A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data

Yizhou Zhang [a,b], Torsten Wik [a], John Bergström [b], Michael Pecht [c], Changfu Zou [a,*]

[a] *Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden*
[b] *China Euro Vehicle Technology AB, Gothenburg, 41755, Sweden*
[c] *Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD, USA*

HIGHLIGHTS

- A comprehensive feature pool based on the battery usage histogram data is formed.
- A systematic machine learning framework based on histogram data is introduced.
- A comparison of the four most commonly used algorithms on three datasets is shown.
- An individualisation process is developed during the online deployment.
- A real-world PHEV fleet is used to verify the efficacy of the proposed framework.

GRAPHICAL ABSTRACT
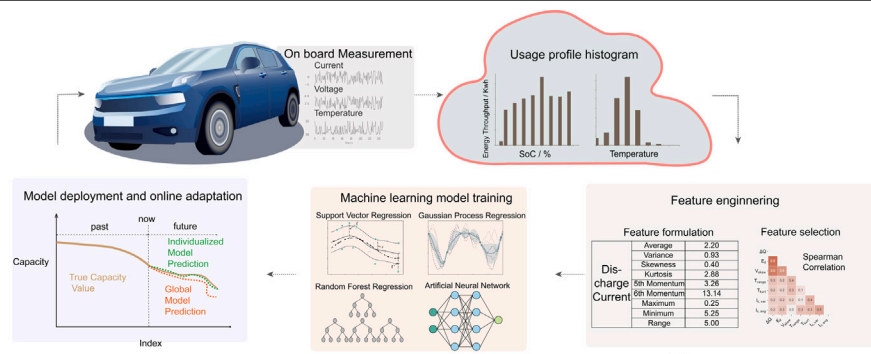
ARTICLE INFO

ABSTRACT

Accurately predicting batteries' ageing trajectory and remaining useful life is not only required to ensure safe and reliable operation of electric vehicles (EVs) but is also the fundamental step towards health-conscious use and residual value assessment of the battery. The non-linearity, wide range of operating conditions, and cell to cell variations make battery health prediction challenging. This paper proposes a prediction framework that is based on a combination of global models offline developed by different machine learning methods and cell individualised models that are online adapted. For any format of raw data collected under diverse operating conditions, statistic properties of histograms can be still extracted and used as features to learn battery ageing. Our framework is trained and tested on three large datasets, one being retrieved from 7296 plug-in hybrid EVs. While the best global models achieve 0.93% mean absolute percentage error (MAPE) on laboratory data and 1.41% MAPE on the real-world fleet data, the adaptation algorithm further reduced the errors by up to 13.7%, all requiring low computational power and memory. Overall, this work proves the feasibility and benefits of using histogram data and also highlights how online adaptation can be used to improve predictions.

1. Introduction

Current transport heavily relies on fossil fuels and has caused serious public concerns about air quality and global warming [1]. In the foreseeable future, the most promising way to address the problem and thus fulfil a carbon-neutral vision is by electrification of vehicle powertrains combined with the use of renewable energy [2]. Batteries,

* Corresponding author.
E-mail addresses: yizhou@chalmers.se (Y. Zhang), tw@chalmers.se (T. Wik), john.bergstrom@cevt.se (J. Bergström), pecht@umd.edu (M. Pecht), changfu.zou@chalmers.se (C. Zou).

today normally lithium-ion batteries, possess high energy density and reasonable cycle life, thus playing a crucial role in driving the electric revolution [3–5]. However, batteries are still the most expensive vehicle component, and their production consumes considerable energy and metals, such as lithium, cobalt, and nickel—all of which, unfortunately, are limited and problematic [6]. Consequently, no matter from the perspectives of energy and resource efficiency or economy, it is vital to prolong battery life. While major progress in chemistry and materials may occur in the long run [7], there are also significant advances to make by optimal battery utilisation, such as health-conscious management, timely maintenance, and judicious second-life applications, for which accurate knowledge of the future ageing behaviour and remaining useful life (RUL) is essential [8–13].

There have been many research attempts to predict the future state of health (SoH) for lithium-ion batteries, which were initially mainly dedicated to the development of empirical [14–17], semi-empirical [18], or physical-based ageing models [19–22]. As reviewed in [23–25], the empirical/semi-empirical models were designed to fit the degradation curves using a predefined parametric function and have shown good performance for laboratory data generated under well-controlled conditions. However, extrapolating these models to complex and time-varying real-world battery usage will inevitably lose guarantees of accuracy and can result in large deviations in the prediction. The physics-based modelling has achieved excellent progress in capturing battery dynamics in a fast timescale [26–28], but their application to online prediction of battery ageing is very challenging. The obstacles stem from two main causes. First, battery degradation results from diverse, interlaced, and nonlinear degradation mechanisms [29–31]. There is no single model to describe all the mechanisms precisely, nor an effective way to quantify them separately. Second, electrochemical states evolve nearly unobservable by the existing sensing technology and involve a set of state-dependent parameters which are difficult to calibrate [32].

With the recent emergence of big data in the battery community and the penetration of artificial intelligence, a growing body of data-driven models have been developed to predict the battery SoH profile or RUL. For example, Bloom et al. [33] have applied incremental capacity analysis, Dubarry et al. [34] differential voltage analysis, and Shibagaki et al. [35] differential thermal voltammetry to battery degradation prognosis. The related health indicators have been widely adopted in battery ageing prognosis as summarised in [10]. However, the repeatable and specific operating profiles required by these models are impractical to achieve in many real-world applications, such as hybrid electric vehicles and stationary energy storage systems. For more data-driven models using different features and machine learning methods, see [36–39] and reference therein. The models were generally designed based on time series data. Their size for a large-scale battery system over hundreds of cycles can easily reach gigabytes or even terabytes, and it is unrealistic to save them in onboard electronic control units (ECUs) [11]. An alternative way is to transmit the data from vehicles to data centres, which enables cloud battery management with the battery digital twin [40,41]. The cloud system's high computation capability and enormous storage space render advanced high-performance algorithms for smart battery usage and life prognostic possible. However, frequent time series data collection and remote transmission of a large amount of data are not only confronted with issues, such as data latency, loss, mismatch, or breach [42], but are also expensive and energy-consuming. For the vehicle on-board data storage, battery usage data are widely stored as histograms to save power and memory. This strongly solicits reliable modelling frameworks and methodologies dealing with histogram data directly. A few studies on laboratory data have recently formulated features using histogram-like data for modelling battery health and lifetime. Richardson et al. [43] applied capacity throughput and time duration as the input, while Greenbank et al. [44] mainly used calendar time and time spent in specific voltage regions as inputs. These methods have good generalisability, but with a limited number of features, their prediction capability can be restricted, especially when dealing with the richness of field data. They also lack a systematic pipeline to learn battery degradation from various histogram data of electric vehicle usage.

The existing data-driven models were commonly trained offline and later deployed to online usage. This means that when battery cells start with the same capacity and then operate under the same cycling conditions, the predicted ageing trajectories for them will always be the same. Unfortunately, this will contradict battery ageing characteristics in the real world due to the inherent cell inconsistency [45,46]. Furthermore, from an economic standpoint, only a very limited number of sensors are deployed in traction battery systems [47,48]. These sensors may not be able to capture the local differences between battery cells with respect to the selected features, resulting in major prediction errors. Accordingly, it is critical to develop individualised models for battery ageing prediction that are online adaptive to cell variations stemming from manufacturing deviations and unbalanced operating conditions. Several attempts have been made in the literature to address such a problem. Li et al. [49] proposed a recurrent neural network-based sequence-to-sequence model to predict the future capacity trajectory in one shot and used historical capacity measurements as the input. Hu et al. [37] extracted several health indicators from multi-stage constant current charging curves to estimate battery SoH in a battery module with four series-connected cells. However, this class of methods requires repeatable cyclic charging or discharging profiles, which contradict the inherently stochastic battery usage in the real world. Moreover, these offline trained models were implemented directly and without online adaption. As a result, valuable information of the considered battery cells, e.g., historical SoH values, cannot be incorporated to further improve the prediction model.

In this work, we propose a histogram data-based framework for online adaptive prediction of battery ageing trajectory and lifetime under diverse operating conditions. For the first step, we derive a general procedure for feature construction involving a two-step data compression and using a range of statistical properties of the histogram data. From the constructed comprehensive feature pool, a feature dependence check-and-control scheme is then designed to select the most relevant and independent features. Based on the selected features, different machine learning methods are used to offline build global models. After evaluating their performance, the most suitable global model is selected and then intelligently adapted online for cell individualised prediction. The framework is illustrated by applying it to three different types of data; periodic accelerated ageing test data, laboratory data emulating real-world profiles, and data from a fleet of plug-in hybrid electric vehicles (PHEVs). Examples of four types of machine learning algorithms with different characteristics are then employed; kernel-based, decision tree-based, probabilistic based, and neural network-based. The results show that the framework is effective on all three datasets and the chosen machine learning methods.

