Early prediction of battery life by learning from both time-series and histogram data

Early prediction of battery life by learning from both time-series and histogram data

Yizhou Zhang [*,**] **Torsten Wik** [*] **Yicun Huang** [*]
John Bergström [**] **Changfu Zou** [*]

[*] *Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden*
[**] *China Euro Vehicle Technology AB, Gothenburg, 41755, Sweden*
(e-mail: yizhou@chalmers.se)

Abstract: Due to dynamic operating conditions, random user behaviors, and cell-to-cell variations, accurately predicting battery life is challenging, especially using information from only a few early cycles. This work proposes a data-driven battery early prediction pipeline using both time-series, measurement-related features, and usage-related histogram features. We first investigate the prediction performance of using these two feature sources individually, then two methods of systematically combining these two feature sources are devised. Additionally, four machine learning algorithms with different characteristics are applied to compare their performances on battery prognostic problems. We show that the prediction accuracy of using these two feature sources individually is comparable. Moreover, a systematic combination of these two features considerably improves the prediction performance in terms of accuracy and robustness, achieving excellent prediction results with a root mean square error of around 150 cycles using only the first 100 cycle's data. Finally, experimental data of different cell types and cycling conditions are used to verify the developed method's effectiveness and generality.

Keywords: Machine learning, Lithium-ion battery, Battery life prediction, Remaining useful life, Battery management system.

1. INTRODUCTION

For a more sustainable future, many efforts have been put into different sectors, such as transportation and electricity production, to replace fossil fuel usage with sustainable solutions. Lithium-ion (Li-ion) batteries play an important role in this transition by serving as traction batteries for electric vehicles or energy storage devices for the power grid, due to their lower cost, higher reliability, and longer lifetime compared to other alternative solutions. However, as electrochemical devices, Li-ion batteries are also plagued by their nonlinear, complex, and path-dependent aging characteristics, imposing significant challenges for their broad adoption by end customers (Birkl et al., 2017). Therefore, optimally controlling battery usage to extend their lifetime and enhance performance becomes indispensable. To this end, an accurate and robust early prediction algorithm is pivotal to pave the way for a later health-conscious control strategy.

Battery lifetime prediction approaches can be classified into three main categories, namely empirical, physics-based, and data-driven methods, as discussed by (Li et al., 2019). Empirical models, which were initially popular due to their simplicity and low computational requirements, have limitations in accuracy and reliability when applied to real-world scenarios with varying operating profiles and cell-to-cell differences. In contrast, physics-based models provide detailed insights into battery aging mechanisms, but they require complicated parameterization and high computational resources, making them impractical for online applications. Finally, data-driven methods offer flexibility and can recognize patterns and trends in complex and dynamic situations, making them a promising solution for predicting battery life.

The choice of inputs to a data-driven model, often referred to as features, is of great importance for such a method. Many efforts have been made to extract features from the charge or discharge time-series measurement data to predict battery capacity, lifetime, and knee point (Li et al., 2019; She et al., 2022). Additionally, different features based on electrochemical impedance spectroscopy (EIS), cell pressure change, or acoustic analysis have also been investigated (Yang et al., 2021). However, the cost of instrumentation necessary to collect these measurements may hinder the applicability of such methods. As most of the data-driven methods are developed based on repetitive battery cycling in a laboratory, deploying such a model in real-world applications undergoing diverse cycling profiles and an ever-changing cycling environment is very challenging (Sulzer et al., 2021a). Our recent work has used usage-related histogram data-based features, where only information of how the battery has been used is recorded, to recursively predict the future aging trajectory of traction batteries using their historical capacity data (Zhang et al., 2022). However, to the best of our knowledge, there

⋆ This work was funded by the Swedish Energy Agency under the Vehicle Strategic Research and Innovation Program (Grant No. 50187-1)

is no work using such information to early forecast battery life and investigate its relationship to methods using time-series measurements.

In this work, we aim to fill the identified research gap by proposing a data-driven battery early-life prediction pipeline using both time-series measurement data and usage-related information. The contributions arise from three perspectives. First, we demonstrate that using usage-related features alone can achieve the same prediction performance as using the time-series data, indicating that these two feature sources are effectively interchangeable and also complimentary. Second, we developed two methods to systematically combine the information from these two feature sources, both methods achieving excellent performance with increased accuracy and tightened confidence intervals compared to using them individually. Last, the generality of the developed pipeline is demonstrated on two different data sources of different cell chemistry.

