Uncertainty-aware and explainable machine learning for early prediction of battery degradation

Laura Hannemose Rieger[a], Eibar Flores[a], Kristian Frellesen Nielsen[a], Poul Norby[a], Elixabete Ayerbe[b], Ole Winther[c,d,e], Tejs Vegge[a] and Arghya Bhowmik[a*]

Enhancing cell lifetime is a vital criterion in battery design and development. Because lifetime evaluation requires prolonged cycling experiments, early prediction of cell ageing can significantly accelerate both the autonomous discovery of better battery chemistries and their development into production. We demonstrate an early prediction model with reliable uncertainty estimates, which utilizes an arbitrary number of initial cycles to predict the whole battery degradation trajectory. Our autoregressive model outperforms previous approaches when predicting the cell's end of life (EOL). Beyond being a black-box, we demonstrate through explainaibility analysis that our deep model learns the interplay between multiple cell degradation mechanisms. In this way, the learned patterns align with existing chemical insights into the rationale for early EOL despite not being trained for this or having received prior chemical knowledge.

1 Introduction

The scale of deployment of lithium-ion batteries is expected to grow dramatically over the next decade as the transportation sector gets electrified and grid level battery storage becomes more commonplace to balance the fluctuating renewable energy sources. Designing batteries with higher cycle life directly leads to better economics[1] and lower carbon/ecological footprint from mining and manufacturing of batteries[2]. The capacity of a secondary (rechargeable) battery cell degrades over its lifetime and for many practical purposes such as in the automotive sector, the failure threshold is considered to be when the battery's capacity falls below 80% of the nominal capacity, which is called cycle life. After this point, the capacity degradation will tend to progress at a much faster rate and therefore the cell needs to be replaced[3]. The cycle life of a cell depends on the usage scenarios, internal chemical and structural details[4], which can vary significantly even between cells manufactured from the same production line. However, the process of optimizing batteries (materials and cell design parameters) for improved cycle life is hindered by a slow lifetime evaluation process. Being able to accurately determine the battery lifetime with limited early cycle data would enable us to fast forward the battery development cycle.

The cycle life of a battery is strongly dependent on how it is operated, e.g., the charging C-rate, temperature or cut-off voltage as these conditions determine the insidence of deleterious electrochemical side reactions in anode, electrolyte, and cathode. Balancing between longer cycle life and faster charging time is a major engineering challenge in making electric vehicles (EV) a competitive alternative to conventional cars. Designing charging schedules to achieve this goal can be extremely time consuming, as testing typically extends months before observing any effect that design has on cell cycle life. Early prediction of cycle life would significantly expedite such testing and thus enable designing smarter charging schedules that extend battery life[5]. Early prediction of degradation would also support the design of advanced battery management systems (BMS). Since battery performance at the pack level deteriorates when cells operate at heterogeneous states of health[6], an early prediction model would grant BMSs control over individual cells based on their specific degradation trajectory, and thus warrant significant lifetime improvements at the battery-pack level[7].

Accurately predicting the battery lifetime is challenging because each cell undergoes complex electrochemical processes during operating conditions and nonlinear degradation associated with cycling[8]. Physics-based modelling of battery degradation that captures a plethora of multi-time/length scale electrochemical and mechanical processes would be prohibitively expensive. Instead, parametric models (e.g. P2D (Pseudo-Two-Dimensional) or single particle) approximate cell degradation using simpler governing equations, with a limited ability to capture complex interactions between degradation mechanisms. Data driven models recently have been able to overcome cost-accuracy trade offs in this task by learning high dimensional correlations among system-level observables that might implicitly represent internal electrochemical processes. Recently, hybrid physics and machine learning models, and physics and uncertainty-aware machine learning models have been envisioned as the future direction of research[9,10].

Most data-driven models of cell degradation use online data[11] (battery state of health until the point of prediction) to predict near future behaviour[12–14], but do not model early prediction

[a] Department of Energy Conversion and Storage, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.
[b] CIDETEC, Basque Research and Technology Alliance (BRTA), Pº Miramón 196, 20014 Donostia-San Sebastian, Spain.
[c] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark.
[d] Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark.
[e] Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark.
∗ Correspondence:arbh@dtu.dk

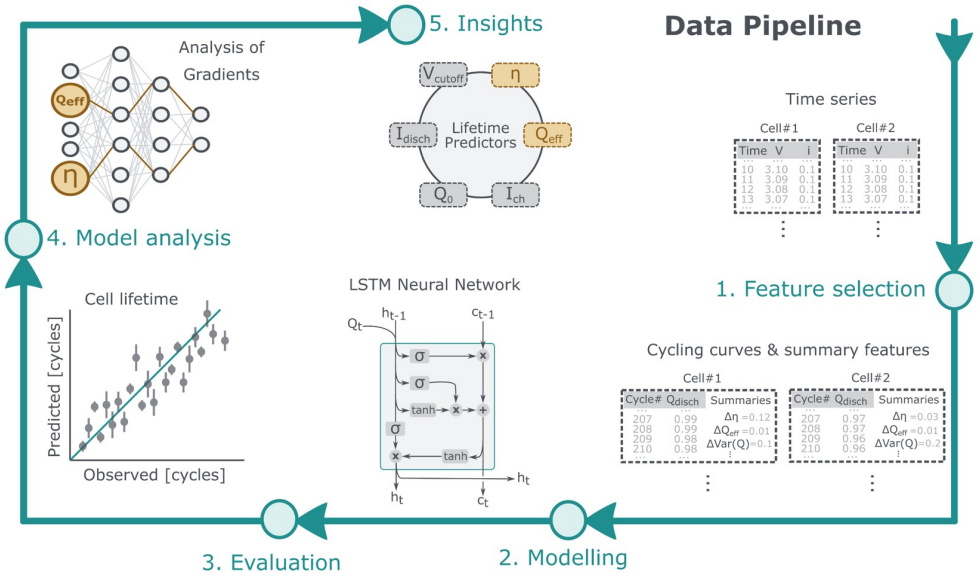

Fig. 1 Schematic overview of the prediction process

(lifetime behaviour from early cycles). Additionally, most published work does not consider the variabilities encountered between similar types of cells[12,15] (even of the same chemistry and design). Pioneering work Severson *et al.*[8] followed by two other articles[16,17] have showcased different approaches towards data driven early degradation prediction considering intra-chemistry variance.