2. Dataset

A myriad of lithium-ion battery datasets are available in the literature, as summarised in [50], and are being collected in laboratories and the ever-growing uptake of electric vehicles (EVs). This work selected three different types of data, as illustrated in Fig. 1, to develop battery ageing prediction algorithms and evaluate their effectiveness. The first type, represented by the Stanford University battery dataset, was collected from 169 cylindrical lithium-iron-phosphate (LFP) cells under periodic fast charging/discharging profiles and a well-controlled laboratory environment [9,51]. The second dataset was generated from tests of 25 cylindrical lithium-cobalt-oxide (LCO) cells at NASA and was emulating random real-world battery usage [52]. The third one was measured directly from the traction battery systems that were installed
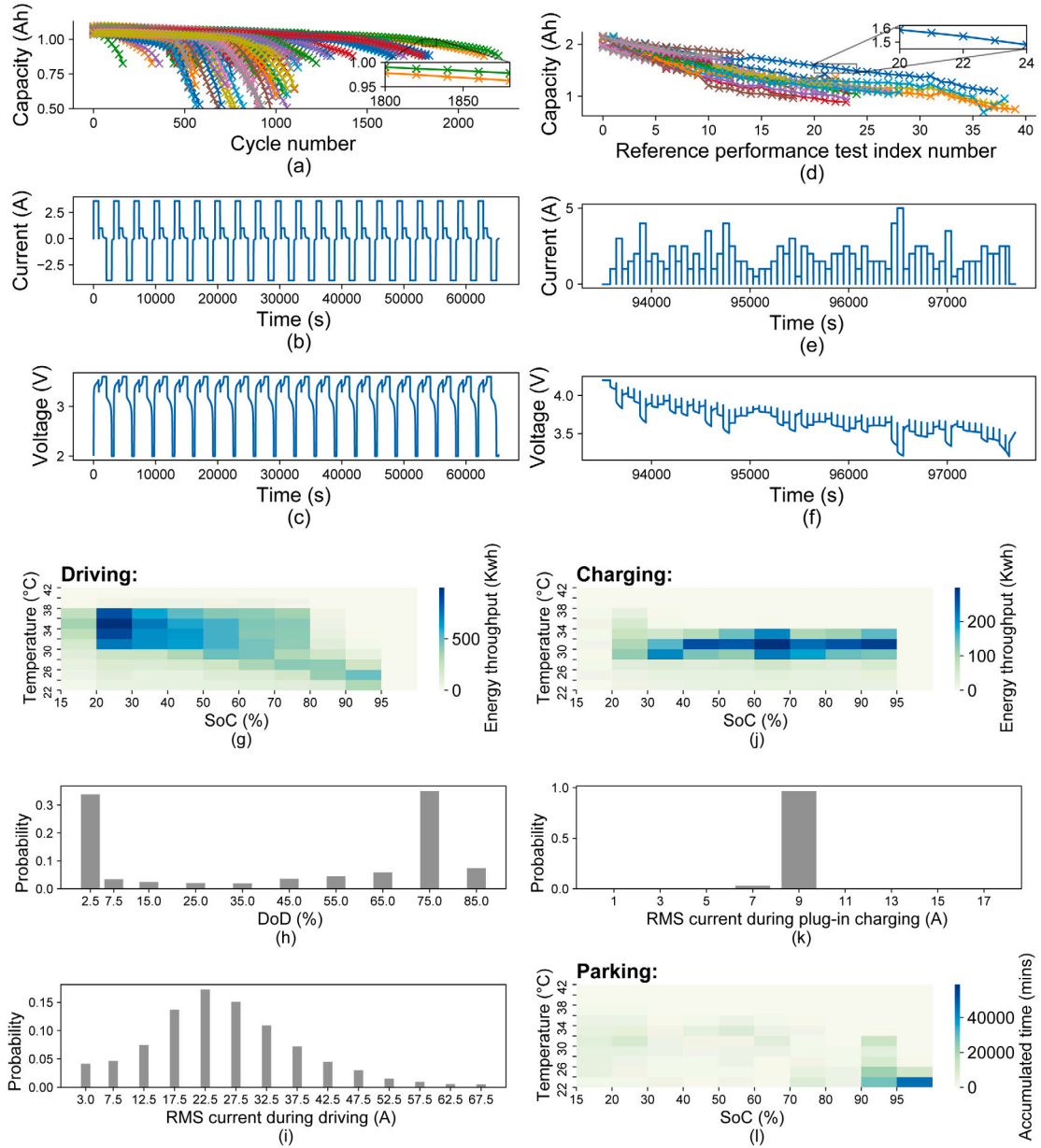
Fig. 1. Illustration of three battery datasets used for algorithm development, validation, and tests. For the Stanford dataset, (a) shows the capacity trajectories of 169 cells, and (b)–(c) are the current and voltage profiles over 20 cycles. (d), (e), and (f) exemplify battery operating profiles in the NASA dataset. (g)–(l) show the usage information of a typical vehicle in the fleet dataset, where (g) and (j) depict the accumulated energy throughput in predefined SoC and temperature window with 5% and 2 °C intervals, respectively. Similarly, (l) represents the accumulated parking time in predefined SoC and temperature windows. (h) shows the probability distribution of the DoD usage frequency. (k) and (i) represent the 3 min RMS current during the plug-in charging and driving modes, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in 7296 PHEVs, each consisting of 90 pouch nickel–manganese–cobalt (NMC) cells. The vehicle fleet dataset is much larger than the two laboratory ones, and to the best of our knowledge, it is the largest real-world PHEV dataset in the literature for battery ageing studies. While the laboratory datasets were stored in time series, the vehicle fleet data was in the form of one- or multiple-dimensional histograms. The selected data sources of high dissimilarity may lead to different features extracted to indicate battery health, though the underlying principle of feature engineering and the pipeline of algorithm development will remain the same. The purpose of combining these representative and complementary datasets is to comprehensively examine the accuracy and applicability of the proposed prognosis framework under generalised operating conditions.

2.1. Data collection and processing

Labelled output. The system output of interest, y, has been defined as the capacity change at a specific time interval. With the time interval as a fixed number of full charge and discharge cycles for the two laboratory datasets, and as the time between two adjacent vehicle visits at workshops for the fleet dataset, the output y can be well labelled for supervised learning. When one knows the present SoH and the future trajectory of y, it is straightforward to predict the battery ageing and RUL.

Stanford dataset. This dataset contains two batches of data, both from experimental tests on A123 cylindrical cells that have LFP as the positive electrode and graphite as the negative electrode. The first batch

was measured from 124 cells cycling under different one- or two-step fast charging policies [9], while the second batch consisted of 45 cells undergoing a four-step fast charging policy [51]. All these cells were discharged at a constant current of 4C rate. The current and voltage profiles of 20 cycles are presented in Fig. 1(b) and (c), respectively. Under the same thermal condition, with a fixed ambient temperature of 30 °C, the cycling tests started with new battery cells and stopped when their capacities degraded to 80% of the nominal value, whereas for the latter, there were certain exceptions. Given that no reference capacity checking was undertaken in the original tests, we calculated the capacity values from the measured discharging current curves and used them as labelled data to develop a framework for battery ageing prognosis. Specifically, for every 20 cycles, the capacity value was derived by integrating the current over time in the last cycle. These capacity 'measurements', plotted as crossing marks in Fig. 1(a) with cells distinguished by colour, are indexed in chronological order.

NASA random walk dataset. At the NASA Ames Prognostic Centre of Excellence Randomised Battery Usage Repository, this dataset was collected from LG Chem cylindrical 18 650 battery cells with LCO positive electrodes, graphite negative electrodes, and a nominal capacity of 2.1 Ah [52]. In total, 28 cells were used in this test campaign, and they were divided into seven subgroups. Each cell was continuously charged to 4.2 V and discharged to 3.2 V under a randomised sequence of currents. As an example, the current and voltage profiles for discharging of cell 13 are depicted in Fig. 1(e) and (f), respectively. For this study, 25 out of the 28 cells were selected, discarding cells 2, 3, and 18 because of their problematic data, such as temperature measurements as low as −4000 °C. Each group underwent different current–voltage cycling profiles to mimic real-world battery usage. The ambient temperature though was maintained at 40 °C for cells in groups 6 and 7 and was 20 °C for all the remaining cells. Fig. 1(d) shows the capacity degradation profiles of all the selected cells, where the crossing marks indicate the reference capacity measurements. Similarly to the Stanford dataset, the sequence of these reference performance test cycles are ordered chronologically and treated as capacity measurement data samples.

Vehicle fleet dataset. The third dataset is composed of aftermarket battery diagnostic data from a running fleet with 7296 PHEVs. The data was initially stored in the onboard battery management system (BMS) and then downloaded to the data centre whenever a vehicle visited service/maintenance workshops. The battery data extracted from the vehicle fleet database can be categorised into four groups: driving-related data, charging-related data, parking-related data, and static data. Fig. 1(g)–(l) present the original data of a vehicle randomly chosen from the fleet. In the driving-related data, the pack-level accumulated energy throughput at different temperatures and state of charge (SoC) intervals was saved in a 2D histogram format, while the depth of discharge (DoD) and three-minute root mean square current of the battery pack were in a 1D histogram format. In each pack, there were ten temperature sensors installed, and their average value was used as the reference for data saving. Meanwhile, the pack SoC was recorded to formulate the histogram. Similarly, the charging-related data recorded a 2D histogram in terms of the accumulated energy throughput, the average charging current, and the charging time. The parking-related data included the accumulated parking time at different temperatures and SoC intervals. The static data kept the SoH of individual battery cells, the battery elapsed time from when the vehicle was produced, the activation time of BMS balancing function, and the maximum capacity difference among battery cells of the PHEVs. Therein, the SoH had been defined as the ratio of the actual battery capacity to the nominal capacity and was saved when performing data readouts in the vehicle workshop. In this study, the capacity values calculated from the recorded SoH trajectory were used as the reference of battery ageing.

Data split. For the Stanford dataset, 40 cells were randomly chosen as the test set, and the remaining 129 cells served as the training set.

For the NASA dataset, as the operation of each group was performed uniquely, we randomly selected one cell from each group to form the test set and kept the rest of the cells in the training set. For the fleet dataset, 20% of the vehicles were randomly chosen as the test set, and the rest of the vehicles formed the training set. To obtain fair and robust evaluation results, the above procedure was repeated ten times for its corresponding dataset. For the two lab datasets, the cell indices for each train and test split are detailed in Supplementary Tables 4–5.