2. BATTERY DATASET INTRODUCTION

Two battery datasets are investigated in this study, one originating from Stanford University (Severson et al., 2019), which will be referred to as the Stanford dataset, and another from Sandia National Laboratories (SNL) (Preger et al., 2020), which will be called the SNL dataset hereafter. For the Stanford dataset, the cells used in the test campaign are of lithium iron phosphate (LFP) type, whereas for the SNL dataset, nickel magnesium cobalt (NMC) type batteries were used. The detailed cell specifications are listed in Table 1. Moreover, the cycling profile in the Stanford dataset is various fast charging policies with identical 4C constant discharge, while in the SNL dataset, more diverse cycling profiles are used to investigate their impact on cell aging performance.

Table 1. Battery cell specifications

Dataset	Stanford	SNL
Battery type	LFP	NMC
Manufacture	A123	LG Chem
Nominal capacity (Ah)	1.1	3
Voltage range (V)	2 to 3.6	2 to 4.2
Max discharge current (A)	30	20
Operating Temperature (°C)	-30 to 60	-5 to 50

3. FEATURE CONSTRUCTION AND ENGINEERING

3.1 Usage-related histogram features

The usage profiles have a profound impact on battery life. Therefore, we presume that incorporating both the historic usage recordings and predicted future usage patterns will considerably improve the prediction accuracy and robustness. As per Woody et al. (2020), the following usage-related features, also known as stress factors, are of interest: the depth of discharge (DoD), charge current rate, discharge current rate, cycling/calendar temperature, voltage, SoC and cycling/calendar time. These stress factors form the foundation for the usage-related feature pool that we will construct for later machine learning (ML) algorithm training. It is worth noting that not all features mentioned are available for extraction due to the limitations of the datasets. For example, in the Stanford dataset, only the charge current and charge time can be used due to other features being identical for all the batteries across the whole dataset.

For the Stanford dataset, we construct our features based on the fact that most cells experience a unique charging policy. First, the complete SoC range is divided into 20 intervals of 5%. Then, the corresponding average charge current in C-rate is assigned to each interval to form a feature vector of 20 elements, e.g., $F = [f_1, \cdots, f_{20}]$. The statistical properties of this vector, e.g., mean(F) or var(F), are then calculated to reduce the feature dimension with the hope of not affecting the prediction performance. Apart from using the charge current information, the charging time of the first cycle is also used as one of the features to indicate how the battery is being charged. As for the SNL dataset, the initial SoC, DoD, discharge current, mean and variance of the ambient temperature, and mean and variance of the cell temperature are selected, all as scalar values.

3.2 Time-series features

With a fixed charge or discharge policy, both the current and voltage curves will gradually change when the battery degrades. Considering the hidden aging mechanisms that can potentially be manifested by the current and voltage profiles, the on-purpose feature construction based on these time-series measurement data can be quite useful in predicting the lifetime of the battery. Here, we first investigate the cycle-to-cycle evolution of the accumulated charge/discharge capacity (Q) as a function of the voltage curve (V), namely $dQ(V)$, which has been demonstrated useful in the prediction of battery lifetime (Severson et al., 2019). Second, we study the effectiveness of using the features constructed from the incremental capacity (IC) curves. IC is the ratio between the changing rates of the accumulated charge capacity and of the voltage, which is a powerful tool for battery aging diagnostics. Thus, we hypothesize that the rich aging information that the IC curve contains can also make IC-related features a good health indicator for battery early prediction. In practice, for the Stanford dataset, we adopt the $dQ(V)$ curve and the IC curve during discharge as the baseline features because the batteries experience the same constant discharge profile across the whole dataset. Conversely, with the same reasoning, we use the charge part of the profile to form the baseline features for the SNL dataset.

4. EARLY PREDICTION MODEL

The early prediction problem of battery life is formulated as a regression problem with the aim of minimizing the measured lifetime and the model predictions for all batteries in the test set. Four ML algorithms are considered in the prediction pipeline to evaluate their performance in such cases. Among them, two are linear models, i.e., elastic net (EN) and Bayesian ridge regression (BRR), and the other two are nonlinear support vector regression (SVR) and random forest regression (RFR) models. Fig. 1 illustrates the overall prediction pipeline for the battery's lifetime prediction.

