Predicting the State of Charge and Health of Batteries using Data-Driven Machine Learning

Man-Fai Ng[1], Jin Zhao[2], Qingyu Yan[2*], Gareth J. Conduit[3*], Zhi Wei Seh[4*]

[1]Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Connexis, Singapore 138632, Singapore

[2]School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

[3]Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom and Intellegens, 28 Chesterton Rd, Cambridge CB4 3AZ, United Kingdom

[4]Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), 2 Fusionopolis way, Innovis, Singapore 138634, Singapore

*Corresponding authors: alexyan@ntu.edu.sg (Q.Y.), gjc29@cam.ac.uk (G.J.C.) and sehzw@imre.a-star.edu.sg (Z.W.S.)

Machine learning is a specific application of artificial intelligence that allows computers to learn and improve from data and experience via sets of algorithms, without the need for reprogramming. In the field of energy storage, machine learning has recently emerged as a promising modelling approach to determine the state of charge, state of health and remaining useful life of batteries. In this review, we first discuss the two most studied types of battery models in the literature for battery state prediction: the equivalent circuit and physics-based models. Based on the current limitations of these models, we showcase the promise of various machine learning techniques for fast and accurate battery state prediction. Finally, we highlight the major challenges involved, especially in accurate modelling over length and time, performing in-situ calculations and high-throughput data generation.

1 Introduction

With rising concerns about global warming, electrification of transport has emerged as an important vision in many countries in recent years. The successful development of electric vehicles (EVs) depends highly on the cycling performance, cost, and safety of the batteries. Rechargeable lithium-ion (Li-ion) batteries are currently the best choice for EVs due to their reasonable energy density and cycle life.[1] Further research and development on Li-ion batteries will lead to even higher energy density and more complicated battery dynamics, where the efficiency and safety of such batteries will become a concern. An advanced battery management system (BMS) that can monitor and optimize battery behavior and safety is thus essential for the entire electrification system.[2]

Today, one of the major barriers to widespread adoption of EVs is range anxiety. The ability of a BMS to accurately determine the state of charge (SOC) and state of health (SOH) of batteries, and hence the estimated driving range, will alleviate this problem. In addition, reliable prediction of remaining useful life (RUL) will allow batteries to be used to their fullest potential and maximum life expectancy before replacement or disposal. Knowledge of the RUL of spent batteries will also enable their re-deployment in less demanding, second life applications such as stationary grid storage. If we are able to sort manufactured cells based on

their expected lifetime using early-cycle data, we can further accelerate the testing, validation, and development process of new batteries. In summary, accurate prediction of the current and future state of batteries will open up vast opportunities in battery manufacturing, usage, and optimization.[3,4]

SOC and SOH are the two most important parameters in battery management and are generally defined as:

$$SOC = \frac{C_{current}}{C_{full}} \times 100\%$$

$$SOH = \frac{C_{full}}{C_{nominal}} \times 100\%$$

where $C_{current}$ is the capacity of the battery in its current state, C_{full} is the capacity of the battery in its fully charged state, $C_{nominal}$ is the nominal capacity of the brand new battery.[2]

In essence, SOC denotes the capacity of the battery in its current state compared to the capacity in its fully charged state (equivalent of a fuel gauge), while SOH describes the capacity of the battery in its fully charged state compared to the nominal capacity when brand new. By convention, SOC is 100% when the battery is fully charged and 0% when it is empty, while SOH is 100% at the time of manufacture and reaches 80% at end of life (EOL). In the battery manufacturing industry, EOL is often defined as the point at which the actual capacity at full charge drops to 80% of its nominal value.[2] The remaining number of charge/discharge cycles until the battery reaches EOL is the RUL of the battery. Current BMSs can determine the SOC of Li-ion batteries within 0.6% to 6.5%,[5] but are unable to predict the SOH and RUL of batteries accurately.[6]

The traditional methods for SOC estimation include ampere-hour counting estimation; open circuit voltage based estimation; impedance based estimation; model-based estimation; Fuzzy logic; Kalman filter and observer.[4–14] Among all these methods, the major advantage of the model-based method is its ability to be used for on-line applications. In fact, equivalent circuit models (ECMs) are currently the main battery models that are widely used in the BMS of EVs for on-line SOC estimations due to their low computational demand. But the accuracy is usually limited to the range that the model has been parameterized. A further improvement on model-based method is to develop physics-based models (PBMs). The most studied PBM model is called the pseudo two-dimensional (P2D) model which provides insights into the internal dynamics of the batteries. However, the governing equations are complicated and require a high computational cost to solve, making it less practical for on-line applications. Moreover, the traditional PBM model does not take into account the details of materials information, which is vital for understanding degradation behavior related to SOC, SOH and RUL of batteries. In section 2, we will discuss in detail the intrinsic characteristics of these two most studied models (ECM and PBM) in the literature.

Despite the progress in developing more accurate and fast models for on-line SOC and SOH estimations, there remains a clear tradeoff between the computational efficiency and the accuracy of model-based predictions. Recently, data-driven models (DDMs) have drawn much attention. Combined with machine learning techniques, these models are able to make predictions without prior knowledge of the system (Figure 1). Machine learning techniques including neural network, support-vector machine, random forest, and regression techniques

have been applied to predict the SOC, SOH, and RUL of batteries. In section 3, we summarize the recent works on how various machine learning techniques can be applied for battery state predictions and provide insight into the predictabilities of these techniques. Machine learning techniques that can accurately model over length and time, and perform in-situ calculations, allow us to incorporate domain knowledge such as materials information into a new explainable model. In addition, the fidelity of the model depends strongly on the size and quality of the data set. High-throughput computation and experimentation is one approach that can produce huge volumes of precise data within well-controlled conditions. The major challenges involved, together with our perspective on the future development of data-driven machine learning for battery state predictions, will be discussed in section 4. For the ease of reference, the common acronyms in battery modelling research are listed in Table 1.

Table 1. Definition of common acronyms for battery modelling used in the main text.

| Abbreviation	Definition	Abbreviation	Definition
EV	electric vehicle	SPM	single particle model
BMS	battery management system	PBM	physics-based model
SOC	state of charge	ROM	reduced order model
SOH	state of health	P2D	pseudo two-dimensional
RUL	remaining useful life	DDM	data-driven model
EOL	end of life	PDE	partial differential equation
ECM	equivalent circuit model	ODE	ordinary differential equation
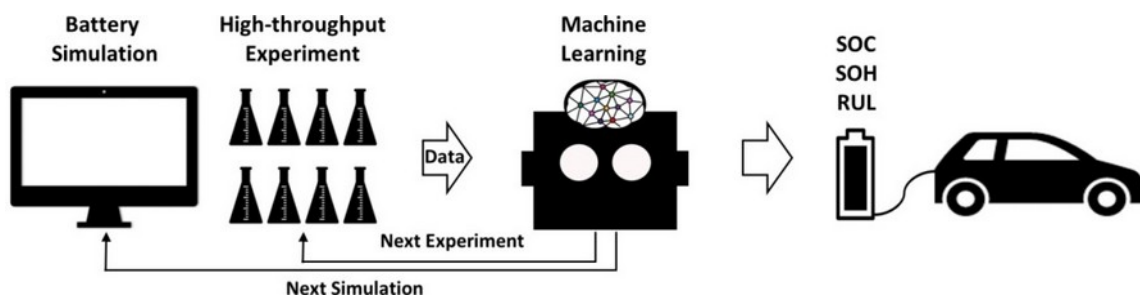
Figure 1. A machine learning approach for state of charge (SOC), state of health (SOH) and remaining useful life (RUL) predictions of Li-ion batteries.

