Chinese Journal of Chemical Engineering 37 (2021) 1–11

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Review

Machine learning of materials design and state prediction for lithium ion batteries

Jiale Mao [1], Jiazhi Miao [2], Yingying Lu [1,*], Zheming Tong [2,*]

[1] State Key Laboratory of Chemical Engineering, Institute of Pharmaceutical Engineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China
[2] State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 24 February 2021
Received in revised form 19 April 2021
Accepted 22 April 2021
Available online 2 May 2021

Keywords:
Lithium ion batteries
Machine learning
Materials design
State prediction

ABSTRACT

With the widespread use of lithium ion batteries in portable electronics and electric vehicles, further improvements in the performance of lithium ion battery materials and accurate prediction of battery state are of increasing interest to battery researchers. Machine learning, one of the core technologies of artificial intelligence, is rapidly changing many fields with its ability to learn from historical data and solve complex tasks, and it has emerged as a new technique for solving current research problems in the field of lithium ion batteries. This review begins with the introduction of the conceptual framework of machine learning and the general process of its application, then reviews some of the progress made by machine learning in both improving battery materials design and accurate prediction of battery state, and finally points out the current application problems of machine learning and future research directions. It is believed that the use of machine learning will further promote the large-scale application and improvement of lithium-ion batteries.

1. Introduction

Today, the emergence of portable electronics and electric vehicles has greatly contributed to the development and application of rechargeable batteries, such as lead acid, nickel cadmium, nickel metal hydride and lithium ion batteries (LIBs) [1–4]. LIBs are gradually become the mainstream of battery development as its high energy density, high energy efficiency, no memory effect, low self-discharge rate and wide operating temperature range, *etc*. As an important part of global new energy vehicles and other electronic products, consumers have increasingly higher requirements on LIBs material performance and efficiency and safety during actual use. Current research issues in the field of LIBs are mainly focused on the following two areas.

On the one hand, there is a need to accelerate the innovation and optimization of battery materials. The research and development of traditional materials are mainly based on experience and repeated 'trial and error' experiments. Which consists of seven stages: discovery, development, property optimization, system design and integration, certification, manufacturing and deployment [5]. Different stages would be carried out by research teams from diverse institutions. As a result, it usually takes 10 ~ 20 years for a new material from discovery to application, which is quite a long time and cannot meet the rapidly developing demand of the market [6]. Since the 1980s, the interdisciplinary integration of materials, physics and computational science has greatly facilitated the development of computational simulation methods. Common simulation methods include first principle calculation, molecular dynamics, quantum mechanics and so on. Compared with 'trial and error' methods, computational simulations allow experiments to be carried out with completely control of the variables involved, accelerating the study of battery materials. However, the cost of a single experiment and related calculation is often very high, and continuous attempts based on limited knowledge or experience will increase the number of experiments or calculations, resulting in a waste of resources [7].

The other aspect is the need for accurate prediction of battery state. With the widespread use of LIBs, the efficiency and safety of LIBs in practical applications is becoming a key concern, which requires the construction of advanced battery management systems (BMS) that can accurately predict the state of charge (SOC), state of health (SOH) and remaining useful life (RUL) of the battery to ensure the continued safe and efficient use of the battery. The most commonly used battery condition prediction methods are equivalent circuit models (ECMs) and physical-based models

* Corresponding authors.
E-mail addresses: yingyinglu@zju.edu.cn (Y. Lu), tzm@zju.edu.cn (Z. Tong).

https://doi.org/10.1016/j.cjche.2021.04.009

(PBMs). ECMs based on circuit elements are the most widely used models to predict battery state online. ECMs are computationally efficient, but the prediction accuracy is limited due to the lack of physical information on system state and parameters [8]. PBMs are based on the electrochemical physics of the battery and although they can accurately predict the battery state, the computational cost is greatly increased due to the number of control equations [9]. The challenge in predicting battery state models is how to balance the fidelity of the model with the computational complexity [10].

With the rapid development of artificial intelligence technology, machine learning (ML) as its core technology has been widely used in biology, medicine and other fields [11–13], which has greatly promoted the development of corresponding industry. ML also offers new ways to solve problems in the field of battery research. At present, the application research of ML in the field of battery is still in the preliminary stage. This paper is mainly based on the application of ML in battery materials and battery state. The application of ML in enhancing the performance of LIBs, the existing problems and research prospects are summarized. This review provides significant references for researchers to understand the importance of ML in the application of lithium batteries.

2. Machine Learning

This section provides a brief introduction to ML and the general process of its algorithmic implementation in LIBs. ML is a multidisciplinary and convergent discipline involving several disciplines such as probability theory, statistics, approximation theory, convex analysis and algorithmic complexity theory. Essentially, ML is a statistical model for data analysis and prediction that enables computers to automatically learn knowledge from existing datasets and make predictions for specific learning tasks. Currently, the ML algorithms commonly used in battery research are supervised learning and unsupervised learning. Supervised learning refers to the ML problem of learning predictive models from labelled data and includes algorithms such as support vector machines, Naive Bayes, linear regression and decision trees [14]. Unsupervised learning, on the other hand, refers to the ML problem of learning a training model from unlabeled data, used to achieve, for example, the identification and classification of data trends, commonly used in battery research as clustering algorithms [15].

The general process of implementing ML algorithms for practical problems includes: datasets, feature engineering, training algorithmic models and model evaluation [16–18]. The application of ML starts with the construction of a complete ML dataset, the quality of which directly influences the quality of the data analysis and mining. After building the dataset, we need to transform the material features into feature parameters or descriptors that can be recognized by programs and algorithms. Primitive features or descriptors usually have sparse, poorly correlated and redundant information. Evaluation of sparsity, relevance and redundancy of raw features and integration of domain expert knowledge to optimize feature selection will further improve model prediction performance [19]. Again, depending on the characteristics of the data, one or more ML algorithms need to be selected to implement the mapping between the input and output variables. Usually the more complex the model selected for training, the better the model fits the training data, but when the model is too complex, the generalization ability of the model decreases, a situation known as overfitting, and vice versa known as underfitting. Finally, the trained models are evaluated using test data to verify the validity of the models, and the best model is selected based on the scores of multiple models. A good understanding of the algorithm and the problem will help in the selection of the algorithm and the function to be used according to the actual research problem. As shown in Table 1, we have summarized the advantages and disadvantages of commonly used ML algorithms and the basic functions they involve in lithium batteries and other fields.

3. Machine Learning in Battery Materials

3.1. Screening the battery materials

All solid state lithium ion batteries (ASSLBs) are considered to be the next generation of energy storage devices with advantages such as high safety and high energy density, and are currently a popular research direction in the field of LIBs [20–22]. Solid state electrolytes are a key part of ASSLBs, for which they need to meet: high ionic conductivity, stable interfacial properties and other requirements. The selection of suitable materials is currently the main technical bottleneck in the development of solid-state battery technology. ML can rapidly screen out suitable candidates from large material databases based on the required electrolyte properties, greatly reducing experimental cycles and costs, and can effectively solve the problem of difficult material selection [23,24]. The importance of materials data for science and technology innovation, smart manufacturing and other fields is becoming increasingly evident in the data age. Many open-source material databases such as the Inorganic Crystal Structure Database, Materials Project and Total Materia have been established by major research institutes, which include energy bands, energy gaps, crystal structures and fundamental physical properties data. Zhang *et al.* [25] used unsupervised learning to screen out all Li^+-containing compounds from a database of inorganic crystal structures, and based on the anion lattice structure, which is one of the key factors affecting ionic conductivity. The more than 12,000 Li^+-containing compounds from the Inorganic Crystal Structure Database were divided into two groups with high ionic conductivity and low ionic conductivity, and then combined with ab initio molecular dynamics (AIMD) simulations to quantifyσ_{RT}, 16 solid-state electrolyte candidates with σ_{RT} exceeding 10^{-4} $S{\cdot}cm^{-1}$ as shown in the solid symbols in Fig. 1 (a) were identified, three of which even exceeded 10^{-2} $S{\cdot}cm^{-1}$, which is comparable to the highest known solid state electrolyte materials in terms of ionic conductivity. Sendek *et al.* [26] found that ML-based guided search is 2.7 times more likely to identify materials with high ionic conductivity than a random search, and the logistic regression model F1 score was 3.5 times higher than a random search. Identifying high ionic conductivity materials from their atomic structure such as Li − Li bond number per Li, average sublattice bond ionicity and the average coordination of the anions, the ML model identified them 1000 times faster than the materials science PhD. The experiment is shown in Fig. 1 (b), which shows that a logistic regression model trained using only a small amount of sample data can achieve fast and accurate screening of high ionic conductivity materials. In a subsequent study, it was found that the ML based model accuracy, recall and F1 scores were all much higher than the material space random search and the Materials Science Ph.D., which is a strong indication that ML models can greatly speed up the high ionic conductivity material selection process.