Using hand-engineered features that incorporate domain knowledge, Severson *et al.*[8] trained a linear regression model with elastic net regularization to predict the total lifetime of the battery cell. The model does not consider uncertainty over the output, making it hard to detect when the battery is out of the training distribution or the prediction is otherwise unreliable. With a dilated CNN (Convolutional neural network) model, Hong *et al.*[17] predicted the remaining useful lifetime based on in-cycle time series data from early cycles. Although this model provides an uncertainty estimate for the EOL it does not predict the entire degradation trajectory, rendering the model uninterpretable in regards to the degradation mechanism. Li *et al.*[16] with a sequence-to-sequence LSTM (Long short-term memory neural network) recently predicted the full degradation trajectory. Gaussian Process Regression (GPR) have been used to detect failure of deployed batteries as well[14]. Jiang *et al.*[18] have employed an uncertainty-aware Hierarchical Bayesian Model (HBM) to predict the quality of a battery cycling protocol independent of the intra-chemistry variance between batteries cycled with the same protocol and only focused on lifetime. In another very recent work, Paulson *et al.*[19] used an extensive feature selection process, considering 396 features, to predict battery lifetime for a multi-chemistry dataset.

However, currently, no existing model can predict full degradation trajectories with uncertainty estimates both in and out of distribution (both near and far from training data distribution) ideally capturing both epistemic and aleatoric uncertainties. Having uncertainty estimates allows us to e.g. recognize when a prediction is unreliable because the data point lies too far from the training distribution such as would happen with a different battery chemistry and perform on the fly control of how much cycling is done. Prediction of full trajectory helps understanding causation of degradation. To learn a universal function of how cells degrade across cell chemistries requires an expressive and complex model[20].

Uncertainty-aware, predictive, explainable and flexible models to predict battery degradation also represent essential building blocks to accelerate the development of new sustainable high-performance batteries.[20–22]. More specifically, the ability to predict the EOL and degradation trajectory of any cell from its initial cycles, would shrink the costs/time of experimentally testing it for hundreds of cycles; such a model would effectively accelerate the search of promising battery chemistries. Battery research acceleration would greatly benefit from models capable of making accurate and certain predictions on cells similar to those used for training, already from few initial cycles. On the other hand, such models must be aware when they are wrong, i.e. express high uncertainty when, for instance, they attempt to make predictions on cells with significant different chemistry (e.g. new electrolyte formulation). If in addition these models are explainable, they would enable scientists to gain insights into degradation mechanism and thus have trustworthy extrapolatable models. However, achieving model expressiveness that allows learning of different degradation mechanism purely from easy to access data across cell formats and cell chemistry requires complex deep learning models, which thwart understanding of how the model learns.

Although recently published models cover one or the other desirable aspects, none of them can serve the purpose of a autonomous battery development acceleration platform. So we focused on developing a model that can be trained with different chemistries/formats, can handle varying cycling parameters, dis-

play uncertainty over the prediction and can on-the-fly adjust the trajectory rollout while being explainable. Towards this challenge, our contributions can be summarized as such:

- We introduce a deep recurrent neural network architecture (LSTM) that can be trained to predict future capacity and EOL (End of Life) without requiring insight into the battery chemistry. It can be trained with datasets of different battery chemistry together.
- The trained model outputs full degradation trajectory as not only the mean but also the variance, allowing for uncertainty-aware prediction. Ensemble models are built in addition to uncertainty as direct model output.
- We evaluate this architecture on available battery datasets, showing that it outperforms multiple competitive baselines. We also show how robust uncertainty capability lets the model adjust input cycle information on the fly to lower uncertainty.
- We consider the explainability of the trained model on LFP battery data. Our findings show that the model has learnt to capture underlying physical phenomena without being trained specifically.

To our knowledge, our model is the first that is able to model the entire capacity fade trajectory from the early cycles without a fixed limit on the maximum lifetime. Using an LSTM allows us to visualize the influence of different inputs over the lifetime of the battery and draw chemical insights from the data-driven model.

Preliminary results indicate our architecture to be chemistry neutral, based on few openly available but limited cycling datasets from commercial cells using nickel cobalt aluminum oxide (NCA) and nickel manganese cobalt oxide (NMC) as positive electrodes. While the architecture itself is chemistry neutral, i.e. not restricted to use on a specific chemistry, it requires a dataset of batteries with the specific chemistry that we want to predict capacity for. Models that can simultaneously predict capacity for f.e. LFP, NCA and NMC batteries at the same time, would require a dataset that allows to generalize over chemistries by including batteries with different chemistries that were cycled under comparable conditions. We leave this work to future research.

2 Methods

2.1 Data resources

For this work, we use a previously available battery cycling dataset[8]. The reader is referred to the original publication for more detailed information. The dataset originally consists of 135 commercial LFP/graphite cells, each with a nominal capacity of 1.1 Ah and cycled in a temperature-controlled chamber at 30 °C. Each cell was operated at one of 72 different fast-charging protocols. Discharging was identical for all cells at 4 C to 2.0 V. Varying the charging conditions resulted in a wide range of cycle lives, ranging from 150 to 2,300 cycles. The dataset consists of three batches, referring to a selection of cells that were simultaneously cycled inside the chamber under different testing conditions and dates. We removed cells with experimental errors, as suggested by the authors in their published code, resulting in 124 usable cells for training and testing[8].

In contrast to the original work from Severson et. al., we analyze the performance on calendar-aged and non-calendar-aged data separately. To do this, we split up the first two (non-aged) batches of the dataset into 50% training data and 25% respectively for validation and test, resulting in the same number of training points as in the original paper. The performance on the calendar-aged batteries of the third batch is analyzed separately.

