July 2022; Accepted: 02 July 2022
Please cite this article as: Lim L.D., Tan A.F.J., Tomacruz J.G.T., Castro M.T., Remolona M.F.M., Ocon J.D., 2022, Cyclic Degradation Prediction of Lithium-Ion Batteries using Data-Driven Machine Learning, Chemical Engineering Transactions, 94, 787-792 DOI:10.3303/CET2294131

However, ML models typically rely on large data and multiple input features to accurately estimate the capacity of the battery. In this study, the potential of using a single-feature approach was explored. While this would limit the descriptors of the dataset, this would also greatly reduce the amount of data required to conduct capacity estimation. A study by Wang et al. (2018) extracted the geometrical feature in the constant voltage profile and suggested a strong relationship between the remaining capacity of the battery and the extracted aging profile, which was mapped out using support vector regression. In this work, battery capacity degradation was performed on the Oxford Battery Dataset using the Gaussian process regression (GPR) and recurrent neural network – long short-term memory (RNN-LSTM) models. The findings suggest the competitiveness of using a single input feature in accurately estimating the capacity of the battery with existing benchmark studies utilizing multiple input features. It highlights the potential of conducting similar studies requiring the minimum amount of data and computational power.

2. Methodology

This work (Figure 1) demonstrates the potential of single-feature capacity estimation. The cycle index and charge capacity of the chosen dataset were extracted from the available cycle data and were used in the ML models. At a specified train to test split, the generated models predicted the future capacity of the cell. The results were then evaluated and compared with those of related publications.

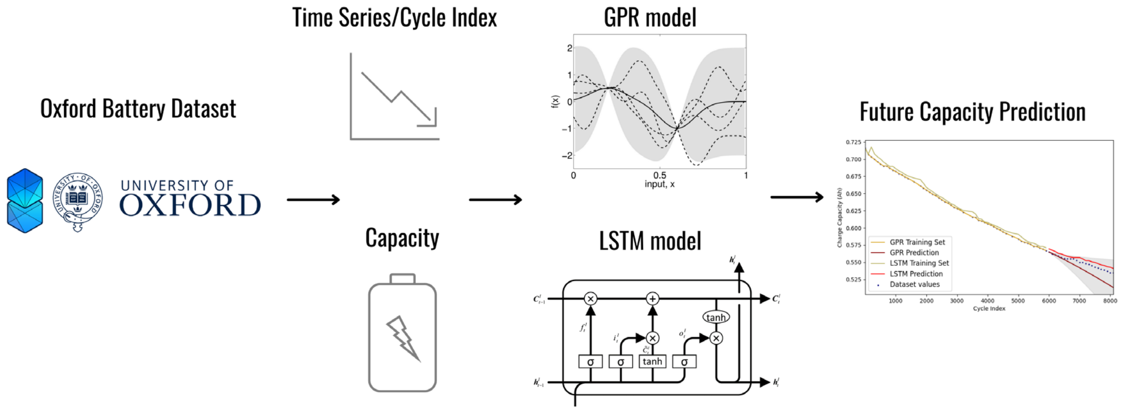

Figure 1: Summary of the methodology flow in this work

2.1 Battery Dataset

The Oxford Battery Degradation Dataset was taken from the Battery Archive website (2021) which contains long-term battery aging data from eight Kokam (SLPB533459H4) 740 mAh lithium cobalt oxide (LCO) pouch cells tested using the Bio-Logic MPG-205, 8-channel battery tester. The cells were tested in a Binder thermal chamber at 40 ºC and were exposed to a constant-current-constant-voltage charging profile followed by a drive cycle discharging profile. Cycle-aging measurements were taken every 100 cycles (Birkl et al., 2017). The repository included cycle data and time-series data. From the cycle data, the cycle index and charge capacity were taken to be the input and output of the generated model.