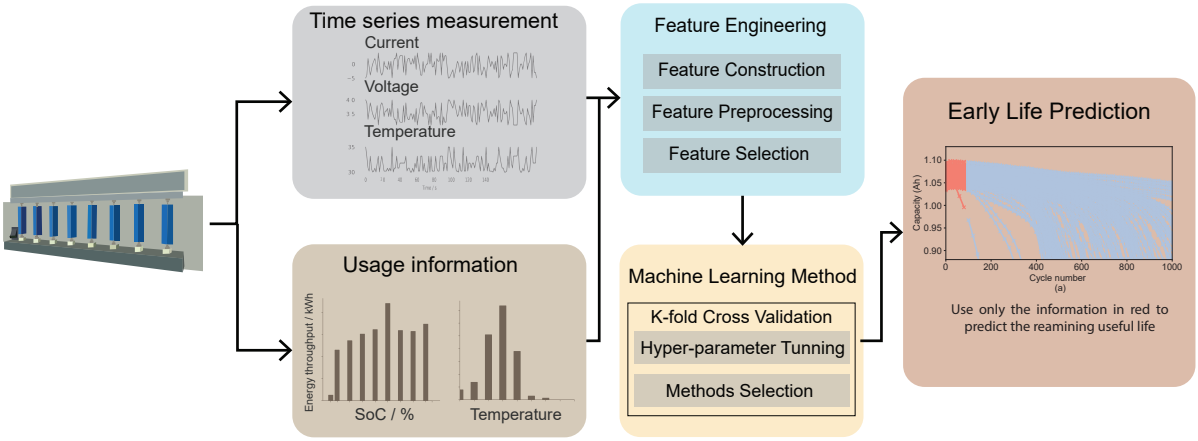

Fig. 1. The overall battery early life prediction pipeline.

Linear models have advantages, such as efficient training and a relatively low computational cost during online deployment, making them the first choice when the accuracy of the model output can meet the requirement of the desired applications. Additionally, the simple and neat model structure makes the final prediction result highly interpretable. EN consists of an ordinary least square formulation together with two regularization terms, e.g., L1 and L2 regularization, with the goal of solving the following optimization problem:

$$\hat{\theta} = \arg\min_{\theta} \|y - X\theta\|_2^2 + \lambda \frac{1-\alpha}{2} \|\theta\|_2^2 + \alpha \|\theta\|_1, \qquad (1)$$

where y is the prediction target, in this case, the measured lifetime of the battery, X is the feature matrix, θ is the model coefficient vector, and both λ and α are hyperparameters that are tuned using cross-validation. Similarly, applying a Bayesian approach to the ordinary least square problem leads to the BRR. The algorithm can incorporate the regularization parameters in the estimation procedure and naturally and systematically propagate its prediction uncertainty.

A nonlinear model, such as SVR and RFR, may require a complex calculation procedure for both the training and testing but may increase the prediction accuracy and robustness for certain applications. Mathematically, an SVR algorithm tries to minimize the ϵ-insensitive loss:

$$L(y, \hat{y}) = \begin{cases} 0 & \text{if } |y - \hat{y}| < \epsilon \\ |y - \hat{y}| - \epsilon & \text{otherwise,} \end{cases} \qquad (2)$$

where $\hat{y}$ is the predicted lifetime value. By introducing the slack variables and constructing the Lagrangian of the primal formulation, we can find the solution to such an optimization problem. The built-in sparsity characteristics and outstanding fitting performance for nonlinear systems make SVR a popular algorithm for battery prognostics. RFR is an ensemble learning method, which is used through bootstrapping and data aggregation to achieve a good trade-off between bias and variance. Due to its ensemble characteristics, RFR is a robust and accurate supervised learning algorithm for highly nonlinear systems with complicated data and features. To quantify the estimation uncertainty, we adopt the method from (Wager et al., 2014) that uses the bootstrap replicates during the training process to calculate the confidence interval of the prediction result for RFR.