For training on non-LFP cell chemistries we used a dataset of 40 cells (22 NMC, 18 NCA) published by the Sandia National Labs and show results in the supplements[23].

Inconsistency in data generation and insufficient documentation presented a large issue when choosing datasets for this work. For example, testing cycles were often not annotated and many datasets contained several discontinuities such as outliers and unexplained steps in capacity (examples are seen in Figure S2). The general quality of the data presented an issue when finding datasets for this work. For one, measurement inconsistencies meant that not all cells reported the voltage profiles necessary to calculate covariates. Often, the degradation trajectories showed outliers or unexplained jumps in capacity (examples are seen in Figure S2) which, without appropriate annotation, cannot be automatically removed or imputed. As in any other field of research, the development and testing of new machine learning models for cell degradation is limited by the amont and quality of available data.

2.2 Data processing

For each cell we use an input trajectory, i.e. the degradation trajectory up to an arbitrarily-chosen number of initial cycles, as the basis for predicting the full degradation trajectory. For instance, an input trajectory of 20 cycles $Q[0-20]$ uses the degradation trajectory of the first 20th cycles to predict the remaining trajectory until the EOL. We report results for a range of input trajectories from 20 to 100 in order to explore how many initial cycles are needed for accurate EOL predictions.

For the network training, we used the degradation trajectory, i.e. the trajectory of discharge capacity vs. cycle, the charging schedule and a set of three covariates described in Section 2.5. We preprocessed the data by removing obvious outliers and replacing them with the mean value over the dataset. Subsequently, we standardized the covariates to have a mean of zero and a variance of one. We also use the logarithm of the current cycle number as a supplementary covariate. This does not contain any information about the internal state of the battery in itself. However, in practice, we found that this improves the training process for the EOL prediction because it facilitates an easier comparison of how fast capacity degrades.

To calculate the variance between capacity-voltage curves, we follow Severson *et al.*[8] and fit the discharge capacity as a function of voltage which is evaluated at 1000 linearly spaced points between 2 and 3.5 Volt. We calculate the variance between the resulting vectors for the tenth cycle and the last input cycle. The charging schedule of each cell is expressed as a three-dimensional

vector containing the minimum, maximum and average charging rate throughout the cycling for the LFP dataset.

The discharge capacity at every cycle Q_n (the subscript denotes the cycle number), is predicted as the remaining proportion of the capacity in the previous cycle. The degradation trajectories often contain a lot of noise. For the target (the next cycle capacity) during training we therefore preprocess the capacity trajectories for training with a simple moving average filter (MAF) over twenty cycles centered around the current cycle.

2.3 LSTM architecture

We use an LSTM to process the capacity trajectories[24]. In extension of a traditional neuron, an LSTM neuron contains a memory state that is updated in each time step. Time sequences such as the capacity trajectory are fed into the LSTM neural network concurrently. For multi-step prediction, the predicted output is appended onto the input to create the next input. The NN proposed consists of one LSTM layer, one fully connected hidden layer, and one fully connected output layer to predict the loss in capacity in the current cycle based on previous capacities as shown in Figure S1. To choose the optimal number of neurons, cross-validation was performed, resulting in 32 neurons for each layer.

To capture uncertainty, instead of predicting a single value our NN outputs both the predicted mean capacity at the next step y_{pred} and its expected variance σ^2; the variance allows us to express uncertainty in the prediction. To train the NN, we use the negative log-likelihood (NLL) as a loss function over the Gaussian distribution output by the NN and the true next step value. With the NN outputting the mean y_{pred} and variance σ^2 this is

$$NLL = \frac{1}{2}\left(\frac{(y_{pred}-\hat{y})^2}{\sigma^2} + log(\sigma^2)\right)$$

This corresponds to maximizing the probability that the true next step value comes from the probability distribution the NN predicted.

When predicting the trajectory for a new battery, we obtain the next time step value by sampling from a Gaussian distribution with the predicted mean and variance. Since we are interested in obtaining uncertainty over the entire trajectory, we sample multiple independent trajectories from each NN in the ensemble during test time. For each trajectory, we concurrently obtain the next time step in the manner just described until the predicted trajectory reaches its EOL. The mean and variance of time step t for K neural networks in the ensemble with L trajectories sampled from each become

$$\mu_t = \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L} y_{t,l,k}$$

$$\sigma_t^2 = \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L} (y_{t,l,k}-\mu_t)^2$$

The NN was trained with the Adam optimizer with the default learning rate of 0.001[25]. The training is stopped once validation loss no longer improves for three concurrent epochs. At testing time, it is required to roll-out multiple capacity trajectories to obtain an accurate measure of the uncertainty over the output. We use an ensemble of five neural networks (trained with different random seeds) and sample ten trajectories from each network. During roll-out, we concurrently sample the next value in the capacity trajectory from the output mean and variance predicted by the neural network in the current step. The trajectories of all neural networks in the ensemble are concatenated and the distribution of trajectories is calculated.

Unless otherwise noted, all performance metrics are averaged over five random seeds. All experiments were done with Pytorch on an Nvidia RTX 3090[26]. The code used to process the datasets, train the models and create the results presented in this study will be released on acceptance.

2.4 Saliency analysis

To analyze how important input parameters change during the trajectory prediction, we apply a saliency analysis to the LSTM on the test data. Neural networks are trained with gradient descent, i.e. computing the gradient of the loss function over the weights and taking a step in the negative direction. We use a similar approach and take the absolute gradient of the output over the input. Intuitively, this highlights input dimensions where a slight change of the input will result in a large change of the output.

To be able to compare between different batteries and cycles, we normalize these values such that the gradients for one cycle always sum up to one. By taking the average importance of the inputs over different subsets of batteries and cycles, we can extract information about the general importance of e.g. the coulombic efficiency for prediction. We discuss the results in Section 3.5.