2 Current Battery Models

Battery modelling is the core part of a BMS and is vital for maintaining safe and optimal operation of the battery pack. A battery model combining various estimation techniques can be used not only to determine the current state of an operating battery (e.g. SOC) but also predict its 'future' state (e.g. SOH and RUL). In the literature, the most studied battery models for Li-ion batteries are ECMs, PBMs and, more recently, DDMs with machine

learning techniques. Each model has its own merits and challenges. For example, ECMs are computationally efficient and thus suitable for online battery status predictions (e.g. SOC), but attaining high accuracy remains a challenge. PBMs provide internal information about a battery such as the Li-ion concentration within the electrodes and electrolytes, but solving the governing partial differential equations (PDEs) requires significant computational resources and a large number of input parameters. Also, a battery model needs to work with sufficient random access memory which is used to store the instant data for a BMS. The memory requirement depends highly on the complexity of the modelling equations. In this section, the intrinsic characteristics of ECMs and PBMs, and the strategies commonly used to improve their adaptability and predictability will be discussed.

ECMs[15–25] are currently the major models that are widely used in the BMS of EVs for online SOC estimations due to their ability to predict battery behavior in real time. The models are essentially derived from empirical knowledge and experimental data in which the batteries are represented by groups of electrical components such as resistors and capacitors, forming resistor-capacitor networks (Figure 2) that are used to monitor the battery's behavior at different time constants associated with the diffusion and charge-transfer processes.[15] Typical ECMs are the Rint models,[16] the hysteresis models,[17,18] the Randles models,[19–21] and the resistor-capacitor or Thevenin models.[22–25] Despite their computational efficiency, ECMs generally show limited accuracy in predicting battery characteristics across a range of operation conditions such as ageing and dynamics environments in real-life applications, due to parameterization of model parameters based on laboratory conditions. In addition, the lack of physics-based information of the system states and parameters makes it hard to predict the SOH and RUL of batteries precisely.

PBMs should offer more accurate battery models. The pioneering work of full physics-based Li-ion battery models is the development of a P2D porous electrode model, which is based on porous electrode theory, concentrated solution theory, and the Bulter-Volmer kinetic equations (Figure 2).[26,27] The model delivers insights into the internal dynamics of batteries such as lithium ion diffusion, Ohmic effects, and electrochemical kinetics. This opens the possibility of analyzing the battery's degradation mechanisms, predicting the SOC and SOH with ageing effects, and designing optimal charging strategies. However, the P2D model is generally described by a number of PDEs and is considered a full order PBM. Solving the PDEs require intensive computations which makes it impractical to embed the P2D model into a controller of a BMS for real-time applications.[28]

The bottleneck of applying the full PBM in the BMS for EVs lies in the computational complexity. As such, simplifying the PBMs is the main strategy to reduce the computation demand, but approximations must retain sufficient physical information to accurately predict battery behavior. One of the most studied simplified models is the single particle model (SPM) (Figure 2).[29–31] The key assumptions of the model are that a spherical particle represents each electrode, and the concentration and potential effects in the solution phase are neglected. With such approximations, the computational time is reduced significantly. However, the SPM model is inaccurate for high-rate simulations,[32] though efforts to improve this limit are ongoing.[33–36]

The PDEs that govern battery behavior in the P2D model are non-linear, so reducing the order of the equations is another approach to build a practical PBM. The models are commonly known as reduced order models (ROMs) which comprise fewer ordinary differential equations (ODEs). In addition, reformulation of P2D model[37] is another approach to develop a more efficient yet accurate model. Typical approaches to construct ROMs or reformulated P2D model are through mathematical techniques such as parabolic profiles

approximations,[38,39] proper orthogonal decomposition;[40] residue grouping technique;[41] the Padé approximations;[42] or polynomial profiles.[43] Using polynomial profiles for solid concentration is the most common method; it is mathematically simple and computationally fast, but prediction accuracy is reduced by the assumption that the profile coefficients are temperature and age independent.

In addition, degradation of materials within a battery is closely related to the SOC, SOH and RUL. Exploring the degradation behavior of batteries hence requires simulations at the materials level. Multiscale modelling that includes density functional theory, molecular dynamics, and phase field method can be used to study the degradation mechanisms of batteries. Incorporating these physics information in a battery model is challenging, but it can significantly improve the accuracy and explainability of battery state predictions.

In summary, the main challenge of current battery models lies in achieving an appropriate balance between model fidelity and computational complexity, as shown in the plot of accuracy vs. CPU time proposed by Subramanian and co-workers (Figure 2).[32] Recently, DDMs with machine learning techniques are gaining importance due to their immense potential in achieving high accuracy with low computational cost. In the next section, we will discuss state-of-the-art machine learning techniques for battery state prediction.
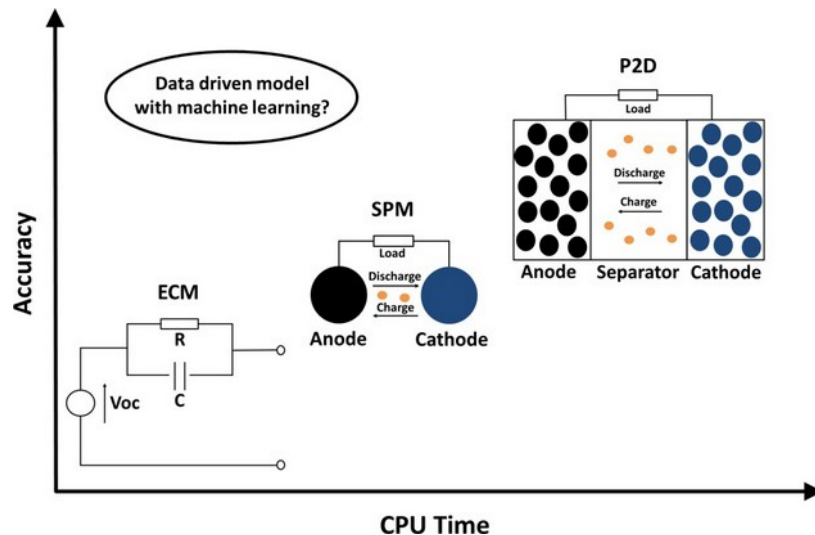

Figure 2. Accuracy vs. CPU time for equivalent circuit model (ECM), single particle model (SPM) and pseudo two-dimensional (P2D) model. Data-driven model (DDM) with machine learning is a type of battery model that is promising for fast and accurate battery state predictions.

3 Machine Learning for Battery State Prediction

We often want to predict the future behavior of a battery, for example to understand how much further an EV can drive, or how to design a battery that will have the best behavior in the field. Often, we are interested in the SOC of the battery within a single charge/discharge cycle, or the SOH of the battery spanning many charge/discharge cycles. Having two relevant time-scales will make predictions particularly challenging. All of these problems can be summarized as the fact that we need a function that inputs the current state of the battery to predict future behavior. A promising approach is machine learning - a flexible but efficient fitting function with no underlying physical knowledge. Table 2 summarizes the approaches taken by a range of authors over the past few years.[3,4,44–69] We first summarize the input and

output parameters captured by the different modelling approaches and battery systems analyzed, before we focus on the advantages and disadvantages of the various machine learning techniques for predictive analytics of batteries. We then summarize the most useful machine learning algorithms to predict SOH, SOC, and RUL. Finally, we offer a perspective on the future outlook and opportunities in modern machine learning and data generation to better understand and predict battery behavior.

3.1 Battery parameters: inputs & outputs

In order to understand, design, and predict battery properties, a range of variables that captures their full behavior must be incorporated. Usually some variables are either ignored or held constant to simplify the model. The possible input variables for a machine learning model can be split into continuous, integer, and categorical. Continuous variables can take any value and include the current flow, the internal structure, the geometry, and the temperature. Integer variables include the number of charge/discharge cycles that the battery has gone through. Categorical variables take particular values that cannot be sorted into a list, examples include the type of battery: Li-ion, nickel-metal hydride, or lead-acid. A machine learning method should ideally be able to input continuous, integer, and categorical variables in order to make predictions.

