Energy Advances

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Energy Adv.*, 2023, 2, 615

Artificial intelligence-navigated development of high-performance electrochemical energy storage systems through feature engineering of multiple descriptor families of materials

Haruna Adamu,[abc] Sani Isah Abba,[d] Paul Betiang Anyin,[e] Yusuf Sani[f] and Mohammad Qamar *[a]

Received 7th March 2023,
Accepted 27th March 2023

DOI: 10.1039/d3ya00104k

rsc.li/energy-advances

With the increased and rapid development of artificial intelligence-based algorithms coupled with the non-stop creation of material databases, artificial intelligence (AI) has played a great role in the development of high-performance electrochemical energy storage systems (EESSs). The development of high-performance EESSs requires the alignment of multiple properties or features of active materials of EESSs, which is currently achieved through experimental trial and error approaches that are tedious and laborious. In addition, they are considered costly, time-consuming and destructive. Hence, machine learning (ML), a crucial segment of AI, can readily accelerate the processing of feature- or property–performance characteristics of the existing and emerging chemistries and physics of active materials for the development of high-performance EESSs. Towards this direction, in this perspective, we present insight into how feature engineering can handle multiple feature/descriptor families of active materials of EESSs.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1. Introduction

To harvest energy from renewable energy sources effectively and for widespread electrification, electrochemical energy storage is necessary to overcome the inherent intermittency nature of renewable energy generation and mitigate the destabilization of the environment by climate change catastrophes through the reduction of CO_2 emissions from fossil fuel consumption.[1] Therefore, renewable energy (such as solar, wind, geothermal, biomass, and hydro) sources are critical for creating a clean and sustainable future. However, these renewables are habitually intermittent, unpredictable, and poorly distributed globally, and therefore when integrated directly in the energy mix without effective and efficient energy storage systems, they significantly disrupt the global energy supply network. Consequently, the development of efficient and reliable energy storage systems is essential for the effective utilization of renewable energies. Thus, electrochemical energy storage systems (EESSs) are an integral part in the development of sustainable energy technologies.

In efforts to reduce greenhouse gas emission, while simultaneously meeting the growing global energy consumption, more research attention has been given to renewable energy sources such as solar and wind. However, given that renewable energy sources are not spontaneous in nature, energy storage plays a key role in storing available energy resources and is a requirement for their future use. Therefore, electrochemical energy storage systems are the main technologies that can address the renewable energy demand and the need to eradicate or reduce CO_2 emissions. In this context, electrochemical energy storage systems include various short- and long-term energy storage technologies that allow energy to be save in ample quantities over different periods. Since the discovery of electricity, man has continuously sought for effective ways to store this type of energy on demand. Thus, besides energy storage, the large-scale generation of electrical energy is necessary to meet the global energy demand in modern societies. However, as technology advances and the energy demand keeps changing, the issue of energy storage has continued to evolved, adapt, and revolutionise. Therefore, researchers have devoted their efforts to the development of clean and renewable energy storage systems as alternatives to the consumption of fossil fuels, which has detrimental effects on society.

[a] *Interdisciplinary Research Center for Hydrogen and Energy Storage (IRC-HES), King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia. E-mail: qamar@kfupm.edu.sa*
[b] *Department of Environmental Management Technology, ATBU, Bauchi, Nigeria*
[c] *Department of Chemistry, ATBU, Bauchi, Nigeria*
[d] *Interdisciplinary Research Center for Membrane and Water Security, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia*
[e] *Faculty of Engineering, Department of Computer Engineering, Ahmadu Bello University, Zaria, Nigeria*
[f] *School of Computing and Communications InfoLab 21, Lancaster University, Lancaster LA1 4WA, UK*

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

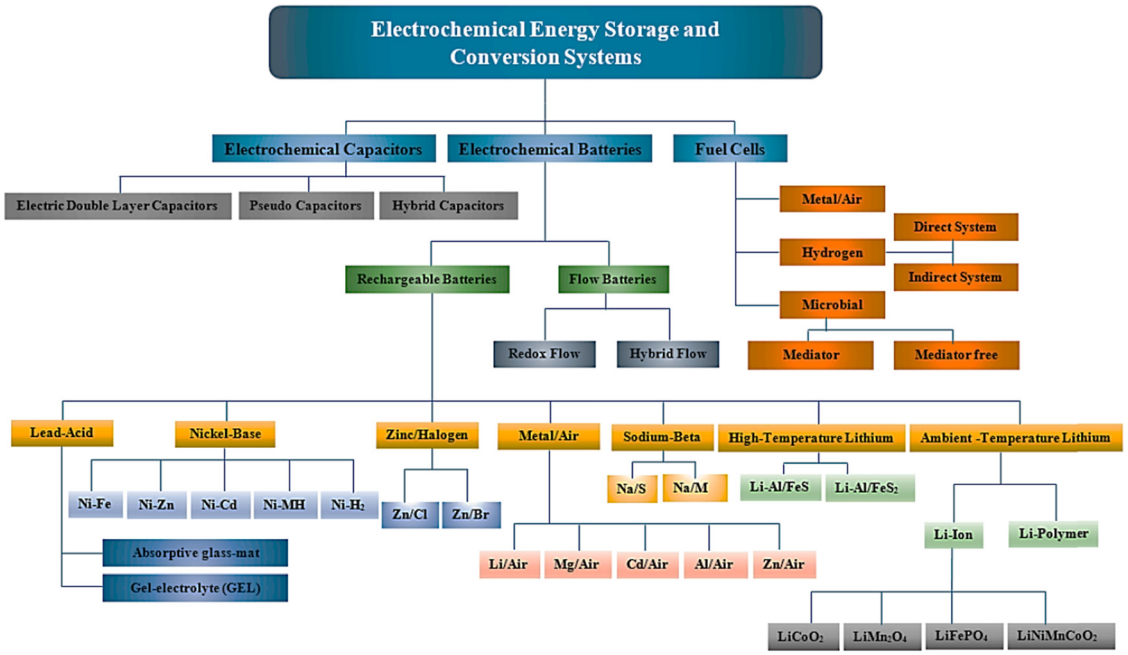

Fig. 1 General classification of electrochemical energy storage systems. Reproduced with permission from ref. 2. Copyright 2020, MDPI.

Although a broad array of energy storage technologies is available including mechanical, electrical, chemical, thermochemical, and electrochemical storage systems, electrochemical energy storage systems have received the most attention and been adopted in everyday applications due to their high coulombic efficiencies and environment-friendliness.[2,3] In this case, electrochemical energy storage systems offer a wide range of technological approaches to manage the energy supply and their comprehensive classification is shown in Fig. 1. This figure presents the diversity of electrochemical energy storage systems based on their basic working principles (mechanism of action in energy storing ability), technical characteristics, and distinctive properties or features.