2.5 Feature selection

The goal of the ML algorithm is to predict the lifetime of LiFePO/graphite cells from a given number of initial cycles. Accordingly, we train the algorithm with cycling trajectories (i.e. discharge capacity vs. cycle) and a set of additional electrochemical features from the initial cycles. These features are selected based on being both i) informative, i.e. known to be correlated with the cell's lifetime, ii) accessible, i.e. available from most common cycling experiments and iii) generalizable between experiments. For instance, while the cycling temperature affects capacity fade, we disregard it as a feature because the impact of temperature is highly dependent on e.g. the environment's temperature and the form factor of the cell; hence, incorporating the temperature as feature in the model restricts its ability to generalize to other cell designs.

The cycling charge rate affects the degradation of LiFePO4/graphite cells[27–29] and is always recorded in cycling experiments; therefore, we include the maximum, minimum and mean charging rate as features to account for the cycling conditions. The discharge rates might also be considered, but we do not use them because all cells in the dataset are discharged at the same rate[8]. In addition to the charging rates, we select three electrochemical metrics as features. The Coulombic efficiency, the charge-discharge voltage gap and differential capacity trajectories reflect the loss of active Li+, the build-up of internal

Table 1 RMSE prediction error in number of cycles on the prediction of EOL for baselines and LSTM based on the first hundred cycles

	Non-aged batteries	Calendar aged batteries
Linear Regression *	151	202
LSTM (Ensemble)	110	184
LSTM	172	243
DNN (no capacity)	207	402
LSTM (no covariates)	587	384

resistance and the electrochemical reaction mechanisms of a cell during a cycle, respectively[30–32]. Given the Coulombic efficiency is a scalar value, it can be directly used as a feature. On the other hand, the voltage gaps and the differential charge curves are voltage (or state-of-charge)-dependent vectors that need to be encoded as single scalar features per cycle. As a simple approximation, we describe the voltage gap as the difference between the mean voltage during charge and the mean voltage during discharge. Finally, we inherit the differential charge curve feature engineered from Severson et al., who demonstrated that the variance of the difference between charge and discharge capacity vectors correlated well to the cell's lifetime.[8]

Using 6 features for every cycle would result in hundreds of inputs to characterize the degradation behaviour of a single cell, which is impractical. Training on more features than examples would render the model not only larger, but also ill-posed to generalize. Fortunately, these features vary very little from cycle to cycle for a single cell, so it is sufficient to summarize them by their variations within the initial cycles. For consistency with Severson et al[8] , we use the feature difference between the 10th cycle and the last cycle available for prediction, represented as Xn-10. In the following, we refer to these battery cell-specific, time-independent features as covariates, which are described in Table S1 and illustrated in Figure S6.

In addition to being readily accessible from electrochemical time series, we believe these features implicitly capture i) the influence of uncontrolled experimental conditions (e.g. cell manufacturing, geometry, preconditioning) and ii) the cell's state of health during cycling independently from the chemistry of the electrodes . Chemistry-neutrality ensures that the features can be readily used to train on datasets from other cell chemistries and pave the way for high accuracy chemistry-neutral models to predict degradation with additional data.

3 Results

3.1 LSTM performance compared to baseline models

LSTMs are well-suited for modelling sequential data as they do not have constraints on the total time series length and can model complex correlations and features in sequential data streams. To demonstrate the advantage of LSTM operating with electrochemically inspired features, we compare the LSTM with three baselines, a simple Linear Regression (LR) with Elastic Net regularization based only on the covariates (inspired from and similar to[8]), a Dense Neural Network (DNN) based only on the covariates and an LSTM based on only the capacity trajectory. In contrast to the neural networks, the LR model can only estimate the End of Life (EOL) but not model the complete capacity trajectory.

To compare the modeling approaches with the linear regression, we report the RMSE (Root Mean Square Error) on the predicted total lifetime in Table 1. The RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=0}^{N}(y_n - y_{n,pred})^2}$$

The results show that an ensemble LSTM (comprised of five NN initialized with different random weights) has a lower Root Mean Square Error (RMSE) when predicting the EOL than the neural network baselines and the LR, indicating that it better captures the relationship between cycling patterns within the first few cycles and the total lifetime. In contrast to previous work[8], we separately evaluate the prediction performance on calendar-aged and non-calendar-aged batteries as we found significant differences in the behavior of the battery cells depending on their storage history (see Figure S4) .

Calendar-aged cells seem to last longer compared to non-aged cells under the same cycling conditions. Prolonged storage of cells might influence, for instance, electrode passivation in a way that results in improved lifetime compared to non-aged cells, similar to the outcomes of performing formation cycles[33]. Consequently, the model struggles to predict the degradation of aged cells which it has not seen during training.

Notably, the improvement in performance also holds for calendar-aged batteries. In Table 1 we see that the ensemble LSTM is more accurate in predicting EOL for calendar-aged batteries as indicated by the lower RMSE, implying that some of the chemical processes happening during the calendar ageing are implicitly captured in the early cycles as well and are learnt by the LSTM.

To show that in addition to accurate EOL prediction our model also matches the capacity trajectories, we show capacity trajectories on the test set of non-calendar-aged batteries in Fig. 3. We provide capacity trajectory predictions for calendar-aged batteries in Figure S6. In contrast to only predicting EOL, modeling the entire capacity trajectory allows detailed analysis of the degradation pattern and helps us gain understanding on possible electrochemical phenomena causing it.The LSTM ensemble matches the actual trajectory closely, accurately predicting the knee point, i.e. the cycle number where the trajectory visibly bends to an accelerated degradation.