In lithium metal batteries, most safety issues are caused by uncontrolled lithium dendrite generation, which can lead to short circuits within the cell [27,28]. Studies have shown that once lithium dendrites start to grow, it is difficult to retard their growth tendency [29], and therefore it is important to prevent dendrite generation to ensure smooth electrodeposition throughout the cell cycle. Starting from the perspective of stable electrodeposition, Ahmad *et al.* [30] input the isotropy and anisotropy of the material interface, *etc.* as stability parameters into a crystal graph

Table 1
Summary of common ML algorithms

ML algorithm	Function	Advantages	Disadvantages
Naïve Bayes	Bayes' Theorem	Stable classification efficiency, good performance for small data sizes, less sensitive to real data, simpler algorithms	Need to calculate prior probabilities, classification decisions have error rates and are sensitive to the form of representation of the input data
Logistic regression	Sigmoid function	Simple implementation, low computational effort, high speed and low storage resources	Poor performance when feature space is large; easily under-fitted and generally less accurate
Linear regression	$Y = \omega X + b$	Simple to implement, simple to calculate	Not suitable for fitting non-linear data
K-nearest neighbor	Euclidean distance	No assumptions on data, high accuracy, can be used for non-linear classification; mature theory	Computationally intensive, prone to sample imbalance problems and requires large amounts of memory
Decision trees	Information gain, Gini coefficient	Computationally simple, easy to understand and highly interpretable; able to handle uncorrelated features and better suited to handle uncorrelated features	Tends to over-fit and ignore correlations between data
Support vector machine	Kernel function	Can solve high-dimensional problems; can handle interactions of non-linear features; can improve generalization	Less efficient with larger sample sizes; finding the right kernel function; sensitive to missing data
Artificial neural networks	Sigmoid function, tanh function, ReLU	High classification accuracy, robust and fault-tolerant to noisy nerves, and able to adequately approximate complex non-linear relationships	Requires a large number of parameters; cannot observe the learning process inside the network, output is difficult to interpret; long learning time
Random forests	Multiple decision trees bagging	Regression and classification can be performed. No need to adjust parameters repeatedly. No scaling of data required	For sparse data with very high dimensionality (*e.g.* text data), random forests often do not perform very well
Convolutional neural networks	Convolutional layers, Pooling layers	Parameter sharing and sparse connections result in a significant reduction in training parameters. With translational invariance	The presence of pooling layers can lead to the loss of a lot of very valuable information, as well as ignoring the connections between the whole and the parts
Recurrent neural networks	$s_t = f(s_{t-1}, x_t, \theta)$	Deep models in the time dimension that can model sequence content	More parameters to train, prone to gradient disappearance and gradient explosion problems; no feature learning capability

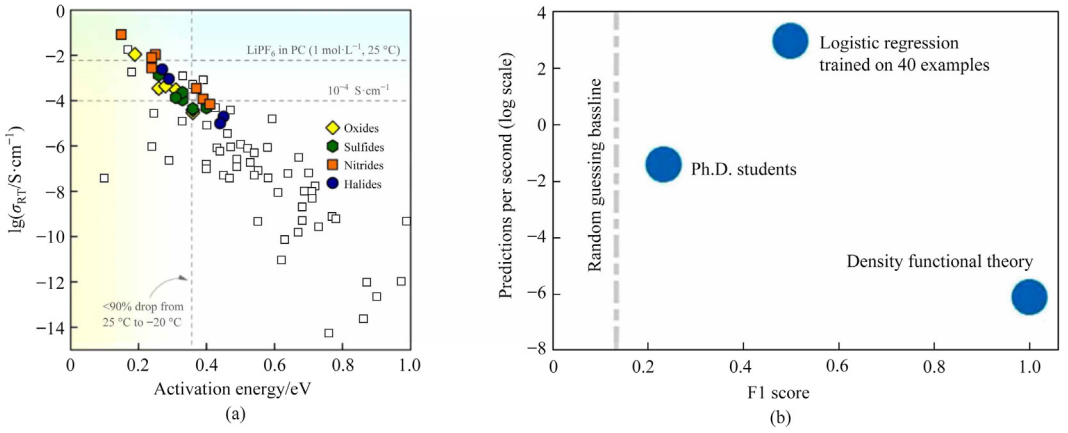

Fig. 1. (a) Unsupervised learning finds σ_{RT} and activation energy calculated for electrolyte materials (filled symbols) compared to past reported materials (hollow symbols) [25]. (b) Comparison of F1 scores and prediction per second for screening materials with high ionic conductivity from Li^+-containing compounds for logistic regression trained on 40 examples and materials science Ph.D. students, DFT calculations [26].

convolutional neural network prediction model based on shear modulus and bulk modulus and the elastic constants of the cubic material were trained using gradient augmented regression and kernel ridge regression. The model uses isotropic stability criteria and anisotropic stability criteria to screen six solid materials and over 20 interfaces from 12,950 solid materials and over 15,000 interfaces, respectively, that have the potential to inhibit lithium dendrite generation. Considering how to achieve a stable solid electrolyte interphase (SEI) between electrolyte and electrode, Liu *et al.* [31] developed a ML model based on support vector machine and kernel ridge regression to screen and evaluate the possible reactions and thermodynamic stability of the Li|LLZOM interface under different chemical conditions. The results show that the stability of LLZOM to Li metal depends only on the dopant and that strong hetero-oxygen bonding incorporation can improve the thermodynamic stability of the interface, which is confirmed by DFT calculations.

In addition, ML has been applied in the screening of electrode coating materials. The coating effectively reduces the interfacial impedance between the electrodes and the electrolyte, increasing the cyclability of the battery [32]. Wang *et al.* [33] proposed a workflow combining a real-time dynamic ML model based on moment tensor potentials with molecular dynamics (MD) simulations to calculate the diffusion mobility of lithium ions, which gradually screened two candidate coating materials with excellent performance based on properties such as Li molar fraction, oxidation potential, interfacial stability, ionic conductivity, etc. The workflow progressively identifies $Li_3Sc_2(PO_4)_3$ and $Li_3B_7O_{12}$ as the top performance candidates. In addition, the average absolute error of the migration energy calculated by this 'learning on the fly'-MD has been reduced from 0.32 eV to 0.13 eV compared with that of high-temperature AIMD, which has significantly improved the calculation accuracy and reduced the cost. The ability of $Li_3B_7O_{12}$ to

provide excellent interfacial stability has been demonstrated in the study by Xiao *et al* [34].

3.2. Prediction of material properties

For battery materials, there are potentially thousands of compounds that can be synthesized by inserting metal ions (Li, Na, *etc.*) or changing the molar fraction of the constituent elements, many of which have not yet been synthesized, and it is even more difficult to characterize the materials, so it is extremely important for battery research to combine previous experimental data to predict the properties of new materials more accurately [35]. Using a ML model based on support vector regression and Gaussian kernel function algorithm, Fujimura *et al.* [36] investigated theoretical data D_{1600}, T_C, and V_{dis} (D_{1600} is the diffusion coefficient at 1600 K, T_C is the transition temperatures, V_{dis} is average volume of disordered structures) for 72 different compositions of LISICON-type electrolytes and experimental data for 95 different ionic conductivity measurements at different temperatures to predict σ_{373} with a prediction error of 0.373. The experimental prediction of σ_{373} is the highest for Li_4GeO_4 among the 72 components, which is better than the already synthesized $Li_{3.5}Zn_{0.25}GeO_4$. Jalem *et al.* [37] further introduced complex material microstructure parameters such as born effective charge, average bond length, bond angle, *etc.* as input feature variables and used a combined DFT and neural network based model to predict the Li diffusion potential and cohesive energy of $LiMXO_4$ (M-main group elements, X-group 14 and 15 elements) structured electrolyte materials, which was shown to be more accurate in capturing the nonlinear relationship between input and output variables than the DFT-PLS (multivariate partial least squares regression) prediction model.

In addition to using ML to predict the ionic conductivity of electrolyte materials, Joshi *et al.* [38] also attempted to predict the electrode voltage of lithium-ion batteries. The study compared three ML algorithms of deep neural networks, support vector machine and kernel ridge regression with models that characterized a particular electrode material with 80 characteristic parameters including: working ions within the cell, concentration of active metal ions within a given fraction, lattice type and space group numbers. An error analysis of the three models found that the model that bases on deep neural network predicted electrode voltages with better accuracy than support vector machine and kernel ridge regression. Joshi *et al.* also provide a web-based tool in the battery community that can quickly and accurately predict electrode voltages for li-, Na- and Al-ions, requiring only the stoichiometry of the material at low concentrations, metal ion cell type, lattice type, *etc.* parameters. In addition, ML models can obtain the same trends as DFT calculations in fitting material properties such as thermodynamic stability of arbitrary chemicals [23], binding energies of multiple compounds [39], melting temperatures of simple component solids [40] and coordination energies of alkali group metals in battery electrolytes to achieve accurate predictions [41].

