Supplements

1 Training details

We used a neural network with one LSTM layer and two dense layers, containing 32 neurons each. The input to the LSTM layer is one-dimensional, containing only the current capacity. The covariate input is concatenated with the output of the LSTM layer for input into the dense layers. Since the covariate inputs are not time-dependent, we do not need to input them into the LSTM layer.

We model the output with a tanh unit for the output of the mean and a softplus unit for the output of the variance. Since the variance is always positive, we predict the output as the logarithm of the variance with an added constant (10e-6) for stability. Since we predict the output as the percentage remaining of the last capacity, we restrict the NN to predict values between 85% and 100%. The output is scaled to be between -1 and 1 (the range of the tanh function). As an example, if the capacity of the previous cycle was 1.05 Ah and in the current cycle is 1.00 Ah, the target is $y_{pred} = \frac{1.00}{1.05} \approx 0.9524$. Since the output non-linearity of the LSTM is a tanh unit with a range between -1 and 1, we set -1 and 1 to 85% and 100% respectively as the expected range. The prediction target would then become $\frac{0.9524-0.85}{1-0.85} * 2 - 1 \approx 0,365$

For the covariates, we transform the data to have a mean of 0 and a standard deviation of 1 based on the training set. For the ensembles, we combine five neural networks with identical architecture, identical training procedures and a different random seed for the initialization of the weight and bias terms of the neural network.

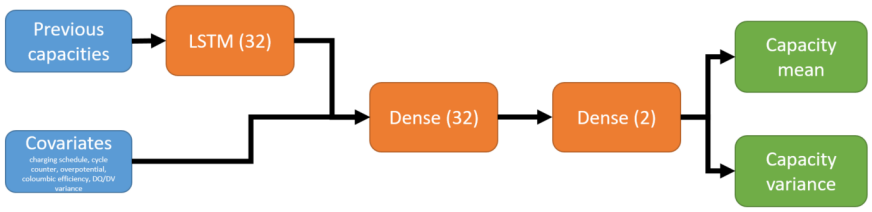

Figure S1: Overview of LSTM architecture.

2 Baseline models

We compare the LSTM with several ML models, a linear regression (with elastic net regularization as used by Severson et. al), an LSTM with no covariate features and a DNN (densely connected neural network) that outputs the entire capacity trajectory as an output. For the linear regression, we evaluated elastic net and ridge regularization, choosing the regularization parameters via 4-fold cross-validation based on the average R2 value and retraining the model with the entire training set with the best parameter.

For the LSTM model with no covariates we chose an architecture consisting of one LSTM layer and two dense layers, equivalent to the LSTM model with covariates. The training process was equivalent to the LSTM model with covariates. For the DNN, we use a three-layer neural network. The number of neurons was chosen via cross-validation on the validation set. The training process was equivalent to the training of the LSTM. Since we output the entire capacity degradation trajectory at once, we use the sum of L1 and L2 loss as the loss function (this generated better results as measured by the validation RMSE than either L1 or L2 loss separately). To calculate the RMSE, we took the first time the capacity dipped below 80% capacity as the predicted EOL.

3 Applying our model to different chemistries

We present the majority of our results on a dataset composed of LFP batteries. To show that our model is not inherently restricted to LFP, we present preliminary results on NMC and NCA batteries from a dataset created by the Sandia National Laboratories in Fig. S2. The dataset contains 40 batteries, (22 NMC batteries, 18 NCA batteries; 21 LFP batteries present in the dataset were excluded as the goal of this study was to show results on non-LFP batteries) which were randomly assigned to the training, validation and testing set. We split the data into 50:25:25 for the training, validation and test set respectively.

We use an identical model to the one described in the main manuscript with the only difference being the encoding of the charging schedule. All batteries in this dataset were cycled with a one-step charging schedule but were also discharged using different C-rates. We encode the charging strategy as a two-dimensional vector with the charge and discharge rate. Otherwise, no changes were made to the ML model.

Compared to the results on LFP batteries, we see in Fig. S2 that the model learned to predict the rate of degradation but is less accurate in prediction. This is likely due to the noise, including lasting upwards jumps of the capacity, present in the data that can readily be seen from the true degradation trajectories in Fig. S2 as well as the two different chemistries in the dataset.

In Fig. S2 we show the prediction of a single LSTM model on the test set batteries. As the batteries were cycled to varying SOH end states and frequently display 'dips' in the discharge capacity far below 80%, it is not possible to calculate an RMSE on the EOL prediction.