3.2 Capturing uncertainty

Highly parameterized deep learning models like ours are prone to fail when generalizing to datasets that are very dissimilar to the dataset that the model is trained with, such as new battery chemistries or ageing processes that dramatically alter the degradation pattern. The trustworthiness of the model can be questioned if the model is overconfident in predictions i.e. it does not know when it is wrong. To provide reliability to our approach, we model the uncertainty of the output trajectory (the shaded area in Fig. 3 encompasses the 5th to the 95th percentile). Information about the certainty of the prediction is important for risk assessment during model deployment and can be used for active

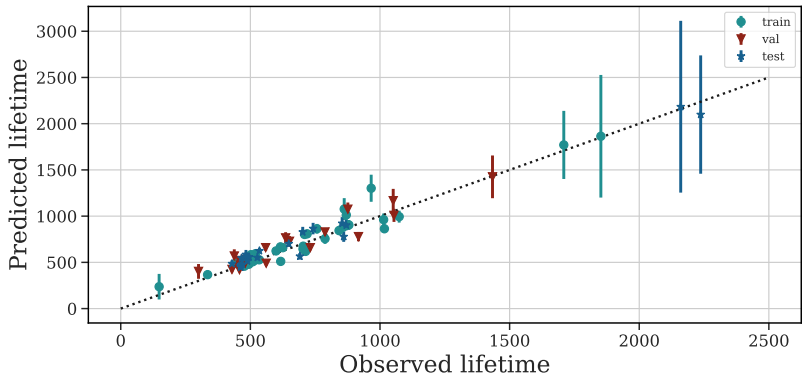

Fig. 2 LSTM ensemble prediction vs ground truth based on 100 cycles. Error bars indicate SD.

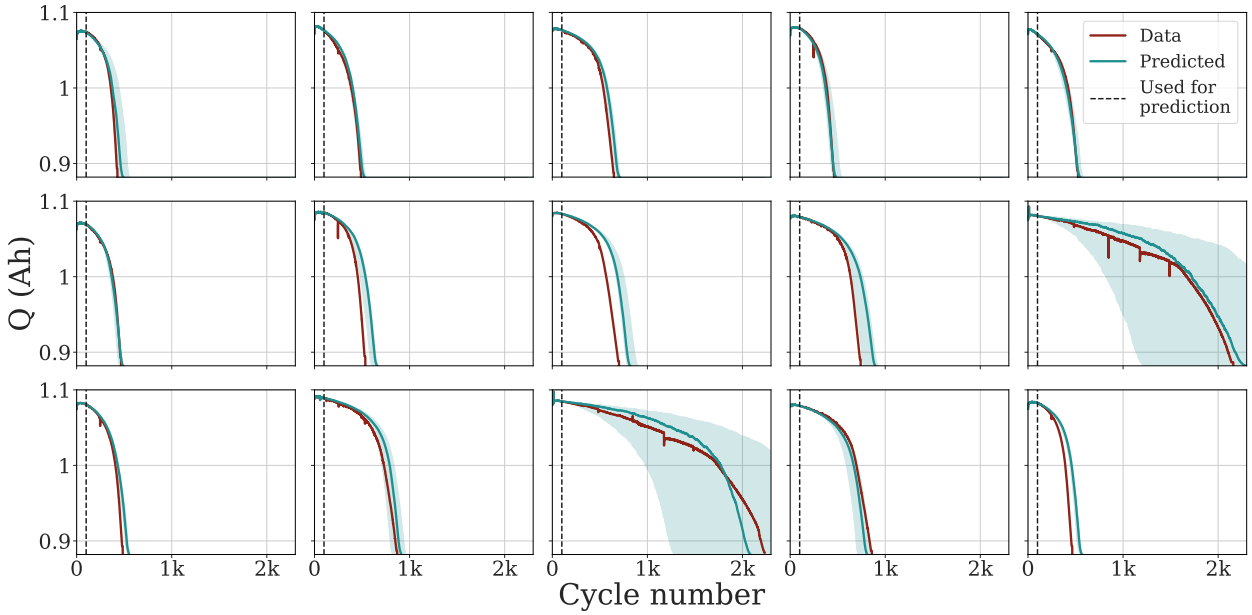

Fig. 3 Predicted trajectories of unseen batteries. Prediction is the 50th percentile, the shaded area is the uncertainty estimate (from the 5th to the 95th percentile). For the two batteries with a higher lifetime than any battery seen in the training set, the uncertainty is high.

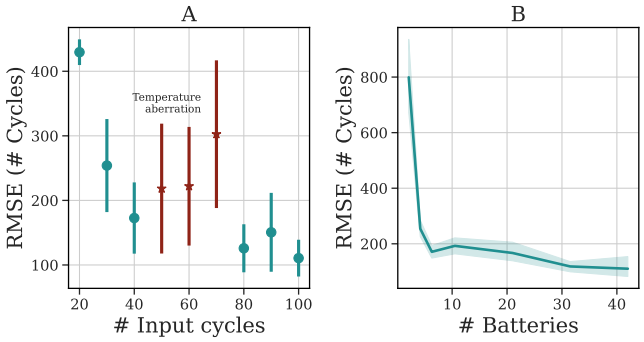

Fig. 4 A Accuracy of LSTM ensemble plotted against the number of cycles used as input. The error bars indicate SD. According to[8] there was a temperature aberration in the test chamber around 55 and 70 cycles, causing a larger error for 50,60, and 70 cycles. B Accuracy of LSTM plotted against the number of batteries used for training. The shading indicates SD. For accurate learning, only a small number of batteries is needed.

learning based data collection from experiments.

For most batteries, uncertainty over the trajectory is low. Notably, two batteries in the test set have a lifetime beyond any lifetime seen in the training set. The predictions for these batteries are associated with a high uncertainty over the capacity trajectory and EOL. In support, Fig. 2 shows that uncertainty over the EOL is low for batteries with a lifetime of up to 1000 cycles (as expected given that the majority of battery cells in the training data set lie in this range) and high above that. This is expected and desired behavior, as the ML model has less information about the data distribution in this range.

In Figure S4 we show example trajectories for the calendar-aged batteries. The model clearly predicts comparatively higher uncertainty over the trajectory for aged cells. Since the predictions become more uncertain on the trajectories of aged cells, we conclude our model has learned data patterns -possibly electrochemical signatures- differentiating aged from non-aged cells.

