Comparative Analysis of Battery Cycle Life Early Prediction Using Machine Learning Pipeline

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 3757–3763

Comparative Analysis of Battery Cycle Life Early Prediction Using Machine Learning Pipeline ⋆

Huang Zhang * **Faisal Altaf** ** **Torsten Wik** *** **Sébastien Gros** ****

* *Department of Electromobility, Volvo Group Trucks Technology, Gothenburg, Sweden (e-mail: huang.zhang@volvo.com) and Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden (email: huangz@chalmers.se).*
** *Department of Electromobility, Volvo Group Trucks Technology, Gothenburg, Sweden (e-mail: faisal.altaf@volvo.com)*
*** *Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden, (e-mail: torsten.wik@chalmers.se)*
**** *Department of Engineering Cybernetic, Norwegian University of Science and Technology, Trondheim, Norway, (e-mail: sebastien.gros@ntnu.no)*

Abstract: Lithium-Ion battery system is one of the most critical but expensive components for both electric vehicles and stationary energy storage applications. In this regard, accurate and reliable early prediction of battery lifetime is important for optimizing life cycle management of batteries from cradle to grave. In particular, accurate aging diagnostics and prognostics is crucial for ensuring longevity, performance, safety, uptime, productivity, and profitability over a battery's lifetime. However, current state-of-art methods do not provide satisfactory prediction performance (lack of uncertainty quantification) using early degradation data. In the present work, to produce the best model for both battery cycle life point prediction and range prediction (i.e., confidence intervals or prediction intervals), a pipeline-based approach is proposed, in which a full 33-feature set is generated manually based on battery degradation knowledge, and then used to learn the best model among five machine learning (ML) models that have been reported in the battery lifetime prediction literature, and two quantile regression models for battery cycle life prediction. The calibration and sharpness property of battery cycle life range prediction is properly evaluated by their coverage probability and width respectively. The experimental results show that the gradient boosting regression tree model provides the best point prediction performance, while the quantile regression forest model provides the best range prediction performance with both full 33-feature set and the MIT 6-feature set.

Keywords: Computational intelligence in control, lithium-ion battery, cycle life early prediction, uncertainty quantification, prediction intervals

1. INTRODUCTION

Lithium-ion batteries have been widely used as energy storage systems in various applications, such as, mobile devices, electric vehicles, and microgrids, due to their high energy and power density, low cost and long lifetime characteristics Schmuch et al. (2018). However, as a result of a complex interplay of different physical and chemical mechanisms, the performance (e.g., available energy, and available power) of lithium-ion batteries gradually degrades over time where the degradation rate is a nonlinear function of storage and cycling conditions (ambient temperature, state-of-charge (SoC) window, charge/discharge current. energy throughput etc) Vetter et al. (2005). In the application of stationary energy storage systems, the battery capacity degradation can be converted into the replacement cost of energy storage systems, which increases with battery degradation Wang et al. (2020). Battery degradation may also incur safety issues (e.g., due to excessive lithium plating and dendrite growth leading to internal short-circuit) during their service over lifetime Han et al. (2019). Therefore, to ensure profitable and safe battery usage throughout their whole lifetime, accurate and reliable battery lifetime prediction using early degradation data is of great importance. Moreover, optimizing a parameter space of fast-charging protocols for maximizing battery lifetime can become tractable by a significant reduction of the number of testing cycles per battery cell Attia et al. (2020). However, the lithium-ion battery degradation process exhibits high nonlinearities and depends on both external factors (e.g., storage conditions Grolleau et al. (2014), cycling conditions and their sequences Raj

⋆ The authors would like to thank Volvo Group and Swedish Energy Agency for funding this work.

10.1016/j.ifacol.2023.10.1545

et al. (2020)), and internal factors (e.g., variances due to manufacturing tolerances Baumhöfer et al. (2014)). All these factors complicate the battery lifetime prediction.