We note that in two cases, the prediction deviates significantly from the actual trajectory from the start. In both cases, the battery experienced significant degradation in the first 100 cycles (down to 70% of the initial capacity) and based on visual inspection seems to differ from the rest of the distribution.

If this is the case or if the uncertainty is very large, our model can easily incorporate cycling the battery for longer to reduce the uncertainty.

We note that the accuracy of this model is unsatisfying for any serious deployment. As is visible from the true capacity decline, the data is characterized by many irregular outliers, 'jumps', etc., likely due to experimental conditions, resulting in reduced accuracy for the trained model. As such, the experiments shown here are preliminary.

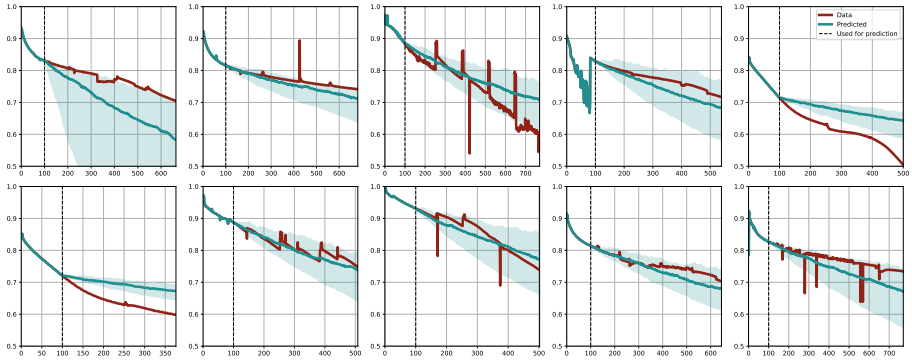

Figure S2: Results on NMC and NCA batteries. The shaded area indicates the uncertainty as the 5th to the 95th percentile.

4 Calendar-aged data

We use the battery cycling dataset from Severson et. al.[7] The batteries were procured at the same time and cycled in three batches. The third batch was cycled approximately one year after the first two batches. Upon inspection of the dataset, the calendar-aged batteries behave differently from the non-calendar-aged batteries under the same environmental conditions, i.e. when cycled with the same cycler and the same charging profile as shown in Fig. S3.

We trained and evaluated the models on the data from the first two batches. The third batch was used as an additional test set to evaluate the model's capacity to adapt to data from a different distribution. We report the RMSE for EOL prediction in the main manuscript. In Fig. S4, we also show degradation patterns for the calendar-aged batteries. We note that the uncertainty predicted by the model in most cases is higher than for the non-calendar-aged data and encompasses the actual trajectory.

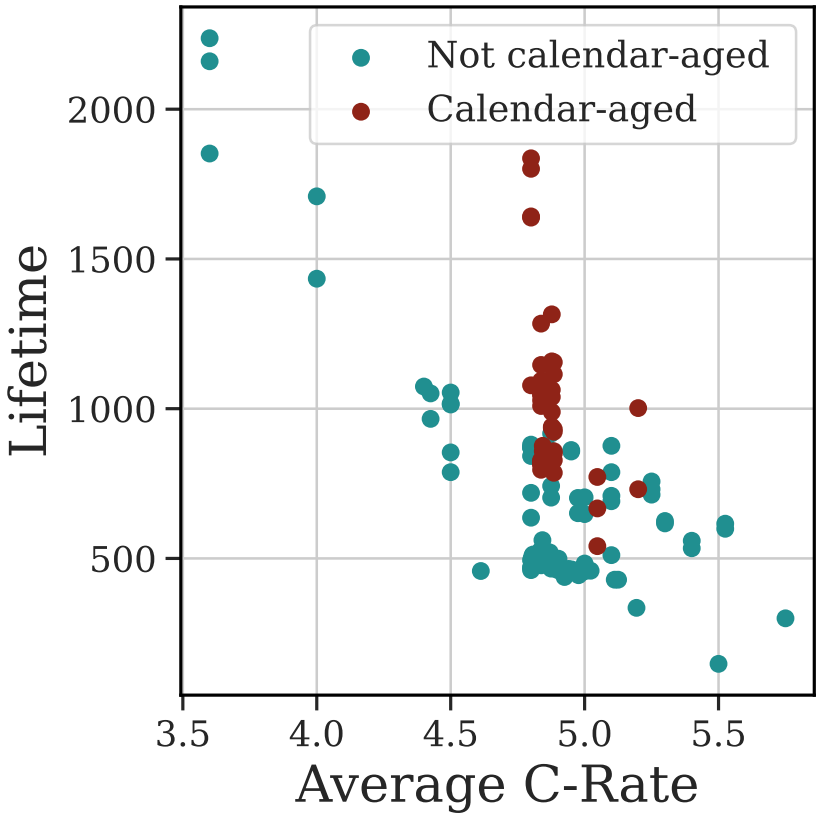

Figure S3: Calendar-aged and non calendar-aged batteries behave differently under the same environmental conditions.

