Preprints, 10th IFAC International Symposium on Advances in Automotive Control Columbus, Ohio, USA, August 28-30, 2022

Lithium-ion Cell Ageing Prediction with Automated Feature Extraction

Jose Genario de Oliveira Jr. * **Cisel Aras** ** **Thyagesh Sivaraman** *** **Christoph Hametner** *

* *Christian Doppler Laboratory for Innovative Control and Monitoring of Automotive Powertrain Systems TU Wien, Vienna, Austria (e-mail: genario.oliveira,christoph.hametner@tuwien.ac.at).*
** *AVL Research and Engineering, Istanbul, Turkey, (e-mail: cisel.aras@avl.com)*
*** *AVL List GmbH, Graz, Austria, (e-mail: thyagesh.sivaraman@avl.com)*

Abstract: This paper aims to investigate how some features commonly associated with more generic time-series analysis are associated with capacity fade in lithium-ion cells and how they can be used to create simple but effective machine-learning models. This is done by processing the current, voltage, and temperature measurements, which span around two hundred cells for roughly two years, with a popular automated time-series analysis routine that extracts a significant number of different characteristics from the dataset for each signal. The most promising factors associated with the capacity fade are obtained by using a feature selection technique that is simple, quick and does not depend on a specific model structure. An analysis of the most relevant results is done, together with a standard hyperparameter search strategy using bayesian optimization for different classical regression models. With this step-by-step approach, the most promising features were investigated and an average error smaller than 5% was obtained on previously unseen validation data.

Keywords: Battery management systems; Energy storage systems: electrochemical systems, supercapacitators, fuel cells

1. INTRODUCTION

With the widespread usage of Lithium-ion batteries currently in almost every electronic device and more recently, in the automotive industry, it is paramount to have models that predict battery degradation in an accurate fashion. This is due to the significant monetary and economical costs associated with replacing these battery packs, making it vital to understand how the capacity fade process occurs and what factors have the most impact on it. The main usage of these models is to correctly predict the remaining useful life of a given cell, which is directly related to the safety and maintenance of such devices. Usually, the ageing models are either physics-based, purely data-driven or hybrid. The former presents a significant challenge from the modelling perspective due to a wide variety of underlying failure modes that cause ageing from mechanical, electrochemical or chemical phenomena. In physics-based approaches, ideally each side-reaction causing the capacity to decrease is modelled, usually by simulating them on an electrochemical battery model such as the one shown in Doyle et al. (1992). A good overview of this is given in Reniers et al. (2019) and more generally also for other approaches in Hu et al. (2020). The modelling of all these effects requires a deep knowledge of the cell itself and depends on its geometry and chemistry. Additionally, due to the complexity of such approaches, sometimes only one of the several processes related to age is modeled, e.g. the solid electrolyte interface (SEI) growth or loss of cyclable lithium. On the other hand, data-driven approaches normally uses curve-fitting tools on a certain amount of data to obtain the model parameterisation. A large-scale dataset of LFP cells was aged using various charging profiles in Severson et al. (2019) and ML tools were used to predict and classify cells by cycle life. In Aykol et al. (2021), several architectures consisting of physics-based and machine learning models are investigated in order to improve the forecast of battery life. A holistic data-driven ageing model is developed under a Gaussian Process framework in Lucu et al. (2020) and capacity degradation under storage conditions is investigated with a limited number of features.

This paper uses a data-driven method due it being simpler, yet still achieving enough accuracy for the objective, which is prediction of remaining useful life. An important distinction should be made here between ageing estimation and prediction. For ageing estimation, the task is to estimate the current cell capacity given a snapshot of operational data. On the other hand, the focus here is on ageing prediction, where the goal is to predict ahead of time what will happen with the capacity given a certain predetermined usage.

Several battery datasets that can be used for ageing prediction, such as Saha and Goebel (2007) and Luzi (2018), are becoming increasingly more available, enabling the usage

of models that require more testing data than the classic ageing modelling approaches.

This paper presents an approach in which the feature extraction step is ideally as fully automated as possible and different model structures are tested to achieve a low validation error on previously unseen data. The main contributions are both the analysis of the extracted features in the context of battery ageing modelling, giving some insights on potential candidates for other data-driven approaches and how they perform to predict the capacity fade when tested with different model structures.