Combining features from different sources during the training process is one way to systematically incorporate information from both time-series measurements and battery usage. Alternatively, fusing the prediction results of independent models trained with different feature inputs is also a plausible candidate. One of the motivations for considering fusing the independent model result is to have the flexibility to choose which model to use depending on the availability of the feature input, as practical data collection challenges (communication delays, data corruption, or memory shortage) may hinder the applicability of a certain type of feature input. Herein, we adopt an inverse-variance weighting method to aggregate the prediction results, for which the fused prediction result is given by:

$$\hat{y}_f = \frac{\sum_m \hat{y}_m / \sigma_m^2}{\sum_m 1/\sigma_m^2}, \qquad (3)$$

where $\hat{y}_f$ is the fused result, $\hat{y}_m$ is the individual prediction result using model m, and σ_m is the standard deviation of the prediction result from model m. Correspondingly, the prediction variance of the fused result is calculated as

$$\mathrm{Var}(\hat{y}_f) = \frac{1}{\sum_m 1/\sigma_m^2}. \qquad (4)$$

5. EVALUATION MATRICES

For the train-test set split, a ratio of 8/2 is adopted, where 80% of the batteries are randomly selected as the train set, and the remaining unseen data are assigned to the test set. Three error evaluation matrices are chosen to quantify the performance of the developed method, i.e., the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute percentage error (MAPE)

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \qquad (5)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \qquad (6)$$

$$\mathrm{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \qquad (7)$$

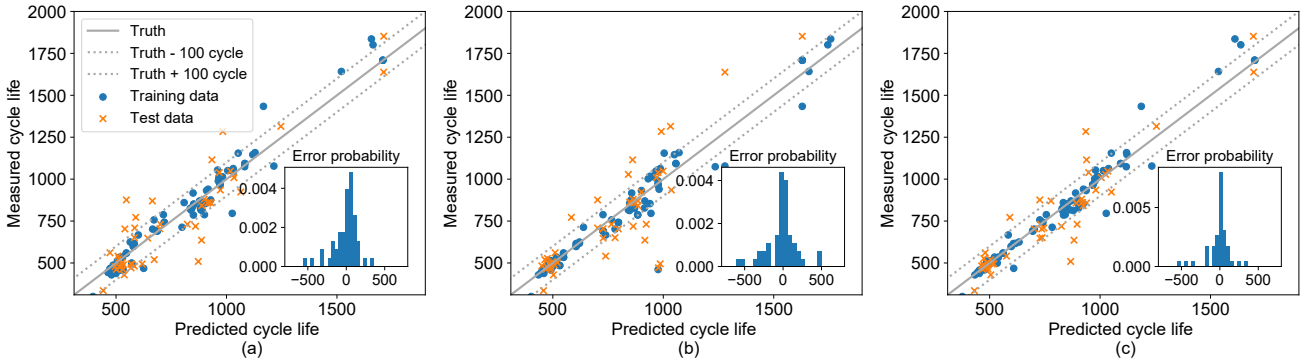

Fig. 2. Prediction results using different feature inputs extracted from the Stanford dataset. (a) Using time-series features only. (b) Using usage-related histogram features only. (c) Using both feature sources.

where $\bar{y}$ is the mean value of the measured lifetime, and N is the number of data points in the test set.

6. RESULTS AND DISCUSSIONS

We first discuss the results obtained from the Stanford dataset, followed by the SNL dataset. In the beginning, only the cycling information from the first 100 cycles is used for prediction. Then, we perturb the number of early life cycles to study its sensitivity to the prediction results.

To compare the prediction performance using different feature sources, we applied the same ML algorithm, RFR in this case, but trained it with different feature inputs. The results are summarized in Table 2. It can be seen that the prediction performance of using time-series features or usage-related histogram features alone are similar in terms of the MAPE, with the histogram feature-based algorithm slightly outperforming that of the time-series features. Based on this, we could conclude that the two feature sources are effectively interchangeable for battery life prediction. However, it is worth highlighting that when both feature sources are used, significant performance improvement is achieved, showing that the two feature sources are complementary and should be used together when possible. By examining the detailed prediction result shown in Fig. 2, the model trained with combined features tends to violate less the ±100 cycles prediction boundary of the measured lifetime when compared to using individual feature sources. Additionally, the zoom-in figures show the error histogram of the prediction results using different feature sources. This result again shows the superiority of combining both feature sources, including less extreme predictions and narrower error distribution. The superiority stems from the fact that the usage-related feature can indicate how the cycling profiles affect the battery life in an average sense, while the time-series features are able to identify the cell-to-cell variations.