3.3. Calculating the optimum composition of composite battery materials

Pure component battery materials are generally difficult to meet the needs of battery design, so they need to be used in the form of multiple components or additives, which in turn improve the electrochemical performance of the material. Experiments and DFT calculations are difficult to search for the multi-component ratio or the amount of additives with the best performance from a complex composition space, and the introduction of ML can significantly reduce the number of calculation iterations and effectively solve the material improvement problem. Sumita *et al.* [42] found that oxide electrolytes are more stable in the high voltage cathode region compared to sulfide electrolytes, but this may also lead to a significant reduction in the ionic conductivity of the oxide. It has been shown that mixing lithium salts containing oxygen can effectively increase the ionic conductivity of the electrolyte. So Sumita *et al.* [43] selected 15 groups of electrolytes with different composition ratios from the Li_3PO_4-Li_3BO_3-Li_2SO_4 ternary hybrid system and calculated their ion conductivity values at 600 K using density functional molecular dynamics. By using a ML model based on Gaussian process and Bayesian optimization algorithm, the conductivity values of electrolyte materials with different hybrid ratios were interpolated to simulate the conductivity values, and the possibility of high ionic conductivity was investigated based on the z-score values, and it was found that the conductivity of the ternary system was significantly higher than that of the pure component system and the binary system. Where $z-\text{score} = [\sigma(x) - \sigma_{max}]/\delta(x)$, σ_{max} = 193.0 S·cm^{-1} is the calculated maximum value of ionic conductivity for 15 composition ratios; $\delta(x)$ is the standard deviation at point x; $\sigma(x)$ is the predicted value of lithium ion conductivity prediction as a function of component composition ratios; x = (A, B, C) where A, B and C are the mass molar fractions of each component.

The combination of high-throughput experiments and ML offers the possibility to explore the optimal solution for the design space of complex systems. Whitacre *et al.* [44] designed a fully automated controlled test stand based on ML to automatically mix and measure electrochemical properties such as conductivity and electrochemical stability windows for binary electrolytes of sulfate and nitrate, which greatly reduces the time required to search for the design space of binary mixed electrolytes. On this basis, Dave *et al.* [45] constructed an automated experimental platform based on Bayesian optimization algorithm in Fig. 2 to search the experimental design space for ternary lithium salt ($LiNO_3$, $LiClO_4$, Li_2SO_4) and quaternary sodium salt ($NaClO_4$, $NaNO_3$, Na_2SO_4 and NaBr) electrolyte systems. The Bayesian optimization algorithm provides real-time feedback on the results of each experiment to achieve reverse material design during the experiment, reducing the original need to search the ternary electrolyte design space for all 62,000 experiments to more than 250. van Duong *et al.* [46] also applied ML to explore the effect of adding additives vinyl carbonate (VC) and lithium dioxalate borate to a carbonate-based electrolyte on the cycling performance of lithium-ion batteries ($LiNi_{0.5}Mn_{1.5}O_4$ oxide and graphite cathode), effectively reducing the experimental cycle time, and the experimentally designed artificial neural network model can be further extended to optimize other parameters of LNMO/graphite and LNMC622/graphite batteries.

3.4. Material microstructure analysis

The microstructure of materials, such as crystal structure [47] and polymer molecular weight [48], also has an important influence on the physical and chemical properties of materials. Using ML to assist in the statistical analysis of the patterns between microstructure and material physical and chemical properties can further guide the design of batteries and the improvement of battery performance. Hatakeyama-sato *et al.* [49] constructed the largest ML database of Li^+ conducting polymers, which contains all the information on the relationship between chemical structure and conductivity, covering most of the basic chemical structures of solid electrolytes: polyanions, anions, neutral polymers, *etc.* A ML model based on a Gaussian process model and a transfer-learned graph neural network was designed to discover the relationship between chemical structure, temperature, composition ratio, *etc.* and ionic conductivity in Fig. 3. Charge-transfer complexes of polyphenylene sulfide (PPS) derivatives and aromatic oxidants were

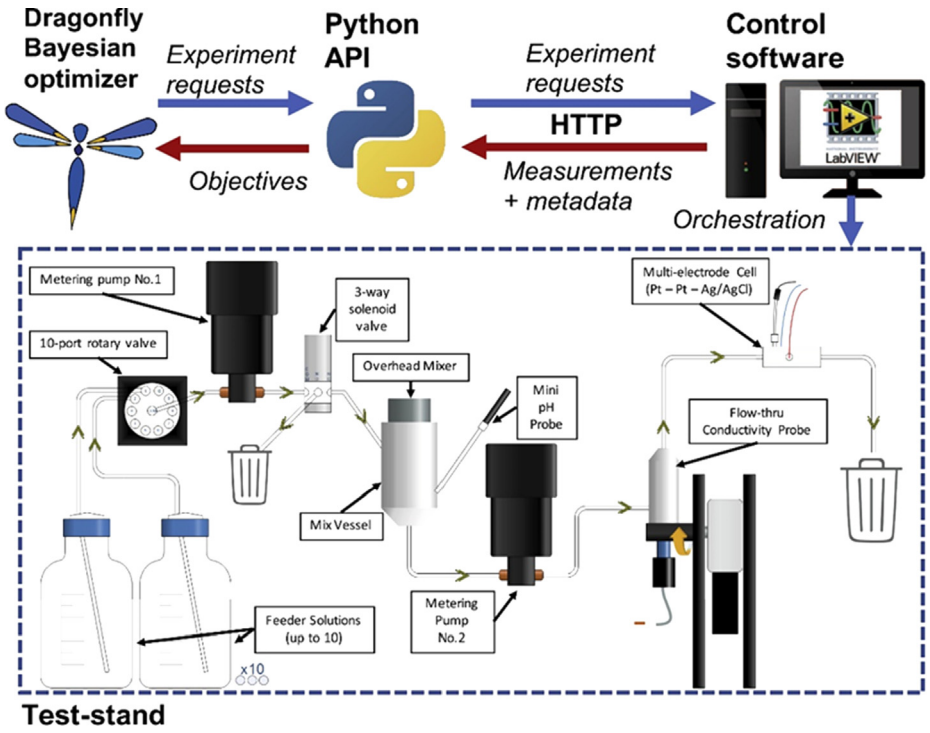

Fig. 2. Schematic diagram of the robot automated experimental platform and Bayesian optimization combined to achieve closed-loop inverse material design. The robot platform automatically runs the experiments, stores and transmits the experimental data to the Bayesian optimization model, and the ML model provides real-time feedback to adjust the direction of the experiments [45].

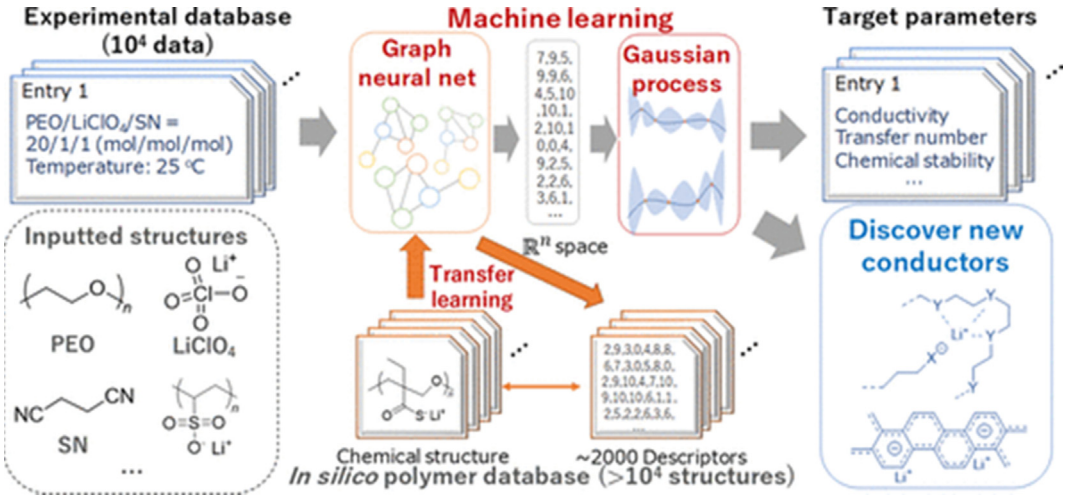

Fig. 3. A transfer-learned graph neural network and gaussian process based on ML model to predict polymer electrolyte material physicochemical properties and discover new material scenarios from within an experimentally constructed material structure databases [50].

found to have better lithium ion transport (>10^{-4} S·cm^{-1}), but PPS has previously been considered as an insulating engineering material. In addition, a more important finding is that the charge transfer complexes dimethyl-substituted PPS in the glassy state also have a higher ionic conductivity of 10^{-3} S·cm^{-1}, better than some of the crystalline polymers, and are stable over a wide temperature range. As conventional rubbery electrolyte designs typically require lower transition temperatures, this also leads to the need to reconsider the design strategy for glassy electrolyte materials. ML based unbiased predictions to obtain glassy polymer electrolyte materials with high ionic conductivity will require more experiments in the future to explore their application prospects. Wang *et al.* [50] combined coarse-grained molecular dynamics simulation and Bayesian optimization to predict the relationship between molecular-level material properties such as molecular size and intermolecular interactions with electrolyte performance. Coarse-grained can retain information at the molecular level of the material, and Bayesian algorithms can efficiently explore the high-dimensional design space. The model further proposes the principles of altering TFSI-, introducing secondary sites and replacing PEO backbones to further improve the ionic conductivity of PEO-LiTFSI materials.