This work is organized as follows: Section 2 gives a brief explanation on the dataset that was utilized. Section 3 describes the feature extraction procedure. Section 4 consists of the feature selection problem with a discussion on the most relevant features that were found. Finally, Section 5 shows a comparison of the results obtained with different model structures, combined with a bayesian hyperparameter search strategy.

2. FEATURE EXTRACTION PROCEDURE

2.1 Battery ageing data

For this work, the dataset consists of measurements from about two hundred NMC 18650 cells with graphite anodes. These were aged under diverse conditions, with both accelerated ageing tests and calendar ageing experiments present, in which the main design variables changed were storage state of charge(SoC), ranging from 5 to 95% and chamber temperature, from -10 to 60°C. For the accelerated ageing data, the conditions that were changed are: Temperature(T), constant charging current(CC), peak discharge current(PDC), average SoC(SOC) and delta depth of discharge(dDoD). For a more detailed analysis of the dataset, the testing equipment and procedures, please refer to de Oliveira et al. (2021).

From time to time, a reference test cycle is conducted at $25°C$, aiming to estimate the cell capacity at the same temperature and irrespective of the accelerated ageing profile conducted on that specific cell. By associating a capacity fade ΔQ to the interval between two consecutive reference test cycles, the modelling challenge is then to map the set of inputs time-series, namely current I, voltage V and temperature T to the capacity fade output. A visual explanation of this is shown in Fig. 1. Finally, for the methods outlined in this work, the cells were sorted according to the total capacity lost, being evenly assigned to training and validation datasets.

2.2 Automatic Feature Extraction

In order to solve the problem at hand, relevant characteristics need to be extracted from the input signals to obtain a standard regression problem. To find out which of the characteristics, or features, are meaningful, the strategy here is to use an automated tool that extracts a high amount of features and then filters the irrelevant ones out. The feature extraction is done with the time-series analyzer tsfresh presented in Christ et al. (2016). Some of the original features had to be removed for this work due to the high number of points of some I, V, T signals when evenly sampled (around 10 million, one second sampling time), making the computation time impractical. Table 1 displays some of the features, computed for each input signal, that might be relevant for ageing, the complete list of features is available in the documentation of tsfresh. It is interesting to note that some features have a physical interpretation, such as average temperature, which is used in plenty of previous works to explain an Arrhenius dependence associated with the capacity fade, or even the signal energy of the current, which is associated with the total dissipated thermal power in the cell according to Ohm's Law. Most of the features however, do not have a direct physical interpretation, e.g kurtosis of the voltage, longest strike above mean temperature. This is the price to be paid by the utilization of such purely data-driven methods, in which it might be the case that the best model uses some features that were expected and known to have an impact on the capacity fade, together with a feature representing e.g. the first coefficient of a wavelet transform, which is harder to interpret. It is also important to mention that there are some features that depend only on the time-series, and others that depends on some user-defined parameters. An example of this is the autocorrelation, that depends both on the signal and on the number of lags, making it harder to automate this step entirely. This problem was addressed here by making a different feature for each predefined parameter, in the case of the autocorrelation for example, n different features are created, each associated with a different lag.

2.3 Feature Computation

The features associated with each of these data intervals, as shown in Fig. 1, are then computed by a weighted average of the features extracted using tsfresh from the individual tests that are in said interval, with the weight being the length of the individual tests, no windowing strategy is being used. Given that for basically all the tests, the duration of the accelerated ageing cycles, which were done at several temperature levels, are much larger than for the reference cycles, in the end they are the main drive of the feature computation in a data interval.

Table 1. Some of the extracted features

Feature and Description
Signal Energy - $\sum_1^n x_i^2$
Absolute Sum of Changes - $\sum_1^{n-1} \lvert x_{i+1} - x_i \rvert$
Mean absolute differences - $\frac{1}{n}\sum_1^{n-1} \lvert x_{i+1} - x_i \rvert$
Complexity measure cid - $\sqrt{\sum_{i=0}^{n-2lag}(x_i - x_{i+1})^2}$
Nonlinearity measure c3 - $\frac{1}{n-lag}\sum_{i=0}^{n-2lag} x_{i+2lag}^2 x_{i+lag} x_i$
Number of points - N
Number of points below average
Energy Ratio by chunks of the signal
FFT coefficients - $A_k = \sum_{m=0}^{n-1} a_m exp(-2\pi \frac{mk}{n})$
First location of maximum/minimum value
Maximum/minimum value of the time series
Number of duplicates of maximum/minimum value
Index of mass quantiles
Kurtosis of the time-series distribution
Sample Entropy