3.3 How early can we predict with how little data?

Getting accurate information about a battery's future degradation pattern early into its lifetime is vital. Each potential application of our model might require a different trade-off balance between the accuracy of prediction and how much cycling data the model would need as input. We characterize such trade-off by evaluating how the accuracy of our algorithm changes with the number of cycles input into the neural network in Fig. 4.

The accuracy of prediction (in terms of the RMSE on the test data) improves as more of the initial cycling data is used as input to the model as shown in Fig. 4. With a larger part of the degradation trajectory visible, forecasting for future degradation becomes easier as more information about the degradation process becomes available. Additionally, the results in Fig. 4 show that the error rapidly decreases with as few as thirty cycles available, demonstrating that the LSTM can robustly predict the EOL early into the lifetime; more specifically, the model predicts the EOL within 200 cycles of accuracy, using only the first 40 cycles of the trajectory.

Testing conditions like operating temperature directly affect the internal electrochemistry of the battery. Fluctuations in these variables, if not taken into account in the modelling, can cause larger errors. Even if our model is not trained using temperature as predictor, it still captures some of these effects implicitly through other observables and provide a high level of accuracy when it is trained on data with and without those aberrations (e.g. predicting degradation from 100th cycle onwards while data from 55th to 70th cycle was noisy).

Another important factor in training and using ML models is how much data is required to obtain a robust model. Battery cycling data covering full capacity degradation is expensive to acquire and the maximum dataset size may be limited by other factors as well. We examined the model accuracy dependent on the number of batteries in the training dataset in Fig. 4. Again, we observe that the error decreases with increasing size of the dataset but rapidly levels off, implying that the model can generalize about the degradation process already from a dataset with as few as six batteries. A recent paper from Dechtent et al. came to a similar conclusion, showing that a simple linear model that captures cell-to-cell variability can be fit with as few as nine batteries[34].

The data efficiency of our model opens up the possibility of rapid prototyping of models for completely new battery chemistries. This enables much expedited lifetime optimization of the new class of batteries without performing full life cycle tests.

3.4 Projecting forward

Our model can integrate longer input trajectories without needing to retrain. This allows us to flexibly decide how many cycles the cell should go through and predict the remaining capacity trajectory with our model. If a higher accuracy is desired, the battery can be cycled for longer, resulting in more information available for prediction and consequently higher accuracy. Importantly, this is a different scenario than the one presented in Fig. 4. In Fig. 4 we train multiple models while varying the initial number of cycles that the model is trained with. In contrast, in Fig. 5 we use a fixed model and simply append cycles to show how the accuracy and projected uncertainty of the prediction changes, allowing a flexible trade-off between cycling time on one side and accuracy as well as uncertainty on the other.

In Fig. 5 A we show the percentage of batteries for which the prediction for EOL is less than fifty cycles off from the true EOL dependent on the number of cycles used as input for the model. Expectedly, the proportion of accurate predictions goes up as more cycling information becomes available.

We can use the uncertainty over the EOL as a proxy criterium (see Figure S7 in the supplements where we show that accuracy and uncertainty are strongly inversely correlated, making the uncertainty over the output a valid criterium to decide whether a battery needs to be cycled longer). In Figure S3 we additionally show results for the NMC dataset.

In Fig. 5 B we show how the predicted uncertainty changes for

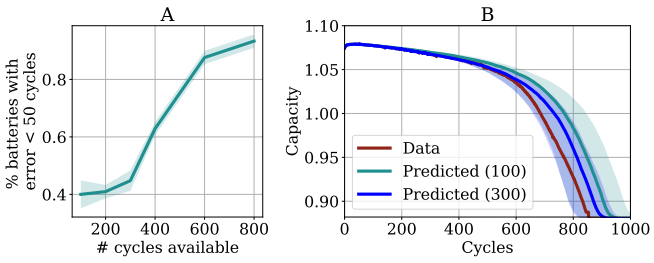

Fig. 5 More cycle information decreases uncertainty and error. A: Percentage of batteries with an error of less than 50 cycles for EOL when more input cycles are available. B: Uncertainty is reduced when more input cycles are available.

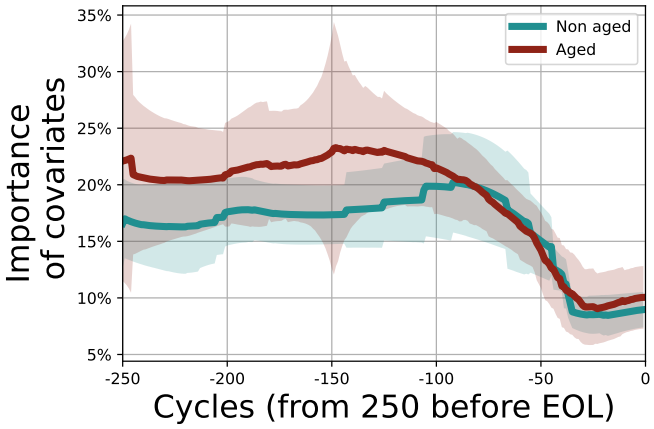

Fig. 6 Importance of in-cycle information for calendar-aged and non calendar aged batteries. The shaded area indicates one standard deviation in each direction.

one exemplary battery. We see that the uncertainty is reduced and the prediction becomes more accurate when more cycle information is input.

3.5 Inspecting what the LSTM has learned

In addition to accurate prediction, we are interested in analyzing what features were important for prediction and whether they reflect prior knowledge about the chemical processes inside the battery cell. We analyze the relative importance of the inputs in the predictions, differences between short- and long-lived batteries as well as between calendar-aged and non-aged batteries.

For this analysis, we compute the gradient of the output in regards to the input. The gradient indicates how fast the output changes with a change in the respective input variable, representing an intuitive measure of feature importance. To make the resulting importance measures comparable between cycles and batteries, we normalize them such that the total importance of all input features for one battery cell sums up to one. We show results of this analysis averaged over the test set Fig. 7. The data in Figure 2 shows that the capacity trajectory of LFP batteries consists of a relatively flat initial regime followed by a phase in which the cell's capacity declines rapidly until the EOL. The point at which the cell enters the rapid decline phase is often visible as a knee point in the trajectory. Since this point is a determining factor of a battery's life, it is of particular interest what triggers this transition. We visualize the gradients over the last 300 cycles (as predicted by an exemplary LSTM for each battery cell).