3. Methods

3.1. Histogram data-based feature construction

Although the battery ageing process is complex, the stress factors contributing to the capacity fade are the same [53]. As per [54–59], the following factors play a significant role in degrading the battery capacity: DoD, charge current rate, discharge current rate, temperature, voltage, accumulated cycling/calendar time, accumulated ampere-hour (Ah) throughput, and SoC. These common stress factors are then used to construct an initial feature pool in a two-step procedure. First, the raw data, which can be in a time series or histograms of any dimension, will be transformed into 1D histograms. With an interval of current of 0.5 A, Fig. 2(a)–(c) demonstrate the transformation process and results for time series laboratory data from the NASA dataset. Analogously, (f)–(h) present the transformation process and results for a 2D histogram from the fleet dataset. The second step is to extract and calculate the statistical properties of the constructed 1D histograms resulting from the first step. Fig. 2(d), (e), (i), and (j) list a part of the statistical properties and their calculated values. The complete feature lists for different datasets and the mathematical definitions are provided in Supplementary Tables 6–9.

The above feature extraction procedure essentially performs two-step data compression to enable histogram data-based modelling. It can significantly reduce the number of data points, thus saving computational power, energy, and memory space for data management, transmission, and storage, respectively. Furthermore, the final data then match the format used in vehicles (as described in Section 2.1). The use of the statistical properties, instead of the direct histograms, may better reveal the correlation between battery usage and capacity degradation. Note that not all the stress factors are suitable for constructing features for the battery ageing prognosis. For example, certain factors, such as the DoD and mean SoC, may be identical for all battery cells in a dataset (e.g., Stanford and NASA datasets), even though the cells exhibit different ageing behaviours. It is noteworthy that this procedure of feature extraction is sufficiently general, irrespective of the battery sizes, shapes, and types of chemistry.

3.2. Feature engineering

With the constructed feature pool, feature engineering is the next step for the development of machine learning methods. In principle, the selected feature set should render the associated methods highly accurate and robust at the expense of reasonable computation. To achieve this, a series of analyses are first conducted for the correlation between each feature and the system output, i.e., the change of battery capacity at a defined time interval and the correlation among different features. While the former correlation analysis paves the way for determining the most relevant features, the latter can be used to avoid selecting strongly interdependent features. Specifically, the Spearman correlation analysis is applied to measure the strength and direction of monotonic association between two variables. The Spearman correlation analysis [60] is adopted to all features in the pool relative to the capacity change, y. The absolute values of Spearman's coefficients are used for feature ranking. Each feature x_i is then assigned a score, $C_{x_i \rightarrow y}$ (a value between 0 and 1), based on the absolute value of their correlation coefficients with respect to y and ranked based on its score. The features with scores
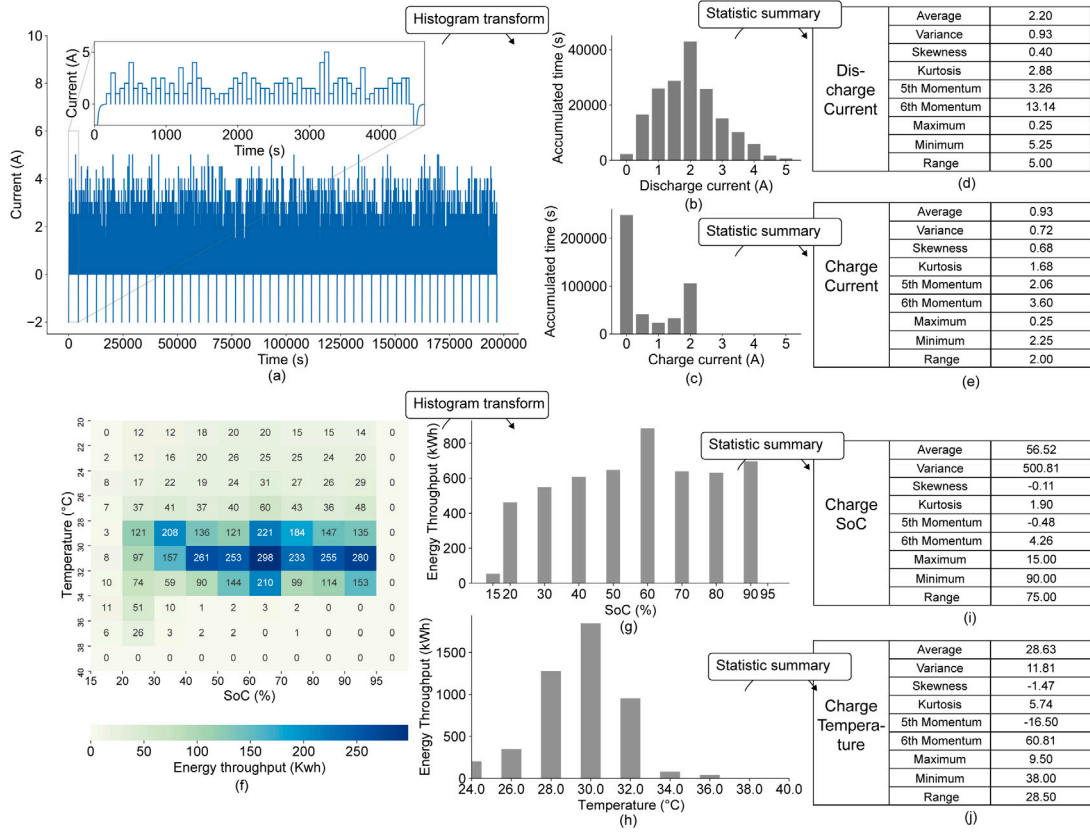
Fig. 2. Illustration of the feature engineering process. (a), (b), and (c) show the transformation process of the NASA dataset from the original time series to histogram. (d) and (e) include a part of the statistical properties of the constructed histogram. (f), (g), and (h) show the transformation process of the vehicle fleet dataset from its original 2D histogram to a 1D histogram. (i) and (j) are the statistical properties of the constructed histogram.

below a low threshold corr_l are removed from the pool. In case some features in the remaining pool are heavily related to each other, it is not necessary to use all of them. By again applying the Spearman correlation analysis, a feature dependence check-and-control scheme is introduced to identify and discard the redundant features. With corr_h denoting an upper threshold of the inter-feature correlation and the following scheme is applied (starts with $i = 1$):

1. In the remaining pool, select the feature with the highest score relative to the capacity change and denote it as x_i.
2. Apply the Spearman correlation analysis to calculate the correlation scores of all the other features x_j $(j > i)$ relative to x_i and denote them as $C_{x_j \to x_i}$.
3. Compare the obtained correlation scores with corr_h and remove the feature x_i from the pool whenever $C_{x_j \to x_i} > \text{corr}_h$.
4. Increase the index from i to $i + 1$ and repeat from 1 again.

For the three considered datasets, we choose corr_l as 0.2 and corr_h as 0.8. After implementing the above scheme, if there are still too many features, random forest regression (RFR)-based repeated k-fold cross-validation [61] can be used to further reduce the number of features. The basic idea is to implement an RFR algorithm with gradually added features and then evaluate the accuracy of the corresponding models on the training data; when the model accuracy cannot be appreciably improved by adding more features, then it is advisable not to add those features.

3.3. Global model and algorithm development

The task of battery ageing prognosis is formulated as a regression problem within the framework of supervised machine learning. The

overall pipeline to accomplish the task is summarised in Fig. 3, which includes an offline path for the global model development and an online path for model adaptation along with the streaming data.

The global models are developed only from the offline training dataset involving a number of battery cells of the same type. From these cells, each model essentially tries to learn the averaged ageing behaviour in response to the selected features. The model development process includes hyperparameter tuning, method selection, model evaluation, and online deployment. With the labelled input–output data samples from Section 2, it is observed that the battery ageing trajectories are highly nonlinear. Among a large toolbox of machine learning methods for nonlinear model regression, support vector regression (SVR), RFR, Gaussian process regression (GPR), and artificial neural network (ANN), as illustrated in Fig. 4, are very powerful and have been widely used in different applications. Although these methods have some disadvantages, they also have advantages and may be good candidates to solve the battery prognostic problem. They are all applied to estimate the model function from the corresponding dataset. The goal is to minimise the average deviations between the measured outputs and model predictions over all data points in the test set.