Fig. 3 illustrates the prediction result on a randomly selected cell in the test set using different feature sources. The prediction result of using combined feature sources is not only better in terms of prediction accuracy, but it also has a narrower prediction confidence interval compared to the other two predictions using only one feature source.

As stated in Section 4, instead of combining the feature sources and then training one ML algorithm, an alternative

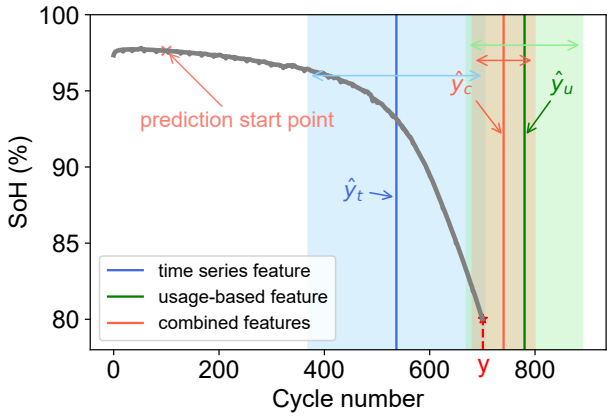

Fig. 3. The prediction result and its confidence interval using different feature sources was demonstrated on a randomly selected cell, where the shaded color illustrates the 95% prediction confidence interval. $\hat{y}_c$, $\hat{y}_t$, and $\hat{y}_u$ indicate the predicated cycle lives, and y is the measured cycle life.

Table 2. Results of different feature inputs and two combination methods for early prediction of battery life using the Stanford dataset

Feature input	R^2	RMSE (cycles)	MAPE (%)
Time-series	0.78	197	15.41
Histogram	0.85	162	14.31
Combined features	**0.87**	**149**	10.51
Combined models	**0.87**	151	**10.11**

option is to train two ML algorithms using all the feature inputs and fuse their prediction results. The last row in Table 2 shows the numerical result of training two RFR prediction models and then fusing the prediction results of the two using (3). Not surprisingly, the obtained results are similar to the ones using combined features to train one ML algorithm. The detailed prediction results in Fig. 2(c) and Fig. 4(a) also verify the similar prediction performance of the two methods. This result can be attributed to the fact that these two feature sources are largely complementary. The histogram of the standard deviation of the prediction uncertainty and their median value for different methods are shown in Fig. 4(b). It can be observed that when using the same number of trees in the RFR algorithm to propagate prediction uncertainty, one of the advantages of training two RFR models and

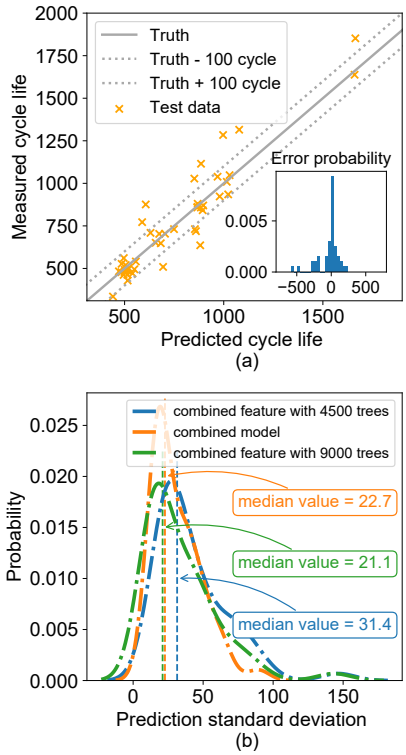

Fig. 4. (a) Detailed prediction result by systematically fusing the individual prediction output of RFR trained with different feature sources. (b) The comparison result of the prediction standard deviation for different combination methods.

fusing the final results is that the confidence interval of the outputs is considerably tightened. Additionally, to achieve a similar confidence interval level, the number of trees needs to be doubled for the model combining the feature sources. Therefore, combining different models can save computational resources.

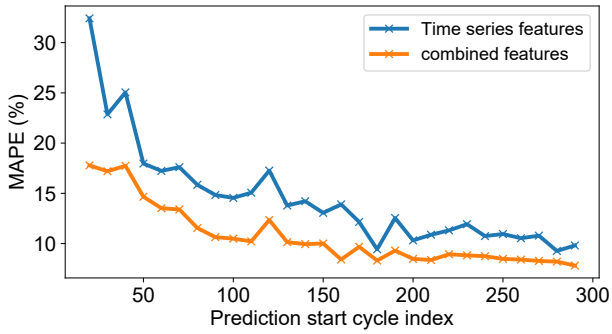

Fig. 5. Prediction error as a function of prediction start cycle number.