Achieving net zero emissions by 2050 is dependent on the production of 92% energy from renewable energy sources.[4] Thus, to support this energy demand with renewable energy sources, electrochemical energy storage systems are required. Also, to adapt to this renewable energy future, electrochemical energy storage systems can be used to balance the increasing global peak energy demands and support the variability in renewable sources. Unfortunately, the current electrochemical energy storage systems have limitations in meeting the global demand.[4,5] Considering the various requirements of energy density, power, response time, storage time and types of applications, it is believed that different electrochemical energy storage systems have different energy and power performance characteristics. Therefore, the application restrictions of different storage technologies are clearly noticeable. Some systems are not suitable for power quality-type application, while others are not suitable for bulk energy long-term storage-type applications. Accordingly, individual electrochemical energy storage systems have different levels of maturity, and thus their current performance characteristics are limited. Consequently, to realize a high return on this global energy mission, it is important to produce electrochemical energy storage systems with desirable high-performance characteristics. In this case, although substantial advancements have been realised in energy storage technology, particularly for large-scale energy storage, the need for high-performance, efficient energy storage systems is greater than ever before. Accordingly, electrochemical energy storage systems are the keys to achieving decarbonisation and net zero emission targets by 2050.[4,6]

However, although research and development in electrochemical energy storage systems have resulted in some improvements, the development progress is not fast enough to support and allow greater penetration of renewable energies in the global energy grid. Therefore, more research needs to be conducted. Obviously, the application of computational chemistry, specifically density functional theory (DFT), compared to traditional trial and error approaches has made significant contributions in providing useful information for the research and development of new materials. These contributions were made with support of some modern chemical simulation toolkits and high-throughput screening methods, which truly speed up the discovery and development of new materials. However, the large-scale screening of novel materials still consumes a lot of time together with the high computational cost of high-precision DFT calculations.[7]

With the increasing and rapid development of computer skill as well as data science and engineering, AI, which simulates human intelligence, has opened another window to modern research, and thus attracted global attention. With recent advancement, AI has been applied in the fields of image recognition[8] and autonomous driving,[9] and thus has great

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

potential to surpass the existing reasoning level of humans in some fields.[10] Accordingly, it can be viewed as disruptive technology that has potential to radically change the way we think and do things. In this case, electrochemical energy research will also follow this trend. Briefly, AI has the potential to transform the way energy is produced and stored in different ways. The future of electrochemical energy storage systems will look very different from that known thus far. Obviously, AI will play a significant role in influencing that future, given that it results in a promising performance in the face of unpredictable settings. Thus, this technology can be employed to improve the performance in a range of challenging fields including the energy sector. Hence, given that machine leaning (ML) algorithms can automatically mine implicit relationships that are hidden behind a large volume of data,[11] AI can have vast applications in the fields of chemistry and materials science, particularly in the area of electrochemical energy storage systems. Consequently, with the huge volume of data on the current performance and lifetime of electrochemical energy storage systems becoming available owing to the advent of artificial intelligence (AI), the fast development of AI can open a new way to address the performance limitations suffered by the current electrochemical energy storage systems. Although AI-based approaches have been applied to infer multiple aspects of electrochemical energy storage systems including the state of health (HOS) with the aim to estimate the remaining useful life (RUL) of energy storage systems,[12] they have also been employed in the discovery of energy storage systems.[13–16] Techniques such regression and neural networks have been applied to predict the properties of energy storage materials such as electric conductivity[17–19] and screening conducting solid-state electrolyte materials for lithium ions[14,20,21] electrolyte materials that suppress the growth of lithium dendrites.[22] However, the major challenge is to distil and obtain useful information from a large dataset, and therefore it is often difficult to directly measure and make predictions.[23] In addition, the traditional trial and error methods employed to optimize the practical properties of materials used in electrochemical energy storage systems are costly and time-consuming. Therefore, feature engineering can be a key step in solving this challenge, given that this approach eliminates irrelevant or redundant features (feature selection) or/and reduces the number of features (feature extraction), thus improving the prediction accuracy and reducing the training time for prediction. This implies that feature engineering is capable of unravelling the interdependence between features and parameters in multi-dimensional datasets. Consequently, this approach can be an excellent tool to compliment or support traditional research methodologies in predicting behaviour and performance patterns. In this direction, large-scale data on the performance features or characteristics generated by energy storage systems can support the development of AI-based approaches, thereby leading to the creation and development of a new set of high performance electrochemical energy storage systems. However, attempts to achieve this with feature engineering techniques have not been reported to date.

In this perspective, we provide insight on how the basic workflow of feature engineering, which includes data collection, pre-processing, feature selection and extraction, employing relevant technical know-how in data science can create features that enable ML algorithms to predict the best and high-performance electrochemical energy storage systems. This implies the revolutionisation and substitution of physical experiment-based approaches with the data-driven nature of ML towards decreasing experimental efforts, while improving the predictive accuracy and minimising waste of efforts and resources. The goal is to speed up the scientific discovery of electrochemical features or characteristics economically through the judicious combination of existing experimental data and data science. It is believed that the analytics of ML helps in offering rational responses or guidelines for designing high-performance electrochemical energy storage systems, given that its combination with experiments offers complementary information and facilitates predictability.[12,24–27]

2. Current challenges and development of electrochemical energy storage systems

Electrochemical energy storage systems including batteries, flow batteries, capacitors/supercapacitors, and fuel cells store energy in various forms.[28] These systems are promising technologies to address some of the most urgent global challenges such as development of clean and sustainable energy and reduction of CO_2 emission and other associated air pollution problems. Electrochemical energy storage systems function through the interconversion of chemical species and electric charges. Therefore, their typical functionality parameters include energy density, power density, storage capacity, response time, efficiency, charge–discharge rate, lifetime, heat sensitivity, environmental and safety considerations and operational cost,[29] which amongst other practical properties strongly depend on their materials. This implies that these systems are typically expected to possess multiple functionality parameters. Now, the challenge is not only to find system(s) with appropriate functionality but also ones exhibiting specific requirements. Consequently, there is still a lack of electrochemical energy storage system(s) that exhibit the desired performance and longevity. For example, the performances of electrochemical energy storage systems can be compared in the Ragone plot, as illustrated in Fig. 2. Due to the physical difference in the energy storage capacity of capacitors compared to batteries and fuel cells, they feature relatively large power densities but poor energy densities. Alternatively, batteries and fuel cells have high energy densities but low power densities as a result of their sluggish reaction kinetics. Although electrochemical capacitors or supercapacitors connect the gap between capacitors and batteries/fuel cells, the performance of the system is limited by low energy density. At the present, the storage capabilities of energy storage systems are inversely equal. where the strength of one storage technology is the weakness in

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

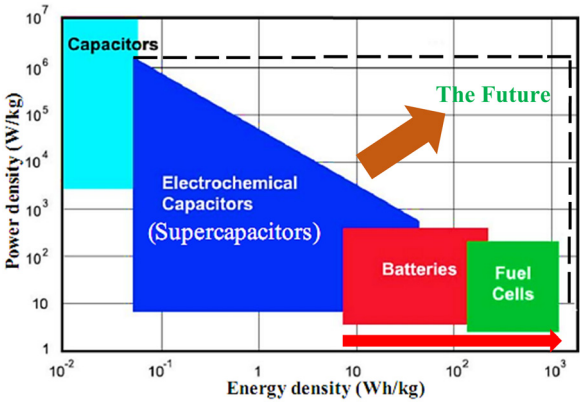

Fig. 2 Illustration of the variation in energy density and power density of various electrochemical energy storage systems and envisaged future research direction and development for electrochemical energy storage systems.

another technology. Accordingly, the future research and development of electrochemical energy storage system(s) should focus on retaining the high energy density of batteries and fuel cells without compromising the high power density of capacitors, as marked with red arrow (close to the energy density axis) and dash-lines at the top in Fig. 2, respectively.

The development cycle of any electrochemical energy storage system depends on its structure–property relationship and property–performance relationship. Consequently, the current developments can be grouped into five (5) categories, as follows: (i) to improve the performance and efficiency, (ii) to promote the reliability and prolong the durability, (iii) to ensure usage safety, (iv) to reduce the environmental impact, and (v) to reduce the cost.[30] Achieving these goals is associated with the identification and selection of features or parameters, including selection of materials, choice of structural features or parameters (such as durability, cost, safety and environmental considerations), performance and efficiency parameters. In addition, these goals are associated with challenges in the optimization and development of efficient energy storage systems, given that the materials used for the design and construction of storage systems possess unique features. Specifically, the primary goals in the development of electrochemical energy storage systems are achieving high energy storage capacity, high power density, long duration, and low cost systems. All these are dependent on the in-depth understanding of the chemistries of the materials, which requires a decade or much longer for the experiment-to-commercialisation transition.[31,32] Consequently, the development of electrochemical energy storage systems has been historically very long. Traditionally, this development has been by trial and error, followed by sequential processes of understanding the individual and combined electrochemical responses. Currently, with the high global demands, it is important to shorten this time-frame by coming up with feasible solutions to overcome the related challenges.