SVR. As per [62], SVR utilises a nonlinear mapping function $\Phi(\cdot)$ to transform the data from a low-dimensional space in terms of x_p to a high-dimensional space in terms of $\Phi(x_p)$, and after the transformation, the function to be estimated, $f(\cdot)$, becomes linear in $\Phi(\cdot)$, namely $f(x_p) = w^T \Phi(x_p) + b$. In Fig. 4(a), there is a tube with the radius ϵ around a hyper-plane $f(\Phi(x_p))$, and the goal is to let as many training samples as possible fall into the tube and to keep the hyper-plane as flat as possible. The modelling problem can be formulated as a constrained
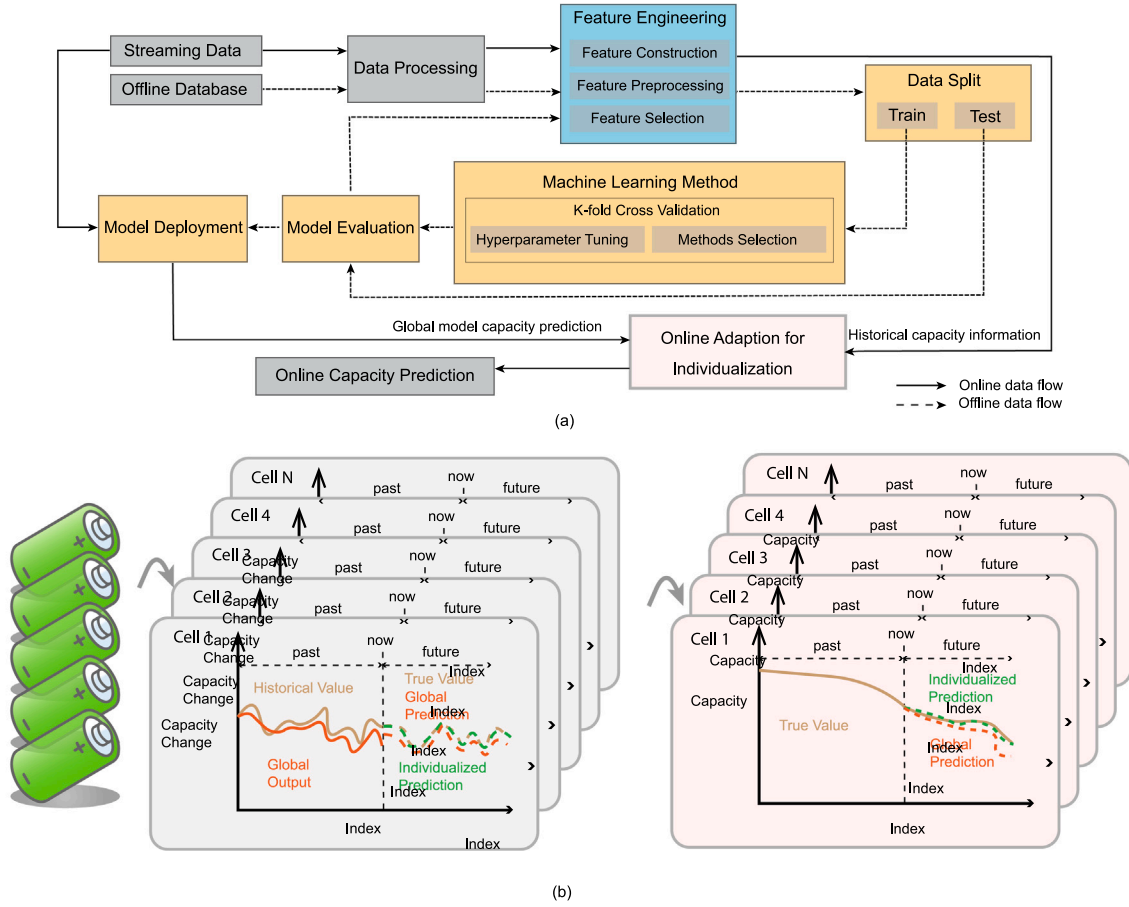
(a)

(b)

Fig. 3. Pipeline to develop data-driven algorithms for battery ageing prognosis. (a) summarises all the required modules and their connections. (b) zooms in on the online adaptation module, where both the global model-based predictions and the individual cell's historical information are utilised.

linear regression, [63]

$$\min_{w,b,\xi,\xi^*} \quad \frac{1}{2}\|w\|^2 + C \sum_{p}^{P} (\xi_p + \xi_p^*)$$

$$\text{subject to} \quad \begin{cases} y_p - w^T \Phi(x_p) - b \le \epsilon + \xi_p \\ w^T \Phi(x_p) + b - y_p \le \epsilon + \xi_p^* \\ \xi_p, \xi_p^* \ge 0 \end{cases} \quad (1)$$

where ξ_p and ξ_p^* are positive slack variables, C is the regularisation coefficient, and ϵ is the error tolerance coefficient. Both C and ϵ are hyperparameters to be determined using cross-validation. To avoid explicitly computing $\Phi(x_p)$, the kernel trick can be used, e.g., the radial basis function (RBF) kernel that has been widely used and is implemented in this work.

Due to the use of support vectors, SVR has the built-in characteristic of sparsity and can also effectively handle sparse data. Furthermore, enabled by data transformation and the linear regression formulation in (1), SVR can fit various nonlinear systems accurately and robustly. However, the required computational time increases substantially with the number of data samples and is a significant concern when deploying SVR on large-scale battery problems.

RFR. With the structure illustrated in Fig. 4(c), RFR is a supervised ensemble learning method consisting of two core steps, i.e., "*bootstrap*" and data *agg*regation, whose combination is called "bagging" [64]. The bootstrap technique is applied in the sample collection process, in which each of the uncorrelated trees randomly selects a subset of training data and features to perform prediction. The samples which

are not selected are called out-of-bag samples and can be used for validation or performing feature importance analysis. After obtaining the result from each tree, a final prediction is made by aggregating the individual outputs of all the trees. The total number of decision trees and the maximum depth of each individual tree are two important hyperparameters to trade-off the model performance, computational resources, and training time.

RFR leverages the "bagging" technique to gain a much better bias–variance balance than single decision tree methods, particularly in the presence of missing data and outliers. Additionally, RFR is easy to train and relatively interpretable compared to SVR and ANN. However, the high modelling accuracy entails a large number of trees and levels within each tree, requiring a substantial amount of computational power and resulting in a long training period.

GPR. Instead of representing the training data using a predefined, parametric function as in ANN and SVR, the estimated function, $f(\cdot)$, for GPR is distributed according to a Gaussian process (GP) given by [65]

$$f(x) \sim \text{GP}\left(m(x), \kappa(x, x')\right) \quad (2)$$

where x and x' are two arbitrary data samples, $m(x)$ represents the mean value of $f(x)$, and $\kappa(x, x')$ is the covariance of $f(\cdot)$ between the points x and x'. Both the mean and covariance functions can incorporate prior knowledge about the shape of $f(\cdot)$. By defining $\mathbf{X}$ as a vector of all the P samples, $\mathbf{X} = [x_1, \ldots, x_P]$, and then conditioning f on the training dataset $(\mathbf{X}, f(\mathbf{X}))$, the posterior probability of the predicted output $f(x^*)$ at any point x^* in the test set or from real-world prediction

applications [65], we have

$$p\left(f(x^*)|x^*, \mathbf{X}, f(\mathbf{X})\right) \sim \mathcal{N}\left(f(x^*); m^*, \Sigma^*\right) \tag{3}$$

$$m^* = \mathcal{K}(\mathbf{X}, x^*)^T \mathcal{K}(\mathbf{X}, x^*)^{-1} f(\mathbf{X}) \tag{4}$$

$$\Sigma^* = \mathcal{K}(\mathbf{X}, x^*) - \mathcal{K}(\mathbf{X}, x^*)^T \mathcal{K}(\mathbf{X}, x^*)^{-1} \mathcal{K}(\mathbf{X}, x^*) \tag{5}$$

where $\mathcal{K}$ is the covariance kernel matrix having κ as elements. A zero mean, i.e., $m(x) = 0$, works well in most practical cases and is adopted in this work. Then $\kappa(x, x')$ indicates the similarity between the samples x and x' and will heavily impact the prediction result. Three different kernel functions, including RBF kernel, Matérn kernel, and rational quadratic kernel [66], will be used for the training of battery models, and the best function will be selected for online applications.

When only a small amount of data is available, GPR is expected to be superior to the other three methods in the prediction accuracy and can even learn from new samples to extend the operating window of the designed model. Moreover, different from tree frequentist-based methods, GPR possesses probabilistic characteristics, through which confidence intervals can be derived for model-based predictions, making it very attractive to safety-critical prognostic problems. It is noteworthy that due to the need for matrix inversion in (4) and (5), when the number of data samples is large, GPR becomes computationally very expensive.

ANN. ANN is a modelling technique conceptually inspired by the human brain's cognitive process to learn from data and find the best model mapping from the input to the output. With the structure sketched in Fig. 4(d), a typical ANN has multiple layers, each further consisting of a number of units. For a unit $j \in \{1, \dots, J_l\}$ in an arbitrary hidden layer $l \in \{0, 1, \dots, L\}$, the mathematics behind the propagation from one unit to another can be generally formulated as

$$z_{j,l} = \sum_{j=1}^{l_J} w_{j,l} h_{j,l-1} + b_{j,l} \tag{6}$$

$$h_{j,l} = a_l(z_{j,l}) \tag{7}$$

where w, b, and a are the weight of the unit, the bias factor, and the activation function, respectively. Explicit forms of the activation function and the loss function to minimise have been well described in [67] and are not repeated here for brevity. With all the weights and biases as decision variables, gradient descent-based algorithms are often applied to solve the optimisation problem.

ANN has a remarkable ability to detect patterns and identify trends from complex and highly nonlinear systems with complicated or imprecise data. Furthermore, its excellent performance in handling big data and the flexibility to accommodate parallel computation make ANN a preferred choice for many applications, such as image classification and voice recognition. ANN also has some downfalls, including a high required tuning effort, risk of overfitting, and relatively long training time.

3.4. Individualised model and algorithm development

When predicting the capacity change using a global model, only the future model inputs (features) of the considered cell are employed, and the prediction executed is, in essence, an open-loop model-based simulation. More precisely, the predictor is blind to any cell-specific ageing behaviour, not even if a cell deviates completely from the others. The historical capacity profile of a cell may certainly contain valuable information for learning its future ageing peculiarity. With this in mind, we will derive an individualised prediction model for each battery cell by directly adjusting the results of the global model.