Takagishi *et al.* [51] attempted to use ML to study the mesoscale structure of LIBs electrodes. Experiments were carried out by Matrixlaboratory to generate 2100 3D artificial electrode structure models in 50 μm × 50 μm × 50 μm space based on the volume

fraction and radius of active component particles and binder volume fraction, *etc*. A neural network with three hidden layers and a Bayesian optimization algorithm were constructed to inverse analyze the process parameters of total specific resistance to obtain the experimental results in Fig. 4. It was found that the total specific resistance decreases and then increases as the volume ratio increases, so that there is a minimum value of total specific resistance, similarly the total specific resistance has the same result for the diameter of the active component particles and the pressure during compaction. Attarian Shandiz *et al.* [52] have also effectively demonstrated a strong correlation between the physical properties (formation energy, band gap, *etc*.) of the crystal system and the lithium ion silicate cathode using a ML based statistical model. Wang *et al.* [53] combined *ab initio* calculations and machine learning to investigate the effect of discharge product structure and solvation effects on the reaction kinetics of lithium-air batteries. They chose the Gradient Boosting Decision Tree algorithm, which performs well in small sample size and gives feature importance, to predict the binding energy between solvent molecules and LiOH. The input features of the model include parameters such as molecular properties, dipole moments and atomic properties of functional groups. It was found that the functional group of the solvent molecule is the key to modulating the solvation ability and that phosphate solvents are the most effective in promoting the kinetics of LiOH decomposition.

4. Machine Learning in Battery State Prediction

The BMS is the vital link used to connect and manage lithium-ion batteries and other electrical components such as electric vehicles. One of the key functions of BMS is to predict the SOC, SOH and RUL of lithium-ion batteries [10]. The following Eqs. (1) and (2) are commonly used in the industry to define SOC and SOH, while RUL is the remaining time or number of cycles during which the maximum capacity of the battery decays to 70%–80% of the battery's rated capacity during the current cycle [54]. Battery state prediction can significantly improve the stability of battery use, which is currently a hot topic in battery research. However, as Li-ion batteries are a complex system, the irreversible change in battery state is caused by several interacting processes such as lithium inhomogeneous deposition and side reactions, so it is not possible to measure battery state directly and accurately. Data-based ML models are able to fit functions of input and output variables flexibly and efficiently (Fig. 5) and have become an effective way to solve battery condition prediction problems in recent years [55,56].

$$\mathrm{SOC} = \frac{C_{\mathrm{curr}}}{C_{\mathrm{full}}} \times 100\% \tag{1}$$

$$\mathrm{SOH} = \frac{C_{\mathrm{full}}}{C_{\mathrm{nom}}} \times 100\% \tag{2}$$

where C_{curr} refers to the remaining capacity of the battery within the current cycle, C_{full} refers to the maximum capacity of the battery for the current cycle and C_{nom} refers to the rated capacity of the new battery.

4.1. SOC predictions

Battery SOC is used to describe how much capacity is left in the battery for the current cycle. SOC prediction is similar to the function of a car's fuel gauge. Accurate estimation of SOC protects the battery from overcharging and discharging, thus extending battery life. Methods commonly used to estimate the SOC of Li-ion batteries include the open circuit voltage method, the ampere-time integration method, model-based methods and data-based methods. In the estimation of battery SOC, open circuit voltage requires the battery to be left to stand for a period of time for each measurement [57] and ampere-time integration method can suffer from initial and cumulative errors that are difficult to eliminate [58], making it difficult for both methods to accurately predict battery condition directly online. Model-based methods such as the single particle model and pseudo two-dimensional model are limited by the number of partial differential equations describing the internal state of the battery, which significantly increase the computational

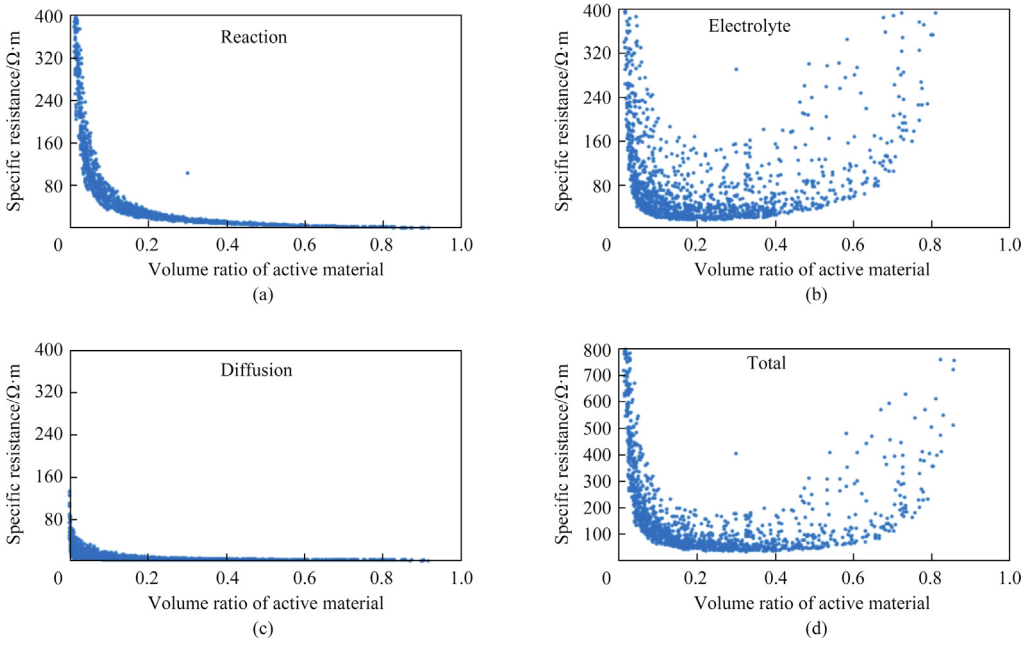

Fig. 4. Scatter plot of fitted resistance data for 2100 different volume fractions of active components, $R_{\mathrm{tot}} = R_{\mathrm{l}} + R_{\mathrm{reac}} + R_{\mathrm{diff}}$,$R_{\mathrm{tot}}$ is the total resistance, R_{l} is the electrolyte resistance, R_{reac} is the reaction resistance and R_{diff} is the lithium ion diffusion resistance. The electrode resistance is much less than the electrolyte resistance and can be ignored in the calculation of the total specific resistance [52].

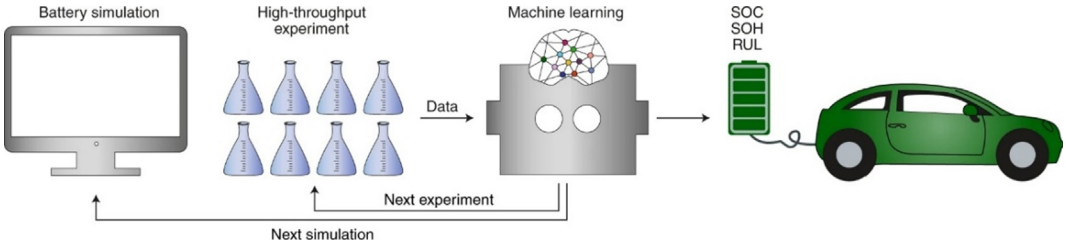

Fig. 5. Data-driven ML models to predict battery SOC, SOH and RUL scenarios, with the data required to train the models coming from simulations of multi-scale battery modelling and high-throughput automated experiments [10].

cost [59]. The accuracy of filtering algorithms such as Kalman filter [60] and particle filter [61] depends on the quality of the battery model, and the simplified models of batteries proposed at this stage are only applicable to some specific conditions such as isothermal. ML, on the other hand, does not require a priori knowledge of the battery and can be trained with large amounts of data to achieve fast and accurate prediction of the battery state. Many researchers have considered the introduction of ML algorithms to solve the SOC prediction problem.