In Fig. 7 we show the importance of the input features averaged over the test set along with the capacity trajectories. We observe that the importance of the previous capacity stays relatively constant over the lifetime of the battery until approximately 200 cycles before the EOL. At this point, there is a dip in importance for the capacity trajectory and the covariates gain more relevance. Our interpretation is that the cycling conditions such as the charge and discharge rate are more relevant in predicting the 'point of descent' and thus the EOL, while the capacity in previous cycles is more relevant for predicting the shape of the capacity trajectory.

In the phase of quick degradation, we note that the importance of previous capacities increases again, becoming the almost sole determining factor for the output. We hypothesize that once degradation enters an accelerated phase, the only relevant factor for prediction seems the current rate of degradation, which is encoded in the previous values of capacity.

In the lowest row of Fig. 7 we additionally visualize the mean importance of coulombic efficiency over the last 250 cycles, as calculated for long- (red) and short-lasting (green) batteries; since the average lifetime of the dataset is 691 cycles, we use 700 cycles as the cut-off value. There is a qualitative difference in the importance of coulombic efficiency for long- and short lasting batteries, both in the absolute value but also in the proportional increase of importance as a battery approaches its EOL. For short-lived batteries, the importance increases slightly but steadily as the batteries approach EOL. Low coulombic efficiency is possibly due to high SEI forming exchange current density that leads to loss of active lithium[34]. Expedited loss of cyclable lithium leads to rapid capacity loss. Thus, low coloumbic efficiency can be an indicator for a shortened cell life. The importance of coulombic efficiency increasing proportionally more for short-lived than long-lived batteries is therefore in line with our understanding of the internal state of the battery.

In Fig. 6 we show the importance of the in-cycle info (Overpotential, Coulombic Efficiency, and variance in the difference between charge and discharge) for the last 250 cycles for calendar aged and non-calendar-aged batteries. Since calendar aged batteries have a much higher mean lifetime than non-aged batteries, we consider only non-aged batteries with a lifetime longer than 600 cycles. In this way, both sets of batteries have approximately the same average lifetime. We observe that there is a quantitative difference in the gradients, i.e. the importance of in-cycle information differs for calendar aged and non-aged batteries from the 250th to last to the 100th to the last cycle. The lifetime of calendar-aged cells is more affected by the cycling conditions. The quantitative difference indicates that the LSTM discriminates already from the initial cycles that calendar-aged and non-aged cells belong to different data distributions. Such difference is also manifested in higher uncertainty for the calendar-aged cells.

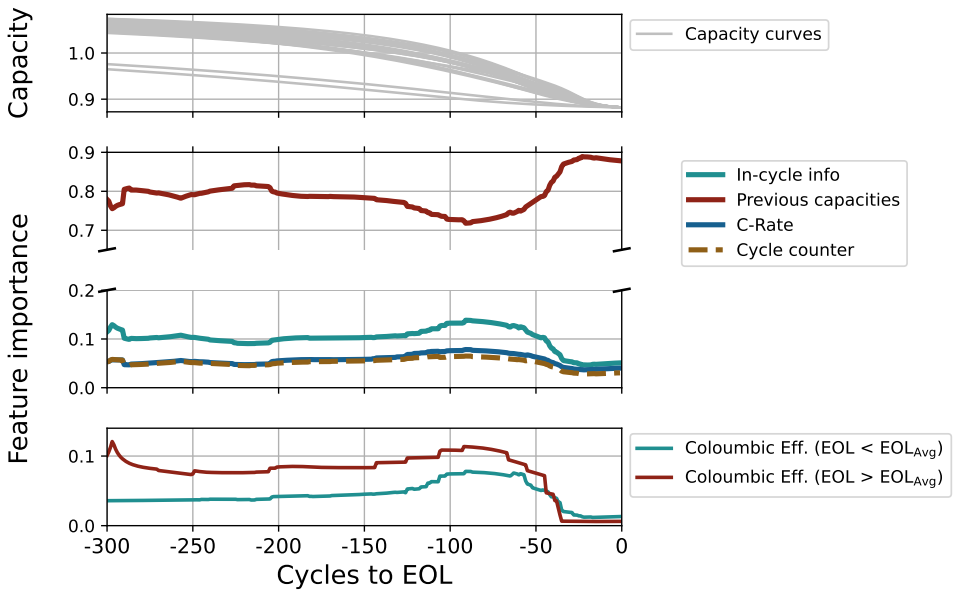

Fig. 7 Sensitivity to input features over the last 300 cycles. The importance of Coulombic efficiency increases around the knee point. Upper row: Capacity trajectories (for orientation) Middle row: Absolute gradients of all input-features Lower row: Absolute gradients of covariates

4 Conclusions

We have proposed and demonstrated a data-efficient autoregressive model for early prediction of battery degradation that not only improves upon the state of the art in accuracy but also supports uncertainty awareness (both epistemic and aleatoric), explainability and chemistry agnostic modelling while predicting the whole degradation trajectory. It relies on features that are easy to obtain from simple charge discharge curves during the early cycling of any battery chemistry. With explainability analysis we show evidence that the model learns important chemical descriptors and their evolution during the degradation process in a completely data-driven fashion. As we model uncertainty directly and the model is chemistry agnostic, such a model can be reliably trained and deployed for existing and new classes of cell chemistries in the future. We show the potential with a preliminary model trained with a small dataset of openly available but limited NCA and NMC cell cycling data.

Early prediction with our uncertainty aware model will create the basis for an accelerated autonomous battery design platform by shortening the time consuming life cycle assessment tests. As our model predicts the whole trajectory with uncertainty and is flexible towards the length of initial input cycles, it is well suited towards such use cases where the actual length of cycling tests varies. Access to model uncertainty opens up the possibility of building accurate models for new cell chemistries with limited amounts of data collection via active learning.