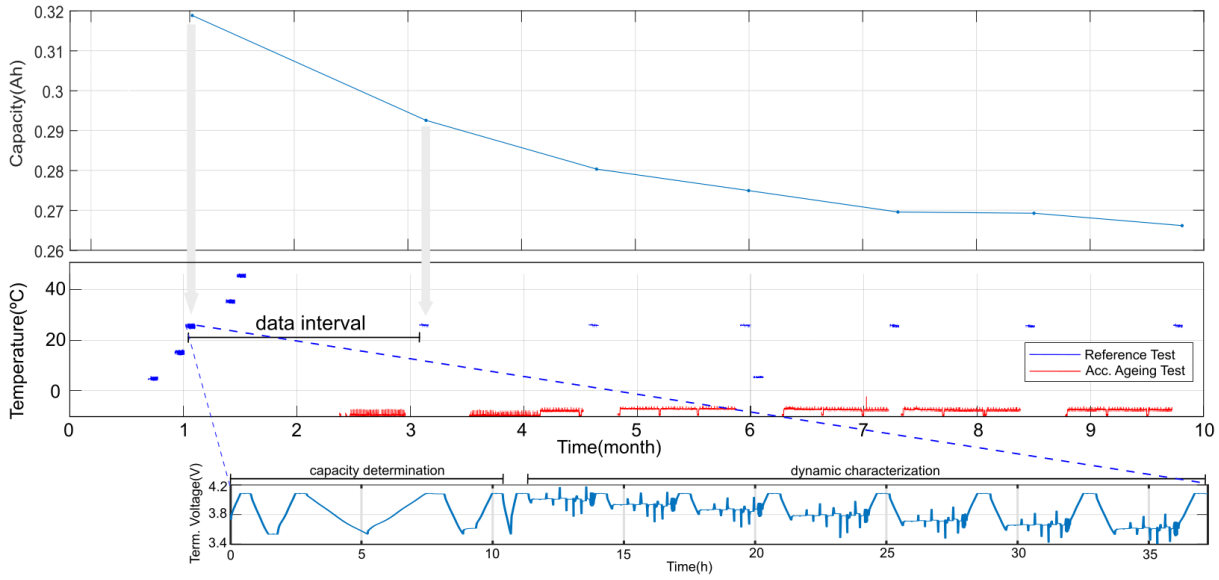

Fig. 1. Data intervals shown over a cell ageing tests for 10 months, associated capacity differences with each interval and reference test at $25°C$

3. FEATURE SELECTION

Once the features extraction step is done, it is necessary to choose among those which ones are the most relevant for the model. Usually these feature selection tools are split into supervised and unsupervised methods i.e. if the regression output should be considered in the selection process. Also there is the division between wrapper and filter strategies, where the former basically train the model for each combination of features and according to some metric, select the best performing feature combination and the latter is not dependent on a specific model structure or training, presenting a trade-off between model performance and simplicity. The methodology used here was presented in Ferreira and Figueiredo (2012) and is an unsupervised filter method. The reason behind the choice is that it might happen that two or more different features are good predictors together for the capacity fade and poor predictors individually, the disadvantage of this approach is that the number of selected features might be overestimated. Also, a filter method is considered, this is because wrapper methods can take a large computational time to obtain the optimal feature combination e.g. if $f_{number} = 300$ features are extracted from the data, there will be a total of $2^{f_{number}}$ feature combinations that the wrapper method could have to search theoretically to find the optimal features. While usually wrapper approaches use a heuristic search algorithm to reduce the number of feature combinations that are tried, the point that it is model dependent and time-consuming still holds. Given that the hyperparameters of different model structures are also being optimized at a later stage, this would ideally have to be done for each set of features, resulting in an enormous computational burden.