In addition to the prediction results obtained above using the first 100 cycles, we evaluate the robustness of the developed methods by varying the early life data from 20 to 300 cycles, with the result shown in Fig. 5. It can be noted that incorporating usage-related features can dramatically increase prediction accuracy relative to the case of time-series features, especially when performing early prediction. For example, when we utilize the first 20 cycles, the prediction accuracy can be improved by around 45%. Furthermore, the superiority of using combined features is consistent over the whole examined range of $[20, 300]$ cycles, verifying the importance of having such information included in the feature construction step.

Finally, the prediction results of different ML algorithms trained with combined feature input are reported in Table 3. The best-performing algorithm, in this case, is RFR with a MAPE of around 10%. However, we want to emphasize the importance of having several ML algorithms of different characteristics in the prediction pipeline, as which algorithm performs the best may depend on the dataset.

Table 3. Results of different ML algorithms for battery life prediction using Stanford dataset

ML algorithm	R^2	RMSE (cycles)	MAPE (%)
EN	0.83	175	17.91
BR	0.81	181	14.73
SVR	0.61	263	20.62
RFR	**0.87**	**149**	**10.51**

The model parameter θ of the EN model can be used to evaluate how much each feature contributes to the final prediction result. Fig. 6 shows the absolute value of the weighting coefficient for the selected features. Among these, the variance and the minimum value of the $dQ(V)$ curve are the most important features with the highest contribution to the final prediction result, which is in line with the conclusion drawn in (Severson et al., 2019). Other than that, the next most important features are extracted from the charging time and the mean charge current, which demonstrates the effectiveness of including usage-related information in the feature pool, i.e., the increased prediction accuracy and robustness.

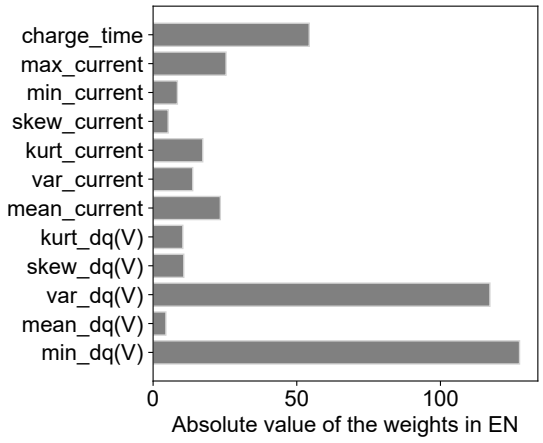

Fig. 6. Feature importance analysis.

To verify the generalizability of the proposed method, we apply our approach to the SNL dataset with NMC-type chemistry. The numerical prediction results of various kinds of feature inputs are shown in Table 4, and the detailed prediction results are illustrated in Fig. 7. Generally, the conclusions that we drew previously on the Stanford dataset are also valid here. Comparatively, the prediction performance discrepancy between using the time-series features and the usage-related histogram features is larger in the SNL dataset. The reason could be that the cells