Author Contributions

A.B. conceptualized and supervised the project; L.R. with input from A.B. and O.W. designed and developed the model and method; E.F. collected, processed and featurized the data; E.F. and L.R did the workflow and result visualizations; L.R. carried out the software engineering, including implementation of model and training framework, implementation of data workflows for the datasets, and analysis of data. All authors participated in the discussion and formal analysis of results; L.R., E.F., A.B. prepared the first draft and all authors contributed to review and editing; T.V. and A.B. did funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge the European Union's Horizon 2020 research and innovation program under grant agreement No 957189 (BIG-MAP) and No 957213 (BATTERY2030PLUS).

Notes and references

1 O. Schmidt, S. Melchior, A. Hawkes and I. Staffell, *Joule*, 2019, **3**, 81–100.

2 T. Le Varlet, O. Schmidt, A. Gambhir, S. Few and I. Staffell, *Journal of Energy Storage*, 2020, **28**, 101230.

3 J. Zhang and J. Lee, **196**, 6007–6014.

4 J. S. Edge, S. O'Kane, R. Prosser, N. D. Kirkaldy, A. N. Patel, A. Hales, A. Ghosh, W. Ai, J. Chen, J. Yang *et al.*, *Physical Chemistry Chemical Physics*, 2021, **23**, 8200–8221.

5 P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang and others, 2020, **578**, 397–402.

6 X. Liu, W. Ai, M. N. Marlow, Y. Patel and B. Wu, *Applied Energy*, 2019, **248**, 489–499.

7 J. T. Warner, *The handbook of lithium-ion battery pack design: chemistry, components, types and terminology*, Elsevier, 2015, pp. 91–101.

8 K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang,

Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis and others, 2019, **4**, 383–391.

9 T. Vegge, J.-M. Tarascon and K. Edström, *Advanced Energy Materials*, 2021, **11**, 2100362.

10 M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. v. Vlijmen, M. D. Berliner, M. Z. Bazant, R. D. Braatz, W. C. Chueh and B. D. Storey, 2021-03, **168**, 030525.

11 X. Hu, L. Xu, X. Lin and M. Pecht, *Joule*, 2020, **4**, 310–346.

12 J. Guo, Z. Li and M. Pecht, *Journal of Power Sources*, 2015, **281**, 173–184.

13 Y. Zhang, R. Xiong, H. He and M. G. Pecht, *IEEE Transactions on Vehicular Technology*, 2018, **67**, 5695–5705.

14 A. Aitio and D. A. Howey, *Joule*, 2021, **5**, 3204–3220.

15 T. Baumhöfer, M. Brühl, S. Rothgang and D. U. Sauer, *Journal of Power Sources*, 2014, **247**, 332–338.

16 W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy and D. U. Sauer, *Journal of Power Sources*, 2021, **506**, 230024.

17 J. Hong, D. Lee, E.-R. Jeong and Y. Yi, *Applied energy*, 2020, **278**, 115646.

18 B. Jiang, W. E. Gent, F. Mohr, S. Das, M. D. Berliner, M. Forsuelo, H. Zhao, P. M. Attia, A. Grover, P. K. Herring *et al.*, *Joule*, 2021, **5**, 3187–3203.

19 N. H. Paulson, J. Kubal, L. Ward, S. Saxena, W. Lu and S. J. Babinec, *Journal of Power Sources*, 2022, **527**, 231127.

20 A. Bhowmik, M. Berecibar, M. Casas-Cabanas, G. Csanyi, R. Dominko, K. Hermansson, M. R. Palacin, H. S. Stein and T. Vegge, *Advanced Energy Materials*, 2021, 2102698.

21 J. Amici, P. Asinari, E. Ayerbe, P. Barboux, P. Bayle-Guillemaud, R. J. Behm, M. Berecibar, E. Berg, A. Bhowmik, S. Bodoardo *et al.*, *Advanced Energy Materials*, 2102785.

22 M. Fichtner, K. Edström, E. Ayerbe, M. Berecibar, A. Bhowmik, I. E. Castelli, S. Clark, R. Dominko, M. Erakca, A. A. Franco *et al.*, *Advanced Energy Materials*, 2102904.

23 Y. Preger, H. M. Barkholtz, A. Fresquez, D. L. Campbell, B. W. Juba, J. Romàn-Kustas, S. R. Ferreira and B. Chalamala, 2020-09, **167**, 120532.

24 S. Hochreiter and J. Schmidhuber, *Neural computation*, 1997, **9**, 1735–1780.

25 D. P. Kingma and J. Ba, 2014.

26 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019, pp. 8026–8037.

27 S. Sun, T. Guan, P. Zuo, Y. Gao, X. Cheng, C. Du and G. Yin, *ChemElectroChem*, 2018, **5**, 2301–2309.

28 P. Zhang, T. Yuan, Y. Pang, C. Peng, J. Yang, Z.-F. Ma and S. Zheng, *Journal of The Electrochemical Society*, 2019, **166**, A5489.

29 Y. Abe, N. Hori and S. Kumagai, *Energies*, 2019, **12**, 4507.

30 M. Dubarry, C. Truchot and B. Y. Liaw, 2012-12-01, **219**, 204–216.

31 X. Han, M. Ouyang, L. Lu, J. Li, Y. Zheng and Z. Li, 2014-04-01, **251**, 38–54.

32 I. Bloom, A. N. Jansen, D. P. Abraham, J. Knuth, S. A. Jones, V. S. Battaglia and G. L. Henriksen, *Journal of Power Sources*, 2005, **139**, 295–303.

33 A. Moretti, V. Sharova, D. V. Carvalho, A. Boulineau, W. Porcher, I. de Meatza and S. Passerini, *Batteries & Supercaps*, 2019, **2**, 240–247.

34 P. Dechent, S. Greenbank, F. Hildenbrand, S. Jbabdi, D. U. Sauer and D. A. Howey, *Batteries & Supercaps*, 2021, **4**, 1821–1829.