The adjustment factor changes with time and is determined online based on the difference between the historically predicted output trajectory by the global model and the measured one from the cell under consideration. When the cell is relatively new, only a small amount

of capacity information is available, based on which the obtained individualised prediction may become biased. Moreover, the capacity measurements are inevitably polluted by noise and/or disturbances. Then, the research problem essentially becomes how much we should trust the individualised model. To address this problem, we introduce a weighting factor to optimally balance the global and the individualised predictions. Two optimisation problems are formulated to solve the adjustment and weighting factors, respectively. By defining the output of the global model for cell $n \in \{1, \dots, N\}$ as $y_{\text{global, n}} = f(x_n)$, a time-varying factor λ_n is introduced to online adjust the global prediction giving an individualised prediction for the capacity change at time t on the form

$$y_{\text{indiv},n,t} = \lambda_{n,t} y_{\text{global},n,t} = \lambda_{n,t} f(x_{n,t}). \tag{8}$$

Based on the employed global model and the historical capacity values of cell n, the optimal value of $\lambda_{n,k}$ can be derived by minimising the difference between the measured outputs and the adjusted global prediction values over the entire history. This optimisation problem is posed as

$$\lambda_{n,k}^* = \underset{\lambda_{n,k}}{\text{argmin}} \quad \frac{1}{k} \sum_{t=1}^{k} \alpha^{k-t} \left\| y_{n,t} - \lambda_{n,k} f(x_{n,t}) \right\|^2 \tag{9}$$

where α is a forgetting factor to exponentially decrease the importance of old measurements. The recursive least squares (RLS) estimation algorithm (see, e.g., [68]) can be applied to solve the optimisation problem and obtain the correction factor $\lambda_{n,k}^*$, sequentially at each time step k.

When the cell is relatively new, only a small amount of historical capacity information is available, based on which the individualised model for online adaptive prediction may be biased because of the inherent measurement noise and potential disturbances. To suppress the effect of this, we further introduce a weighting factor $w_{k,t}$ to trade-off the global and individualised predictions, i.e.,

$$y_{\text{indiv},n,t}^w = (1 - w_{k,t}) y_{\text{global},n,t} + w_{k,t} y_{\text{indiv},n,t}, \quad t \in \{k, \dots, k_{\text{end}}\} \tag{10}$$

where k is the present time step when the prediction starts and k_{end} is the final time step in the prediction horizon. To reduce the number of variables for online applications, all battery cells are supposed to have the same weighting factors at given k and t. The optimal weighting factor $w_{k,t}^*$ can be found by minimising the prediction error. As stated earlier, RUL plays a crucial role in battery management (e.g., for reliability evaluation, maintenance, and replacement). It is thus of great importance to predict RUL accurately and, therefore, we choose to optimise $w_{k,t}$ by minimising the RUL prediction error for all battery cells, i.e.,

$$\min_{w_{k,k}, \dots, w_{k,k_{\text{end}}}} \sum_{n=1}^{N} \left(Q_{n,k} + \sum_{t=k}^{k_{\text{end}}} \hat{y}_{\text{indiv},n,t}^w - Q_{n,k_{\text{end}}} \right)^2 + \beta \sum_{t=k}^{k_{\text{end}}} w_{k,t}^2 \tag{11}$$

where β is a regularisation coefficient, and the first two terms in the bracket represent the predicted capacity in the final time step. By iterating k from 1 to k_{end}, the optimal weighting factors for each prediction starting point k can be calculated by solving the above regularised linear regression. At each time step k, a fixed individualisation factor is then assumed for the prediction over $[k, \dots, k_{\text{end}}]$, namely $\lambda_{n,t}^* = \lambda_{n,k}^*$.

Finally, by combining (8), (10), the online optimised factor $\lambda_{n,k}^*$ and the offline optimised weight $w_{k,t}^*$, the online adaptive algorithm predicts the capacity change trajectory according to

$$y_{\text{indiv},n,t}^* = (1 - w_{k,t}^*) f(x_{n,t}) + w_{k,t}^* \lambda_{n,k}^* f(x_{n,t}). \tag{12}$$

The complete workflow is illustrated in Fig. 5. Overall, this proposed adaptation algorithm operates in a closed loop that not only takes into account the ageing characteristics of battery cells in the database but also explicitly includes the considered cell's property and operating conditions. As a result, issues like cell variations, measurement noise, and disturbances are systematically handled.
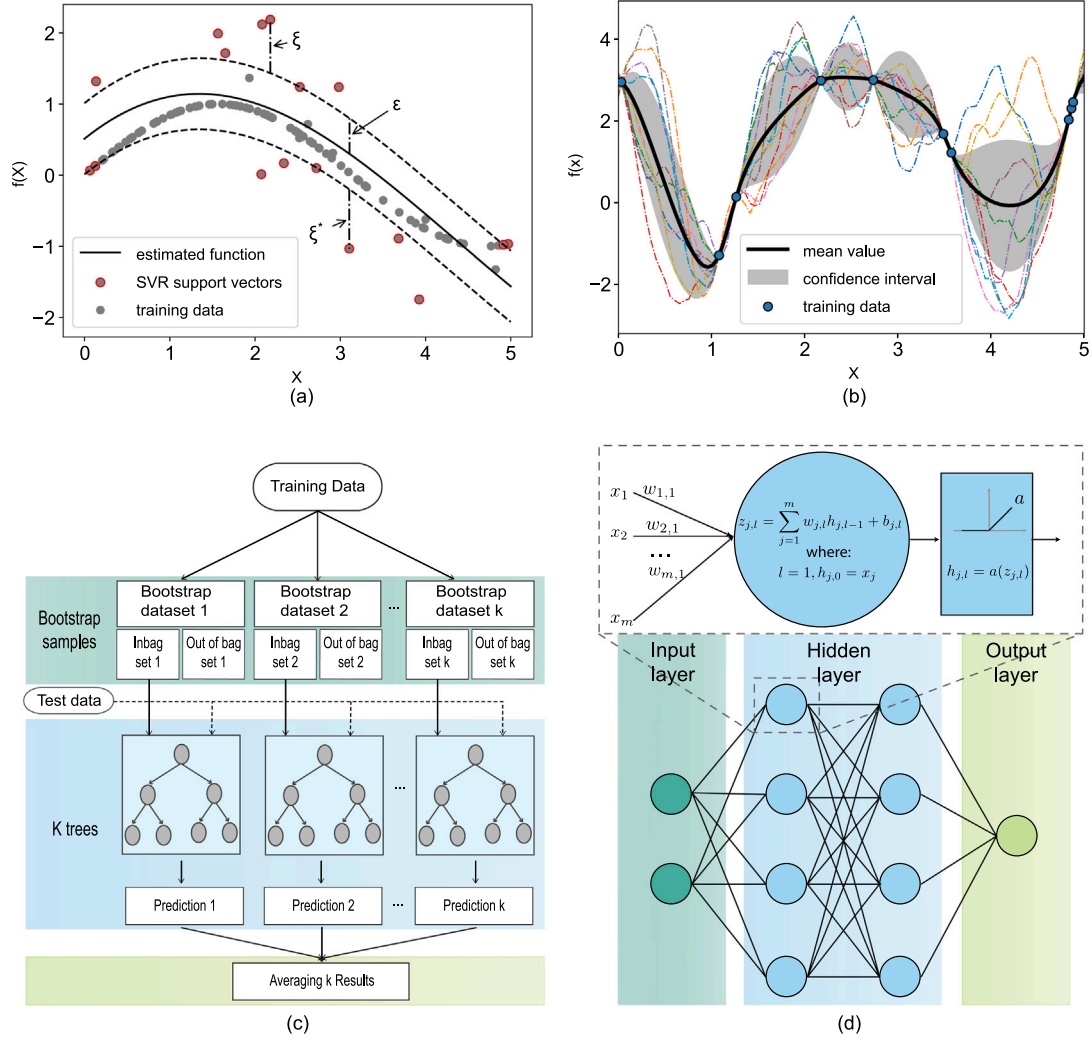
Fig. 4. Illustration of four powerful and popular machine learning methods: (a) support vector regression, (b) Gaussian process regression, (c) random forest regression, and (d) artificial neural network. (a) and (b) assume a single feature case to ease the demonstration.
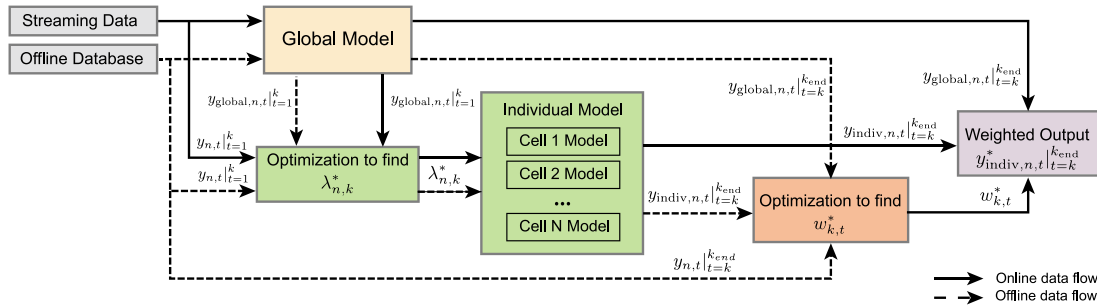

Fig. 5. Workflow of the proposed online adaptation algorithm for battery capacity prediction. The individualised model makes use of the predicted result $y_{\text{global},n,t}$ from a global model, the measured signal $y_{n,t}$, the optimised adjustment factor λ^*, and the optimal weighting factor $w^*_{k,t}$.