Given that these challenges are transdisciplinary, the digitalisation of research can play a crucial role in the acceleration of the optimization and discovery of new high-performance systems. Hence, it is expected that AI-based approaches can reduce the number of redundant experiments and duration required, given that they can promote the efficiency of development (*e.g.*, materials selection, choice of structural features or parameters, and extension of lifetime, cost). In addition, now that the world is entering exciting times in research digitalisation, particularly based on AI and ML, ideas are becoming reality. In short, AI is spreading globally, thereby influencing the development of new and existing technologies. In the area of energy storage, AI-based approaches are promising in accelerating the discovery of new materials and interfaces, as well as the optimization of their performance.[33] Therefore, it has become necessary to understand this ongoing data-based science revolution to realize the full potential of feature engineering as a tool to improve the speed and efficiency of electrochemical science.

3. Survey on the applications of artificial intelligence in the performance prediction of electrochemical energy storage systems

In the development of electrochemical energy storage systems (EESSs), from the discovery of new materials to the stages of testing their performance, each stage takes several months or even years of evaluation. This has been the limiting factor in the development of EESSs. The application of AI in the development of EESSs has greatly alleviated this problem.[34] The most common problems that AI-based approaches are employed to address are state estimation and prediction, lifetime prediction, analysis and classification of properties, fault discovery and diagnosis, modelling, design, and optimization.[35] In this case, choosing appropriate AI-based strategies to improve performance is also significant.

3.1 Application of AI-based approaches of performance prediction for batteries

Artificial intelligence-based approaches have been applied to infer multiple aspects of battery development.[12] Consequently, nowadays there is high interest in developing robust and accurate AI-based models for allowing the prediction of the performance of EESSs. Currently, the emergence of AI-based tools has focused on estimating the electrochemical state, the degree of degradation, and/or the prediction of new materials. Specifically, the state of charge (SOC), state of health (SOH), remaining useful life (RUL) and optimization of the operating conditions of batteries are the common goal.[36,37] In fact, both academia and industry have devoted major efforts to these areas, particularly for automotive applications.[38,39]

In contrast, this perspective calls on the development of AI-based techniques to infer from the available experimental data in the literature and other means the battery performance

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

characteristics within the shortest possible time, instead of the months or years using the current tedious, laborious, destructive, and costly methods. This will set the stage to accelerating structure–property and property–performance categorisation and prediction for the development of high-performance EESSs from the existing chemistries. In this respect, well-trained ML approaches can potentially combine both accuracy and low computational cost, making them interesting for the accurate prediction of battery performance.

In the implementation of AI-based approaches to establish cell performance prediction, an extreme learning machine (ELM) model was proposed to predict the evolution of battery temperature, voltage, and power.[40] The group was able to compare the predicted values with experimentally observed values, and thus proposed replacing the function of activation with a set of models targeted to further improving long-term performance prediction. Interestingly, this approach allowed the enhancement of the ELM model in terms of current at varying temperatures.

It was also reported that the multidimensional multiphysics (MDMP) model can simulate and predict cell performance by inputting battery design parameters under the observed operating conditions.[41] It worth mentioning that the consideration of the design parameters under realistic operating conditions made the reported results more reliable. Similarly, MDMP models were also used to predict the surface temperature distribution (suggested performance) with response to lithium concentration in large format prismatic cells.[42,43] In the construction of the models, the parameters of the cells used included cell design (heterogeneity and nonlinearity of the real cell components and geometry) and the physical properties of the materials used. Although intensive study of parameter sensitivity is essential for an in-depth understanding of cell performance among several important characteristics, the sensitivity parameter analysis of cells mostly relies only on minor parameter changes and avoiding the entire parameter space. This is related with the high computational cost of large numerical quantification of parameter sensitivity prediction. Therefore, feature engineering can be a solution and useful to perform wide-ranging parameter sensitivity analysis with reasonable computational cost without compromising the entire parameter space. Within this context, the available database of battery design parameters and physical properties of materials used should be made features or descriptors, which are then be used to predict performances (predictors). With ML algorithms, features or descriptors can be identified and categorised, followed by selection or extraction of the best features or descriptors that would offer a guide in the design of high-performance batteries.

In the implementation of AI-based approaches in redox flow batteries, a computational workflow was proposed, which coupled a data-driven model with physical parameter model, providing insights into the relationship between the pore-scale electrode structure reaction and the device-scale electrochemical reaction homogeneity inside a redox flow battery,[44] as presented in Fig. 3. This group succeeded in training and validating a deep neural network (DNN) model with more than 100 pore-scale, consequently establishing a quantitative relationship between redox flow battery operating conditions (such as electrolyte inlet velocity, current density, and electrolyte concentration) and surface reaction at the pore scale. Regrettably, although a significant reduction in pump power consumption for targeted surface reaction homogeneity was recorded, only a slight decline in electric power output for discharging was also witnessed. This calls for further improvement or development of another approach.

Considering that the performance optimization of redox flow battery systems is essential for their application and commercialisation in large-scale energy storage, the development of innovative AI-based methodologies to predict performances and efficiencies with great accuracy is urgent. The performance of vanadium redox flow batteries (VFBs) is directly related to their stack and electrolyte. Thus, RL models were developed based on a database of over 100 stacks with varying power outputs to predict the performance of VFBs.[45] The models successfully optimized the VFB materials and structure. They also predicted future performance development with emphasis on reducing the electrochemical polarisation and ohmic polarisation at high current densities, as well as decreasing the concentration polarisation of the FB stacks. In redox flow batteries (FBs), as the medium where the electrochemical reaction takes place, the pore structure and surface area of the electrode affect the performance of FBs. Thus, an ML algorithm coupled with data generation method was constructed to predict porous electrodes with large surface area and high hydraulic permeability for FBs.[46] In this approach, the stochastic reconstruction method, morphological algorithm and lattice Boltzmann method were used to build the dataset, consequently generating 2275 fibrous structures. For the ML algorithms, logistic regression (LR), artificial neural network (ANN), and random forest (RF) were employed for the construction of the models and used for surface area and hydraulic permeability prediction of the porous electrodes. In this study, more than 700 promising porous electrode materials were screened through the combination of genetic algorithm (GA) and ANN. Similarly, as an important component of VFBs, the membrane directly affects their performance of. Recently, LR and ANN were employed to predict the performance of a polybenzimidazole (PBI) porous membrane treated with different solvents, as illustrated in Fig. 4. For the modelling, 9 different solvent properties and 5 experimental parameters were used for the prediction of the most appropriate solvent and alcohols were found as the most appropriate solvent for regulation of the membrane porous structure.[47]

Lithium-ion batteries (LIBs), as well-known energy storage technology, power a wide range of large-scale applications, including electric vehicles and grid storage, as well as small-scale applications including mobile devices such as smartphones and laptops. In addition, due to the rising global demand for clean and renewable energy to eliminate greenhouse gases, the LIB markets is expanding continually.[48,49] Recently, the LIB technological advancement has greatly