Over the past decade, there has been significant research work on battery lifetime prediction, which is also known as remaining useful life (RUL) prediction. Based on a measure of battery lifetime (e.g., cycle number, Ah-throughput) together with an end of life (EoL) threshold, various methods have been developed in studies to estimate the lifetime remaining until EoL. The methods developed in these studies can be roughly divided into two categories, model-based methods and data-driven methods. Model-based methods typically start with developing a mathematical model that captures the battery degradation dynamics, for example, a physics-based model Sadabadi et al. (2021), a semi-empirical model Lian et al. (2020) or an empirical model Hu et al. (2018). The mathematical model is then incorporated into a recursive Bayesian filter framework, such as a Kalman filter Chang et al. (2017) or a particle filter Miao et al. (2013), in which battery model parameters are recursively updated from measured data. Although the battery lifetime or RUL prediction performance of model-based methods has successfully been demonstrated in the aforementioned studies, these methods still struggle to predict battery lifetime at the early degradation stage because of the typical two-stage nonlinear degradation process as well as limited data collected from a short range of early lifetimes Severson et al. (2019).

Instead of developing an explicit mathematical model to characterize the battery degradation dynamics, data-driven methods are used in some studies to formulate the battery lifetime or RUL prediction problem as a standard regression problem, with the goal of learning a mapping function from input features extracted from battery degradation data to the battery lifetime or RUL directly, given a large set of input-output training pairs. Severson et al. (2019) extracted features mainly from discharge voltage and capacity curves, and then selected three feature subsets to learn three different elastic net models to predict battery lifetime. Zhang et al. (2021) extracted features from discharge voltage curves, discharge capacity, and internal resistance, and then high-importance features were selected to learn a general regression neural network for RUL prediction. In some other studies that also employ data-driven methods, the battery lifetime or RUL prediction problem is formulated as a time-series forecasting problem, in which the future battery state of health (SoH) is estimated first and then the battery lifetime or RUL is predicted as the time that the estimated future SoH reaches a predefined EoL threshold. Richardson et al. (2019) extracted features from elapsed time, charge throughput in each load pattern, elapsed time in the selected voltage, current, temperature windows in each load pattern, and then used a Gaussian process regression (GPR) with a Matérn kernel to predict battery SoH. Ma et al. (2019) used a convolutional neural network (CNN) to automatically extract features from the battery capacity data in a sliding window and then employed long short-term memory (LSTM) to predict battery SoH by using the extracted features. However, the amount of required battery degradation data for accurate RUL prediction by LSTM accumulates up to 25% of the whole lifetime data Zhang et al. (2018).

Although data-driven methods that are employed in the aforementioned studies provide outstanding lifetime prediction accuracy with limited data at early cycles, the uncertainty associated with battery lifetime prediction is either lacking or quantified with Gaussian assumption that is not necessarily hold, especially in the case of a small battery dataset.

Considering the great benefits of battery lifetime prediction using early degradation data for optimization of battery life cycle management from cradle to grave, there is still a need of accurate and reliable battery lifetime prediction using as early degradation data as possible. Here in this work, we propose a pipeline-based approach (illustrated in Fig. 1) to automate the process of producing the best model for both battery cycle life point prediction and battery cycle life range prediction. In the proposed pipeline, 33 input features are first extracted from the first 100 cycles of degradation data. Next, to generate the best model for battery cycle life point prediction provided by conditional mean, five ML models in the battery lifetime prediction literature and two quantile regression models are compared with respect to their point prediction performance. Alternatively, to generate the best model for battery cycle life range prediction provided by either confidence intervals or prediction intervals, two quantile models and Gaussian process regression are compared with respect to their range prediction performance. The effectiveness of the proposed pipeline for battery cycle life point prediction and range prediction are demonstrated by several comparative simulation experiments.

The novelty and contributions of this work are as follows:

- A pipeline-based approach is proposed for automating the process of producing the best model for both battery cycle life point prediction and range prediction using early degradation data, which makes model selection easier prior to its deployment in an application. The proposed pipeline offers several advantages over conventional lifetime prediction methods, such as defined performance evaluation metrics for assessing cycle life point prediction and range prediction, respectively, adaptability to different charging protocols with minimum knowledge of battery degradation, and versatility to other battery state estimation and prediction tasks.
- Two representative quantile regression models are introduced to provide battery cycle life point and range prediction, and their prediction performance is benchmarked to five ML models that have been reported in the battery lifetime prediction literature. Both calibration property referring to the statistical consistency between the constructed prediction interval and the observations and sharpness property referring to the concentration of the predictive distributions of range predictions are evaluated by their coverage probability and width respectively. Simulation results show that quantile regression models are not only capable of providing cycle life point prediction with high accuracy but also cycle life range prediction with high reliability.