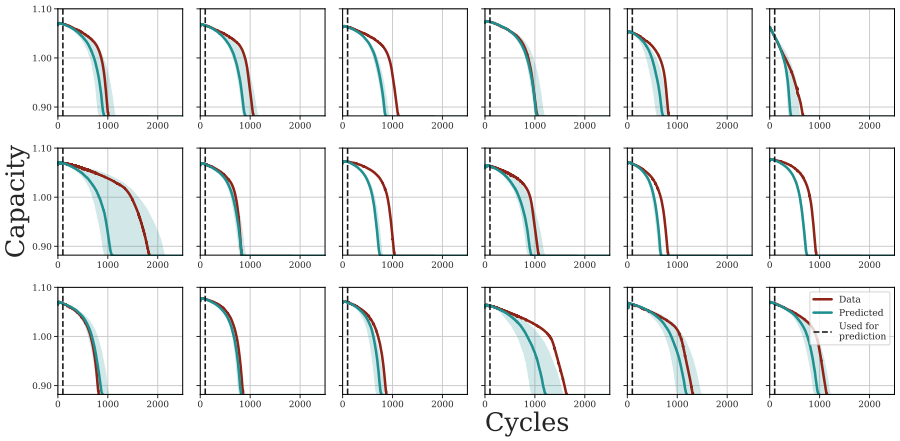

Figure S4: Prediction on previously unseen calendar-aged batteries. On the calendar-aged batteries, the model shows higher uncertainty over the output trajectories than on non-calendar-aged batteries.

5 Input variables

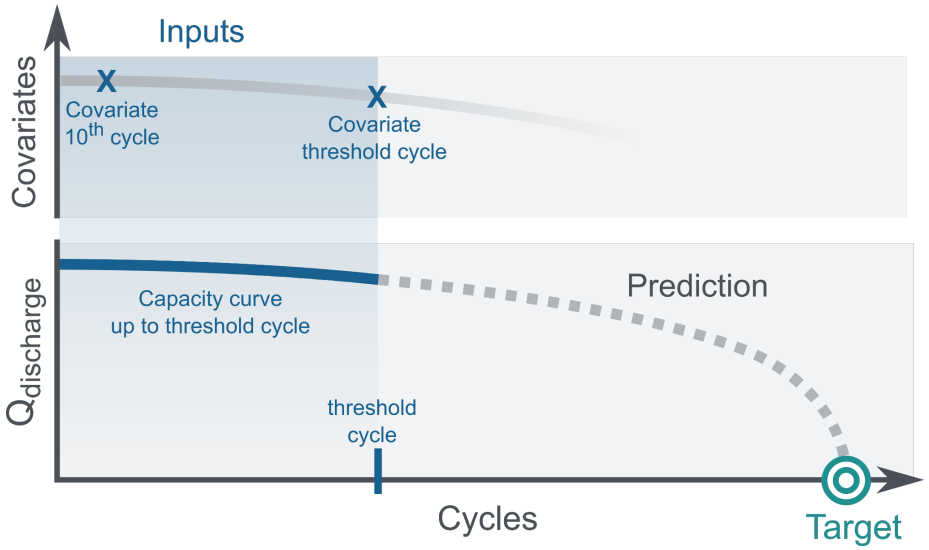

Figure S5: Inputs for each cell used for the LSTM model. During the training phase, the complete degradation curve is used, along with the covariates, i.e. the feature differences between the threshold cycle (last cycle available for prediction) and the 10th cycle. During the prediction phase, the LSTM model uses the first cycles up to the threshold cycle (, along with the covariates, to predict the full degradation curve of an unseen dataset. The RMSE error quantifies how the prediction (target in green) deviates from the experimentally observed end of life.