The outputs can be classed into two main categories: (1) short-time scale over a single charge/discharge cycle to understand the SOC, and (2) long time-scale over many charge/discharge cycles to understand the SOH. The first approach is to predict the evolution of the battery during a single charge/discharge cycle. Endpoints predicted can include the SOC, the current rate, and the concentration and size of defects formed within the battery. By tracking the evolution during a charge/discharge cycle, the model can address any point in the lifetime of a battery and extrapolate forward in time, but it is susceptible to accumulating errors if applied over too many charge/discharge cycles.

The second approach is to predict the evolution of the battery from the same point cycle-to-cycle over many cycles. This approach can be readily applied across hundreds of cycles covering the entire lifetime of the battery, but cannot be applied during a given cycle, and can start from and propagate to only a particular defined point during the cycle, for example when fully charged. In Table 2, machine learning models are seen to successfully predict the evolution of battery properties. The accuracy level attained in terms of the averaged percentage error of these works is 4.1 % for SOC, 3.8 % for SOH, and 4.1 % for RUL.

Table 2. Summary of recent work on machine learning for battery state predictions.

Article		Method						Feature set inputs								Output			
		Neural network	Support-vector machine	Gaussian / Bayesian	Regression	Random forest / Tree	Kalman filter	Voltage	Current	Temperature	Cycle number	Capacity	Power	Geometry	P2D model parameters	SOC	SOH	RUL	Percentage error
Severson et al.	3				✓			✓	✓		✓	✓						✓	9.1
Nuhic et al.	4		✓					✓		✓	✓	✓					✓		6.4
Guo et al.	44			✓				✓	✓		✓							✓	8.7
Wu et al.	45	✓							✓				✓				✓		7
Zahid et al.	46	✓						✓	✓	✓			✓			✓			0.1
Chemali et al.	47	✓						✓	✓	✓						✓			2.4
Jiménez-Bermejo et al.	48	✓						✓	✓	✓						✓			1.3
Mansouri et al.	49	✓	✓		✓	✓		✓										✓	3.3
Donato et al.	50	✓				✓		✓	✓		✓					✓			13
Huang et al.	51						✓	✓	✓	✓						✓			3
Ren et al.	52	✓	✓	✓	✓			✓	✓	✓								✓	6.7
Khumprom et al.	53	✓						✓	✓		✓							✓	0.6
Sahinoglu et al.	54			✓				✓	✓	✓						✓			0.8
Álvarez Antón et al.	55		✓					✓	✓	✓						✓			12
Tong et al.	56	✓						✓	✓	✓						✓			3.8
Kang et al.	57	✓						✓	✓				✓			✓			10
Hu et al.	58	✓						✓	✓							✓			1.7
Wu et al.	59	✓						✓	✓	✓						✓			3
Wu et al.	60	✓						✓			✓						✓		5
Hu et al.	61		✓	✓				✓	✓			✓				✓			2.1
Berecibar et al.	62	✓	✓		✓			✓	✓		✓	✓					✓		1.6
Richardson et al.	63			✓				✓	✓		✓	✓						✓	3.2
Zhang et al.	64	✓						✓			✓	✓						✓	1.2
Hu et al.	65		✓					✓	✓	✓	✓		✓			✓			0.2
Tseng et al.	66				✓			✓	✓		✓							✓	0.2
Hussein et al.	67	✓					✓	✓	✓							✓			1.5
Yang et al.	68	✓						✓	✓								✓		5
Dawson-Elli et al.	69	✓				✓									✓	✓			5

Note: For Refs 4, 44, 45, 48, 49, 50, 54, 55, 58, 62, 63, 65, 67, and 68, the percentage errors are estimated from the root mean square errors using the approach described in section 3 in the main text. For Refs 53 and 64, the percentage error in RUL is estimated from the (error in cycle predicted /actual cycle) ×100%

3.2 Machine learning techniques

Machine learning uses a general fitting function with optimizable parameters tuned to deliver the desired behavior, usually a fit to experimental training data. The function can then make predictions for other battery systems. Below we discuss two main issues: (1) how to validate

the fitting function and (2) the selection of the fitting function. Finally, we summarize the most appropriate machine learning methods to predict different battery properties.

A key question is how to validate the machine learning model once it has been fitted to the data. Like all fitting functions with optimizable parameters, machine learning models can be susceptible to over-fitting – perfectly fitting the training data by introducing unphysical features that would poorly reproduce parts of the underlying function not present in the training data. To properly validate the model, a common procedure is to hold back some of the data, unseen to the model, to later benchmark the fitting function through calculation of an error metric. This metric is most commonly the root mean square error that can itself be recast as the coefficient of determination by dividing by the variance of the data to deliver a metric that is independent of scaling the data, or as a percentage by simply dividing by the range of the data. We outline two standard protocols: in the first, known as "hold out", the model is trained on a fraction of the total available data (typically 80%), and then the accuracy is gauged by testing against the remaining (typically 20%) data that was withheld from the training process. The second protocol is "cross-validation" where the hold-out procedure is repeated on several (typically five) randomly selected fractions of the available data, which provides an averaged measure of performance over several train-test splits. The minimization of the error metric allows the correct hyper-parameters to be obtained, including the number of optimizable parameters, and delineate the performance of different models. With the validation strategy in place, we now review the advantages and disadvantages of the fitting functions previously used to model the behavior of batteries, which are summarized in Table 2.

3.2.1 Linear regression

The straightforward fitting of a straight line (one input dimension) or a hyperplane (multiple input dimensions) to the data is probably the simplest model possible, and consequentially can deliver insights into the underlying physics. The model has the form

$$y = m \cdot x + c$$

where y is the output, x are the vector of input variables, m is the vector of fitting parameters corresponding to the gradients, and c the fitting parameter for the shift. The fitting parameters are often selected by minimizing a mean-square error cost function.

The robustness of the fit can be improved through singular value decomposition, which circumvents singular solutions. This approach is clear, robust, and fast, and furthermore requires a minimal amount of training information to form a model. However, regression is not just limited to fit just a straight line, as many batteries will display non-linear behavior. To capture this behavior, we can expand the method to be non-linear, by including quadratic and higher order terms in the fit, by analogy to a Taylor expansion.

A linear model that combined nine battery descriptors was used by Severson et al[3] to predict the RUL of lithium iron phosphate/graphite cells after 100 charge/discharge cycles. The model input the current cycle number, voltage, current flow, and capacity to predict the RUL with a typical error of 9.1%. The simple linear model allows fast computational time for training and predictions that can be directly deployed in devices.

3.2.2 Random forest / tree and support-vector machine

Random forest involves a set of generalized classification trees, each trained with randomly selected data. The split of each layer is often chosen to maximize the reduction in variance of the remaining training data. A new query passes down the trees to deliver an ensemble of

predictions that are averaged to give the expected value alongside an uncertainty. The average prediction at x is

$$y = \frac{1}{T} \sum \sum \Theta_t(x_i, x) y_i$$

where t index the T trees, (x_i, y_i) is the ith data point, and should x_i and x be in the same leaf then $\Theta_t(x_i, x)$ is the reciprocal of number of entries in that leaf, and is otherwise 0. A random forest is most straightforward to train with categorical data. The random forest is accurate, easy to train, and robust against outliers, but the function delivered is often discrete rather than smooth.

An example of the successful application of a tree method to predict the RUL of a Li-ion battery is demonstrated by Mansourei et al.[49] Focusing on batteries in unmanned aerial vehicles, the authors aimed to extend the flying time window. The authors found that the random forest approach that inputs simply the variation of voltage with time delivered a typical prediction error in the RUL of 3.3%, outperforming linear models, a support-vector machine, and a neural network.