2. BATTERY DATASET AND MACHINE LEARNING PIPELINE

The machine learning pipeline that automatically selects the best model among various ML models for battery cycle life early prediction is presented in Fig. 1.

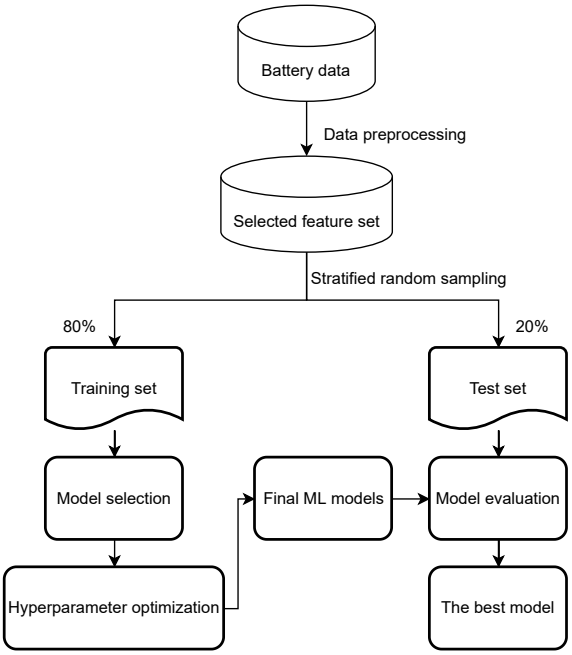

Fig. 1. The proposed machine learning pipeline for selecting the best ML model to predict battery cycle life.

2.1 Battery dataset

The battery dataset generated by Severson et al. (2019) is used to validate the effectiveness of the proposed pipeline. There are 124 lithium iron ferrous phosphate (LFP)/graphite cells with 1.1 Ah nominal capacity. These cells are from 3 different batches (i.e., the "2017-05-12" batch, the "2017-06-30" batch, and the "2018-04-12" batch) where the batch date indicates the date of start of experiment. Each battery cell is charged from 0% to 80% state-of-charge (SoC) with a one-step or two-step fast-charging protocol, and then charged from 80% to 100% SoC with a uniform 1C constant current-constant voltage (CC-CV) charging step to 3.6 V and a current cutoff of C/50. Subsequently, the cells are discharged identically with a 4C CC-CV discharging step to 2.0 V and a current cutoff at C/50. For example, a charging protocol "6C(30%)-3.6C" represents a 6C charging step from 0% to 30% SoC, followed by a 3.6C charging step from 30% to 80% SoC. All cells are cycled until they reach end of life, i.e., 80% of initial nominal capacity.

2.2 Data preprocessing

The data preprocessing can have a significant impact on generalization performance of a ML model, which consists of data cleaning, normalization, transformation, feature engineering and selection, etc Kotsiantis et al. (2006). In the present application of battery cycle life prediction, there are 33 features extracted from the first 100 cycles data with aid of battery domain knowledge. The 33 features are divided into five groups and listed in Table. 1.

2.3 Train-test split

Notably, there are 72 different charging protocols used in this battery dataset, whose nominal charging time from 0% to 80% SoC ranges from 9 to 13.3 min. In order to reduce the possibly large sampling error due to relatively small dataset in this work, the stratified random sampling method Reitermanova et al. (2010) is used to randomly split the dataset, with 80% in a training set (99 samples) and 20% in a test set (25 samples). Equal ratios of fast-charged (i.e., less than 10.5 min) cells , medium-charged (i.e., between 10.5 and 11.7 min) cells, and slow-charged (i.e., greater than 11.7 min) cells are preserved in the training and test set at each split. Moreover, for the purpose of reducing the random effect of the selected split, the stratified random sampling is repeated 5 times, and the experimental results of 5 train-test splits are averaged.

2.4 Model selection

Quantile regression forest Quantile regession is one type of regression method which provides conditional quantiles of battery cycle life given input feature values from the first 100 cycles degradation data Koenker (2005). Comparing with the classical least squares estimation, quantile regression estimates are more robust against possible outliers in battery cycle life observations. Moreover, distribution assumptions of battery cycle life as output variable, such as Gaussian, are not required for quantile regression models.