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

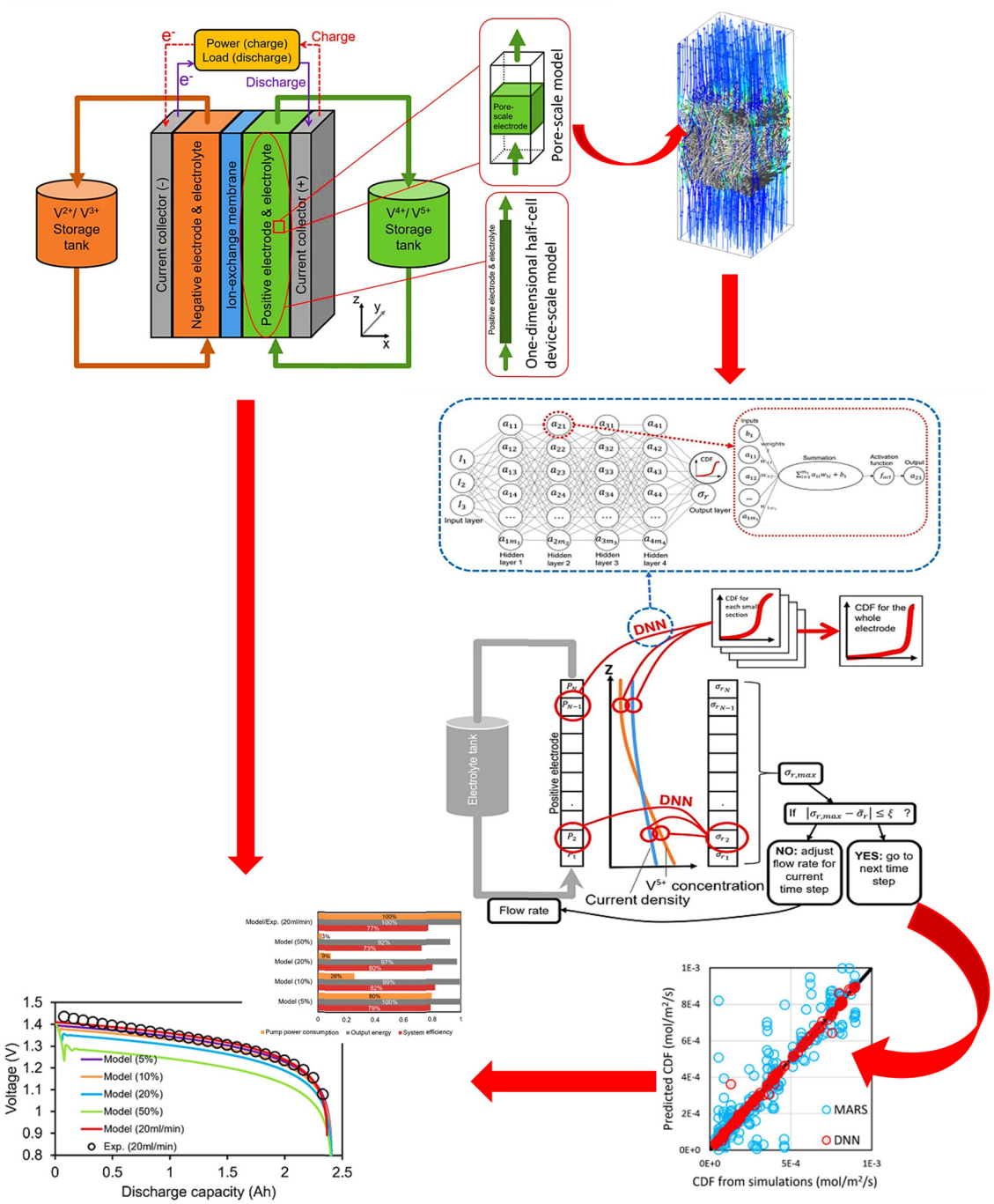

Fig. 3 Graphical illustration of DNN model used to predict the relationship between flow battery operating conditions in VFB and surface reaction homogeneity. Insert is pump power consumption, output energy, and system efficiency for different targeted surface reaction and constant flow rate (20 mL min^{-1}). Adapted with permission from ref. 44. Copyright 2020, Wiley.

evolved through artificial intelligence-based approaches by the creation of new chemistries and architectures that accelerate the property–performance development from existing LIB chemistries to emerging LIBs in cost-effective ways. In this effort, the performance of the battery is directly influenced by its microstructural features such as active material particle size, shape, alignment, and distribution.[50,51] Alternatively, these microstructural features are typically measured as a function of the reactive area density, which is controlled by the porosity, or volume fraction left by the solid electrode material phase. Similarly, the tortuosity of porous electrodes, which is associated with electrical conductivity and chemical diffusivity of LIBs, is another important microstructural feature linked to battery performance.[52–54] However, despite the significance of the microstructural features and their influence on the performance, cost, and deterioration of LIBs, it is still expensive,

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

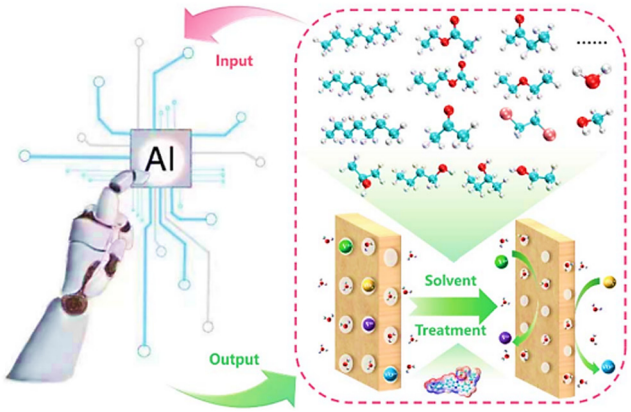

Fig. 4 Schematic illustration of AI-based approach applied to screen suitable solvent for the treatment of PBI membrane. Reproduced with permission from ref. 47. Copyright 2021, The Royal Society of Chemistry.

time-consuming, and difficult to quantitatively estimate the battery properties. Currently, experimental inferences of battery microstructural features are made from the reconstruction of tomographic images.[50,52,55–57] However, although tomography experiments are useful for determining and identifying microstructural features, processing the electrode layers, preparing the samples for imaging, and processing the resulting images require significant economic and computational resources.[56,58–60] AC impedance-based methods, the polarization-interrupt method, which infers tortuosity from the effective chemical diffusivity,[61] and the blocking electrolyte method,[62] which infers tortuosity from the effective electrical conductivity, are further experimental methodologies that can be employed. In these techniques, the investigational electrode sample has to be processed experimentally, and some of them require additional data fitting with electrochemical models.[63] Overall, it appears impracticable to run existing non-destructive microstructural quality battery evaluation methods concurrently and in real time with the production line.

In contrast, machine learning (ML) approaches have been used to infer various features of battery technology,[64] such as the state of health (SOH), with the goal of estimating the remaining usable life (RUL) and improving the operating conditions of the battery.[65,66] In recent years, real-time SOH and RUL monitoring methods based on neural networks have become increasingly popular.[66] Zhang *et al.*[67] employed recurrent neural networks with long short-term memory (RNNs) to learn the long-term dependencies in degradation data and predicted the RUL. RNN and its variants have typically been used to predict SOH and RUL, where they were trained using data from cell voltage, current, and temperature.[64–66] In particular, the state of charge (SOC) of LIBs during voltage discharge was estimated using deep learning techniques and correlations established among the voltage, current, temperature, power, and energy of LIBs during voltage discharge.[68–72]

Battery material discovery has also made use of ML techniques.[73–76] Regression and neural network methods were utilized to estimate the electrical conductivity and reaction rates of materials.[77–79] To identify promising lithium ion conducting solid-state electrolyte materials, Sendek *et al.*[74,80] used a regression model, while Jalem *et al.*[81] used a neural network. To find electrolyte components that suppress the formation of lithium dendrites, a trained graph of convolutional neural network (CNN) has been developed thus far.[82] Accordingly, ML has now received significant attention in accelerating the property–performance characterization process in the development of high-performance EESSs.