2.2 Prediction Method

Future capacity estimation, or direct learning, uses data from early cycles to estimate the capacity of the same cell for future cycles. The method splits data in a singular cell into training and testing sets. To observe the performance of the models at different train to test splits, sequential fitting was performed. This enabled errors to be obtained per split from a 5 %: 95 % to a 90 %: 10 % train to test split. The main objective of direct learning is to achieve low errors despite small train sets. This will make capacity estimation before degradation within the cell occurs possible. For purposes of comparability, the results shown are at 70 %: 30 % and 15 %: 85 % train to test splits to determine the effectivity and capacity of the models for early capacity prediction. Specifically, the training set was reduced to 15 % of the dataset to show the capability of the model for early capacity prediction.

2.3 Prediction Models

A Gaussian process model is a probabilistic supervised machine learning method that is capable of performing either regression or classification tasks. Gaussian processes define a probability distribution over a set of functions and are fully specified by a mean function, $m(x)$, and a covariance function, $\kappa(x, x')$, as shown in Eq(1).

$$F(x) = GP\big(m(x), k(x, x')\big) \qquad (1)$$

GPR models are capable of making predictions and providing uncertainty measures over these predictions. It is a non-parametric Bayesian approach to solving regression problems. To select an appropriate model, non-parametric methods allow models to grow with an increase or decrease with the available data, as opposed to parametric models which have a fixed number of parameters (Orbanz and Teh, 2010). Additionally, as with any Bayesian method, Gaussian processes begin with a prior distribution, or a probability distribution of possible values for a certain prediction and update this distribution as more data points are introduced and observed. The non-parametric and probabilistic approach of GPR are factors that cause it to be an ideal prediction method to approximate non-linear systems, where the parametric forms of unknown processes are difficult to obtain. In this paper, the Gaussian process regression was implemented using the scikit-learn package. The Gaussian process regressor application programming interface (API) is used to implement regression methods involving Gaussian processes. For this API, the prior mean, $m(x)$ is assumed to be zero or equal to the mean of the training data. The covariance, $f(x, x')$ can be specified using a kernel. Table 1 summarizes the settings used in this paper to perform the capacity estimation using GPR.

Table 1: Summary of the settings used for GPR

Parameter	Setting
Kernel	Matern (v = 3/2)
Restarts optimizer	10 to 100
Confidence interval	95 %

A long short-term memory network is a type of recurrent neural network that determines the order dependence in problems involving sequence predictions, making it a suitable tool for time-series predictions. An advantage of using deep-learning algorithms such as LSTM for complex problems requiring real-time calculations such as battery capacity estimation is that they are capable of efficiently extracting features directly from raw data even in relatively smaller datasets. LSTM is described by the Eqs (2-7), see (Park et al., 2020).

$$f_t = \sigma\big(w_f \cdot [h_{t-1}, x_t] + b_f\big) \qquad (2)$$
$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \qquad (3)$$
$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \qquad (4)$$
$$\tilde{C}_t = \tanh(w_C \cdot [h_{t-1}, x_t] + b_C) \qquad (5)$$
$$C_t = f_t^* C_{t-1} + i_t^* \tilde{C}_t \qquad (6)$$
$$h_t = o_t^* \tanh(C_t) \qquad (7)$$

where σ is the sigmoid activation function, $tanh$ is the hyperbolic tangent activation function, w consists of a set of matrices containing the different weights, x_t is the input at the current timestep. The LSTM architecture used in this paper was developed using Python 3 environment, using TensorFlow 1.14 as the backend, and using the Keras API to create the layers for the deep learning environment. The API used in this paper utilizes the “Adam” optimizer to make the predictions more accurate. Additionally, the mean absolute error (MAE) is used to compile the algorithm. Other parameters such as learning rates, train to test splits of the training dataset, number of epochs, and dropout value are pre-configured and are summarized in Table 2.