3.5. Evaluation metrics for model fidelity

Two different evaluation metrics are used to quantify the performance of the proposed prognosis algorithms. The first one is the root mean squared percentage error (RMSPE) between the measured capacity Q and the predicted capacity $\hat{Q}$ over all $\mathcal{P}_n$ data points for cell n in the test set, and is mathematically defined as

$$\text{RMSPE}_Q = \sqrt{\sum_{n=1}^{N} \left(\frac{1}{\mathcal{P}_n} \sum_{p=1}^{\mathcal{P}_n} \left(\frac{Q_{n,p} - \hat{Q}_{n,p}}{Q_{n,p}} \right)^2 \right)} \tag{13}$$

where, without loss of generality, we assume the initial capacity is known and $\hat{Q}_{n,1} = Q_{n,1}$. The subsequent capacity estimates are calculated recursively as $\hat{Q}_{n,p+1} = \hat{Q}_{n,p} + \hat{y}_{\text{global},n,p+1}$. In contrast to RMSPE,

which penalises larger deviations more heavily, the second evaluation metric is the MAPE of the predicted capacity trajectory, and it emphasises the main trend of the predictions

$$\text{MAPE}_Q = \sum_{n=1}^{N} \left(\frac{1}{\mathcal{P}_n} \sum_{p=1}^{\mathcal{P}_n} \frac{\left| Q_{n,p} - \hat{Q}_{n,p} \right|}{Q_{n,p}} \right). \tag{14}$$

To evaluate the performance of online adaptation for cell n, the error throughout the prediction range $\{k + 1, \ldots, \mathcal{P}_n\}$ is taken into account, where k is the time step where the prediction starts. With this in mind, RMSPE defined in (13) is slightly modified as

$$\text{RMSPE}_Q^k = \sqrt{\sum_{n=1}^{N} \left(\frac{1}{\mathcal{P}_n - k} \sum_{p=k+1}^{\mathcal{P}_n} \left(\frac{Q_{n,p} - \hat{Q}_{n,p}}{Q_{n,p}} \right)^2 \right)}. \tag{15}$$

3.6. Machine learning software and libraries

All the presented results were obtained in Python 3.8.5 with an Intel i7 CPU and 32 GB RAM. Publicly available libraries were used for the training and testing of machine learning methods. Among others, scikit-learn [69] was adopted for SVR, RFR, and GPR, while ANN was implemented using Keras with Tensorflow as the backend [70].

4. Results and discussion

4.1. Results of feature engineering

The selected features and their corresponding Spearman correlation matrix for each training dataset are presented in Fig. 6(a)–(c). The most suitable feature numbers are unique for the considered three datasets, and their best feature sets have few overlaps. Except for the feature of the temperature range (i.e., T_{range}) overlapping between the Stanford and NASA datasets, and the time duration (i.e., $\varDelta t$) overlapping between the NASA and vehicle fleet datasets, all the other features picked up for these datasets are different. The reasons behind this can be explained from several aspects. Intuitively, the reason is that the initial feature pools for the datasets differ from each other because of their different data characteristics. For example, the statistical properties of DoD and SoC are adopted as features for the fleet dataset, but they are hardly distinguishable among the well-defined cycles in the two lab datasets, and are accordingly not included in their corresponding feature pools. Physically, the battery systems targeted in each dataset have experienced distinctively different cycling conditions and usage behaviours, and could consequently exhibit different ageing features because of that. When more randomness is introduced into the cycling profiles, battery systems generally tends to exhibit more ageing features. This is well supported by the fact that the Stanford dataset has the fewest features, while the fleet dataset has the most. Finally, data size may also influence the selection of features. When there are more independent data samples, a relatively large feature set may be used to achieve a better bias–variance trade-off. This is consistent with observations in this study. For the fleet dataset, which is the biggest, a large set of 25 features was initially derived from the correlation analysis and the feature dependence check-and-control scheme. Then, the RFR-based repeated k-fold cross-validation method was applied to further reduce the number of features, with the result demonstrated in Fig. 6(d), which indicates that it would be judicious to select 15 features for machine learning. Complete lists of the investigated and selected features for the three datasets are given in the Supplementary materials (Tables 6, 7, and 8).

4.2. Results of battery ageing prognosis

Periodic and accelerated ageing test. For its well-controlled operating profiles and environment with high-precision measurements, the Stanford dataset serves as the first choice for model and algorithm validation. To test the global models, the predictors always stand at the first data sample, while for online usage, the prediction starting points are arbitrary and depend on the considered scenario. Under such circumstances, we investigate the adaptation algorithm at a wide range of starting points. The results are illustrated in Fig. 7, where (a)–(e) cover the global model-based predictions, and (f)–(g) compare the global and individualised models. The numerical accuracy for the global models to predict the entire ageing trajectory is presented in Table 1.

It can be seen from Table 1 that the mean absolute percentage error (MAPE) of all four machine learning methods is less than 1.7%, and their root mean squared percentage error (RMSPE) is less than 3.3%. Furthermore, the vast majority of the predicted capacity values for all the 40 LFP cells in the test set fall within ±5% error bounds, as demonstrated in Fig. 7(a)–(e). These results validate that the proposed global models are all able to offer a reliable prediction of the lifelong capacity profile for the battery cells that they have never seen before. At the same time, both our constructed histogram-based features and the proposed feature engineering method have been verified to be effective for a range of different types of machine learning methods, represented by SVR, RFR, GPR, and ANN. Specifically, RFR and ANN outperform their alternatives in terms of MAPE and achieve an error of 0.93% and 1.13%, respectively. Such smallness of prediction errors and strong resilience over the whole capacity range make the proposed models very competitive to the prevalent models developed directly upon time series data, e.g., [9,71].

When we look into the details, the predicted points in Fig. 7(b) and (d) are located in the bottom triangle slightly more often than in the top triangle, and the orange and red lines in Fig. 7(e) show lower values in the right-hand side of the positive half-plane. This implies that ANN and RFR tend to less often overpredict the true capacity, which is beneficial to traction battery applications in which safety and reliability are of paramount importance. One advantage of GPR is the provided confidence intervals for its predictions. The two-sigma confidence interval has been highlighted as grey areas in Supplementary Fig. 1(c). However, the covered areas are too large to guide our prediction task. This is potentially due to high variations among different battery cells and because the uncertainties are quite large when applying a global model to predict a specific cell online. One example is cell 'b2c47' in the Stanford dataset that was aged extremely slowly compared to other cells. Its capacity trajectory has been emphasised as dark grey in Fig. 7(a)–(d). Not surprisingly, all the developed global models fail to capture its behaviour and result in an unacceptably large prediction error. This echoes our earlier statement and emphasises the importance of online adaptive prediction.

The prediction performance of the individualised models, with RFR as an example, is depicted in Fig. 7(f). It can be calculated that the individualised model is capable of reducing the prediction error by 13.7% in the best case and by 8.6% on average. These improvements are significant in battery ageing and lifetime prognosis, particularly considering their subsequent applications to health optimisation and lifetime extension of a large number of battery cells. Aside from the case of handling all the testing cells in a batch, we also examine the performance of online adaptation for individual cells. It can be seen from Fig. 7(f)–(h) that for both the abnormal cell 'b2c47' and a normal cell 'b2c35', the individualised model effectively learns from the historical ageing information and continuously adjusts the global predictions along the ageing trajectories to approach the ground truth, leading to more accurate and robust predictions.
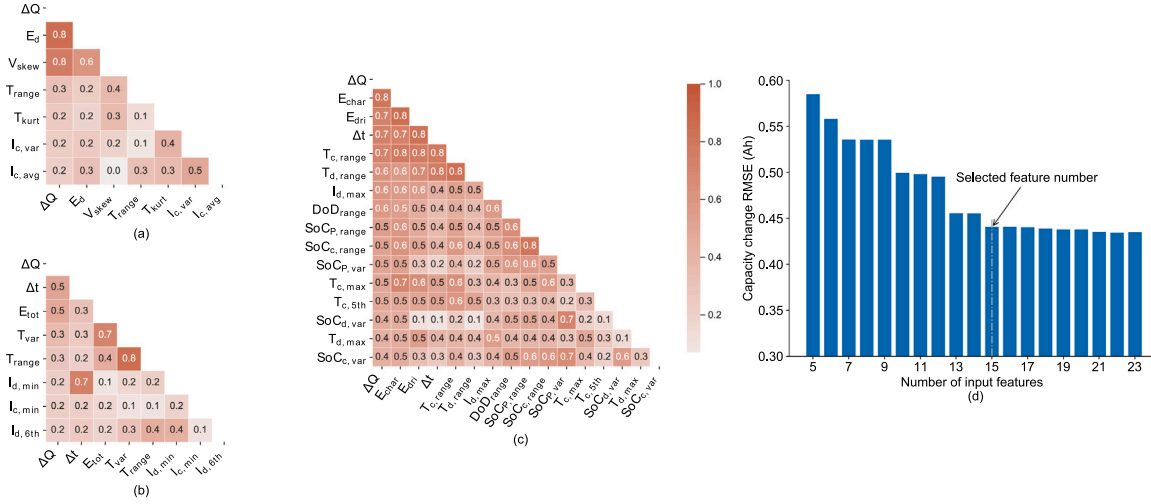
Fig. 6. Feature selection results and Spearman correlation matrices for the considered three training sets. (a)–(c) correspond to the Stanford, NASA, and vehicle fleet datasets, respectively. (d) presents the results of RFR-based repeated k-fold cross-validation for the features obtained for the fleet data by Spearman correlation analysis and the proposed feature dependence check-and-control scheme.
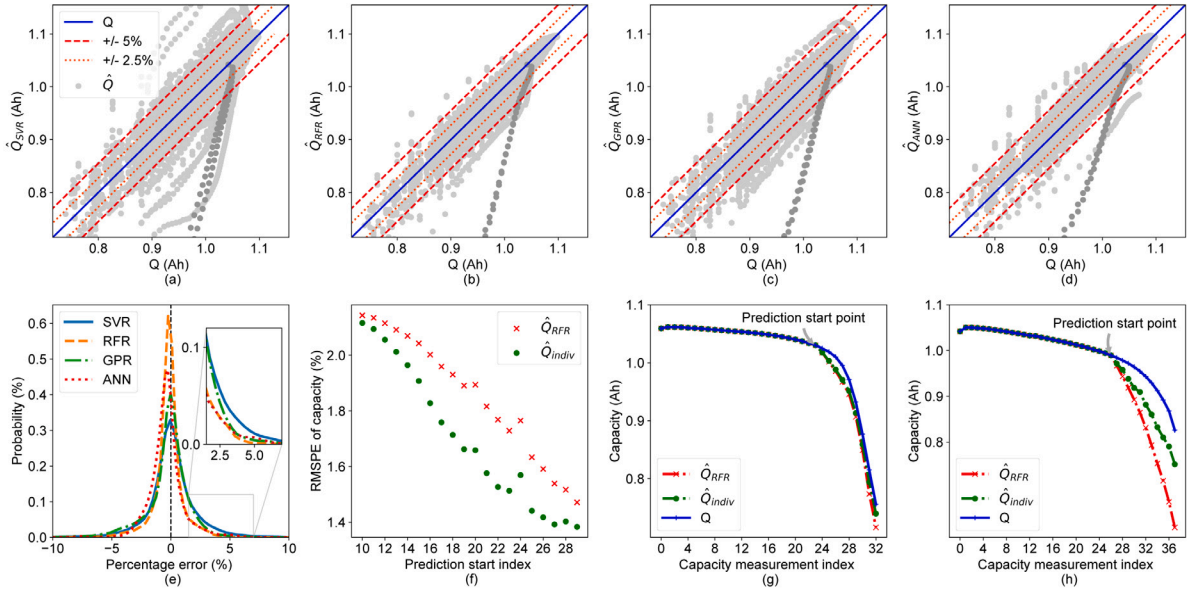