Generally, the battery state is estimated using ANN-based models. To estimate the SOC of Li-ion batteries, the deep neutral network (DNN) was utilized.[83] In this practice, the voltage, temperature, average current, and average voltage of the battery at time t were used as the inputs of the DNN and the output was the SOC value at time t. Additionally, a load-classifying neural network (NN) model was developed to determine the SOC.[84] Based on this ML procedure, the NN divides the input vectors (which include the extracted features of current, voltage, and time) into three categories (charging, idling, and discharging, as illustrated in Fig. 5), and trains the three sub-NNs concurrently. Hence, the load-classifying NN has a more flexible selection of training data, simpler training process, and lower computing cost. The estimated SOC was the result of the load-classifying NN (after filtering). The load profile of the driving cycle of a vehicle was used to train the model.

In a different approach, based on complex feature datasets with time-series characteristics, CNN and RNN were employed to estimate the battery state. To estimate the SOC of Li-ion batteries, CNN and LSTM networks were combined.[85] In the CNN-LSTM structure, the CNN was employed for pattern recognition and spatial feature distillation, while the LSTM processed time-series data by learning the temporal properties of the battery dynamic evolution. To train the CNN-LSTM system, a total of 24 815 sets of data points from experiments were collected. The output is the time-dependent SOC estimation, and the inputs for each group are the current, voltage,

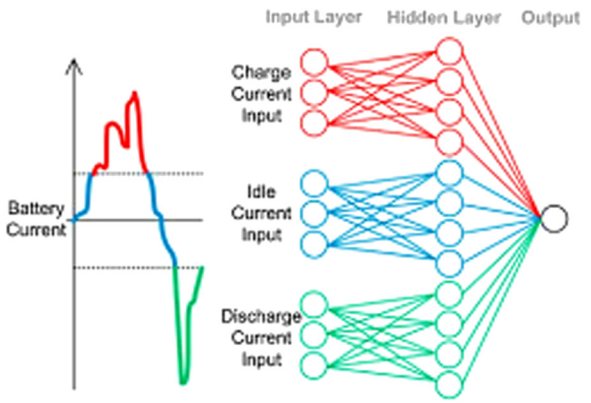

Fig. 5 Illustration of the assembly of the load-classification of neutral network with presentation of the input data divided into three subsets based on the three types of battery behaviour (charging, idling, and discharging) segregated by three different colours. Reproduced with permission from ref. 84. Copyright 2016, Elsevier.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

temperature, average current, and voltage. Based on the results, the trained CNN-LSTM can estimate the SOC with an overall root-mean-square error (RMSE) and mean absolute error (MAE) of $<2\%$ and 1%, respectively. Therefore, transfer learning and ensemble learning strategies are receiving increasing attention. To calculate the capacity of Li-ion batteries, deep convolutional neural networks with ensemble learning and transfer learning (DCNN-ETL) were utilized.[86] The output was capacity, while the input data groups were each designed to include 25 segments and represent current, voltage, and charge capacity. The dataset of 25 338 sample group divisions was utilized to pre-train the n DCNN sub-systems, and 525 sample group divisions were used to re-train the DCNN-ETL system. Although these findings demonstrated that DCNN-ETL has a lower training efficiency than other machine learning algorithms, it has a greater training accuracy than DCNN, DCNN-TL (DCNN with transfer learning), and DCNN-EL (DCNN with ensemble learning).

To predict the RUL of Li ion batteries for the input dataset with complex time-series characteristics, both CNN and recurrent neural network (RNN) models were reported. For example, a convolutional neural network with long short-term memory (CNN-LSTM) hybrid neural network was used.[87] Although the LSTM predicted the unknown sequence of capacity data based on the features recovered by CNN, the CNN was used to extract the relevant information. To train the hybrid neural network, a section of the discharge capacity degradation curve was employed. The size of the input vector (the segment length of the capacity-cycle curve chosen for the input) and the suitable sliding window size were both determined using the false closest neighbours (FNN) technique. The predicted capacity cycle curve was the output. In a similar development, to predict RUL, an LSTM-DNN deep learning model was also used.[88] The deep learning model was trained using the battery capacity decaying pattern discovered by virtual testing conducted under various C-rates (based on a standard procedure).[89] The current, voltage, battery working temperature, and battery capacity from previous results were the input features for the deep learning model. The capacity at the present time is the production. Similarly, other research has employed ML algorithms with LSTM to predict the SOH or RUL of batteries.[90–92]

In the case of degradation analysis, regression using the Gaussian process regression is also receiving significant interest. For instance, GPR was used to build a calendar capacity loss model to examine the Li-ion battery aging characteristics.[93] The kernel function in the model is customized. Specifically, they specified the capacity loss as the output of the GPR model and the storage period for which the aging was predicted, the reciprocal temperature corresponding to this storage time, and the SOC level corresponding to this storage time as the inputs. According to their findings, the mean-absolute-error of the model for predicting the capacity loss and capacity is 0.31% and 0.53%, respectively, when it was trained using data from just 18 cells that were tested under 6 storage conditions.

The detection and classification of battery cells with abnormal behaviour are another common use of the deep learning methodology. Using a deep belief network (DBN) model, storage battery voltage anomalies can be identified.[94] The probability distribution of the input data can be learned by the model, which can then utilize the estimated probability to determine the active state of each node. Thus, the model outperforms the conventional back-propagation neural network in terms of training speed and convergence. After parameter extraction, the voltage–time and current–time curves can be used to train the DBN (with 15 hidden layers).[94] The 9 features included in the input were the charge/discharge current, time, and temperature, while the output was the voltage. In addition, several ML algorithms were applied to classify the unbalance and degradation status of Ni-MH, Na-ion, and Mg-ion battery cells, including logistic regression,[95] kernel neural network (k-NN), kernel-support vector machine (KSVM with a Gaussian radial base function kernel),[96] Gaussian naive Bayes (GNB), and NN with only one hidden layer.

Overall, in each of the above-mentioned scenarios, the data availability and generation are among the major concerns for ML modelling in LIB technology.

3.2 Application of AI-based approaches of performance prediction for capacitors/supercapacitors

Recently, double layer capacitors, commonly known as supercapacitors, have attracted attention from both researchers and industry due to their massive energy storage potential.[97,98] Therefore, they play vital roles in storing energy, particularly in electric vehicles and other battery-based products. However, their low energy density has been a limiting factor in their widespread acceptability and usage. Meanwhile, the research and development of materials strongly rely on scientific intuition through trial and error experiments, which are generally time-consuming. Consequently, the development cycle of the device is quite long. Therefore, at the moment, to quickly reduce this period, one possible solution is to use AI-based approaches to boost the development of new materials for high performance supercapacitors.