Álvarez Antón *et al.* [62] constructed a support vector regression-based ML model to estimate battery SOC. The model chose a kernel function (radial basis function) to implement higher dimensional eigenspace regression and used voltage, current and temperature data as input variables. The support vector machine model fits very well on the test dataset, with the maximum error remaining below 6% and an RMSE of 0.71%. Considering the significant influence of different operation modes (idling, charging, and discharging) on the battery behavior, Tong *et al.* [63] divided the input layer of the neural network into three layers corresponding to the three operation models of the battery according to the input current threshold, and the average error of the SOC estimation results reached 3.8%. This scheme improves the adaptability of ML models to different sources of data. He *et al.* [64] tried to combine the filtering algorithm and ML algorithm to further improve the prediction accuracy. The experiment uses an unscented Kalman filter to eliminate the error of neural network estimation. After unscented Kalman filter filtering, the RMSE error of the SOC prediction value is reduced to within 2.5%, and the maximum error at multiple temperatures is also reduced to within 3.5%. In addition to combining filtering algorithms, Song *et al.* [65] combined two ML algorithms, convolutional neural network (CNN) and long short term memory (LSTM), where CNN was used to extract high-level spatial features from the original data and LSTM was used to model the relationship between SOC and historical input data. The overall prediction results had an RMSE of less than 2% and MAE of less than 1%. The experimental results of the combined CNN-LSTM model and the CNN and LSTM alone for predicting different initial SOC batteries are shown in Fig. 6, which shows that the proposed model outperforms the LSTM and CNN alone for estimating cell SOC, and the CNN-LSTM model can better respond to the effect of temperature on battery SOC. In addition, with the popularity of optimization techniques, the combination of ML and optimization techniques is becoming an effective solution to the computational complexity of ML algorithms. M. A. Hannan *et al.* [66] combined the Recurrent nonlinear autoregressive with exogenous inputs neural network algorithm (RNARX) and lightning search algorithm (LSA), LSA significantly reduces the number of iterations for ML algorithms to obtain the optimal hyperparameters, and the RMSE of the optimized ML for SOC prediction results under different battery test experiments (constant discharge test, hybrid pulse power characterization test and dynamic stress test) is between 0.4%–0.9% and MAE between 0.3%–0.6%. The model outperforms current state-of-the-art ML methods in terms of RMSE and MAE reduction, while RNARX-LSA is also able to predict SOC accurately under different operating conditions.

4.2. SOH, RUL predictions

LIBs are subject to irreversible ageing as they are used more and more, with consequent safety and performance issues. Therefore there is a need to accurately predict battery SOH in real time during the life of the battery. In contrast to SOC prediction over a short time span of a single charge/discharge cycle, SOH prediction is performed over multiple charge/discharge cycles over a longer time span and therefore SOH prediction requires data covering hundreds or thousands of cycles over the entire battery life. Data-based ML methods are an effective way to solve the SOH prediction problem. Landi *et al.* [67] constructed neural network models with different numbers of neurons in the hidden layer and obtained the best predictive neural network model with 12 neurons in the hidden layer after training. This model uses discharge depth, temperature and current as input parameters, and the error in testing the model is less than 5%. The predictions of the neural network model are still a good fit for the actual values on a subset of experimental data with different discharge depths. Since the neural network algorithm is based on gradient learning, the problem of convergence results to local extremes and estimation errors may occur. Pan *et al.* [68] used the extreme learning machine algorithm with a simpler model structure to predict SOH, and experimentally found that the maximum error of extreme learning machine was 2.22%, MAE was 1.72% and RMSE was 0.0109, which were all better than back propagation neural network. In addition, the prediction time of extreme learning machine was reduced to 0.0136 s. In addition, algorithms such as deep neural network [69] and ensemble learning [70] have also been applied to battery SOH prediction, further improving the accuracy of SOH prediction. Dong *et al.* [71] considered combining ML and filter algorithms to improve the original SOH estimation method based on filtering algorithms, and experimentally implemented a support vector regression-particle filter algorithm, the introduction of support vector regression effectively avoids the degeneracy phenomenon problem caused by using particle filter alone, maintains the diversity of particles, and support vector regression-particle filter shows higher prediction accuracy and robustness than particle filter. Since ML can fit more accurate battery equations of state from historical data, Michel *et al.* [72] found that the introduction of equations of state derived from ML made the Kalman filter estimation method more adaptive.

In essence, RUL prediction for LIBs is similar to SOH estimation studies, because when SOH is reduced to 70%–80% of rated capacity, the battery is also considered to have reached retirement, so predicting the remaining life of LIBs is also predicting how many more cycles the battery will go through before SOH will drop to the specified capacity. Therefore this section focuses on the

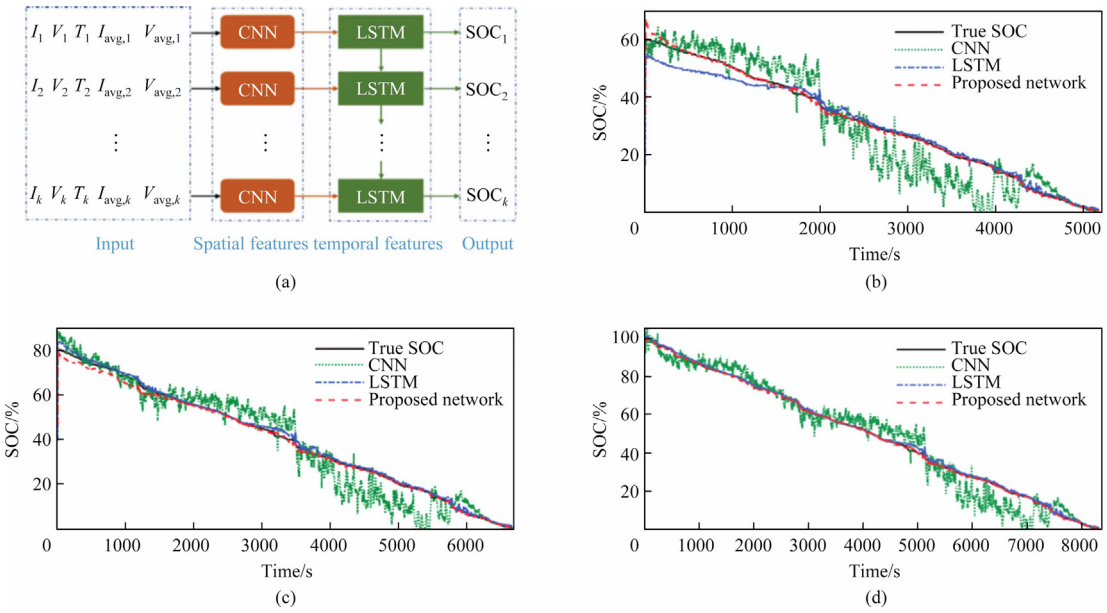

Fig. 6. CNN-LSTM combined model structure and prediction results for batteries with different initial SOC. (a) structure of the experimentally proposed CNN-LSTM network model, (b) prediction results for an initial battery SOC of 60%, (c) prediction results for an initial battery SOC of 80%, (d) prediction results for an initial battery SOC of 100% [65].

application of ML in RUL prediction from the perspective of input variables. The choice of input variables can seriously affect the accuracy of RUL prediction, and the choice of input variables is generally divided into direct variables and processed variables. Direct variables include external features such as capacity, terminal voltage, charge/discharge current [73], cycle time and temperature [74]. Although direct variables are easy to obtain, accurate prediction will require a large amount of external feature data, increasing the computational cost. It is therefore necessary to select variables with a high correlation to the battery RUL or to process direct variables such as equivalent voltage rise charge/discharge intervals and equivalent current fall charge/discharge intervals from the battery charge/discharge curve for a specific cycle in order to improve prediction accuracy and reduce calculation costs. In Patil *et al.*'s study, in addition to the NASA database including direct features such as capacity, voltage, and current in each discharge cycle, Patil *et al.* [75] further obtained 13 processed variables such as energy of signal, fluctuation index of signal, and concave convex index from the voltage and temperature profiles of each discharge cycle. The study used principal component analysis to extract key parameters that represent most of the information in the dataset to reduce the dimensionality of the input variable space, followed by visualization techniques to extract highly sensitive variables, and finally obtained the optimized parameters for voltage curves of energy of signal and fluctuation index of signal (the calculation Eqs. (3) and (4) are shown below). Zheng *et al.* [76] also noted that the energy of voltage signal and the fluctuation index of voltage signal are strong correlated with battery RUL, with correlation coefficients ρ = 0.7 ~ 0.8.

$$\text{Energy of signal} = \int_{-\infty}^{\infty} |x(t)|^2 \mathrm{d}t \tag{3}$$

$$\text{Fluctuation index of signal} = \frac{\sqrt{\sum (y_i - \mu)^2}}{\omega} \tag{4}$$

where $x(t)$ denotes the voltage or temperature signal corresponding to moment t, y_i is the signal at sampling point i, μ denotes the mean value of the signal, and ω denotes the sampling frequency.