Fig. 7. Validation of the developed global models and online adaptation algorithm using the Stanford dataset. (a)–(d) show the predicted capacity by SVR, RFR, GPR, and ANN, respectively, versus the measured capacity Q. The orange dotted lines are the bounds of $\pm2.5\%$ prediction errors, and the red dashed lines correspond to $\pm5\%$ error bounds. One specific cell, 'b2c47', shows an abnormal long lifetime compared to others with similar cycling conditions and is highlighted in dark grey. (e) presents the percentage error histogram of the four machine learning methods to predict the capacity trajectory, in which the predictors stand at the first data sample. (f) shows RMSPE_Q^k of (15) using the RFR-based global model and the individualised model. (g) represents the prediction results for a randomly selected cell, 'b4c36', and (h) for the abnormal cell 'b2c47'. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
MAPE and RMSPE of the predicted capacity change by four machine learning methods.

Algorithms	MAPE (%)			RMSPE (%)		
	Stanford	NASA	Vehicle fleet	Stanford	NASA	Vehicle fleet
SVR	1.61	4.41	1.46	3.22	6.33	2.18
RFR	0.93	4.21	1.43	2.12	5.76	2.16
GPR	1.35	3.23	1.44	2.58	4.45	2.11
ANN	1.13	3.64	1.41	1.92	5.08	2.12

Vehicle driving schedule test. With proven effectiveness against the Stanford dataset, we further assess the designed algorithms for battery cells covered in the NASA and vehicle fleet datasets. The calibrated capacity profiles for the NASA dataset occasionally have some local peaks where the measurements may have drifted from the actual capacity. For the fleet data, the capacity measurements were generated by onboard ECUs, and their accuracy is unknown. In such cases, we assume all the measurements as the ground truth, though this will to some degree affect the numerical results, which is a common situation for studies on real-world battery data. With this in mind, the obtained results are presented in Table 1 and Fig. 8.

For the fleet dataset, the global models based on different machine learning methods all have nearly the same prediction errors, which are around 1.45% in MAPE and 2.15% in RMSPE. For the NASA dataset, although the prediction performance cannot match those obtained with the other two datasets, the MAPE of 3.23% can still satisfy many industrial applications that typically require the error to be less than 5%.
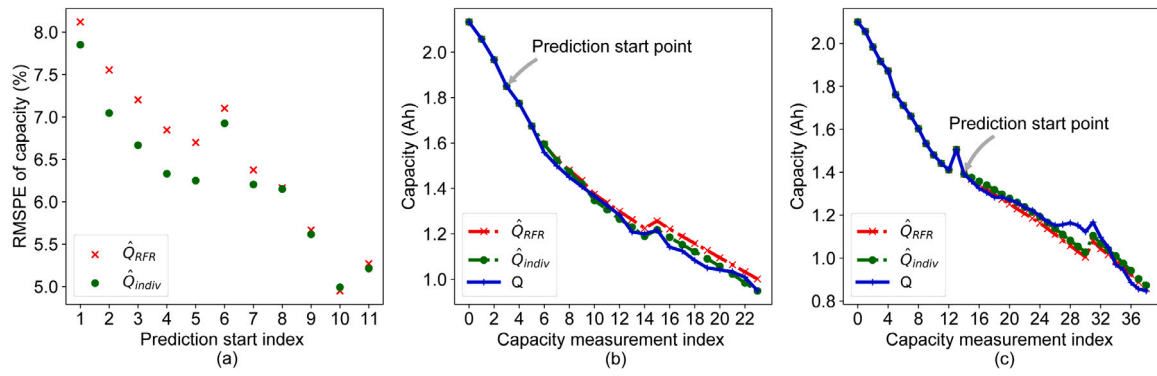
Fig. 8. Comparison of the RFR-based global model and the individual model for predicting battery capacity of the NASA dataset. (a) presents RMSPE_Q^k for each prediction start index k. (b) and (c) show the predicted and measured capacity trajectory of cell No. 15 and No. 10, respectively.

These results further corroborate the effectiveness and practicability of the histogram-based models for battery ageing prognosis. For the NASA dataset having the fewest data, GPR outperforms all its alternatives in both MAPE and RMSPE, which aligns well with GPR's advantages discussed in Section 3.3. On the other hand, for each method, the prediction accuracy for battery cells installed in the vehicle fleet is much higher than that for the NASA ones. This is mainly attributed to the substantially increased number of features and data samples in the vehicle fleet dataset. The numerical relationship between model accuracy and data size is referred to the learning curve presented in Supplementary Fig. 2. These results confirm the importance of having a sufficiently large dataset.

The online adaptation algorithm is also tested on the NASA dataset. Its efficacy in enhancing model accuracy and robustness in the presence of cell variations is again verified. Specifically, as demonstrated in Fig. 8, the individualised model is able to decrease the prediction errors at almost every prediction start index, with a maximum reduction of 7.5% in RMSPE. Similar to the results of Fig. 7(g)–(h), the individualised model follows the measured capacity profile more closely than the global model. Moreover, this is achieved no matter whether the global model overpredicts or underpredicts the measured values, and the performance is maintained even though battery degradation exceeds 50% of the nominal capacity. Additional results are demonstrated in Supplementary Figs. 1–6.

4.3. Requirements for computation and memory

The online prediction algorithm associated with different machine learning methods was sequentially implemented on a typical laptop computer with specifications, software, and machine learning libraries introduced in Section 3.6. The hyperparameters for the employed machine learning methods are provided in Supplementary Tables 10–12. To minimise the stochasticity in quantifying the computational time, with each machine learning method the algorithm has been tested ten times under the same settings, and then the average time has been recorded. Table 2 summarises the time spent in training and testing. It can be first observed that the time to train the most demanding machine learning method associated with the largest dataset and the most features was less than 1150 s. In this regard, we can conclude that the time required for offline model development is very low and thus can be readily satisfied. Second, the online prediction algorithm embedded with any of the developed global models accomplished the task at each sample for all cells in the corresponding dataset within 3.2 s. By using the most suitable machine learning method for online usage, the computational efficiency can be improved by 15–200 times, depending on the specified dataset, feature number, and hyperparameters. Along with the batteries' age, the prediction horizon decreases, and the computational time can be further shortened. In either case, the required time is negligibly short compared to the timescale of battery

Table 2
Computational time and data size to train and implement the battery ageing predictor.

Methods	Training time (s)[a]			Testing time (µs)[b]		
	NASA	Stanford	Vehicle fleet	NASA	Stanford	Vehicle fleet
SVR	0.01	2.4	6.9	18.9	144.5	593.0
RFR	0.2	1.1	29.5	70.9	10.5	108.3
GPR	1.4	291.7	1140.1	53.5	1933.5	1073.0
ANN	2.3	15.1	11.1	282.2	37.0	40.0

[a]The average time in seconds (s) to run the predictor on the complete training set.
[b]The average time in microseconds (µs) to run the predictor on one data sample in the test set.

degradation and the time interval between two adjacent data samples, which are several weeks or even months for real-world vehicle applications. It is notable that the high prediction performance presented in Section 4.2 was achieved at very small data. With up to 10,000 data samples for training, the needed memory resources were extremely limited and could be easily covered by today's cloud system or even a desktop computer. This is a great advantage of the proposed histogram-based algorithm compared to its time series counterparts that are either oversized for onboard ECUs to store or plagued by a range of problems in data quality, security, cost, and energy consumption for transmission to data centres.

5. Conclusions

Data-driven modelling has been widely cited as the most promising approach for battery ageing prediction, but existing models commonly require specific operating profiles, a substantial amount of time series data and lack online adaptability. To address these problems, we have introduced a novel machine learning-based prediction framework. The novelties first come from using features based on data in histograms instead of time series, thereby significantly saving computational power and memory and allowing predictions under generalised operating conditions. Second, the framework is equipped with an online model adaptation algorithm that systematically handles cell variations, measurement noise, and disturbances.