Traditionally, in supercapacitors, a high-surface area electrode results in greater capacitance. Therefore, electrodes with micropores (<2 nm) are preferred. However, it has been proven that micropores in the electrodes are responsible for a decrease in power density and capacitance (performance) due to the sluggish electrochemical kinetics and unsatisfactory volumetric performance.[99–102] Therefore, to accelerate the discovery of materials and establishment of new understanding of materials behaviours, AI-based approaches can extract new knowledge or build predictive models using existing materials database. In an attempt to reduce the proportion of human influence in experiments, several ML studies were conducted in predicting the performance of supercapacitors. For example, an AI-based approach was employed, which quantitatively correlated the structural characteristics of electrodes with power densities for porosity optimization, targeting high-performance supercapacitors.[103] In this work, the input parameters used were the surface area of micropores and mesopores and the scan rate, while the specific capacitance and power density were the output. Four different ML algorithms, namely, LR, LC, RF,

 © 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

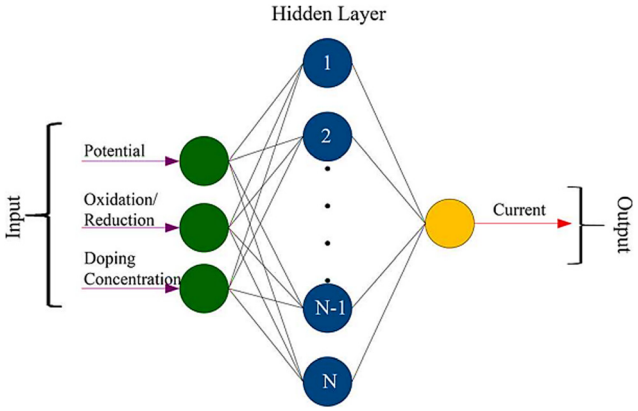

Fig. 6 Illustrative setup of the input, hidden, and output layers in ANN modelling for the prediction of performance behaviour of $CoCeO_2$/rGO nanocomposite supercapacitors. Reproduced with permission from ref. 105. Copyright 2018, the American Chemical Society.

and ANN, were used, and among them only ANN was the best fit in representing the experimental results. Accordingly, the new insights from this approach can boost and shorten research time for the synthesis and preparation of materials and improve the performance of supercapacitors. In a different effort, the performance capacity of supercapacitors based on the physical properties of 13 different electrolytes was predicted by the widely used ANN, SVM, and LR algorithm models.[104] The physical properties used as input for the ML algorithms were diameter (solvent molecular size), dipole moments, viscosities, boiling temperature, and dielectric constant of the electrolyte. The dielectric constant and diameter (solvent molecular size) were labelled as the properties affecting the capacity.

Besides the above-mentioned efforts, the performance of $CoCeO_2$/rGO nanocomposite supercapacitors was predicted by ANN and RF.[105] The datasets for ML training were sourced from experiments. The input parameters were potential, oxidation/reduction, and doping concentration, while the output was the current, as illustrated in Fig. 6. It was found that ANN performed the best with better interpretability and higher accuracy compared to RF. This was determined by the variation of coefficient of determination (R^2) and corresponding root mean square error (RMSE).

3.3 Application of AI-based approaches of performance prediction for fuel cells

AI-Based techniques have also been found useful in the prediction of the performance of fuel cells, which are embodied mostly by the construction of ML models through input and output variables in specific algorithms. In general, the right fuel cell evaluation technology and simulation processes are essentially performance models, which allow the investigation of operating parameters either on a stand-alone basis or as part of *in situ* reaction.[106] Therefore, for the prediction of fuel cell performance, ANN was applied to predict the solid oxide fuel cell (SOFC).[106] With the ANN model, 7 input and 4 output variables were considered. The scale conjugate algorithm was employed for data training, while sigmoid transfer function (hyperbolic tangent) was used on all layers. The trained ANN model simulated the performance of the fuel cell with great accuracy, and thus was helpful in the general effort for better power generation. However, as ANN is structured with multi-hidden layers, the approach needs to be tested with big data comprising of complex chemistries and physics.

Generally, for the implementation of performance prediction of proton exchange membrane fuel cells (PEMFCs), their *I–V* polarization curves are used as an important metric. They represent many important properties of PEMFCs such as cell dimensions, material properties, operating conditions, and physical/electrochemical characteristics including current density, and specific power.[107–110] In this context, RF and CNN were employed to design the performance prediction of PEMFC, which was primarily utilised to reduce unnecessary experiments for the development of fuel cells.[111] In this case, RF was used to select important parameters as the input features, while CNN was adopted for the performance prediction with the *I–V* polarization curve as the output of the model. Due to the effectiveness of the model, the CNN-based prediction curves were in good agreement with the real curves. However, despite the success of this approach, it is important to note that before the selection of important features, the main influence features or factors need to be identified and categorised into descriptor families from physical/structural, electrochemical, and operating condition parameters. Thus, in an effort to solve this issue, a new approach that can combine the three descriptor families to produce subsets or different features (feature engineering) would significantly fast track the process of yielding high-performance cells. In another effort to find the optimum output of solid oxide fuel cells (SOFCs) within a wide range of parameters, ANN was designed to predict the performance of SOFCs using polarisation curves and electrochemical impedance spectra.[112] The results of the ANN models were in good agreement with the measured data. However, very low current densities below 20 mA cm^{-2} were observed. Besides, this approach needs to move towards more complex operating parameters with increased number of available datasets. In short, all the past studies focused on predicting performance under limited features or parameters, ignoring other important performance metrics. The studies mostly predicted performance with operating conditions and/or some few electrochemical features as target metric(s) for evaluation.[107–110,113–118]

DNN containing convolutional layers was used to predict the water coverage ratio for PEMFCs as a metric for performance.[119] The prediction results showed that the trained DNN achieved 94.23% accuracy in the identification of the water coverage ratio. As important as the feed substrate type, six different ML algorithms, namely, LRM, RF, STBS, NN, KNN, and SVM, with radial kernel were used to predict the feed substrates (including acetate, carbohydrate, and wastewater) for microbial fuel cells (MFCs) based on genomic data.[120] In this study, only four different input variables were trained, but considering the discrepancies in microbial resistance to various types of toxic pollutants, more input parameters or features

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

need to be considered for much better performance evaluation. In a similar work, an ANN-based model was applied to predict the performance of MFCs.[121] The input parameters for the construction of the ANN model were the wastewater characteristics and the output was the performance parameters. However, although this approach was successful in predicting the performance of the bioelectrochemical system of the MFC, more inclusive data must be considered to improve its reliability and applicability. Particularly, this study was conducted under a controlled environment, which did not capture the wide range of parameters and real-time behaviour of the natural environment.

As part of the metrics for performance evaluation, AI-based approaches are useful for material selection and property prediction through the use of existing databases. This is advantageous in accounting for the unknown chemistries and physics of materials, and thus can greatly improve the efficiency of performance prediction. For example, features that are mostly related to catalytic performance in the order of importance were highlighted, including the electron affinity, sum of the van der Waals radius, difference in the Pauling electronegativity, product of ionization energy, and distance between two metal atoms. The others are the sum of Pauling electronegativity of two metal atoms, average distance between the first metal atom (TM1) and the second one (TM2) and the surrounding N number of atoms.[122] However, this has been rarely applied to the electrode materials and membranes of PEMFCs, and consequently is a future research direction in the development of high-performance fuel cells, particularly SOFCs and PEMFCs.

4. Materials screening for electrochemical energy storage systems and performance prediction

In view of the escalating environmental challenges and reliance on portable and uninterruptible power sources, sustainable energy and efficient and affordable energy conversion and storage technology have become vital segments in energy technology. Researchers are concentrating on capturing and turning solar and wind energy into electricity. Therefore, the demand for green and sustainable energy is significantly influenced by electrochemical energy storage technology.