Support-vector machine is a generalization of the random forest where the functions trained are simultaneously classified in a multidimensional space rather than split along one input direction. Where the training data is scarce, this approach can improve the quality of the fit, but it comes at the cost of significantly increased computational demands. The support-vector machine protocol is effective on sparse data,[70] particularly when augmented by factorization machines.[71]

Nuhic et al[4] used a support-vector machine to predict both the SOH and RUL of lithium ion batteries. The support-vector machine took account of the voltage, capacity, cycle number, and temperature to estimate SOH between successive cycles within 6.4%, and showed that the SOH and RUL was strongly influenced by environmental and load conditions.

3.2.3 Gaussian processes

This is a stochastic method that delivers a probability distribution of possible predictions

$$p(y) = N(\mu(x), \sigma^2(x))$$

where $p(y) = N$ is a normal distribution with modified mean $\mu(x)$ and variance $\sigma^2(x)$ derived from the covariance of the training data.[72] Although there is no cost function, the covariance itself comprises a statistical prior distribution, often taken to be a normal distribution in the distance of predicted point from the training data. This necessitates storing all of the training data as the foundation of the model. At run-time, once given the input parameters, it calculates the joint probability distribution of the underlying fitting functions, usually Gaussian distributions, and the training data. Furthermore, the approach captures the higher certainty in our knowledge when making predictions near to known training data, and increased uncertainty in the function when making predictions further from the training data or when the data is noisy. However, this increased level of insights often means that the approach is prohibitively expensive.

Sahinoglu et al[54] used Gaussian process regression to estimate the SOC of Li-ion batteries. The model uses battery parameters, including voltage, current, and temperature, as inputs. The Gaussian processes are shown to deliver predictions for SOC within 0.8% and outperform support-vector machine and neural network predictions.

3.2.4 Neural network

The linear fitting method could be extended with a Taylor expansion to capture non-linear behavior. However, a more efficient approach is to use several locally non-linear basis functions to build a composite function in a neural network. The mathematical form of a neural network with a single layer of hidden nodes is

$$y = D + \sum C_i \tanh(A_i \cdot x + B_i)$$

where y is the output, x are the vector of input variables A_i is a vector of fitting parameters, B_i, C_i, and D are further fitting parameters, with a sum over hidden nodes denoted by i. The fitting parameters are often selected by optimizing a mean-square error cost function. The neural network is more expensive to train, but when a large amount of data is available, it often gives the highest possible quality fitting function, hence its widespread use in industry.

In batteries, we are often interested in predicting the evolution of the SOC over a single charge/discharge cycle or the evolution of the SOH over several cycles. For these problems that focus on the passage of time, a convolutional neural network is helpful. This is a specialist fitting function useful on systems that display temporal invariance, fundamentally capturing that, for example, the behavior of a battery is independent of the time of day that it was used.

Yang et al[68] used a neural network to predict the SOH of Li-ion batteries for EVs. Taking in the voltage and current through a first-order ECM, a three-layer neural network could predict the SOH within 5%. In fact, the majority of studies so far focus on the most commercially important system of Li-ion batteries, with only a couple addressing nickel-metal hydride and lead-acid batteries. A single study by Zahid et al[46] presents a generalized neural network model that can address all three battery families. That model inputs voltage, current, power dissipation and power to predict the SOC within 0.1%. This is a valuable direction as it allows information on one battery system, e.g. Li-ion, to inform the behavior of other systems that are less well-studied, and furthermore provide guidance for future possible battery families.

3.3 Selection of machine learning approach

The selection of the most appropriate machine learning approach is a multifaceted problem, depending on the amount of data available, the quality of results desired, and physical interpretability of model required.

Neural networks are probably the industry leading technique in machine learning competitions due to the high levels of accuracy that they can attain, so it is not a surprise that they are the most widely used approach for predicting battery properties. This is especially apparent in the prediction of SOC, where neural networks are adopted and were the preferred approach in 10 out of 15 studies (Table 2). SOC systems are characterized by having access to a large amount of training data that can be collected at small time steps throughout the evolution of the battery, and the neural network performs well on data rich systems. In addition, a hybrid optimization technique for the P2D battery model inspired by the neural network-based chess engine DeepChess has also been proposed.[73]

However, the preferred machine learning approach is more nuanced when predicting either SOH or RUL. These are typically measured once per cycle, so with training data collected once per charge/discharge cycle, the typical datasets are often smaller. Therefore, for predicting SOH and RUL, the preferable machine learning techniques adopted are varied,

with the data heavy requirement of neural networks meaning that they were adopted in just 8 of 13 studies (Table 2), and a variety of other machine learning techniques were used in the others. Here the preferable technique is decided on a case-by-case basis, for example Gaussian processes are used[44,52,63] not only due to the relative lack of data, but more importantly because Gaussian processes can intrinsically predict uncertainties, which is vital for making possibly safety-critical health diagnostics.[63] The RUL can be expressed in terms of number of remaining charge/discharge cycles, an integer rather than continuous quantity, making random forest a suitable method. On the other hand, though many machine learning approaches are black boxes, a physical understanding of the predictions can be more important for safety-critical and scientific applications, meaning that the straightforward nature of linear regression can be preferred.[62]

4 Future Outlook and Opportunities

Here we highlight three long-standing "holy grail" problems for battery state prediction where machine learning has the potential to make significant inroads: (1) holistic battery modelling that transcends time, length, and mechanism scales, (2) accelerating and simplifying calculations to enable them to be done in-situ on the battery itself, and (3) high-throughput computational and experimental data generation.

4.1 Accurate modelling over length and time

The battery models presented have been fragmented: one works at long and another at short length scales, some over one charge/discharge cycle and others over many cycles. This means that each model captures one particular degradation mechanism well, but neglects other factors. The practical use of battery models requires all factors to be captured, with machine learning well-positioned to replace each individual model and merge their predictions together.

Machine learning models are best used when the underlying functional dependence is not known from a PBM. Because of this, machine learning is often referred to as a 'black box', where datasets enter and predictions emerge, but the process between input and output is opaque. The incorporation of domain knowledge into machine learning will help in the development of models that are more explainable and interpretable. Moreover, if a PBM is available, then machine learning can be applied to capture the remaining difference from the experimental data. Although this hybrid approach introduces additional computational cost, it can deliver more accurate and insightful models with less risk of over-fitting the training data.

The models presented from the literature focus either on the prediction of SOC within a charge/discharge cycle, or the SOH/RUL over many cycles. There is however a more general problem: to predict the long-term SOH, but starting from an arbitrary point in the charge/discharge cycle. Machine learning could first use a detailed model to predict until a fixed point in a cycle, for example, the state of being fully charged. Next, a SOH model that covers integer cycles could be applied to predict the final SOH. This hybrid approach would achieve the best of both worlds, and as both the short and long-term behavior models have now been developed, there is an opportunity to juxtapose them into a holistic model of battery evolution.

4.2 Performing in-situ calculations

Improved battery modelling means that we now have accurate predictions of some battery properties. However, these calculations need to be performed using large computing

resources. In practice, if these calculations were available on the battery itself, then the battery could adapt and optimize itself to its use case. This requires much lighter calculations that could be achieved through machine learning and can be done on light embedded devices on the battery.

First-principles simulations such as density functional theory and molecular dynamics have been widely used to study the degradation of materials, e.g. solid electrolyte interphase formation and decomposition of electrolytes.[74,75] The phase field approach is another physics-based simulation mainly for studying the evolution of microstructures with various morphologies, including lithium dendrite growth[76,78] and phase separation of active electrode materials.[79,81] Although phase-field modelling has not yet approached full cell simulations, the simulated results generally match well with experimental data.