Table 2: Summary of the settings used for LSTM using TensorFlow and Keras API

Parameter	Setting
Training loss	MAE
Learning rate	0.0001
Number of epochs	75 %
Dropout value	20 %

2.4 Performance Evaluation Tools

To quantitatively evaluate the performance of the ML model, it is necessary to ensure the comparability of the results with other similar studies using error metrics and performance evaluation tools. The mean absolute percentage error (MAPE) is an indicator used as a loss function for regression problems. It is given by Eq(8).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{\hat{y}_i}\right| \tag{8}$$

where n is the number of observations, y_i and $\hat{y}_i$ are the actual and the predicted values. The root mean square error (RMSE) is the difference between the actual and predicted values where larger absolute values are penalized as they contribute more weight to the RMSE (Li et al., 2020). The RMSE is given by Eq(9).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

Additionally, the predicted data is plotted against the actual data to easily visualize the predictive performance of each machine learning model. The correlation of these data will be measured using the coefficient of determination, R^2. Given the actual capacity y_i, the coefficient of determination is given by Eq(10).

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(\bar{y} - y_i)^2} \tag{10}$$

3. Results and Discussions

With a 70 %: 30 % train to test split, the GPR model accurately predicts the future capacity trajectory of each cell using only a single feature as the input data as shown in Figure 2a. Regression analyses show an average RMSE and MAPE across all cells of 0.0133 Ah and 1.633 %, indicating a good prediction performance. Similarly, the LSTM model using the same split accurately predicts the future capacity of each cell. The model obtained an average RMSE and MAPE of 0.0074 Ah and 1.134 %. Both prediction errors for the GPR and LSTM models are within the range of errors from various benchmark studies, which use multiple features to predict the battery capacity. These findings present the capability of the model to predict the battery capacity using minimal data.

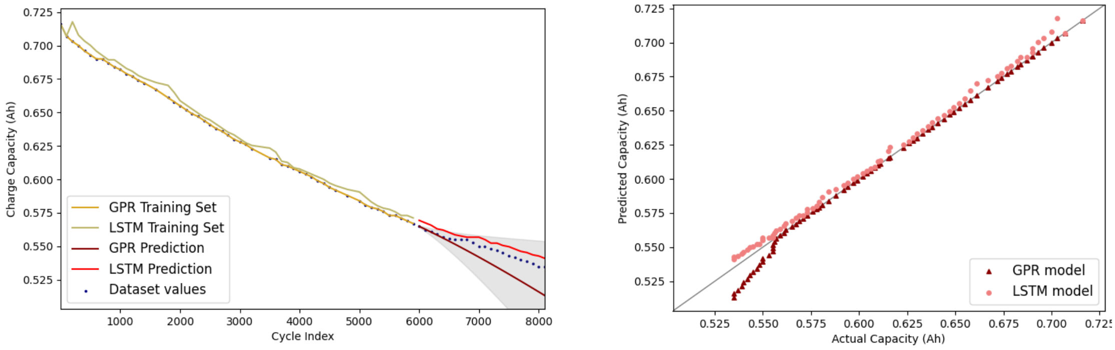

Figure 2: (a) Future capacity prediction for GPR and LSTM and (b) Predicted vs actual capacity plot for 70 %: 30 % train to test split

Moreover, the predicted capacities were plotted against the actual measured capacities available in the dataset to determine the correlation between the two capacities. Good prediction performance is indicated by a good fit between the predicted and actual capacities (i.e., with a R^2 value close to 1). At the same split, the average calculated R^2 for all cells using the GPR and LSTM model are 0.9948 and 0.9943, indicating that the predictions highly correlate to the measured capacities in all cells (Figure 2b).

The training set was further reduced to 15 % of the dataset, as shown in Figures 3a and 3b to determine how changing the length of the training set affects the prediction performance of the model. The anticipated trend is that decreasing the training set will increase the errors in the model due to a higher possibility of overfitting. Regression analysis of the GPR model revealed an average RMSE and MAPE of 0.0598 Ah and 8.355 %. Similarly, the LSTM model obtained an average RMSE and MAPE of 0.0625 Ah and 9.894 % (Figure 3a). Figure 3b confirms that a smaller training set provides a lesser accurate prediction of the battery capacity, as their errors are relatively larger than those from their 70 %: 30 % train to test split counterpart. The R^2 of the GPR and LSTM models for this split also reveal a relatively weaker correlation between the predicted and the actual capacities, which further supports the trend for each split. This is anticipated as the length of the training set increases the ability of the model to map out the relationship between the battery capacity and the different health indicators. Despite this, the values of R^2 achieved by the models, 0.9755 and 0.9898 by the GPR and LSTM models still indicate a good fit.