Severson *et al.* [77] used 124 commercial LIBs (LFP/graphite) to generate data sets under the same discharge conditions and different fast charging conditions and found a strong correlation between the logarithm of the variance of $\Delta Q_{100\text{-}10}(V)$ extracted from the discharge voltage curve and the logarithm of the battery cycle life with a correlation coefficient of $\rho = -0.92$. The study It was found that a simple linear regression model trained on $\Delta Q_{100\text{-}10}(V)$ as an input variable alone was able to predict battery life relatively accurately (with an average error of approximately 15% on the first test set and 11% on the second test set). In addition it was possible to achieve a 4.9% error in classifying the long and short battery lifetimes based on the first 5 cycles of data. Battery electrochemical impedance spectroscopy (EIS) reflects rich material information, but it is difficult to quantitatively select features that are relevant to battery life. Zhang *et al.* [78] collected over 20,000 EIS spectra measured at different temperatures and SOCs and fed the dataset directly into Gaussian process regression to predict RUL. Experiments (Fig. 7) found that Gaussian process regression with an automatic relevance determination kernel automatically assigned the maximum prediction weight to the low-frequency region of the EIS, suggesting that the low-frequency region of the EIS could be used for accurate prediction. The operating principle of LIBs is the embedding and de-embedding of lithium ions between the positive and negative electrodes, and the potential difference between the electrodes determines the potential of the battery, so the analysis of the battery voltage is also the key to understand the change of the battery state. Bereciba *et al.* [79] extracted the differential voltage curve (dV/dQ) and the incremental capacity curve from the battery charge and discharge curves, the peak in the differential voltage curve responds to the phase change on the voltage curve position, while the incremental capacity peak corresponds to the phase equilibrium position on the battery voltage curve. The peak value, peak position and peak area of the incremental capacity curve and the valley value and valley position of the differential voltage curve change during battery ageing, and it was found that there is a strong correlation between the two characteristics and the battery state, with a correlation coefficient ρ = 0.97. Further research into capturing valuable features from historical battery data is important in order to improve ML prediction accuracy and simplify computation.

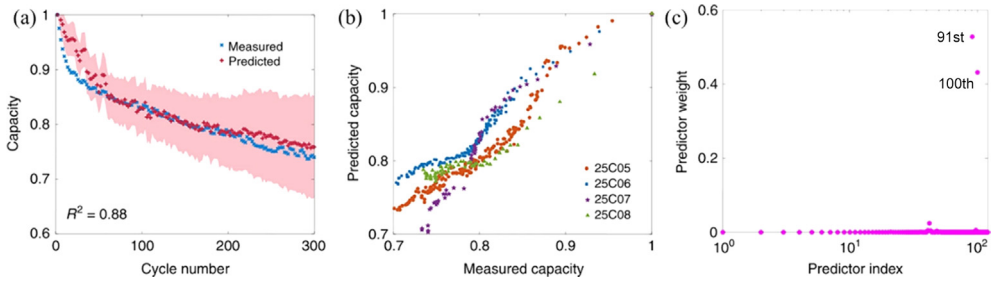

Fig. 7. Battery capacity predictions from EIS with input variables extracted. (a) Blue points are battery capacity measurements for the corresponding cycles, red points are predicted values, and the coefficient of determination R^2 = 0.88. (b) Estimated capacity to tested capacity ratio for each cycle of the four battery types at 25 °C. (c) Weight assignments within the Gaussian process regression model corresponding to 120 frequencies in the range 0.02 Hz–20 kHz, with the maximum weights assigned to 2.16 Hz and 17.80 Hz, with other frequencies weighted approximately at 0 [78].

5. Conclusions

LIBs are becoming the mainstream battery product due to their high energy density and long cycle life. The improvement of battery performance has been slowed by the difficulty of efficiently exploring the high-dimensional material space through traditional material design and calculations. The complexity of the battery degradation mechanism also makes it difficult to accurately predict the battery state. The introduction of data-driven ML may solve the current dilemma in LIBs research in the future. At present, the application of ML in the field of battery is still in its infancy, there are still many directions to think about and explore, such as deep learning algorithms in the field of image recognition for the analysis of advanced characterization pictures of batteries, fault detection of batteries, laddering and so on. In the design of battery materials such as electrolyte screening, the ionic conductivity is still mainly used as the evaluation criterion, however, the performance of electrolyte materials still needs to be considered in terms of electrochemical window, compatibility with electrode materials and thermal stability. In terms of battery state prediction, the current battery charge and discharge data set is simple and has few samples, while the actual working environment is more complex, and further research is needed to determine whether the health factors and algorithm models obtained now are applicable to the actual working conditions. Combined with the current research state of ML in lithium battery applications, the following aspects need to be further addressed in the future:

(1) ML requires sufficient and high quality data. Accurately fitting the curves of the input and output variables requires the introduction of a sufficient amount of data. The database constructed in the current study has limited capacity and the quality of the data is affected by different experimental conditions and measurement errors, which in turn reduce the accuracy of the prediction model [80]. In addition, due to the current experimental conditions, it is difficult to collect experimental data such as SEM maps and TEM maps in large quantities, which also greatly limits the application of ML. In the future, there is a need to share and standardize electrochemical experimental data across research institutions to facilitate systematic collection and application [81].

(2) Advances in ML model architectures. More advanced ML algorithms such as neural Turing machines [82], generative adversarial networks [83] *etc*. need to be introduced to address the lack of data within the electrochemical database. In addition, uncertainty analysis of the model must be considered, and quantification of uncertainty can guide model construction and data selection, *etc*. ML and other experimental methods are combined to further increase the credibility of the model predictions [84].

(3) Interpretability of ML parameters. ML models are often viewed as 'black box' functions between input and output variables, and the choice of parameters within the model lacks physical or chemical interpretability, making it difficult to generalize scientific laws from ML models. Interpretable parameters can also simplify the selection of data features and significantly reduce the amount of data required and computational costs.

This article reviews the achievements of ML in materials and state prediction research for LIBs in recent years, which also fully illustrates the practicality, effectiveness and inevitability of the intersection between ML and the battery field, which will inevitably become an indispensable complementary tool for materials science experiments and computational simulation techniques in the future. However, the process of crossover is still very slow due to problems such as sparse data, learning model selection and lack of inter-disciplinary talents. This review hopes to provide many battery researchers with a new way of thinking to address the challenges within the current research field and actively promote the integration of ML and LIBs. It is believed that with the concerted efforts of many researchers, ML will drive a shift in research and manufacturing in the battery industry in the future and the performance of lithium-ion batteries will be further improved.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge financial supports from the National Key Research and Development Program of China (2018YFA0209600) and the Natural Science Foundation of China (22022813, 21878268, 52075481).

References

[1] P.Z. Lyu, X.J. Liu, J. Qu, J.T. Zhao, Y.T. Huo, Z.G. Qu, Z.H. Rao, Recent advances of thermal safety of lithium ion battery for energy storage, *Energy Storage Mater.* 31 (2020) 195–220.
[2] J.B. Dunn, L. Gaines, J.C. Kelly, C. James, K.G. Gallagher, The significance of Li-ion batteries in electric vehicle life-cycle energy and emissions and recycling's role in its reduction, *Energy Environ. Sci.* 8 (1) (2015) 158–168.
[3] M.S. Whittingham, Ultimate limits to intercalation reactions for lithium batteries, *Chem. Rev.* 114 (23) (2014) 11414–11443.
[4] M. Li, J. Lu, Z.W. Chen, K. Amine, 30 years of lithium-ion batteries, *Adv. Mater.* 30 (33) (2018) e1800561.
[5] Y. Liu, T.L. Zhao, W.W. Ju, S.Q. Shi, Materials discovery and design using machine learning, *J. Materiomics* 3 (3) (2017) 159–177.

[6] D.C. Elton, Z. Boukouvalas, M.S. Butrico, M.D. Fuge, P.W. Chung, Applying machine learning techniques to predict the properties of energetic materials, *Sci. Rep.* 8 (1) (2018) 9059.

[7] Y. Liu, B.R. Guo, X.X. Zou, Y.J. Li, S.Q. Shi, Machine learning assisted materials design and discovery for rechargeable batteries, *Energy Storage Mater.* 31 (2020) 434–450.

[8] B. Xia, X. Zhao, R. de Callafon, H. Garnier, T. Nguyen, C. Mi, Accurate Lithium-ion battery parameter estimation with continuous-time system identification methods, *Appl. Energy* 179 (2016) 426–436.

[9] E. Namor, D. Torregrossa, R. Cherkaoui, M. Paolone, Parameter identification of a lithium-ion cell single-particle model through non-invasive testing, *J. Energy Storage* 12 (2017) 138–148.

[10] M.F. Ng, J. Zhao, Q.Y. Yan, G.J. Conduit, Z.W. Seh, Predicting the state of charge and health of batteries using data-driven machine learning, *Nat. Mach. Intell.* 2 (3) (2020) 161–170.