To perform in-situ calculations, the first step is to build a database of historic results from the multiscale first-principles and phase field simulations and then train a machine learning model. This model is then used as a proxy for the simulations, except if the machine learning reports a large uncertainty, when an additional simulation is performed, added to the database, and the machine learning model retrained. This cycle of active learning can significantly reduce the number of simulations required to understand a system. Machine learning can be used in a similar way for experimental design and to shortcut costly experiments. Further evidence of the potential for machine learning to shortcut simulations comes from studies regarding the mechanical properties of solid electrolytes[82] and voltage.[83]

In fact, every battery in service is different. Due to its particular usage, the behavior of a certain battery is unique, and evolves throughout the battery's service. Therefore, one could also develop a bespoke machine learning model for that particular battery, perturbed from the default, refined by data gathered in service to capture that particular battery's characteristics for accurate on-line predictions.

4.3 High-throughput data generation

Databases underlie all machine learning and data-driven approaches. Compared to the traditional one-by-one approach, high-throughput technologies can generate a large but high quality database in a short time at low cost. Today, high-throughput technologies have been widely employed in various fields, e.g. biological and medical sciences, due to the rapid progress in automation, robotics, and computational technologies.[84–90]

4.3.1 High-throughput computations

With the advancements in computer technology, the discovery of new materials can now be performed using high-throughput computations, material databases and machine learning. The Materials Project[91] and the Open Quantum Materials Database (OQMD)[92] offer valuable database of simulation results for many battery properties, spanning both electrodes and electrolytes. Although initially aimed at screening studies,[93] the resource also provides an ideal platform for machine learning.

In general, computational battery material research focuses on two areas: the electrode and electrolyte. The development of new electrodes with high voltage, high capacity, high chemical stability and low cost are desired. For liquid electrolytes, the general concerns include viscosity, volatility, flammability, compatibility with the separator and electrode. As such, the focus is on developing new electrolytes with better organic molecules and additives that fulfill the requisite properties. For solid-state electrolytes, a faster Li-ion transport property and a more stable electrochemical window are desired.

The combined database and machine learning approach have been applied to design and predict the material properties of electrodes such as voltage, crystallinity and chemical stability, from atomic scale to mesoscale.[83,94-99] In addition, such approach has been applied to design new liquid electrolytes and additives,[100-105] and solid-state electrolytes with fast Li-ion transport[106-108] and mechanical[82] properties. Such computational techniques provide an opportunity for exploring material properties at a lower cost and accelerating the material discovery processes.

Overall, concerted international efforts are needed to promote the sharing of meaningful battery data in a standardized and machine-readable form, in order to continue expanding our valuable material databases. Such meaningful data should include both positive and negative results, as machine learning algorithms need to be able to distinguish between batteries/battery materials that work well and those that do not. This culture of sharing and collaboration between the computer science and battery communities should be strongly encouraged and is expected to drive accelerated scientific research.

4.3.2 High-throughput experiments

In the field of Li-ion batteries, high-throughput experimental data generation involves several aspects, namely material synthesis, material characterization, battery fabrication, and electrochemical testing.[109–114] With high-throughput synthesis, electrode materials with different and optimized compositions can be rapidly prepared for subsequent structural and electrochemical analysis, which can speed up the discovery and optimization of electrode materials. Currently, thin-film sputtering, pulsed laser deposition, combinatorial robotic, and microplate techniques have been developed to synthesize/screen electrode materials and electrolytes, as well as optimize the content of additives, in a high-throughput manner.[109,115] These methods are still limited to modulation and tuning of material compositions, which is below the typical requirement for diverse machine learning datasets. Currently, high-throughput material synthesis remains to be explored for diverse materials with novel microstructures, different crystal structures, controllable dopants, interfaces, and defects.

Material characterization is usually carried out sequentially to measure and ascertain the key properties of the electrode materials before incorporation into batteries. In recent years, many high-throughput diagnostic tools across a range of methodologies have been invented to perform routine characterization tasks. For example, high-throughput X-ray diffraction and X-ray fluorescence techniques have been developed to collect the crystalline phase and elemental information of electrode materials.[110,116] However, ex-situ characterization may not reflect the true information of the actual charge/discharge states due to changes in the thermodynamic non-equilibrium phase after relaxation. With in-situ techniques such as X-ray diffraction and X-ray absorption spectroscopy, the chemical, structural and thermal evolution of materials, as well as the pressure within Li-ion batteries can now be properly monitored.[113,114,117] Overall, further work is required to create fully automated and continuous processes, e.g. by developing advanced robotics and algorithms to directly couple the synthesis and characterization tools.

Automated battery fabrication is crucial as it can accelerate subsequent battery optimization and testing based on realistic operating conditions.[118] Conventional battery development usually starts with small-scale, simplified and discontinuous laboratory equipment, as well as manual processes. The electrodes, associated components and entire cells generally have non-optimized internal structures and often low mass loading. This is far below the requirements of real-life commercial applications and cannot provide reliable data for machine learning. Hence, highly integrated Li-ion battery production processes from the initial automatic electrode synthesis to the final battery testing systems are necessary for high-throughput precise data generation

(Figure 3). Ideally, the anode, cathode, binder, and conductive additives can be selected and optimized automatically, followed by automatic mixing to prepare electrode slurries. The slurries are then coated onto current collectors, calendared into electrode films, and slit into appropriate dimensions for automated assembly into batteries, together with electrolyte and separator.

Finally, high-throughput electrochemical testing of batteries holds the key to generating huge and reliable datasets for machine learning. A variety of electrochemical techniques, including cyclic voltammetry, galvanostatic charge/discharge, and electrochemical impedance spectroscopy, can be used to measure the cycle life, rate capability, capacity and impedance of batteries with high precision and accuracy (Figure 3). Batteries should be screened quickly in parallel based on realistic working conditions (e.g. different current, voltage, power, temperature, mass loading, and cell design) to generate huge volumes of meaningful data. Once the machine learning models are trained with these data, they can further accelerate the process of battery testing, by weeding out potential poor-performing batteries based on their initial cycles. For instance, by using the first five cycles, Severson et al[3] managed to use a trained machine learning model to classify cells into two groups: a 'low-lifetime' and a 'high-lifetime' group, with 4.9% test error.
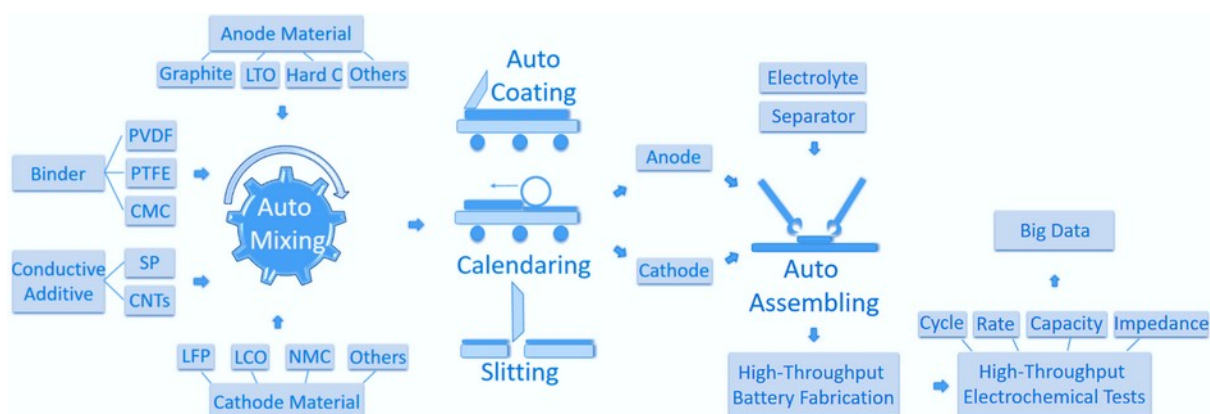

Figure 3. High-throughput battery fabrication and testing. Abbreviations: PVDF: polyvinylidene fluoride, PTFE: polytetrafluoroethylene, CMC: carboxymethyl cellulose, SP: Super P, CNTs: carbon nanotubes, LTO: lithium titanate, Hard C: hard carbon, LFP: lithium iron phosphate, LCO: lithium cobalt oxide, NMC: lithium nickel manganese cobalt oxide.