3.1 Dispersion and similarity metrics

The proposed filters in Ferreira and Figueiredo (2012) are divided into dispersion and similarity metrics. The idea behind the dispersion metric is to assume that the feature importance depends on its dispersion measure, which is presented next. The reasoning behind it is that the important features without dispersion will not make it to the final model regardless of importance, e.g. modelling temperature dependence on a dataset in which the temperature is constant, and later a statistical test is conducted to further filter out the selected features which are uncorrelated with the output.
The dispersion metric of choice here is the the mean-median (MM), defined as

$$MM_i = |\bar{X}_i - median(X_i)|, \tag{1}$$

where X_i is the i-th feature of the feature matrix X and MM_i the mean-median associated with that feature. This dispersion measure was selected because it shows good results on a dense data matrix, another alternative is the mean-absolute-deviation(MAD), but the mean-median was marginally better on the datasets reported in Ferreira and Figueiredo (2012). Since this metric is scale-variant, the dataset was normalized to zero mean and standard deviation one before this step. For the similarity measure, the metric of choice was the sample correlation coefficient $\rho(X_i, X_j)$, mostly due to its simplicity, robustness and ease of computation.

3.2 Feature selection algorithm

Here a small recapitulation of the algorithm presented in Ferreira and Figueiredo (2012) is given:

(1) Sort the features in a descending order regarding a dispersion measure, here the mean-median was chosen.
(2) Always include the first ranked feature in the group of selected features.
(3) Compare the similarity of the i-th ranked feature with the previously included feature using a similarity measure and if this value is lower than a threshold, include that feature and repeat this step until you have a predetermined number of features m in the dataset.

The method of choosing m is heuristic and it is chosen as the value which

$$\frac{\sum_{i=1}^{m} MM_i}{\sum_{i=1}^{n} MM_i} \leq L, \tag{2}$$

with L being fixed here to 0.95.

3.3 Feature Selection Results

After normalizing and processing the data to treat the outliers and removing features with standard deviation 0, there are 409 remaining features in the dataset, 1046 data points for training and 1072 for validation. From the heuristic in (2), the maximum number of features here is $m = 300$ features. Since it is known that there is a significant amount of collinear features in the dataset e.g. length identical for I,V and T and that corresponds to three different features in the dataset. The maximum value for the correlation is set to 0.05. This is done to avoid collinearity in the features. There was 219 selected candidate features from the original 400 at this stage. Note that m denotes the maximum number of features selected, but does not impose a lower bound on this number. At this stage the assumption that the dispersion measure is related to feature importance still holds. One way to account for this limitation is by doing an additional step to filter the results that actually have non-zero correlation with the outputs. A common approach for regression is by considering the correlation between the individual features and the output. There are different correlation coefficients to chose, with the Pearson's Correlation Coefficient being the most popular approach. However, a linear relationship between the variables is considered and this is not necessarily true. Due to that, the Kendall's rank correlation coefficient is used

$$\tau_k = \frac{n_c - n_d}{\frac{1}{2}N(N-1)}, \tag{3}$$

where n_c and n_d denotes here the number of concordances and discordances between two variables and N is the number of data points. For reasonably large N, the τ_k coefficient distribution under H_0, the null hypothesis that both variables are uncorrelated, have $E(\tau) = 0$ and $Var(\tau) = \frac{2(2N+5)}{9N(N-1)}$. Additionally, the variable z defined as

$$z = \frac{3\tau_k\sqrt{N(N-1)}}{\sqrt{2(2N+5)}} \tag{4}$$

approximately follows a standard normal distribution, as shown in P.Sprent and N.C.Smeeton (2001). The null hypothesis H_0 presented above is considered for each feature X_i with the output variable Y, the capacity fade. This procedure reduces the number of selected features even more, to 157 values. This number can change depending on the threshold of the similarity metric, which was not done here, the maximum cosine distance was set to 0.05. Table 2 shows 14 features for this dataset when sorted by dispersion. Table 3 shows the 17 most correlated features with the capacity loss. It is worth mentioning that the two tables showing the feature importance are widely different depending on how they are ranked. Some of these results have an immediate physical interpretation, such as the ratio of the I outside $1A$, which is directly linked to how harsh the cell is being cycled, the "Count of zero values for I" is basically how much the cell is at rest at a given time window. Others are harder to interpret or might have no possible physical explanation at all, such as the complexity measure for the signal or the Variance of the FFT coefficients of I. Additionally, some of the features shown in table 2, such as "Ratio of T greater than mean $\pm 10°C$" are linked directly with how much thermal stress is present on the cell during a given interval, possibly being an indicator of other relevant phenomena related to the capacity fade, such as the internal resistance increase.