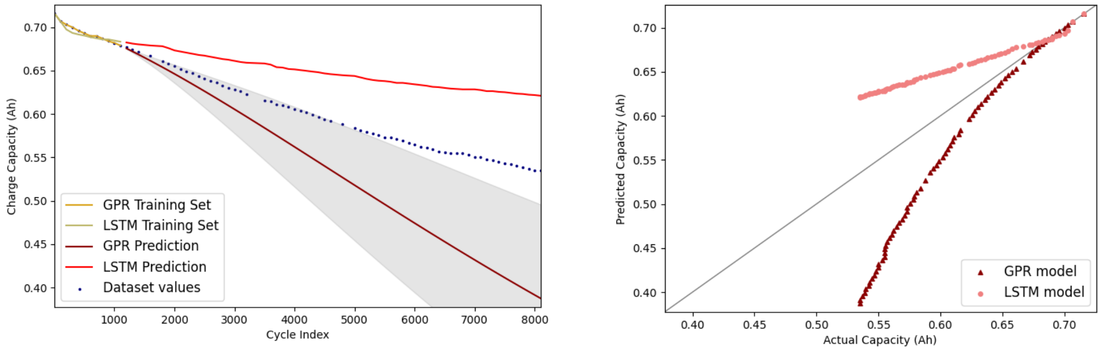

Figure 3: (a) Future capacity prediction for GPR and LSTM and (b) Predicted vs actual capacity plot for 15 %: 85 % train to test split

Summarized in Table 3 are the average errors using various metrics for each prediction method at 70 %: 30 % and 15 %: 85 % train to test splits. It can be observed that the errors garnered by the two ML models do not vary much and similar trends can be observed even with the use of different error metrics. This validates the performance of the models and their efficiency through different train to test splits. To further confirm the validity of the results, benchmarking was also performed.

Table 3: Summary of errors and benchmark studies for comparison of errors

Authors	Battery Dataset	Regression method	Train to Test Split (%)	RMSE (Ah)	MAPE (%)	R^2 (predicted vs actual)
This work (2022)	Oxford Dataset	GPR	70:30	0.0133	1.633	0.9948
			15:85	0.0598	8.355	0.9755
		LSTM	70:30	0.0074	1.134	0.9943
			15:85	0.0625	9.894	0.9898
Guo et al. (2019)	NASA Dataset	RVM Regression	70:30	0.010222	-	-
Garg et al. (2018)	18650 Li-ion batteries	GP hybrid with NN, SVM	70:30	-	1.96 to 6.0	-
Severson et al. (2019)	Severson et al. (2019) 124 Li-ion cells	Elastic Net	12.5:87.5	-	9.1	-
Attia et al. (2018)		Elastic Net, RFR, AdaBoost regression	12.5:87.5	-	10 to 15	-

Overall, the predictions from the generated models are comparable with results found in other published works. At a 70 %: 30 % train to test split, the group of Guo et al. (2019) and Garg et al. (2018) achieved an RMSE of 0.0102 Ah and a MAPE of 1.96 to 6.0 %. At the same split, the generated models in this work achieved an RMSE and MAPE as low as 0.0074 Ah and 1.134 %. Meanwhile, the group of Severson et al. (2019) and Attia et al. (2018) worked on the same dataset with a training set of 12.5 % and achieved a MAPE of 9.1 % and 10 to 15 %. With a training set of 15 %, the generated models using GPR and LSTM were able to achieve an average MAPE of 8.355 % and 9.894 %. The comparability of the obtained results with that of other published works confirms the validity of the models and their generated results. This also suggests the competitiveness of using only a single-feature approach in estimating battery capacity with ML methods, as it minimizes the input data required to accurately estimate the capacity of the battery.