5 Conclusion

Currently, the two most studied models for battery state prediction are the ECMs and PBMs. Despite their popularity and continuous development, there remains a clear tradeoff between computational efficiency and accuracy when using these models for on-line battery state prediction. DDMs with machine learning is a promising way to model batteries that can potentially address the dilemma faced by traditional modelling using ECMs or PBMs. Currently, most of the machine learning models give 'black box' battery state predictions, which makes it difficult to generalize to other battery chemistries. The incorporation of domain knowledge paves the way for explainable 'white box' prediction. Moreover, high-throughput experimentation, perhaps guided by preliminary machine learning results, is the key to provide real-life and high-quality datasets on battery performance for machine learning. With the advancement of computational technologies and mathematical algorithms, together with the reduced costs of data storage devices and high-throughput experiments, we

envision data-driven machine learning to be a promising technique for advanced battery modelling in the future.

Acknowledgements

Author contributions

This work was written through the contributions of all authors.

References

1.   Whittingham, M. S. Ultimate limits to intercalation reactions for lithium batteries. *Chem. Rev.* **114**, 11414–11443 (2014).

2.   Li, Y. et al. Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. *Renew. Sust. Energ. Rev.* **113**, 109254 (2019).

3.   Severson, K. A. et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **4**, 383–391 (2019).

4.   Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M. & Dietmayer K. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *J. Power Sources* **239**, 680–688 (2013).

5.   Cuma, M. C. & Koroglu, T. A. A comprehensive review on estimation strategies used in hybrid and battery electric vehicles. *Renew. Sust. Energ. Rev.* **42**, 517–531 (2015).

6.   Waag, W., Fleischer, C. & Sauer, D. U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. *J. Power Sources* **258**, 321–339 (2014).

7.   Hannan, M. A., Lipu, M. S. H., Hussain, A. & Mohamed, A. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: challenges and recommendations. *Renew. Sust. Energ. Rev.* **78**, 834–854 (2017).

8.   Zheng, Y., Ouyang, M., Han, X., Lu, L. & Li, J. Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles. *J. Power Sources* **377**, 161–188 (2018).

9.   Xiong, R., Cao, J., Yu, Q., He, H. & Sun, F. Critical review on the battery state of charge estimation methods for electric vehicles. *IEEE Access* **6**, 1832–1843 (2017).

10.   Xiong, R., Li, L. & Tian, J. Towards a smarter battery management system: a critical review on battery state of health monitoring methods. *J. Power Sources* **405**, 18–29 (2018).

11.   Zou, Y., Hu, X., Ma, H. & Li, S. E. Combined state of charge and state of health estimation over lithium-ion battery cell cycle lifespan for electric vehicles. *J. Power Sources* **273**, 793–803 (2015).

12.   Zhang, Y., Song, W., Lin, S., Lv, J. & Feng, Z. A critical review on state of charge of batteries. *J. Renew. Sustain. Energy* **5**, 1–11 (2013).

13.   Chang, W. Y. The state of charge estimating methods for battery: a review. *ISRN Applied Mathematics* **2013**, 953792 (2013).

14.   Lu, L., Han, X., Li, J., Hua, J. & Ouyang, M. A review on the key issues for lithium-ion battery management in electric vehicles. *J. Power Sources* **226**, 272–288 (2013).

15.   Nejad, S., Gladwin, D. T. & Stone, D. A. A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states. *J. Power Sources* **316**, 183–196 (2016).

16.   Johnson, V. H. Battery performance models in ADVISOR. *J. Power Sources* **110**, 321–329 (2002).

17.   Huria, T., Ludovici, G. & Lutzemberger, G. State of charge estimation of high power lithium iron phosphate cells. *J. Power Sources* **249**, 92–102 (2014).

18.   Plett, G. L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs Part 2. Modeling and identification. *J. Power Sources* **134**, 262–276 (2004).

19.   Fairweather, A. J., Foster, M. P. & Stone, D. A. Modelling of VRLA batteries over operational temperature range using pseudo random binary sequences. *J. Power Sources* **207**, 56–59 (2012).

20.   Shahriari, M. & Farrokhi, M. Online state-of-health estimation of VRLA batteries using state of charge. *IEEE Trans. Ind. Electron.* **60**, 191–202 (2013).

21.   Bhangu, B. S., Bentley, P., Stone, D. A. & Bingham, C. M. Observer techniques for estimating the state-of-charge and state-of-health of VRLABs for hybrid electric vehicles. *IEEE Vehicle Power and Propulsion Conf.* **10**, 780–789 (2005).

22.   Gould, C. R., Bingham, C. M., Stone, D. A. & Bentley, P. New battery model and state-of-health determination through subspace parameter estimation and state-observer techniques. *IEEE Trans. Veh. Technol.* **58**, 3905–3916 (2009).

23.   Kim, T. & Qiao, W. A hybrid battery model capable of capturing dynamic circuit characteristics and nonlinear capacity effects. *IEEE Trans. Energy Conver.* **26**, 1172–1180 (2011).

24.   Sitterly, M., Wang, L. Y., Yin, G. G. & Wang, C. Enhanced identification of battery models for real-time battery management. *IEEE Trans. Sustain. Energy* **2**, 300–308 (2011).

25.   Hu, X., Li, S. & Peng, H. A comparative study of equivalent circuit models for li-ion batteries. *J. Power Sources* **198**, 359–367 (2012).

26.   Doyle, M., Fuller, T. F. & Newman, J. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *J. Electrochem. Soc.* **140**, 1526–1533 (1993).

27.   Fuller, T. F., Doyle, M. & Newman, J. Simulation and optimization of the dual lithium ion insertion cell. *J. Electrochem. Soc.* **141**, 1–10 (1994).

28. Jokar, A., Rajabloo, B., Désilets, M. & Lacroix, M. Review of simplified pseudo-two dimensional models of lithium-ion batteries. *J. Power Sources* **327**, 44–55 (2016).

29. Santhanagopalan, S., Guo, Q., Ramadass, P. & White, R. E. Review of models for predicting the cycling performance of lithium ion batteries. *J. Power Sources* **156**, 620–628 (2006).

30. Guo, M., Sikha, G. & White, R. E. Single-particle model for a lithium-ion cell: thermal behavior. *J. Electrochem. Soc*. **158**, A122–A132 (2011).

31. Zhang, D., Popov, B. N. & White, R. E. Modeling lithium intercalation of a single spinel particle under potentiodynamic control. *J. Electrochem. Soc*. **147**, 831–838 (2000).

32. Ramadesigan, V. et al. Modeling and simulation of lithium-ion batteries from a systems engineering perspective. *J. Electrochem. Soc*. **159**, R31–R45 (2012).

33. Rahimian, S. K., Rayman, S. & White, R. E. Extension of physics-based single particle model for higher charge–discharge rates. *J. Power Sources* **224**, 180–194 (2013).

34. Luo, W., Lyu, C., Wang, L. & Zhang, L. A new extension of physics-based single particle model for higher charge–discharge rates. *J. Power Sources* **241**, 295–310 (2013).

35. Han, X., Ouyang, M., Lu, L. & Li, J. Simplification of physics-based electrochemical model for lithium ion battery on electric vehicle. Part II: pseudo-two-dimensional model simplification and state of charge estimation. *J. Power Sources* **278**, 814–825 (2015).