Table 2. Fourteen most dispersed features - Mean Median

Feature Description
Ratio of T greater than mean $\pm 10^{o}$C
Max. slope of aggregated linear trends of T in two intervals
Max. p-value of aggregated linear trends of T in ten intervals
Variance of the parameters of a second-order AR model fit to I
Last normalized location of the minimum value of V
Ratio of T greater than mean $\pm 5^{o}$C
Time reversal asymmetry statistic for T
First normalized location of the minimum value of V
Ratio of V outside of 3 standard deviations from mean
Mean p-value of aggregated linear trends of T in ten intervals
Complexity measure cid of I
Skewness of T
97.5% index mass quantile of I
First location of the maximum of I

Table 3. Seventeen most correlated features with the capacity loss - Kendall's Tau

Feature Description
Variance of FFT parameters of I
Ratio of V outside two standard deviations from average
Mean p-value of aggregated linear trends of V in five intervals
Max. offset of aggregated linear trends of I two intervals
Ratio of unique values of T to the signal length
Max. p-value of aggregated linear trends of V in five intervals
Longest strike of V below the mean
Percentage of reoccurring values of I
Time reversal asymmetry statistic for I - 500 lags
Auto-correlation of V - 2000 lags
Mean p-value of aggregated linear trends of I in five intervals
Last location of the minimum of V
Ratio of V outside three standard deviations from average
Last location of the maximum of I
First location of the maximum of I
Ratio of I outside $\pm 1A$
Mean of I

4. MODELLING

Once the features have been obtained, the next step is to model the relationship between them and the capacity fade

$$\Delta\hat{Q} = f(X|\theta, \lambda), \tag{5}$$

where $\Delta\hat{Q}$ denotes the predicted capacity fade, X the regressor matrix, θ the model parameters and λ the hyperparameters associated with the machine learning algorithm utilized e.g. the number of neighbors and the

distance function when using a k-nearest neighbor (KNN). A comparison between different modelling strategies will also be shown here. First there are definitions that need to be made, the prediction of the actual cell capacity is defined for a single cell is

$$\hat{Q}_i = Q_0 - \sum_{j=1}^{i} \Delta\hat{Q}_i, \tag{6}$$

where Q_0 is the initial capacity estimate of the cell and $\Delta\hat{Q}_i$ is the predicted capacity fade at a given sample i. Note that for the predictions it is assumed that only the initial capacity of the cell is known, without any assumptions on intermediate capacity values. Now the $RMSE_Q$ and $NRMSE_Q$ are be defined as

$$RMSE_Q = \frac{1}{N}\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2 \tag{7}$$

and

$$NRMSE_Q = \frac{1}{\bar{Q}} RMSE_Q, \tag{8}$$

with $\bar{Q}$ being the nominal capacity. The metrics $RMSE_{\Delta Q}$ and $NRMSE_{\Delta Q}$ were defined analogously, with the normalization also done with respect to the nominal capacity. While both the $NRMSE_{\Delta Q}$ and $NRMSE_Q$ should be small at the same time, each corresponding to a different type of error. The minimization objective of choice was the $NRMSE_{\Delta Q}$ due to the errors from previous intervals not accumulating over the cell lifetime.

4.1 Hyperparameter Optimization

The data was split into a training and a validation set, so the models were evaluated on previously unseen data. The main goal of the training step is to find the model that has the best generalization capability. There are several common approaches when predicting the model generalization from only the training data, namely holdout, leave-p-out cross-validation and our method of choice, k-fold cross-validation. The k-fold cross-validation implemented here works as follows:

(1) Split the data randomly into k different sets, with approximately the same size
(2) For each fold i, use the k-1 remaining folds to train the model and validate it on fold i, computing the sum of squared errors(SSE) for this fold.
(3) Compute the sum of the SSE and divide by the number of data points to obtain the overall k-fold mean-squared error (MSE)