Different from a random forest, all observations in every leaf of every tree are stored in a QRF, which are used to construct the conditional distribution of the battery cycle life Meinshausen and Ridgeway (2006). To compute the conditional quantile at a given level α by a QRF for a given $\boldsymbol{x} \in \mathbb{R}^p$ represents p features extracted from the first 100 cycles of a battery cell, i.e., $\hat{Q}_\alpha(\boldsymbol{x})$, the empirical conditional cumulative distribution function is first obtained by averaging weights of every observation over all trees of the forest. Then the estimate of the conditional quantile $\hat{Q}_\alpha(\boldsymbol{x})$, is obtained.

Quantile regression gradient boosting By substituting the squared error loss function of the GBRT with a pinball loss function, the GBRT is able to estimate the conditional quantile at a given level α for a given $\boldsymbol{x}$, i.e., $\hat{Q}_\alpha(\boldsymbol{x})$ Friedman (2001).

The prediction intervals are constructed from the predictive quantiles of battery cycle life by quantile regression models. We consider the classical case of central $(1-\alpha)\times 100\%$ prediction interval for battery cycle life, with lower and upper bounds that are the predictive quantiles at level $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$. In this work, the 95% prediction interval for battery cycle life is estimated by

$$\hat{I}(\boldsymbol{x}) = [\hat{Q}_{.025}(\boldsymbol{x}), \hat{Q}_{.975}(\boldsymbol{x})], \tag{1}$$

which means given $\boldsymbol{X} = \boldsymbol{x}$, a new observation of battery cycle life Y is in the interval $\hat{I}(\boldsymbol{x})$ with a probability of 95%.

Table 1. 33 features in 5 groups

Groups	Features
Time-related	Average charge time for the first 5 cycles Severson et al. (2019).
Curve-related	Minimum, variance, skewness, and kurtosis of difference of the discharge voltage curve between cycle 100 and cycle 10 (i.e., $\Delta Q_{100-10}(V)$) Severson et al. (2019). Amplitude and position shift of the highest peak in the discharge incremental capacity curve between cycle 10 and cycle 100 (i.e., $dQdV_{100-10}$). Minimum, variance, skewness, and kurtosis of difference in the discharge cell temperature, as a function of voltage, between cycle 100 and cycle 10 (i.e., $\Delta T_{100-10}(V)$). Minimum, maximum, mean, and variance of discharge cell temperature as a function of voltage at cycle 10 (i.e., $T_{10}(V)$). Minimum, maximum, mean, and variance of discharge cell temperature as a function of voltage at cycle 100 (i.e., $T_{100}(V)$). Difference in minimum, maximum, mean, and variance of discharge cell temperature, as a function of voltage, between cycle 10 and cycle 100.
Capacity-related	Slope of the linear fit to the capacity fade curve from cycle 2 to cycle 100 Severson et al. (2019). Intercept of the linear fit to capacity fade curve from cycle 2 to cycle 100 Severson et al. (2019). Discharge capacity at cycle 2 Severson et al. (2019). Discharge capacity at cycle 100 Severson et al. (2019). Difference between maximum discharge capacity within the first 100 cycles and discharge capacity at cycle 2 Severson et al. (2019).
Internal resistance-related	Minimum internal resistance from cycle 2 to cycle 100 Severson et al. (2019). Maximum internal resistance from cycle 2 to cycle 100. Internal resistance at cycle 2 Severson et al. (2019). Internal resistance at cycle 100. Difference in internal resistance between cycle 100 and cycle 2 Severson et al. (2019).

Additionally, machine learning models, such as elastic net Severson et al. (2019), support vector regression Qin et al. (2015), random forest regression Voronov et al. (2018), gradient boosting regression tree Yang et al. (2020), and Gaussian process regression Richardson et al. (2019), have been reported to be used for battery lifetime prediction, and therefore are included in the proposed pipeline.

2.5 Model performance evaluation

Evaluating averaged performance in LOO-XV To select the best set of hyperparameters for a model that provides point prediction, the averaged squared error (ASE) over each validation sample in the LOO-XV is used as the criterion, and is defined as,

$$\mathrm{ASE}(y_i, \hat{y}_i) = \frac{1}{N_D}\sum_{i=1}^{N_D}(y_i - \hat{y}_i)^2 \tag{2}$$

where N_D denotes the number of samples in the training set, and $y_i \in \mathbb{R}^+$ and $\hat{y}_i$ denote the observed cycle life and the predicted cycle life of validation cell i, respectively.