36. Li, J., Adewuyi, K., Lotfi, N., Landers, R. G. & Park, J. A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (SOH) estimation. *Appl. Energy* **212**, 1178–1190 (2018).

37. Northrop, P. W. C. et al. Efficient simulation and reformulation of lithium-ion battery models for enabling electric transportation. *J. Electrochem. Soc*. **161,** E3149-E3157 (2014).

38. Subramanian, V. R., Ritter, J. A. & White, R. E. Approximate solutions for galvanostatic discharge of spherical particles I. Constant diffusion coefficient. *J. Electrochem. Soc*. **148**, E444–E449 (2001).

39. Subramanian, V. R., Diwakar, V. D. & Tapriyal, D. Efficient macro-micro scale coupled modeling of batteries. *J. Electrochem. Soc*. **152**, A2002–A2008 (2005).

40. Cai, L. & White, R. E. Reduction of model order based on proper orthogonal decomposition for lithium-ion battery simulations. *J. Electrochem. Soc*. **156**, A154–A161 (2009).

41. Smith, K. A., Rahn, C. D. & Wang, C.-Y. Model order reduction of 1D diffusion systems via residue grouping. *ASME J. Dyn. Syst., Meas. Control* **130**, 011012 (2008).

42. Forman, J. C., Bashash, S., Stein, J. L. & Fathy, H. K. Reduction of an electrochemistry based li-ion battery model via quasi-linearization and padé approximation. *J. Electrochem. Soc*. **158**, A93–A101 (2011).

43. Wang, C. Y., Gu, W. B. & Liaw, B. Y. Micro-macroscopic coupled modeling of batteries and fuel cells I. Model development. *J. Electrochem. Soc*. **145**, 3407–3417 (1998).

44. Guo, J., Li, Z. & Pecht, M. A bayesian approach for li-Ion battery capacity fade modeling and cycles to failure prognostics. *J. Power Sources* **281**, 173–184 (2015).

45. Wu, B., Han, S., Shin, K. G. & Lu, W. Application of artificial neural networks in design of lithium-ion batteries. *J. Power Sources* **395**, 128–136 (2018).

46. Zahid, T., Xu, K., Li, W., Li, C. & Li, H. State of charge estimation for electric vehicle power battery using advanced machine learning algorithm under diversified drive cycles. *Energy* **162**, 871–882 (2018).

47. Chemali, E., Kollmeyer, P. J., Preindl, M. & Emadi, A. State-of-charge estimation of li-ion batteries using deep neural networks: a machine learning approach. *J. Power Sources* **400**, 242–255 (2018).

48. Jiménez-Bermejo, D., Fraile-Ardanuy, J., Castaño-Solis, S., Merino, J. & Álvaro-Hermana, R. Using dynamic neural networks for battery state of charge estimation in electric vehicles. *Procedia Comput. Sci.* **130**, 533–540 (2018).

49. Mansouri, S.S., Karvelis, P., Georgoulas, G. & Nikolakopoulos, G. Remaining useful battery life prediction for UAVs based on machine learning. *IFAC-PapersOnLine* **50**, 4727–4732 (2017).

50. Donato, T. H. R. & Quiles, M. G. Machine learning systems based on xgBoost and MLP neural network applied in satellite lithium-ion battery sets impedance estimation. *Advanced Computational Intelligence: An International Journal (ACII)* **5**, 1–20 (2018).

51. Huang, C. et al. Robustness evaluation of extended and unscented Kalman filter for battery state of charge estimation. *IEEE Access* **6**, 27617–27628 (2018).

52. Ren, L. et al. Remaining useful life prediction for lithium-ion battery: a deep learning approach. *IEEE Access* **6**, 50587–50598 (2018).

53. Khumprom, P. & Yodo, N. A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm. *Energies* **12**, 660 (2019).

54. Sahinoglu, G. et al. Battery state-of-charge estimation based on regular/recurrent gaussian process regression. *IEEE Trans. Ind. Electron.* **65**, 4311–4321 (2017).

55. Álvarez Antón, J. C. et al. Battery state-of-charge estimator using the SVM technique. *Appl. Math. Model.* **37**, 6244–6253 (2013).

56. Tong, S., Lacap, J. H. & Park, J. W. Battery state of charge estimation using a load-classifying neural network. *J. Energy Storage* **7**, 236–243 (2016).

57. Kang, L., Zhao, X. & Ma, J. A new neural network model for the state-of-charge estimation in the battery degradation process. *Appl. Energy* **121**, 20–27 (2014).

58. Hu, X., Li, S. E. & Yang, Y. Advanced machine learning approach for lithium-ion battery state estimation in electric vehicles. *IEEE Trans. Transport. Electrific.* **2**, 140–149 (2016).

59.   Wu, T., Wang, M., Xiao, Q. & Wang, X. The SOC estimation of power li-ion battery based on ANFIS model. *Smart Grid and Renewable Energy* **3**, 51–55 (2012).

60.   Wu, J., Wang, Y., Zhang, X. & Chen, Z. A novel state of health estimation method of li-ion battery using group method of data handling. *J. Power Sources* **327**, 457–464 (2016).

61.   Hu, C., Jain, G., Schmidt, C., Strief, C. & Sullivan, M. Online estimation of lithium-ion battery capacity using sparse bayesian learning. *J. Power Sources* **289**, 105–113 (2015).

62.   Berecibar, M. et al. Online state of health estimation on NMC cells based on predictive analytics. *J. Power Sources* **320**, 239–250 (2016).

63.   Richardson, R. R., Osborne, M. A. & Howey, D. A. Gaussian process regression for forecasting battery state of health. *J. Power Sources* **357**, 209–219 (2017).

64.   Zhang, Y., Xiong, R., He, H. & Liu, Z. A LSTM-RNN method for the lithuim-ion battery remaining useful life prediction. *Prognostics and System Health Management Conference (PHM-Harbin) 2017*, 1–4 (2017).

65.   Hu, J. N. et al. State-of-charge estimation for battery management system using optimized support vector machine for regression. *J. Power Sources* **269**, 682–693 (2014).

66.   Tseng, K.-H., Liang, J.-W., Chang, W. & Huang, S.-C. Regression models using fully discharged voltage and internal resistance for state of health estimation of lithium-ion batteries. *Energies* **8**, 2889–2907 (2015).

67.   Hussein, A. A. Kalman filters versus neural networks in battery state-of-charge estimation: a comparative study. *Int. J. Mod. Nonlinear Theor. Appl.* **3**, 199–209 (2014).

68.   Yang, D., Wang, Y., Pan, R., Chen, R. & Chen, Z. A neural network based state-of-health estimation of lithium-ion battery in electric vehicles. *Energy Procedia* **105**, 2059–2064 (2017).

69.   Dawson-Elli, N., Lee, S. B., Pathak, M., Mitra, K. & Subramanian, V. R. Data science approaches for electrochemical engineers: an introduction through surrogate model development for lithium-ion batteries. *J. Electrochem. Soc.* **165**, A1–A15 (2018).

70.   Li, X., Wang, H., Gu, B. & Ling, C. X. Data sparseness in linear SVM. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence,* 3628–3634 (2015).

71.   Rendle, S. Factorization Machines. *Proceedings of the 2010 IEEE International Conference on Data Mining. IEEE Computer Society, Washington, DC*, 995–1000 (2010).

72.   Girard, A. & Murray-Smith, R. Gaussian processes: prediction at a noisy input and application to iterative multiple-step ahead forecasting of time-series. *Proceedings of the Hamilton Summer School on Switching and Learning in Feedback systems, Ed. R. Murray-Smith, R. Shorten, Springer-Verlag, Lecture Notes in Computing Science* **3355**, 158-184 (2005).

73. Dawson-Elli, N., Kolluri, S., Mitra, K. & Subramanian, V. R. On the creation of a chess-AI inspired problem-specific optimizer for the pseudo two-dimensional battery model using neural networks. *J. Electrochem. Soc*. **166,** A886-A896 (2019).