The k-fold MSE computed this way is then used as the generalization metric when comparing different results from the same algorithm with a different set of hyperparameters. It noteworthy that the comparison of the different k-fold MSE between different algorithms is not straightforward. A better comparison of the different models is done by directly checking the performance on the validation set. There are more sophisticated approaches to select the best algorithm for a given problem, such at the one presented in Rice (1976), however this was not considered necessary for the purpose of this work. The formal definition of the optimization problem under k-fold cross-validation is defined in Hutter et al. (2015) as

$$\min_{\lambda} \; V(\lambda) = \frac{1}{k}\sum_{i=1}^{k} \mathcal{L}(A_\lambda, \mathcal{D}^i_{train}, \mathcal{D}^i_{valid}), \tag{9}$$

where A_λ denotes algorithm A with hyperparameter choice λ, $\mathcal{D}^i_{train}$,$\mathcal{D}^i_{valid}$ the i-th fold training and validation sets and $\mathcal{L}$ is the loss function, which in this work is defined as the SSE between the model output and measurement. Equation (9) defined above is a search problem that was solved using a Bayesian Optimization strategy according to Snoek et al. (2012) and implemented in the Statistics Toolbox in MATLAB.

4.2 Modelling Results

Results from five different regression techniques are analyzed in this section, namely: Multilinear Regression (MLR), LASSO, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Random Forests (RF) and Gradient Boosting (GB). They were applied both on the complete and on the reduced features obtained in Section 3.3, to have a baseline comparison and verify how effective the filter strategy was. The performance of the fitted models is shown in table 4 for the reduced features and in table 5 for the full set. It should be mentioned that for the KNN, there is no training error because it uses the training data itself create the model predictions. As expected, the k-fold $NRMSE_{\Delta Q}$ is a better predictor of the validation performance than the training $NRMSE_{\Delta Q}$, especially when comparing methods with a high degree of freedom, such as the RF or the GB, which are much more prone to overfit the data due to a more flexible structure. For the top three best performing structures in terms of validation $NRMSE_{\Delta Q}$ (GB,SVR,KNN), two (KNN and SVR) were obtained using the reduced feature set. This is an indirect indicator of the effectiveness of the feature selection, but given that there was a hyperparameter optimization step in between the feature selection and parameter fitting it is difficult to pinpoint which step had the bigger impact on the model performance. The distribution of the cell-wise $NRMSE_Q$ for the validation set is shown in the histogram depicted in Fig. 2, giving an idea on how the modelling errors vary for different validation cells, showing an accurate model prediction for the majority of these cells. The $NRMSE_Q$ errors were situated around 0.015, which correspond to a small (1-2%) error in estimating the capacity at the end of the tests. Fig. 3 depicts the normalized capacity fade over time of four different cells over time in the validation set, with each plot representing a cell with a different modelling error magnitude, showing the open-loop characteristic of the model and how the errors accumulate over time.

5. CONCLUSION

An investigation on how features commonly extracted for time-series analysis when applied to current, voltage and, temperature signals are related to capacity fade in a relatively large dataset was presented here. Additionally, a step-by-step methodology creating machine-learning models from the extracted features to predict capacity fade in

Table 4. Model results with reduced features

Method	$NRMSE_{\Delta Q}$			$NRMSE_Q$	
	k-fold	train	valid	train	valid
MLR	0.0147	0.0074	0.0095	0.0217	0.0492
LASSO	0.0085	0.0076	0.0096	0.0225	0.0485
KNN	0.0081	-	0.0082	-	0.0395
SVR	0.0071	0.0045	0.0083	0.0147	0.0432
RF	0.0104	0.0063	0.0095	0.0223	0.0525
GB	0.0079	0.0016	0.0083	0.0041	0.0494

Table 5. Model results with all features

Method	$NRMSE_{\Delta Q}$			$NRMSE_Q$	
	k-fold	train	valid	train	valid
MLR	0.0142	0.0058	0.0085	0.0164	0.0465
LASSO	0.0077	0.0064	0.0085	0.0204	0.0426
KNN	0.0084	-	0.0084	-	0.0376
SVR	0.0073	0.0007	0.0085	0.0019	0.0472
RF	0.0095	0.0059	0.0093	0.0194	0.0509
GB	0.0080	0.0012	0.0076	0.0032	0.0368