To select the best set of hyperparameters for a GPR model that provides range prediction, the averaged negative log pseudo-likelihood (ANLL) Williams and Rasmussen (2006) over each validation sample is used as the criterion and is defined as,

$$\begin{aligned}\mathrm{ANLL}(y_i, \hat{\mu}_i, \hat{\sigma}_i) = -\frac{1}{N_D}\sum_{i=1}^{N_D}(-\frac{1}{2}\log\hat{\sigma}_i^2 - \frac{(y_i-\hat{\mu}_i)^2}{2\hat{\sigma}_i^2} \\ -\frac{1}{2}\log 2\pi)\end{aligned} \tag{3}$$

where $\hat{\mu}_i$ is the predictive mean and $\hat{\sigma}_i^2$ is the predictive variance of validation cell i.

To select the best set of hyperparameters for a quantile regression model that provides range prediction, the averaged interval score (AIS) Gneiting and Raftery (2007) over each validation sample is used as the criterion and is defined as,

$$\begin{aligned}\mathrm{AIS} =& \frac{1}{N_D}\sum_{i=1}^{N_D}((u_i - l_i) \\ &+ \frac{2}{\alpha}(l_i - y_i)\mathbb{1}_{\{y_i<l_i\}} \\ &+ \frac{2}{\alpha}(y_i - u_i)\mathbb{1}_{\{y_i>u_i\}}),\end{aligned} \tag{4}$$

where l_i is the lower bound, and u_i is the upper bound of a constructed PI. By rewarding narrow prediction intervals and penalizing intervals missed by the observation depending on the size of α, the AIS defined above can assess both calibration and sharpness properties of PIs.

Evaluating point prediction quality The most common performance evaluation metrics to evaluate the quality of point predictions are root-mean-square error (RMSE), mean absolute percentage error (MAPE) and coefficient of determination (R^2). They are defined as follows:

$$\mathrm{RMSE}(y_j, \hat{y}_j) = \sqrt{\frac{1}{N_T}\sum_{j=1}^{N_T}(y_j - \hat{y}_j)^2} \tag{5}$$

$$\mathrm{MAPE}(y_j, \hat{y}_j) = \frac{1}{N_T}\sum_{j=1}^{N_T}\left|\frac{y_j - \hat{y}_j}{y_j}\right| \tag{6}$$

$$R^2(y_j, \hat{y}_j) = 1 - \frac{\sum_{j=1}^{N_T}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{N_T}(y_j - \bar{y})^2} \tag{7}$$

where N_T denotes the number of samples to be evaluated in the test set, and y_j and $\hat{y}_j$ denote the observed cycle life and the predicted cycle life of cell j, respectively. The average cycle life for in total N_T samples in the test set is calculated as $\bar{y} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j$.

Evaluating range prediction quality Likewise, the two commonly used performance evaluation metrics to evaluate the quality of the range predictions are prediction interval coverage probability (PICP) and mean prediction interval width (MPIW) respectively Pearce et al. (2018).

The PICP assesses the calibration property (the degree of reliability) of the range prediction, which refers to the statistical consistency between the constructed prediction interval (PI) and the observations. The PICP is defined as,

$$\text{PICP} = \frac{1}{N_T}\sum_{j=1}^{N_T} c_j, \tag{8}$$

where c_j is a binary variable and its value can be either 0 or 1. If the observed cycle life of cell j in the test set is within the range constructed by the lower bound l_j and the upper bound u_j, then $c_j = 1$; otherwise, $c_j = 0$.

Since a high PICP value can be easily obtained by simply increasing the width of constructed PIs, which does not make much sense for practical battery range prediction, there is a need to assess the sharpness of the range prediction, which refers to the concentration of the predictive distributions Gneiting and Raftery (2007). The MPIW is defined as

$$\text{MPIW} = \frac{1}{N_T}\sum_{j=1}^{N_T} |u_j - l_j|. \tag{9}$$

Ideally, it is desirable to have PIs with a PICP value close to their nominal coverage (i.e., 95%) and a small MPIW value.