74. Wang, A., Kadam, S., Li, H., Shi, S. & Qi, Y. Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries. *npj Comput. Mater*. **4**, 15 (2018).

75. Kumar, H., Detsi, E., Abraham, D. P. & Shenoy, V. B. Fundamental mechanisms of solvent decomposition involved in solid-electrolyte interphase formation in sodium ion batteries. *Chem. Mater*. **28**, 8930-8941 (2016).

76. Hong, Z. & Viswanathan, V. Prospect of thermal shock induced healing of lithium dendrite. *ACS Energy Lett*. **4**, 1012-1019 (2019).

77. Liang, L. & Chen, L.-Q. Nonlinear phase field model for electrodeposition in electrochemical systems. *Appl. Phys. Lett*. **105**, 263903 (2014).

78. Takaki, T. Phase-field modelling and simluations of dendrite growth. *ISIJ Int.* **54**, 437-444 (2014).

79. Bai, P., Cogswell, D. A. & Bazant, M. Z. Suppression of phase separation in $LiFePO_4$ nanoparticles during battery discharge. *Nano Lett*. **11**, 4890-4896 (2011).

80. Cogswell, D. A. & Bazant, M. Z. Theory of coherent nucleation in phase-separating nanoparticles. *Nano Lett*. **13**, 3036-3041 (2013).

81. Cogswell, D. A. & Bazant, M. Z. Coherency strain and the kinetics of phase separation in $LiFePO_4$ nanoparticles. *ACS Nano* **6**, 2215-2225 (2012).

82. Ahmad, Z., Xie, T., Maheshwari, C., Grossman, J. C. & Viswanathan, V. Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes. *ACS Cent. Sci*. **4**, 996-1006 (2018).

83. Joshi, R. P. et al. Machine learning the voltage of electrode materials in metal-ion batteries. *ACS Appl. Mater. Interfaces* **11**, 18494-18503 (2019).

84. Aspuru-Guzik, A. & Persson, K. Materials acceleration platform: accelerating advanced energy materials discovery by integrating high-throughput methods and artificial intelligence. *Mission Innovation* **6**, 1–100 (2018).

85. Correa-Baena, J.-P. et al. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* **2**, 1410–1420 (2018).

86. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater*. **3**, 5–20 (2018).

87. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* **58**, 586–597 (2015).

88. Ley, S. V., Fitzpatrick, D. E., Ingham, R. J. & Myers, R. M. Organic synthesis: march of the machines. *Angew. Chem. Int. Ed. Engl*. **54**, 3449–3464 (2015).

89. Mannodi-Kanakkithodi, A., Pilania, G., Huan, T.D., Lookman, T. & Ramprasad, R. Machine learning strategy for accelerated design of polymer dielectrics. *Sci. Rep.* **6**, 20952 (2016).

90.   Shevlin, M. Practical high-throughput experimentation for chemists. *ACS Med. Chem. Lett.* **8**, 601–607 (2017).

91.   Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Materials* **1**, 011002 (2013).

92.   Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM*, **65**, 11, 1501-1509 (2013).

93.   Jain, A. et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci*. **50**, 2295-2310 (2011).

94.   Xiao, R. J., Li, H. & Chen, L. Q. Development of new lithium battery materials by material genome initiative. *Acta. Phys. Sin.* **67**, 128801 (2018).

95.   Shandiz, M. A. & Gauvin, R. Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries. *Comput. Mater. Sci*. **117**, 270-278 (2016).

96.   Takagishi, Y., Yamanaka, T. & Yamaue, T. Machine learning approaches for designing mesoscale structure of li-ion battery electrodes. *Batteries* **5**, 54 (2019).

97.   Okamoto, Y. Applying bayesian approach to combinatorial problem in chemistry. *J. Phys. Chem. A* **121**, 3299-3304 (2017).

98.   Allam, O., Cho, B. W., Kim, K. C. & Jang, S. S. Application of DFT-based machine learning for developing molecular electrode materials in li-ion batteries. *RSC Adv*. **8**, 39414 (2018).

99.   Gu, G. H., Noh, J., Kim, I. & Jung, Y. Machine learning for renewable energy materials. *J. Mater. Chem. A* **7**, 17096 (2019).

100.   Cheng, L. et al. Acclerating electrolyte discovery for energy storage with high-throughput screening. *J. Phys. Chem. Lett.* **6**, 283-291 (2015).

101.   Khetan, A., Luntz, A. & Viswanathan, V. Trade-offs in capacity and rechargeability in nonaqueous Li–O_2 batteries: solution-driven growth versus nucleophilic stability. *J. Phys. Chem. Lett.* **6**, 1254-1259 (2015).

102.   Schütter, C. et al. Rational design of new electrolyte materials for electrochemical double layer capacitors. *J. Power Sources* **326**, 541-548 (2016).

103.   Okamoto, Y. & Kubo, Y. Ab initio calculations of the redox potentials of additives for lithium-ion batteries and their prediction through machine learning. *ACS Omega*, **3**, 7868-7874 (2018).

104.   Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater*. **12**, 191-201 (2013).

105.   Qu, X. et al. The electrolyte genome project: a big data approach in battery materials discovery. *Comput. Mater. Sci.* **103**, 56-67 (2015).

106.   Cubuk, E. D., Sendek, A. D. & Reed, E. J. Screening billions of candidates for solid lithium-ion conductors: a transfer learning approach for small data. *J. Chem. Phys*. **150**, 214701 (2019).

107. Jalem, R. et al. Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application. *Sci. Rep.* **8**, 5845 (2018).

108. Sendek, A. D. et al. Machine learning-assisted discovery of solid li-ion conducting materials. *Chem. Mater.* **31**, 342-352 (2019).

109. Liu, P. et al. High throughput materials research and development for lithium ion batteries. *J. Materiomics* **3**, 202–208 (2017).

110. Lyu, Y., Liu, Y., Cheng, T. & Guo, B. High-throughput characterization methods for lithium batteries. *J. Materiomics* **3**, 221–229 (2017).

111. Grey, C. P. & Tarascon, J. M. Sustainability and in situ monitoring in battery development. *Nat. Mater.* **16**, 45–56 (2016).

112. Wang, X., Xiao, R., Li, H. & Chen, L. Discovery and design of lithium battery materials via high-throughput modeling. *Chinese Phys. B.* **27**, 128801 (2018).

113. Schiele, A. et al. High-throughput in situ pressure analysis of lithium-ion batteries. *Anal. Chem.* **89**, 8122–8128 (2017).

114. Roberts, M. & Owen, J. High-throughput method to study the effect of precursors and temperature, applied to the synthesis of $LiNi_{1/3}Co_{1/3}Mn_{1/3}O_2$ for lithium batteries. *ACS Comb. Sci.* **13**, 126–134 (2011).

115. Maruyama, S., Kubokawa, O., Nanbu, K., Fujimoto, K. & Matsumoto, Y. Combinatorial synthesis of epitaxial $LiCoO_2$ thin films on $SrTiO_3(001)$ via on-substrate sintering of Li_2CO_3 and CoO by pulsed laser deposition. *ACS Comb. Sci.* **18**, 343–348 (2016).

116. Vogt, S. et al. Composition characterization of combinatorial materials by scanning X-ray fluorescence microscopy using microfocused synchrotron X-ray beam. *Appl. Surf. Sci.* **223**, 214–219 (2004).

117. Orikasa, Y. et al. Direct observation of a metastable crystal phase of Li_xFePO_4 under electrochemical phase transition. *J .Am. Chem. Soc.* **135**, 5497–5500 (2013).

118. Kwade, A. et al. Current status and challenges for automotive battery production technologies. *Nat. Energy* **3**, 290–300 (2018).