This paper explored four widely adopted machine learning methods for the offline development of global models and evaluated the framework against three large datasets measured from batteries of different sizes, types of chemistry, and usage profiles. This framework has been verified to be effective for each method and dataset, and the computational time to predict the ageing trajectory of all batteries in the corresponding dataset is less than 3.2 s. Specifically, the best global models achieved 0.93% test error on laboratory data and 1.41% test error on the real-world fleet data, and the online algorithm further reduced the errors by up to 13.7%. Overall, this work proves the feasibility and benefits of using histogram data and highlights the importance of online adaptation for data-driven modelling of battery prognostics.

[34] M. Dubarry, B.Y. Liaw, Identify capacity fading mechanism in a commercial LiFePO4 cell, J. Power Sources 194 (1) (2009) 541–549, http://dx.doi.org/10.1016/j.jpowsour.2009.05.036, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775309009604.

[35] T. Shibagaki, Y. Merla, G.J. Offer, Tracking degradation in lithium iron phosphate batteries using differential thermal voltammetry, J. Power Sources 374 (2018) 188–195, http://dx.doi.org/10.1016/j.jpowsour.2017.11.011, URL https://www.sciencedirect.com/science/article/pii/S0378775317314702.

[36] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, D.U. Sauer, Online capacity estimation of lithium-ion batteries with deep long short-term memory networks, J. Power Sources 482 (2021) 228863, http://dx.doi.org/10.1016/j.jpowsour.2020.228863, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775320311678.

[37] X. Hu, Y. Che, X. Lin, Z. Deng, Health prognosis for electric vehicle battery packs: A data-driven approach, IEEE/ASME Trans. Mechatronics (2020) 1, http://dx.doi.org/10.1109/TMECH.2020.2986364, URL https://ieeexplore.ieee.org/document/9064898/.

[38] P. Tagade, K.S. Hariharan, S. Ramachandran, A. Khandelwal, A. Naha, S.M. Kolake, S.H. Han, Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis, J. Power Sources 445 (2020) 227281, http://dx.doi.org/10.1016/j.jpowsour.2019.227281, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775319312741.

[39] X. Li, C. Yuan, Z. Wang, Multi-time-scale framework for prognostic health condition of lithium battery using modified Gaussian process regression and nonlinear regression, J. Power Sources 467 (2020) 228358, http://dx.doi.org/10.1016/j.jpowsour.2020.228358, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775320306625.

[40] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte, D.U. Sauer, Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation, J. Energy Storage 30 (2020) 101557, http://dx.doi.org/10.1016/j.est.2020.101557, URL https://www.sciencedirect.com/science/article/pii/S2352152X20308495.

[41] K. Li, P. Zhou, Y. Lu, X. Han, X. Li, Y. Zheng, Battery life estimation based on cloud data for electric vehicles, J. Power Sources 468 (2020) 228192, http://dx.doi.org/10.1016/j.jpowsour.2020.228192, URL https://www.sciencedirect.com/science/article/pii/S037877532030495X.

[42] N. Subramanian, A. Jeyaraj, Recent security challenges in cloud computing, Comput. Electr. Eng. 71 (2018) 28–42, http://dx.doi.org/10.1016/j.compeleceng.2018.06.006, URL https://www.sciencedirect.com/science/article/pii/S0045790617320724.

[43] R.R. Richardson, M.A. Osborne, D.A. Howey, Battery health prediction under generalized conditions using a Gaussian process transition model, J. Energy Storage 23 (2019) 320–328, http://dx.doi.org/10.1016/j.est.2019.03.022, URL http://www.sciencedirect.com/science/article/pii/S2352152X18307734.

[44] S. Greenbank, D. Howey, Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life, IEEE Trans. Ind. Inf. (2021) 1, http://dx.doi.org/10.1109/TII.2021.3106593, URL https://ieeexplore.ieee.org/document/9520291/.

[45] S.J. Harris, D.J. Harris, C. Li, Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells, J. Power Sources 342 (2017) 589–597, http://dx.doi.org/10.1016/j.jpowsour.2016.12.083, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775316317864.

[46] T. Baumhöfer, M. Brühl, S. Rothgang, D.U. Sauer, Production caused variation in capacity aging trend and correlation to initial cell performance, J. Power Sources 247 (2014) 332–338, http://dx.doi.org/10.1016/j.jpowsour.2013.08.108, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775313014584.

[47] X. Lin, H.E. Perez, J.B. Siegel, A.G. Stefanopoulou, Robust estimation of battery system temperature distribution under sparse sensing and uncertainty, IEEE Trans. Control Syst. Technol. 28 (3) (2019) 753–765.

[48] A. Allam, E. Catenaro, S. Onori, Pushing the envelope in battery estimation algorithms, Iscience 23 (12) (2020) 101847.

[49] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, D.U. Sauer, One-shot battery degradation trajectory prediction with deep learning, J. Power Sources (2021) 230024, http://dx.doi.org/10.1016/j.jpowsour.2021.230024, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775321005528.

[50] G. dos Reis, C. Strange, M. Yadav, S. Li, Lithium-ion battery data and where to find it, Energy and AI (2021) 100081, http://dx.doi.org/10.1016/j.egyai.2021.100081, URL https://linkinghub.elsevier.com/retrieve/pii/S2666546821000355.

[51] 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, P.K. Herring, M. Aykol, S.J. Harris, R.D. Braatz, S. Ermon, W.C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature 578 (7795) (2020) 397–402, http://dx.doi.org/10.1038/s41586-020-1994-5, URL http://www.nature.com/articles/s41586-020-1994-5.

[52] B. Bole, C. Kulkarni, M. Daigle, Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use, in: PHM 2014 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014, 2014.

[53] M. Woody, M. Arbabzadeh, G.M. Lewis, G.A. Keoleian, A. Stefanopoulou, Strategies to limit degradation and maximize Li-ion battery service lifetime - Critical review and guidance for stakeholders, J. Energy Storage 28 (2020) 101231, http://dx.doi.org/10.1016/j.est.2020.101231, URL https://linkinghub.elsevier.com/retrieve/pii/S2352152X19314227.

[54] C.R. Birkl, M.R. Roberts, E. McTurk, P.G. Bruce, D.A. Howey, Degradation diagnostics for lithium ion cells, J. Power Sources 341 (2017) 373–386, http://dx.doi.org/10.1016/j.jpowsour.2016.12.011, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775316316998.

[55] G. Ning, B. Haran, B.N. Popov, Capacity fade study of lithium-ion batteries cycled at high discharge rates, J. Power Sources 117 (1–2) (2003) 160–169, http://dx.doi.org/10.1016/S0378-7753(03)00029-6, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775303000296.

[56] T. Guan, S. Sun, F. Yu, Y. Gao, P. Fan, P. Zuo, C. Du, G. Yin, The degradation of LiCoO2/graphite batteries at different rates, Electrochim. Acta 279 (2018) 204–212, http://dx.doi.org/10.1016/j.electacta.2018.04.197, URL https://linkinghub.elsevier.com/retrieve/pii/S0013468618309800.

[57] T. Guan, P. Zuo, S. Sun, C. Du, L. Zhang, Y. Cui, L. Yang, Y. Gao, G. Yin, F. Wang, Degradation mechanism of LiCoO2/mesocarbon microbeads battery based on accelerated aging tests, J. Power Sources 268 (2014) 816–823, http://dx.doi.org/10.1016/j.jpowsour.2014.06.113, URL https://linkinghub.elsevier.com/retrieve/pii/S037877531400977X.

[58] D. Wong, B. Shrestha, D.A. Wetz, J.M. Heinzel, Impact of high rate discharge on the aging of lithium nickel cobalt aluminum oxide batteries, J. Power Sources 280 (2015) 363–372, http://dx.doi.org/10.1016/j.jpowsour.2015.01.110, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775315001263.

[59] T. Waldmann, M. Wilka, M. Kasper, M. Fleischhammer, M. Wohlfahrt-Mehrens, Temperature dependent ageing mechanisms in Lithium-ion batteries – A post-mortem study, J. Power Sources 262 (2014) 129–135, http://dx.doi.org/10.1016/j.jpowsour.2014.03.112, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775314004352.

[60] Spearman rank correlation coefficient, in: The Concise Encyclopedia of Statistics, Springer New York, New York, NY, 2008, pp. 502–505, http://dx.doi.org/10.1007/978-0-387-32833-1_379.

[61] M. Kuhn, K. Johnson, Applied predictive modeling, 2013, URL http://www.amazon.com/Applied-Predictive-Modeling-Max-Kuhn/dp/1461468485/.

[62] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222, http://dx.doi.org/10.1023/B:STCO.0000035301.49549.88, URL http://link.springer.com/10.1023/B:STCO.0000035301.49549.88.

[63] A. Lindholm, N. Wahlström, F. Lindsten, T.B. Schön, Machine learning - a first course for engineers and scientists, 2021, URL https://smlbook.org.

[64] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[65] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, in: Adaptive Computation and Machine Learning, MIT Press, Cambridge, Mass, 2006, OCLC.

[66] A. Lindholm, F. Lindsten, T.B. Schön, N. Wahlström, An Introduction to Gaussian Process Regression, Tech. Rep., Uppsala University, p. 17, URL http://www.it.uu.se/edu/course/homepage/apml/literature/GP.pdf.

[67] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, http://www.deeplearningbook.org.

[68] L. Ljung, System Identification (2nd Ed.): Theory for the User, Prentice Hall PTR, USA, 1999.

[69] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[70] F. Chollet, et al., Keras, 2015, https://keras.io (Accessed: 02-October-2020).

[71] D. Roman, S. Saxena, V. Robu, M. Pecht, D. Flynn, Machine learning pipeline for battery state-of-health estimation, Nat. Mach. Intell. (2021) http://dx.doi.org/10.1038/s42256-021-00312-3, URL http://www.nature.com/articles/s42256-021-00312-3.