[11] M. Littmann, K. Selig, L. Cohen-Lavi, Y. Frank, P. Hönigschmid, E. Kataka, A. Mösch, K. Qian, A. Ron, S. Schmid, A. Sorbie, L. Szlak, A. Dagan-Wiener, N. Ben-Tal, M.Y. Niv, D. Razansky, B.W. Schuller, D. Ankerst, T. Hertz, B. Rost, Validity of machine learning in biology and medicine increased through collaborations across fields of expertise, *Nat. Mach. Intell.* 2 (1) (2020) 18–24.

[12] T. Radivojević, Z. Costello, K. Workman, H. Garcia Martin, A machine learning automated recommendation tool for synthetic biology, *Nat. Commun.* 11 (1) (2020) 1–14.

[13] J. Goecks, V. Jalili, L.M. Heiser, J.W. Gray, How machine learning will transform biomedicine, *Cell* 181 (1) (2020) 92–101.

[14] C. Crisci, B. Ghattas, G. Perera, A review of supervised machine learning algorithms and their applications to ecological data, *Ecol. Model.* 240 (2012) 113–122.

[15] A.H. Ran, Z.H. Zhou, S.X. Chen, P.B. Nie, K. Qian, Z.L. Li, B.H. Li, H.B. Sun, F.Y. Kang, X. Zhang, G.D. Wei, Data-driven fast clustering of second-life lithium-ion battery: Mechanism and algorithm, *Adv. Theory Simul.* 3 (8) (2020) 2000109.

[16] H.S. Wang, Y.J. Ji, Y.Y. Li, Simulation and design of energy materials accelerated by machine learning, *Wires Comput. Mol. Sci.* 10 (1) (2020) e1421.

[17] M. Wang, T. Wang, P.Q. Cai, X.D. Chen, Nanomaterials discovery and design through machine learning, *Small Methods* 3 (5) (2019) 1900025.

[18] S.Q. Shi, J. Gao, Y. Liu, Y. Zhao, Q. Wu, W.W. Ju, C.Y. Ouyang, R.J. Xiao, Multi-scale computation methods: Their applications in lithium-ion battery research and development, *Chinese Phys. B* 25 (1) (2016) 018212.

[19] Y. Liu, J.M. Wu, M. Avdeev, S.Q. Shi, Multi-layer feature selection incorporating weighted score-based expert knowledge toward modeling materials with targeted properties, *Adv. Theory Simul.* 3 (2) (2020) 1900215.

[20] F. Han, A.S. Westover, J. Yue, X. Fan, F. Wang, M. Chi, D.N. Leonard, N.J. Dudney, H. Wang, C. Wang, High electronic conductivity as the origin of lithium dendrite formation within solid electrolytes, *Nat. Energy* 4 (3) (2019) 187–196.

[21] H. Liu, X.-B. Cheng, J.-Q. Huang, H. Yuan, Y. Lu, C. Yan, G.-L. Zhu, R. Xu, C.-Z. Zhao, L.-P. Hou, C. He, S. Kaskel, Q. Zhang, Controlling dendrite growth in solid-state electrolytes, *ACS Energy Lett.* 5 (3) (2020) 833–843.

[22] S. Randau, D.A. Weber, O. Kötz, R. Koerver, P. Braun, A. Weber, E. Ivers-Tiffée, T. Adermann, J. Kulisch, W.G. Zeier, F.H. Richter, J. Janek, Benchmarking the performance of all-solid-state lithium batteries, *Nat. Energy* 5 (3) (2020) 259–270.

[23] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B* 89 (9) (2014) 094104.

[24] Y. Wang, W.D. Richards, S.P. Ong, L.J. Miara, J.C. Kim, Y.F. Mo, G. Ceder, Design principles for solid-state lithium superionic conductors, *Nat. Mater.* 14 (10) (2015) 1026–1031.

[25] Y. Zhang, X.F. He, Z.Q. Chen, Q. Bai, A.M. Nolan, C.A. Roberts, D. Banerjee, T. Matsunaga, Y.F. Mo, C. Ling, Unsupervised discovery of solid-state lithium ion conductors, *Nat. Commun.* 10 (1) (2019) 5260.

[26] A.D. Sendek, E.D. Cubuk, E.R. Antoniuk, G. Cheon, Y. Cui, E.J. Reed, Machine learning-assisted discovery of solid Li-ion conducting materials, *Chem. Mater.* 31 (2) (2019) 342–352.

[27] D.C. Lin, Y.Y. Liu, Y. Cui, Reviving the lithium metal anode for high-energy batteries, *Nat. Nanotechnol.* 12 (3) (2017) 194–206.

[28] P. Albertus, S. Babinec, S. Litzelman, A. Newman, Status and challenges in enabling the lithium metal electrode for high-energy and low-cost rechargeable batteries, *Nat. Energy* 3 (1) (2018) 16–21.

[29] C. Monroe, J. Newman, Dendrite growth in lithium/polymer systems: A propagation model for liquid electrolytes under galvanostatic conditions, *J. Electrochem. Soc.* 150 (10) (2003) A1377–A1384.

[30] Z. Ahmad, T. Xie, C. Maheshwari, J.C. Grossman, V. Viswanathan, Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes, *ACS Cent. Sci.* 4 (8) (2018) 996–1006.

[31] B. Liu, J. Yang, H. Yang, C. Ye, Y. Mao, J. Wang, S. Shi, J. Yang, W. Zhang, Rationalizing the interphase stability of Li|doped-$Li_7La_3Zr_2O_{12}$ *via* automated reaction screening and machine learning, *J. Mater. Chem. A* 7 (34) (2019) 19961–19969.

[32] W.W. Yan, S.Y. Yang, Y.Y. Huang, Y. Yang, G.H. Yuan, A review on doping/coating of nickel-rich cathode materials for lithium-ion batteries, *J. Alloy. Compd.* 819 (2020) 153048.

[33] C.H. Wang, K. Aoyagi, P. Wisesa, T. Mueller, Lithium ion conduction in cathode coating materials from on-the-fly machine learning, *Chem. Mater.* 32 (9) (2020) 3741–3752.

[34] Y.H. Xiao, L.J. Miara, Y. Wang, G. Ceder, Computational screening of cathode coatings for solid-state batteries, *Joule* 3 (5) (2019) 1252–1275.

[35] G. Pilania, C.C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, *Sci. Rep.* 3 (2013) 2810.

[36] K. Fujimura, A. Seko, Y. Koyama, A. Kuwabara, I. Kishida, K. Shitara, C.A.J. Fisher, H. Moriwake, I. Tanaka, Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms, *Adv. Energy Mater.* 3 (8) (2013) 980–985.

[37] R. Jalem, M. Nakayama, T. Kasuga, An efficient rule-based screening approach for discovering fast lithium ion conductors using density functional theory and artificial neural networks, *J. Mater. Chem. A* 2 (3) (2014) 720–734.

[38] R.P. Joshi, J. Eickholt, L.L. Li, M. Fornari, V. Barone, J.E. Peralta, Machine learning the voltage of electrode materials in metal-ion batteries, *ACS Appl. Mater. Interfaces* 11 (20) (2019) 18494–18503.

[39] A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, I. Tanaka, Representation of compounds for machine-learning prediction of physical properties, *Phys. Rev. B* 95 (14) (2017) 144110.

[40] A. Seko, T. Maekawa, K. Tsuda, I. Tanaka, Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids, *Phys. Rev. B - Condens. Matter Mater. Phys.* 89 (5) (2014), 054303.1-054303.9.

[41] A. Ishikawa, K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama, M. Okada, Machine learning prediction of coordination energies for alkali group elements in battery electrolyte solvents, *Phys. Chem. Chem. Phys.* 21 (48) (2019) 26399–26405.

[42] M. Sumita, Y. Tanaka, T. Ohno, Possible polymerization of PS4 at a Li_3PS_4/$FePO_4$ interface with reduction of the $FePO_4$ phase, *J. Phys. Chem. C* 121 (18) (2017) 9698–9704.

[43] M. Sumita, R. Tamura, K. Homma, C. Kaneta, K. Tsuda, Li-ion conductive Li_3PO_4-Li_3BO_3-Li_2SO_4 mixture: Prevision through densityfunctional molecular dynamics and machine learning, *Bull Chem. Soc. Jpn.* 92 (6) (2019) 1100–1106.

[44] J.F. Whitacre, J. Mitchell, A. Dave, W. Wu, V. Viswanathan, An autonomous electrochemical test stand for machine learning informed electrolyte optimization, *J. Electrochem. Soc.* 166 (16) (2019) A4181–A4187.

[45] A. Dave, J. Mitchell, K. Kandasamy, H. Wang, S. Burke, B. Paria, B. Póczos, J. Whitacre, V. Viswanathan, Autonomous discovery of battery electrolytes with robotic experimentation and machine learning, *Cell Rep. Phys. Sci.* 1 (12) (2020) 1002641.