3. RESULTS AND DISCUSSION

3.1 Battery cycle life prediction using full 33-feature set

The battery cycle life point prediction and range prediction results for all the models are illustrated in Tab. 2. In terms of battery cycle life point prediction, it can be seen from the results in Tab. 2 that the Elastic Net has the worst performance referring to its highest RMSE, MAPE and lowest R^2. Meanwhile, the GBRT exhibits the best battery cycle life point prediction performance, as it has the lowest RMSE and highest R^2. In terms of battery cycle life range prediction, it can also be seen from the results in Tab. 3 that the QRF has the best performance referring to its lowest MPIW and the lowest AIS. Even though the GPR provides prediction with a bit higher confidence (i.e., lower uncertainty) than the QRGB and QRF by exhibiting a larger PICP value, the GPR has lower MPIW than the QRGB. Therefore, the GPR comes as the second best model for battery cycle life range prediction.

In order to statistically compare both the battery cycle life point prediction and range prediction performance among different models, the performance improvement in percentage is calculated with the best performance model (i.e., GBRT) as the reference. As shown in Tab. 2, the performance improvements of the other 6 models are all negative compared to the GBRT model as the reference for the best point prediction, which indicates their worse point prediction performance than the GBRT. Interestingly, the QRGB that is the quantile regression model of the GBRT is the second best point prediction model. Moreover, the SVR has the worst point prediction performance with the largest performance reduction in percentage. Regarding range prediction performance, it can be seen from the Tab. 3 that the performance improvements of the other 2 models are all negative compared to the QRF model as the reference for the best range prediction model except that the GPR has positive performance improvement relative to the QRF model.

Based on the results, it can be concluded that by using the full 33-feature set, the GBRT model provides the best battery cycle life point prediction performance, while the QRF model provides the best battery cycle life range prediction performance.

3.2 Battery cycle life prediction using MIT 6-feature set

In the original work by Severson et al. Severson et al. (2019), a 6-feature set (i.e., minimum, variance, skewness, and kurtosis of difference of the discharge voltage curve between cycle 100 and cycle 10, discharge capacity at cycle 2, and difference between maximum discharge capacity within the first 100 cycles and discharge capacity at cycle 2) is used to learn a "discharge" model for battery cycle life point prediction, which performs the best on the test set among all 3 feature sets (i.e., 1-feature set, 6-feature set, and 9-feature set Severson et al. (2019)). Therefore, the performance of all learned models using the MIT 6-feature set is compared with that of all learned models using the full 33-feature set.

As for battery cycle life point prediction performance among different models, it can be shown in Tab. 4 that the GPR has the worst performance, with the highest RMSE, and the lowest R^2. Meanwhile, the GBRT consistently exhibits the best point prediction performance and the QRGB is the second best after the GBRT. In terms of battery cycle life range prediction, the results coincide with that of using full 33-feature set, i.e., the QRF model performs the best and the GPR comes as the second best. The performance improvements in comparison with using the full 33-feature set are also calculated for all models. The performance improvement results for point prediction and range prediction are shown in Tab. 4 and Tab. 5 respectively. In terms of point prediction, the performance improvements of all seven models are negative (i.e., worse performance) using the MIT 6-feature set. In terms of range prediction, the performance improvements of the GPR for the three metrics are positive (i.e., better performance) using the MIT 6-feature set. Interestingly, the performance improvements of two quantile models for the MPIW metric are positive (i.e., better performance), but for the PICP metric are negative (i.e., worse performance) using the MIT 6-feature set.

These results show that by using the MIT 6-feature set, the GBRT model consistently provides the best battery cycle life point prediction performance, while the QRF model again provides the best battery cycle life range prediction performance. In comparison with the full 33-feature subset, the MIT 6-feature set helps improves range prediction performance of the GPR measured by all three metrics and also range prediction performance of two quantile regression models (i.e., QRF, and QRGB model) measured by MPIW. In terms of battery cycle life point

Table 2. Battery cycle life point prediction performance using full 33-feature set

ML models	**Point prediction evaluation**			**Performance improvement relative to the best model** (%)		
	RMSE (cycles)	MAPE (%)	R^2	RMSE	MAPE	R^2
Elastic net	183	18.5	0.74	-57.8%	-101.1%	-17.8%
SVR	183	13.4	0.73	-57.8%	-45.7%	-18.9%
RF	148	11.4	0.83	-27.6%	-23.9%	-7.8%
GBRT	116	9.2	0.90	/	/	/
GPR	167	16	0.79	-44.0%	-73.9%	-12.2%
QRF	140	10.2	0.85	-20.7%	-10.9%	-5.6%
QRGB	120	9.1	0.89	-3.4%	1.1%	-1.1%

The GBRT model is referenced as the best model. Negative performance improvement percentages indicate worse performance than the GBRT model.