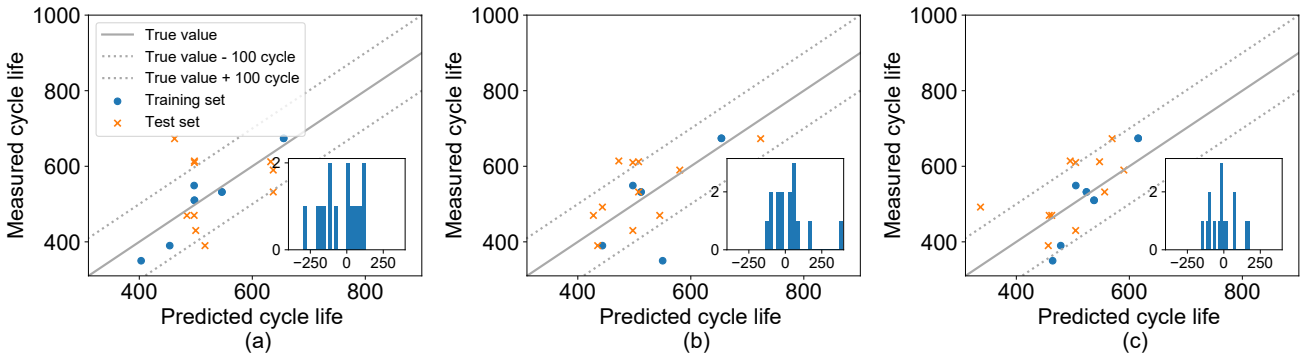

Fig. 7. Prediction results using different feature inputs for SNL dataset. (a) Using time-series, measurement-related features only. (b) Using usage-related features only. (c) Using both feature sources.

in this dataset experience more diverse cycling profiles than those in the Stanford dataset. Moreover, the charging part of the $dQ(V)$ curve does not manifest as much aging information as the discharging part of the $dQ(V)$ curve, which has also been noticed by (Sulzer et al., 2021b). Another interesting finding is that cell consistency, which is how much the battery life varies if the cells experience exactly the same cycling profile, will considerably affect the prediction accuracy using usage-related features. We have noted here that the lifetime variance of the cells that conduct the same cycling profiles inside the Stanford dataset is 78 cycles, whereas, for the SNL dataset, it is 42 cycles. Consequently, the prediction accuracy of the SNL dataset adopting usage-related features is also higher than that of the Stanford dataset.

Table 4. Results of different feature input for battery life prediction using SNL dataset

Feature input	R^2	RMSE (cycles)	MAPE (%)
Time-series	0.95	136	16.38
Usage related	0.96	131	11.32
combined features	**0.98**	**94**	**10.52**

7. CONCLUSION

In this work, we developed a battery lifetime early prediction pipeline that is accurate and robust under various cycling profiles and for different cell chemistry, highlighting the importance and benefits of simultaneously using both time-series features and usage-related histogram features. We foresee such a method should be considered as a standard way to formulate feature pools for battery prognostic problems.

In the future, the efficacy of such a prediction model under more dynamic usage protocols, e.g., the NEDC cycle or even using field data directly, should be investigated. More broadly, further development of the optimal control strategy to extend the lifetime and enhance the battery's performance based on such a lifetime prediction model can bring more value to the battery communities.

REFERENCES

Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. *J. Power Sources*, 341, 373–386.

Li, Y., Liu, K., Foley, A.M., Zülke, A., Berecibar, M., Nanini-Maury, E., Van Mierlo, J., and Hoster, H.E. (2019). Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew. Sustain. Energy Rev.*, 113, 109254.

Preger, Y., Barkholtz, H.M., Fresquez, A., Campbell, D.L., Juba, B.W., Romàn-Kustas, J., Ferreira, S.R., and Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *J. Electrochem. Soc.*, 167(12), 120532.

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. *Nat. Energy*, 4(5), 383–391.

She, C., Li, Y., Zou, C., Wik, T., Wang, Z., and Sun, F. (2022). Offline and online blended machine learning for lithium-ion battery health state estimation. *IEEE Trans. Transp. Electrific.*, 8(2), 1604–1618.

Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., et al. (2021a). The challenge and opportunity of battery lifetime prediction from field data. *Joule*, 5(8), 1934–1955.

Sulzer, V., Mohtat, P., Lee, S., Siegel, J.B., and Stefanopoulou, A.G. (2021b). Promise and challenges of a data-driven approach for battery lifetime prognostics. In *American Control Conference*, 4427–4433.

Wager, S., Hastie, T., and Efron, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.*, 15(1), 1625–1651.

Woody, M., Arbabzadeh, M., Lewis, G.M., Keoleian, G.A., and Stefanopoulou, A. (2020). Strategies to limit degradation and maximize Li-ion battery service lifetime - critical review and guidance for stakeholders. *J. Energy Storage*, 28, 101231.

Yang, S., Zhang, C., Jiang, J., Zhang, W., Zhang, L., and Wang, Y. (2021). Review on state-of-health of lithium-ion batteries: Characterizations, estimations and applications. *J. Cleaner Production*, 314, 128015.

Zhang, Y., Wik, T., Bergström, J., Pecht, M., and Zou, C. (2022). A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data. *J. Power Sources*, 526, 231110.