[46] M.M. van Duongvan Tran, A. Garg, H. van Nguyen, T.T.K.M.L. Huynh, Phung le, Machine learning approach in exploring the electrolyte additives effect on cycling performance of $LiNi_{0.5}Mn_{1.5}O_4$ cathode and graphite anode-based lithium-ion cell, *Int. J. Energy Res.* 45 (3) (2021) 4133–4144.

[47] A.D. Robertson, A.R. West, A.G. Ritchie, Review of crystalline lithium-ion conductors suitable for high temperature battery applications, *Solid State Ionics* 104 (1–2) (1997) 1–11.

[48] B.K. Wheatle, E.F. Fuentes, N.A. Lynd, V. Ganesan, Design of polymer blend electrolytes through a machine learning approach, *Macromolecules* 53 (21) (2020) 9449–9459.

[49] K. Hatakeyama-Sato, T. Tezuka, M. Umeki, K. Oyaizu, AI-assisted exploration of superionic glass-type Li^+ conductors with aromatic structures, *J. Am. Chem. Soc.* 142 (7) (2020) 3301–3305.

[50] Y.M. Wang, T. Xie, A. France-Lanord, A. Berkley, J.A. Johnson, Y. Shao-Horn, J.C. Grossman, Toward designing highly conductive polymer electrolytes by machine learning assisted coarse-grained molecular dynamics, *Chem. Mater.* 32 (10) (2020) 4144–4151.

[51] Y. Takagishi, T. Yamanaka, T. Yamaue, Machine learning approaches for designing mesoscale structure of Li-ion battery electrodes, *Batteries* 5 (3) (2019) 54.

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

[53] A.P. Wang, Z.Y. Zou, D. Wang, Y. Liu, Y.J. Li, J.M. Wu, M. Avdeev, S.Q. Shi, Identifying chemical factors affecting reaction kinetics in Li-air battery *via ab initio* calculations and machine learning, *Energy Storage Mater.* 35 (2021) 595–601.

[54] Y. Xing, E.W.M. Ma, K.L. Tsui, M. Pecht, Battery management systems in electric and hybrid vehicles, *Energies* 4 (11) (2011) 1840–1857.

[55] B. Gou, Y. Xu, X. Feng, State-of-health estimation and remaining-useful-life prediction for lithium-ion battery using a hybrid data-driven method, *IEEE Trans. Veh. Technol.* 69 (10) (2020) 10854–10867.

[56] J. Rivera-Barrera, N. Muñoz-Galeano, H. Sarmiento-Maldonado, SoC estimation for lithium-ion batteries: Review and future challenges, *Electronics* 6 (4) (2017) 102.

[57] H.W. He, X.W. Zhang, R. Xiong, Y.L. Xu, H.Q. Guo, Online model-based estimation of state-of-charge and open-circuit voltage of lithium-ion batteries in electric vehicles, *Energy* 39 (1) (2012) 310–318.

[58] S. Lee, J. Kim, J. Lee, B.H. Cho, State-of-charge and capacity estimation of lithium-ion battery using a new open-circuit voltage versus state-of-charge, *J. Power Sources* 185 (2) (2008) 1367–1373.

[59] S. Santhanagopalan, Q.Z. Guo, P. Ramadass, R.E. White, Review of models for predicting the cycling performance of lithium ion batteries, *J. Power Sources* 156 (2) (2006) 620–628.

[60] H.W. He, R. Xiong, X.W. Zhang, F.C. Sun, J.X. Fan, State-of-charge estimation of the lithium-ion battery using an adaptive extended Kalman filter based on an improved thevenin model, *IEEE Trans. Veh. Technol.* 60 (4) (2011) 1461–1469.

[61] Q. Miao, L. Xie, H.J. Cui, W. Liang, M. Pecht, Remaining useful life prediction of lithium-ion battery with unscented particle filter technique, *Microelectron. Reliab.* 53 (6) (2013) 805–810.
[62] J.C. Álvarez Antón, P.J. García Nieto, C. Blanco Viejo, J.A. Vilán Vilán, Support vector machines used to estimate the battery state of charge, IEEE Trans. Power Electron. 28 (12) (2013) 5919–5926.
[63] S.J. Tong, J.H. Lacap, J.W. Park, Battery state of charge estimation using a load-classifying neural network, *J. Energy Storage* 7 (2016) 236–243.
[64] W. He, N. Williard, C.C. Chen, M. Pecht, State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation, *Int. J. Electr. Power Energy Syst.* 62 (2014) 783–791.
[65] X.B. Song, F.F. Yang, D. Wang, K.L. Tsui, Combined CNN-LSTM network for state-of-charge estimation of lithium-ion batteries, *IEEE Access* 7 (2019) 88894–88902.
[66] M.A. Hannan, M.S.H. Lipu, A. Hussain, P.J. Ker, T.M.I. Mahlia, M. Mansor, A. Ayob, M.H. Saad, Z.Y. Dong, Toward enhanced state of charge estimation of lithium-ion batteries using optimized machine learning techniques, *Sci. Rep.* 10 (1) (2020) 4687.
[67] M. Landi, G. Gross, Measurement techniques for online battery state of health estimation in vehicle-to-grid applications, *IEEE Trans. Instrum. Meas.* 63 (5) (2014) 1224–1234.
[68] H.H. Pan, Z. Lü, H.M. Wang, H.Y. Wei, L. Chen, Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine, *Energy* 160 (2018) 466–477.
[69] P. Khumprom, N. Yodo, A data-driven predictive prognostic model for lithium-ion batteries based on a \r deep learning algorithm, *Energies* 12 (4) (2019) 660.
[70] Y.Y. Li, S.M. Zhong, Q.S. Zhong, K.B. Shi, Lithium-ion battery state of health monitoring based on ensemble learning, *IEEE Access* 7 (2019) 8754–8762.
[71] H.C. Dong, X.N. Jin, Y.B. Lou, C.H. Wang, Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter, *J. Power Sources* 271 (2014) 114–123.
[72] P.-H. Michel, V. Heiries, An adaptive sigma point kalman filter hybridized by support vector machine algorithm for battery-SoC and SoH estimation In: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), 2015.
[73] Y.Z. Zhang, R. Xiong, H.W. He, M.G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries, *IEEE Trans. Veh. Technol.* 67 (7) (2018) 5695–5705.
[74] C. Hu, G. Jain, C. Schmidt, C. Strief, M. Sullivan, Online estimation of lithium-ion battery capacity using sparse Bayesian learning, *J. Power Sources* 289 (2015) 105–113.
[75] M.A. Patil, P. Tagade, K.S. Hariharan, S.M. Kolake, T. Song, T. Yeo, S. Doo, A novel multistage support vector machine based approach for Li ion battery remaining useful life estimation, *Appl. Energy* 159 (2015) 285–297.
[76] Z. Zheng, J. Peng, K. Deng, K. Gao, H. Li, B. Chen, Y. Yang, Z. Huang, A novel method for lithium-ion battery remaining useful life prediction using time window and gradient boosting decision trees, In: 2019 10th International Conference on Power Electronics and ECCE Asia, 2019.
[77] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B.B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, M.Z. Bazant, S.J. Harris, W.C. Chueh, R.D. Braatz, Data-driven prediction of battery cycle life before capacity degradation, *Nat. Energy* 4 (5) (2019) 383–391.
[78] Y.W. Zhang, Q.C. Tang, Y. Zhang, J.B. Wang, U. Stimming, A.A. Lee, Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning, *Nat Commun* 11 (1) (2020) 1706.
[79] M. Berecibar, F. Devriendt, M. Dubarry, I. Villarreal, N. Omar, W. Verbeke, J. van Mierlo, Online state of health estimation on NMC cells based on predictive analytics, *J. Power Sources* 320 (2016) 239–250.
[80] M. Liu, C. Clement, K. Liu, X.M. Wang, T.D. Sparks, A data science approach for advanced solid polymer electrolyte design, *Comput. Mater. Sci.* 187 (2021) 110108.
[81] M. Aykol, P. Herring, A. Anapolsky, Machine learning for continuous innovation in battery technologies, *Nat. Rev. Mater.* 5 (10) (2020) 725–727.
[82] S.M. Faradonbe, F. Safi-Esfahani, A classifier task based on Neural Turing Machine and particle swarm algorithm, *Neurocomputing* 396 (2020) 133–152.
[83] H.L. Zhang, W. Tang, W. Na, P.Y. Lee, J. Kim, Implementation of generative adversarial network-CLS combined with bidirectional long short-term memory for lithium-ion battery state prediction, *J. Energy Storage* 31 (2020) 101489.
[84] C. Chen, Y.X. Zuo, W.K. Ye, X.G. Li, Z. Deng, S.P. Ong, A critical review of machine learning of energy materials, *Adv. Energy Mater.* 10 (8) (2020) 1903242.