Table 3. Battery cycle life range prediction performance using full 33-feature set

ML models	**Range prediction evaluation**			**Performance improvement relative to the best model** (%)		
	PICP (%)	MPIW (cycles)	AIS (cycles)	PICP	MPIW	AIS
GPR	94.4	630	1015	0.9%	-32.1%	-61.4%
QRF	93.6	477	629	/	/	/
QRGB	88.8	709	1132	-5.1%	-48.6%	-80.0%

The QRF model is referenced as the best model. Negative performance improvement percentages indicate worse performance than the QRF model.

Table 4. Battery cycle life point prediction performance using MIT 6-feature set

ML models	**Point prediction evaluation**			**Performance improvement on full 33-feature set** (%)		
	RMSE (cycles)	MAPE (%)	R^2	RMSE	MAPE	R^2
Elastic net	190	20.7	0.72	-3.8%	-11.9%	-2.7%
SVR	188	17.9	0.72	-2.7%	-33.6%	-1.4%
RF	148	12.5	0.83	0%	-9.6%	0%
GBRT	134	11.6	0.86	-15.5%	-26.1%	-4.4%
GPR	210	19.2	0.62	-25.7%	-20%	-21.5%
QRF	145	12.0	0.84	-3.6%	-17.6%	-1.2%
QRGB	135	12.4	0.86	-12.5%	-36.3%	-3.4%

The results using full 33-feature set is the reference. Positive performance improvement percentages indicate better performance using MIT 6-feature set.

Table 5. Battery cycle life range prediction performance using MIT 6-feature set

ML models	**Range prediction evaluation**			**Performance improvement on full 33-feature set** (%)		
	PICP (%)	MPIW (cycles)	AIS (cycles)	PICP	MPIW	AIS
GPR	95.2	623	734	0.8%	1.1%	27.7%
QRF	92.8	455	651	-0.9%	4.6%	-3.5%
QRGB	84.8	563	1163	-4.5%	20.6%	-2.7%

The results using full 33-feature set is the reference. Positive performance improvement percentages indicate better performance using MIT 6-feature set.

prediction, the GBRT model performs the best among all the models thanks to its superior abilities, such as capturing nonlinear relationship, less prone to overfitting, robust against outliers. In terms of battery cycle life range prediction, the reason why the QRF model performs arguably better than GPR and QRGB model may be because of the sensitivity of the kernel selection for the GPR, and the robustness of the QRF model against noise.

4. CONCLUSION

Accurate and reliable early prediction of battery lifetime is important for optimizing life cycle management of batteries from cradle to grave. In the present work, a pipeline-based approach was proposed for automating the process of producing the best model for both battery cycle life point prediction and range prediction. It has been demonstrated that our proposed machine learning pipeline can adapt to different realistic charging protocols in practice. The experimental results illustrated that the gradient boosting regression tree (GBRT) provided the best point prediction performance while the quantile regression forest (QRF) provided the best range prediction performance using both full 33-feature set and the MIT 6-feature set. However, it is important to note that the prediction performance of these models also varies with the quality of the battery data and the input feature set. Therefore, that well motivates the use of our proposed machine learning pipeline with the aim of consistently producing the best model for battery cycle life prediction. In terms of future work, it would be interesting to explore two aspects, i.e., to improve the adaptability of our proposed machine learning pipeline to different battery design, chemistry and usage, a new feature set that is independent of the aforementioned

factors is needed; the effectiveness of the proposed pipeline for battery lifetime prediction in the field would be the second meaningful but challenging future work to explore.

ACKNOWLEDGEMENTS

The authors would like to thank Volvo Group and Swedish Energy Agency for funding this work with project 45540-1.

REFERENCES

Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795), 397–402.

Baumhöfer, T., Brühl, M., Rothgang, S., and Sauer, D.U. (2014). Production caused variation in capacity aging trend and correlation to initial cell performance. *Journal of Power Sources*, 247, 332–338.

Chang, Y., Fang, H., and Zhang, Y. (2017). A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. *Applied energy*, 206, 1564–1578.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477), 359–378.