Redox processes are used in electrochemical energy storage systems to reversibly transform electrical energy into chemical energy, which is then stored as chemical potential in the electrodes. The specific capacity of the electrodes and the working voltage of the cell, which is the difference in the potential between the cathode and the anode, together determine the energy density and power density of an electrochemical energy storage system. Increasing this particular capacity has been the subject of numerous studies over the years; however, there have been difficulties to comprehending and manipulating the electrochemical potential of the electrode materials. In electrochemical energy storage systems, energy density and power density are two parameters that are crucial for assessing their practical performance, which are frequently shown in Ragone plots (Fig. 2). Based on this information, significant studies have been conducted on the synthesis and characterisation of different nanostructured cathode and anode materials to achieve large surface areas and short solid-state transport distances for improving their energy and power densities, but unfortunately the cathode and anode materials are still the limiting factors in the storage capacity performance of EESSs. Consequently, as advanced strategies to ensure the high performance of EESSs, machine learning (ML) and artificial intelligence (AI) are conformally being applied as powerful tools for the selection of materials for design and performance optimization in energy storage technology development. Machine learning, particularly property–performance informed-deep learning and AI can facilitate the development of materials selection in enhancing the performance of EESSs, showing great potential to advance electrochemical energy storage technology.

Electrodes/electrocatalysts are critical in the function and proper operation of EESSs because they facilitate the transport of ions, electrons, and gaseous (or liquid) species. Therefore, the electrodes must have a large surface area, be porous, electrically and ionically conductive, and electrochemically active. However, particularly at low temperatures, it is uncommon for a single material to meet all these requirements. In addition, the electrodes of EESSs should have high activity towards the desired reaction as well as compatibility with the electrolyte on a chemical and thermomechanical level, chemical stability in reducing (anode) or oxidizing (cathode) atmospheres, resistance to poisoning by impurities in the energy storage systems, and structural stability over long periods of operation. Although significant progress has been made in EESSs, they are still limited by materials, particularly the selection of materials for their electrolytes and electrodes/electrocatalysts to ensure high performance. Consequently, here, we highlight a few recent developments in search and advancement of materials for the electrodes and electrolytes of EESSs using ML and AI, which have the potential to considerably simplify the performance optimization.

4.1 Battery materials search and development by artificial intelligence

A potent approach for developing new battery materials and for selecting and extracting key structure–property–performance links is data-driven material discovery. High-performance batteries depend on the design of new electrode and electrolyte materials. The desirable features of electrode and electrolyte materials are listed in Fig. 7, together with the other important properties that should be considered. Conventional trial-and-error experimental techniques are time-consuming, expensive, and ineffective and inefficient.[123–126] Besides, they also rely on prevailing experiences, and thus are uncreative. The ML property predictions for battery materials are summarized and compiled in Fig. 8, which provide guidance on how to select an appropriate model for a particular task.[127,128]

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

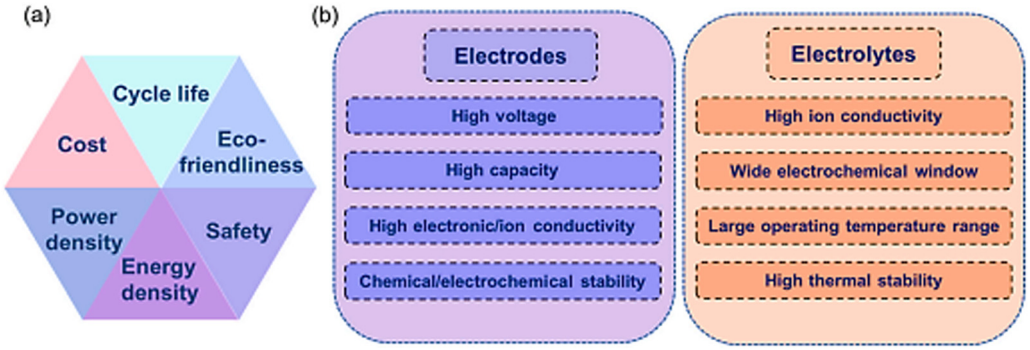

Fig. 7 (a) Main properties of batteries for performance evaluation and (b) appropriate properties of electrode and electrolyte materials. Reproduced with permission from ref. 128. Copyright 2021, Wiley-VCH.

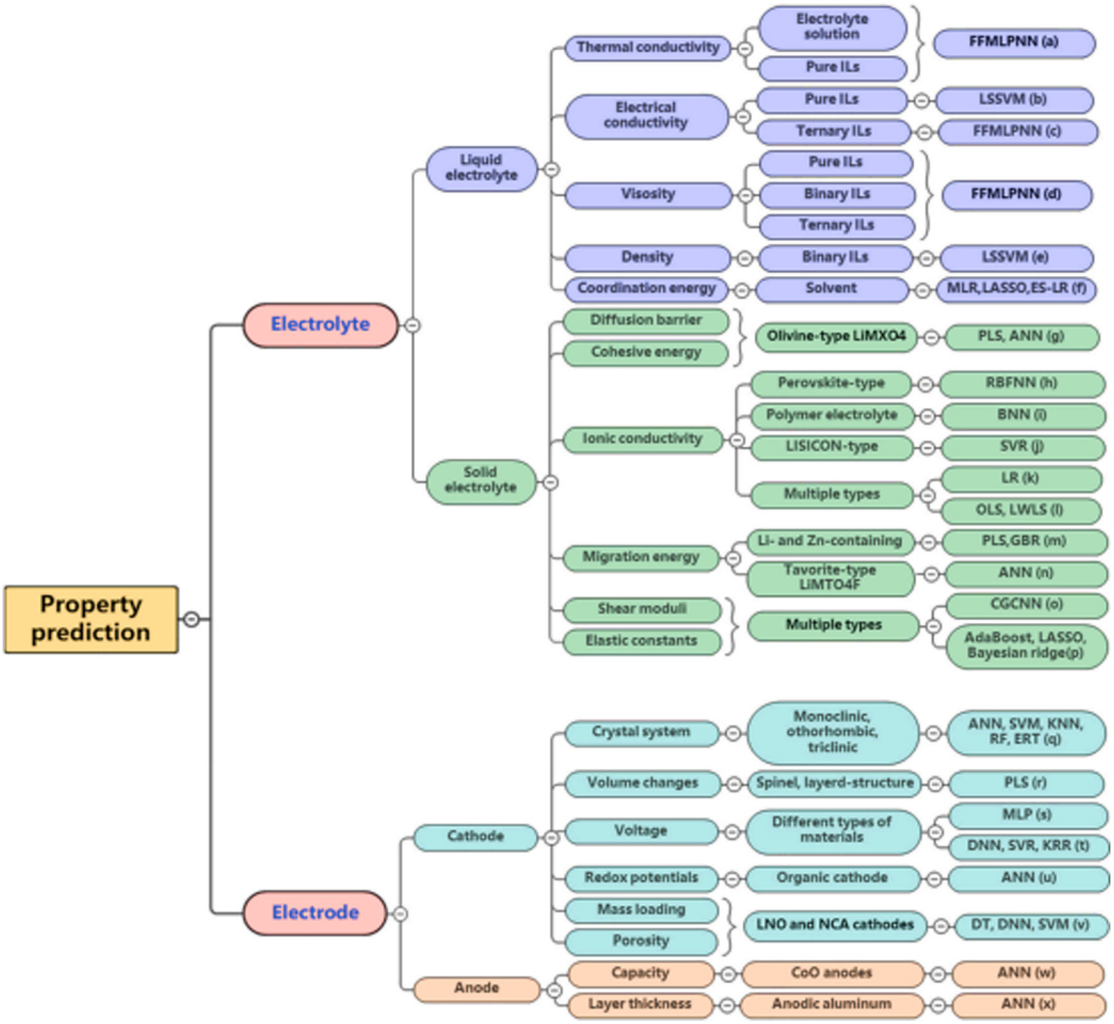

Fig. 8 Classification of machine learning used for battery property prediction. Reproduced with permission from ref. 127. Copyright 2020, Elsevier.