4. Conclusion

Battery capacity estimation is essential in improving the design, safety, performance, and efficiency of Li-ion batteries; however, it requires a large amount of data and several battery features to accurately estimate the capacity of the battery. This paper investigates the potential of ML models with only a single input feature to accurately estimate the capacity of Li-ion batteries. Future capacity estimation was successfully performed on the Oxford Battery Dataset using GPR and RNN-LSTM. The results from the generated models are that of the 70 %: 30 % and 15 %: 85 % train to test splits with errors presented as averages across all cells in the dataset. In terms of the error metrics, the LSTM model performed better with a 70 % training set with an RMSE and MAPE of 0.0074 Ah and 1.134 %. The GPR model produced less error with a 15 % training set as it incurred an RMSE of 0.0598 Ah and a MAPE of 8.355 %. The acceptable errors achieved prove the prospect of early capacity prediction with minimal input data. The results obtained were also found to be comparable and competitive with those of other related published works which validate the potential of ML models to estimate the capacity of the battery using the single-feature approach. These models may be further developed through hyperparameter tuning to improve prediction accuracy. Additionally, the methodology may be applied to different features to help establish the relationship between health indicators and capacity fade.

Nomenclature

API – application programming interface
BESS – battery energy storage systems
GP – Genetic programming
GPR – Gaussian process regression
LCO – lithium cobalt oxide
Li-ion – lithium-ion
MAE – mean absolute error, Ah
MAPE – mean absolute percentage error, %
ML – machine learning
R^2 – coefficient of determination
RFR – random forest regression
RMSE – root mean square error, Ah
RNN-LSTM – recurrent neural network – long short-term memory
RVM – Relevance Vector Machine
SVM – support-vector machine

Acknowledgments

The authors would like to acknowledge the Department of Science and Technology (DOST) through the Niche Centers in the Regions for R&D (NICER) Program and The Commission on Higher Education – Philippine California Advanced Research Institutes (CHED-PCARI) through the CIPHER Project (IIID 2018-008).

References

Attia, P. M., Deetjen, M. E., Witmer, J. D, 2018, Accelerating battery development by early prediction of cell lifetime, <cs229.stanford.edu/proj2018/report/206.pdf>, accessed 14/12/2021

Battery Archive, 2021, Oxford University (OX), BatteryArchive.org <batteryarchive.org/list.html>, accessed 07/09/2021.

Birkl, C. R., Roberts, M. R., McTurk, E., Bruce, P. G., Howey, D. A., 2017, Degradation diagnostics for lithium ion cells, Journal of Power Sources, 341, 373–386, DOI: 10.1016/j.jpowsour.2016.12.011

Garg, A., Peng, X., Le, M. L. P., Pareek, K., Chin, C. M. M., 2018, Design and analysis of capacity models for Lithium-ion battery, Measurement, 120, 114–120, DOI: 10.1016/j.measurement.2018.02.003

Guo, P., Cheng, Z., Yang, L., 2019, A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction, Journal of Power Sources, 412, 442–450, DOI: 10.1016/j.jpowsour.2018.11.072

Li, Y., Li, K., Liu, X., Zhang, L., 2020, Fast battery capacity estimation using convolutional neural networks, Transactions of the Institute of Measurement and Control, 0142331220966425, DOI: 10.1177/0142331220966425

Orbanz, P., Teh, Y. W., 2010, Bayesian Nonparametric Models, Encyclopedia of Machine Learning, <groups.seas.harvard.edu/courses/cs281/papers/orbanz-teh-2010.pdf>, accessed 08/01/2022

Park, K., Choi, Y., Choi, W. J., Ryu, H.-Y., Kim, H., 2020, LSTM-Based Battery Remaining Useful Life Prediction With Multi-Channel Charging Profiles, IEEE Access, 8, 20786–20798, DOI: 10.1109/ACCESS.2020.2968939

Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggedakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., Braatz, R. D, 2019, Data-driven prediction of battery cycle life before capacity degradation. Nature Energy, 4(5), 383–391, DOI: 10.1038/s41560-019-0356-8

Wang, Z., Zeng, S., Guo, J. Qin, T., 2018, Remaining capacity estimation of lithium-ion batteries based on the constant voltage charging profile, PloS one, 13(7), e0200169, DOI: 10.1371/journal.pone.0200169