Grolleau, S., Delaille, A., Gualous, H., Gyan, P., Revel, R., Bernard, J., Redondo-Iglesias, E., Peter, J., and Network, S. (2014). Calendar aging of commercial graphite/lifepo4 cell–predicting capacity fade under time dependent storage conditions. *Journal of Power Sources*, 255, 450–458.

Han, X., Lu, L., Zheng, Y., Feng, X., Li, Z., Li, J., and Ouyang, M. (2019). A review on the key issues of the lithium ion battery degradation among the whole life cycle. *ETransportation*, 1, 100005.

Hu, C., Ye, H., Jain, G., and Schmidt, C. (2018). Remaining useful life assessment of lithium-ion batteries in implantable medical devices. *Journal of Power Sources*, 375, 118–130.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Kotsiantis, S.B., Kanellopoulos, D., and Pintelas, P.E. (2006). Data preprocessing for supervised leaning. *International journal of computer science*, 1(2), 111–117.

Lian, Y., Wang, J.V., Deng, X., Kang, J., Zhu, G., and Xiang, K. (2020). Remaining useful life prediction of lithium-ion batteries using semi-empirical model and bat-based particle filter. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5. IEEE.

Ma, G., Zhang, Y., Cheng, C., Zhou, B., Hu, P., and Yuan, Y. (2019). Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network. *Applied Energy*, 253, 113626.

Meinshausen, N. and Ridgeway, G. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(6).

Miao, Q., Xie, L., Cui, H., Liang, W., and Pecht, M. (2013). Remaining useful life prediction of lithium-ion battery with unscented particle filter technique. *Microelectronics Reliability*, 53(6), 805–810.

Pearce, T., Brintrup, A., Zaki, M., and Neely, A. (2018). High-quality prediction intervals for deep learning: A distribution-free, ensembled approach. In *International conference on machine learning*, 4075–4084. PMLR.

Qin, T., Zeng, S., and Guo, J. (2015). Robust prognostics for state of health estimation of lithium-ion batteries based on an improved pso–svr model. *Microelectronics Reliability*, 55(9-10), 1280–1284.

Raj, T., Wang, A.A., Monroe, C.W., and Howey, D.A. (2020). Investigation of path-dependent degradation in lithium-ion batteries. *Batteries & Supercaps*, 3(12), 1377–1385.

Reitermanova, Z. et al. (2010). Data splitting. In *WDS*, volume 10, 31–36.

Richardson, R.R., Osborne, M.A., and Howey, D.A. (2019). Battery health prediction under generalized conditions using a gaussian process transition model. *Journal of Energy Storage*, 23, 320–328.

Sadabadi, K.K., Jin, X., and Rizzoni, G. (2021). Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. *Journal of Power Sources*, 481, 228861.

Schmuch, R., Wagner, R., Hörpel, G., Placke, T., and Winter, M. (2018). Performance and cost of materials for lithium-based rechargeable automotive batteries. *Nature Energy*, 3(4), 267–278.

Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5), 383–391.

Vetter, J., Novák, P., Wagner, M.R., Veit, C., Möller, K.C., Besenhard, J., Winter, M., Wohlfahrt-Mehrens, M., Vogler, C., and Hammouche, A. (2005). Ageing mechanisms in lithium-ion batteries. *Journal of power sources*, 147(1-2), 269–281.

Voronov, S., Frisk, E., and Krysander, M. (2018). Data-driven battery lifetime prediction and confidence estimation for heavy-duty trucks. *IEEE Transactions on Reliability*, 67(2), 623–639.

Wang, S., Guo, D., Han, X., Lu, L., Sun, K., Li, W., Sauer, D.U., and Ouyang, M. (2020). Impact of battery degradation models on energy management of a grid-connected dc microgrid. *Energy*, 207, 118228.

Williams, C.K. and Rasmussen, C.E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

Yang, F., Wang, D., Xu, F., Huang, Z., and Tsui, K.L. (2020). Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model. *Journal of Power Sources*, 476, 228654.

Zhang, Y., Xiong, R., He, H., and Pecht, M.G. (2018). Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Transactions on Vehicular Technology*, 67(7), 5695–5705.

Zhang, Y., Peng, Z., Guan, Y., and Wu, L. (2021). Prognostics of battery cycle life in the early-cycle stage based on hybrid model. *Energy*, 221, 119901.