Table S1: List of covariates used. I, V and Q correspond to the current, voltage and cumulative capacity vectors of each cycle in A, V and Ah units, respectively. The subscripts indicate whether the vectors span the charge (I>0), discharge (I<0) or full cycle. The subscript $_{th}$ indicates the threshold cycle, i.e. the last cycle available for prediction (e.g. 20, 50, 100).

Feature	Associated covariate	Subtracted between the threshold and the 10th cycle
I_{charge}	$\max(I_{\text{charge}})_{th} - \min(I_{\text{charge}})_{10}$	Maximum current during charge
I_{charge}	$\min(I_{\text{charge}})_{th} - \min(I_{\text{charge}})_{10}$	Minimum current during charge
I_{charge}	$\text{mean}(I_{\text{charge}})_{th} - \text{mean}(I_{\text{charge}})_{10}$	Mean current during charge
V_{gap} [1]	$(V_{\text{gap}})_{th} - (V_{\text{gap}})_{10}$	Voltage gap between charge and discharge
Q_{eff} [2]	$(Q_{\text{eff}})_{th} - (Q_{\text{eff}})_{10}$	Coulombic efficiency
Q_{cycle}	$\text{variance}(Q_{th} - Q_{10})$	Capacity vectors [3]

6 Projecting forward

In Fig. S6 we show the standard deviation over the EOL for each number of cycles dependent on the respective accuracy. Uncertainty over the output can be used to decide whether a battery should be cycled longer to reach the desired accuracy.

[1] $\text{mean}(V_{\text{charge}}) - \text{mean}(V_{\text{discharge}})$

[2] $\max(Q_{\text{discharge}})//\max(Q_{\text{charge}})$

[3] Each vector is sampled at the same regular voltage intervals

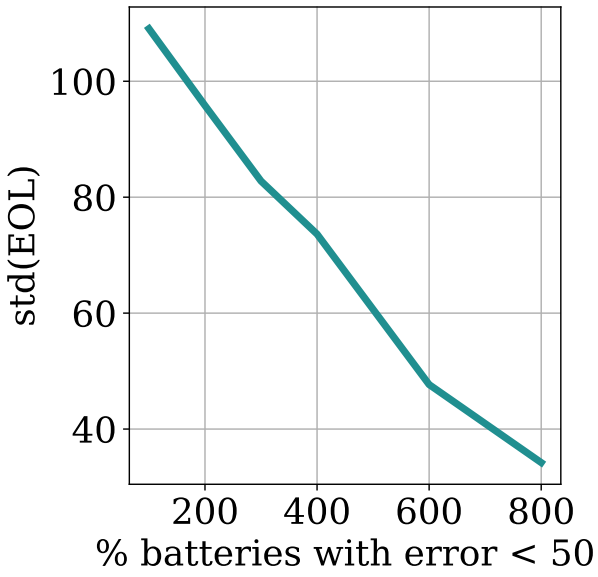

Figure S6: Addition to Section 3.4 - We present the mean standard deviation for each number of cycles depending on the accuracy (presented as the percentage of batteries with an absolute error less than 50.

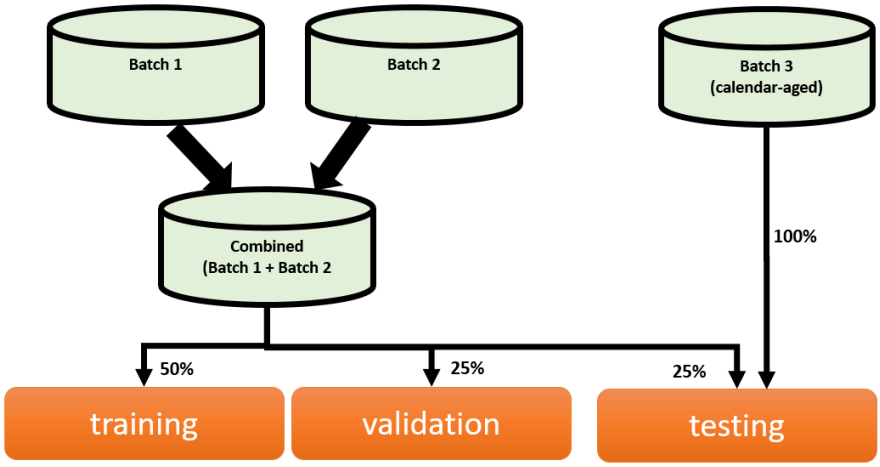

Figure S7: Visualization of the data split for the Severson et. al dataset