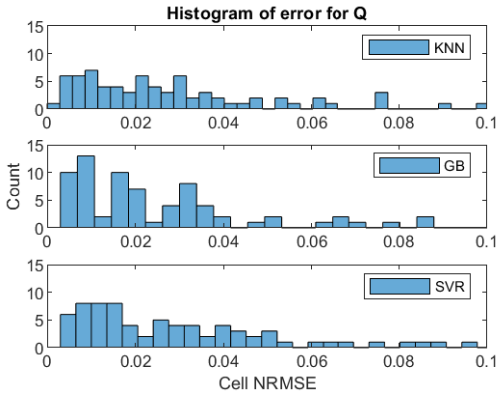

Fig. 2. Model performance breakdown per cell

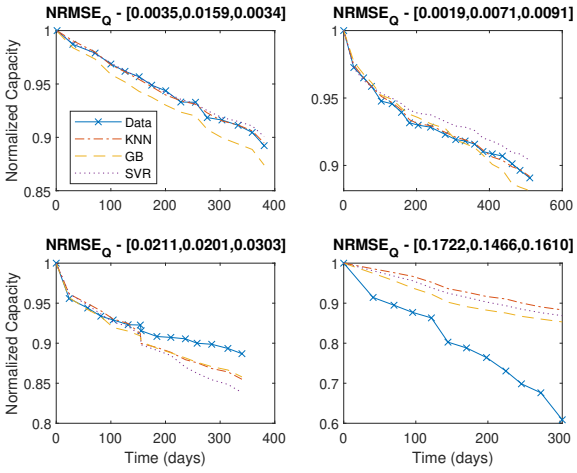

Fig. 3. Capacity fade of 4 different cells with varying error magnitudes

Li-ion cells was also shown, utilizing very minimal prior knowledge of the battery chemistry or underlying phenomena driving the ageing process. Despite the modelling results showing that the extracted features can be used to achieve good results on previously unseen data, probably using some of these features together with chemistry-specific and physically-motivated features would result in a performance increase, especially if a wrapper-based approach is used for the feature selection.

ACKNOWLEDGEMENTS

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs; the National Foundation for Research, Technology and Development; and the Christian Doppler Research Association are gratefully acknowledged.

REFERENCES

Aykol, M., Gopal, C.B., Anapolsky, A., Herring, P.K., van Vlijmen, B., Berliner, M.D., Bazant, M.Z., Braatz, R.D., Chueh, W.C., and Storey, B.D. (2021). Perspective—combining physics and machine learning to predict battery lifetime. *Journal of The Electrochemical Society*, 168, 030525.

Christ, M., Kempa-Liehr, A.W., and Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. *arXiv e-prints*.

de Oliveira, J.G., Dhingra, V., and Hametner, C. (2021). Feature extraction, ageing modelling and information analysis of a large-scale battery ageing experiment. *Energies*, 14(17).

Doyle, M., Fuller, T., and Newman, J. (1992). Modelling the galvanostatic charge and discharge of the lithium/polymer/insertion cell. *Journal of the Electrochemical Society*, 140(6).

Ferreira, A.J. and Figueiredo, M.A. (2012). Efficient feature selection filters for high-dimensional data. *Pattern Recognition Letters*, 33(13), 1794 – 1804.

Hu, X., Xu, L., Lin, X., and Pecht, M. (2020). Battery lifetime prognostics. *Joule*, 4(2), 310–346.

Hutter, F., Lücke, J., and Schmidt-Thieme, L. (2015). Beyond manual tuning of hyperparameters. *KI - Künstliche Intelligenz*, 29(4), 329–337.

Lucu, M., Azkue, M., Camblong, H., and Martinez-Laserna, E. (2020). Data-driven nonparametric li-ion battery ageing model aiming at learning from real operation data: Holistic validation with ev driving profiles. In *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE.

Luzi, M. (2018). Automotive li-ion cell usage data set.

P.Sprent and N.C.Smeeton (2001). *Applied Nonparametric Statistical Methods*. CRC Press, United States, 3 edition.

Reniers, J.M., Mulder, G., and Howey, D.A. (2019). Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. *Journal of The Electrochemical Society*, 166(14), A3189–A3200.

Rice, J.R. (1976). The algorithm selection problem. volume 15 of *Advances in Computers*, 65 – 118. Elsevier.

Saha, B. and Goebel, K. (2007). Battery data set.

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., and Braatz, R.D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4, 383–391.

Snoek, J., Larochelle, H., and Adams, R.P. (2012). Practical bayesian optimization of machine learning algorithms.