In the aspect of electrode materials, voltage, capacity, redox potentials, volume changes, layer thickness, and crystal structure information (space group, formation energy, band gap, number of sites, density, *etc.*) are a few of the properties that can be used to predict the performance of electrode materials. These properties can be used to predict the battery performance indicators such as energy density, output power, and lifespan.[129–131] However, in evaluating the lifetime and cycle performance of batteries, it is vital to consider the average voltage (V_{av}) and volume change ($\Delta V\%$) of the cathode materials. In a different approach, some structure descriptors were selected as ML inputs for the prediction of average voltage (V_{av})

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

and volume change ($\Delta V\%$) of the cathode materials.[132] Based on the 4860 calculated data taken from the materials project database, a DNN regression model was successfully constructed. The best-performing ML model generated an MAE of 0.38 V for V_{av} and 2.0 V for $V\%$ in the Li-only test dataset following ten-fold cross validation. To test the robustness of this ideal model, the group transferred it and produced 22 electrodes with the appropriate energy density and minimal volume change by swapping the lithium ions for sodium ions. It is important to note that a set of 306 features was created to specifically represent each reaction, and the value of each feature was normalized to have a range between −1 and 1 to enhance the fitness of NN easily.

One of the most important aspects of using ML in materials science is identifying features that can accurately represent properties a material. In this context, a machine learning (ML)-based method was constructed to predict the performance of electrode materials in terms of voltage outputs.[133] The materials project database, which contains DFT-predicted voltages for metal ion battery materials was used, which contained 3977 data of intercalation-based electrode materials. The working metal-ions, active metal-ion concentration, lattice type, space group, and other features derived from elemental properties made up the features space. Only 80 principle components were needed in the final model after the 370 redundant features were reduced by 66% using the principle component analysis (PCA) approach. In addition, to predict the performance of electrode materials in terms of the voltage outputs, DNN and SVM were also used.[133]

The electrolyte of a battery is an essential component, where a good electrolyte should conduct ions, while isolating electrons.[134] Even when utilizing computational simulation or experimental approaches, many property measurements, such as measuring the ionic conductivities of solid-state electrolytes (SSEs) and the viscosities and dielectric constants of liquid electrolytes, are challenging, expensive, and time-consuming.[135] Luckily, ML is a viable alternative strategy because it can quickly and cheaply screen an enormous amount of materials.

Quick development of SSEs has now become a hot area of study, given that SSEs have several benefits, including high energy density, high electrochemical stability, and long-term performance.[136–138] A large-scale computational approach was proposed to screen more than 12 000 SSE candidates with good structural and chemical stability, low electrical conductivity, and low cost.[139] The ionic conductivity model was used to apply the logistic regression approach to the structures of the SSE candidates, and 21 promising structures were found. By using ML, the screening process was greatly accelerated.

For the discovery of new materials and the prediction of their properties, multiphysics computational simulation and ML have been extensively used. For example, the conductivity of each composition of LISICON-type materials at 373 K was predicted using first-principles molecular dynamics (FPMD) calculation combined with SVR algorithms.[140]

The properties of a material can be accurately predicted by using the right descriptors for ML algorithms. Zhao *et al.*[141,142] constructed a hierarchically encoding crystal structure-based (HECS) descriptor framework with 32 descriptors in five parts including composition, structure, conduction channels, ion distribution, and special ions, and by using partial least squares (PLS) analysis, they were able to accurately predict the activation energy of cubic-phase Li-argyrodites. It is important to note that the variable significance in projection (VIP) scores has been used to show the synergistic effects of the local and global Li^+ conduction environments in response to Li battery performance. In total, Wang *et al.*[143] retrieved 13 descriptors, including the HOMO, LUMO, dipole moment, and atomic properties of functional groups for the properties of small molecules in organic solvents. To predict the binding energy of the solvent and LiOH molecule, three algorithms, namely, gradient boosted decision trees (GBDT), least absolute shrinkage and selection operator (LASSO), and support vector regression (SVR), that are suitable for short datasets were used. The reaction kinetics of Li–O_2 for enhanced battery performance were discovered to be considerably accelerated by phosphate ester solvents.

Due to the amorphous nature of liquid electrolytes, reports on ML for these materials are scarce. An ML analysis was performed to predict the coordination energy, E_{coord}, of alkali metal ions with liquid electrolyte solvents. The ion transfer process significantly depends on interactions between the ions and electrolyte solvent, and thus E_{coord} is a vital indicator of the ion transfer process at the interface between the electrode and electrolyte. Multiple linear regression (MLR), exhaustive search with linear regression (ES-LR), and LASSO were among the ML models used to predict E_{coord}.[144] Consequently, thus far, the association between the characteristics of alkali metal ions and E_{coord} has been established.

Despite all the above-mentioned factors, many sociological, financial, and technical factors need to be considered when choosing electrode materials for batteries. These factors include their natural abundance, lack of competition from other industrial uses, environment-friendliness in processing and usage, use, and recycling, and low cost. From a technological standpoint, electrode materials must provide a sizable reversible storage capacity at the necessary electrochemical potential. For example, due to their greater specific capacities, lighter elements, which include the majority in the first four periods of the periodic table, are preferred as electrode materials. Also, additional electron-storing sites are made possible by the various valence states of transition metal oxides, and thus the advantages of these materials as cathode materials are unmatched. Moreover, the type of bonding between transition metal ions and ligands is determined by the electronegativity and ionization energies.[145] As shown in Fig. 9, various colors indicate the potential and accessible components for usage as electrode materials. It should be mentioned that other elements, such as Nb[146] and Sn,[147,148] have been confirmed to have significant electrochemical performances in the literature, but are excluded from this chart.

4.2 Supercapacitor materials search and development by artificial intelligence

Supercapacitors have become more popular as efficient energy storage devices in recent years. Accordingly, because of their

© 2023 The Author(s). Published by the Royal Society of Chemistry

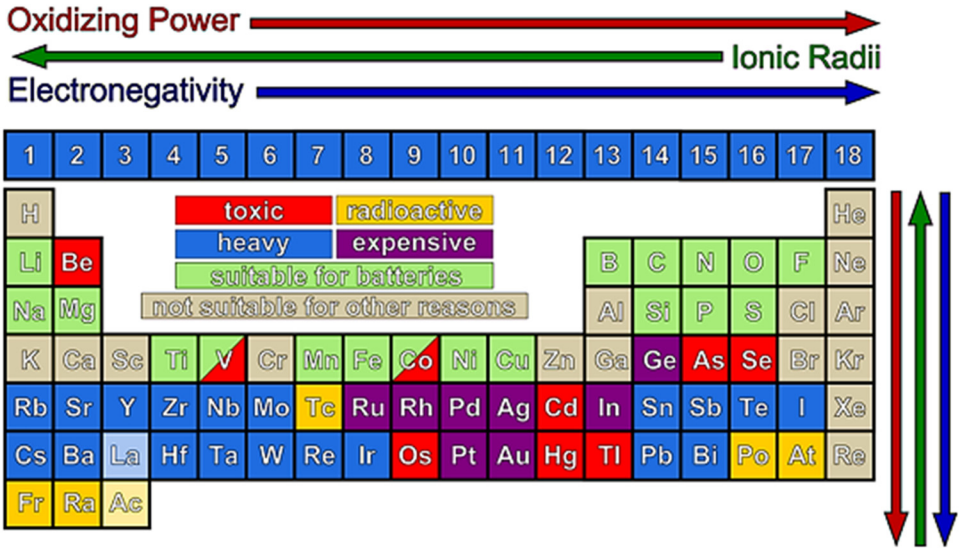

Fig. 9 Illustration of available elements for the fabrication of new electrode materials. Due to their low capacity, high cost, toxicity, or radioactivity, the colored quadrates are not included. Nevertheless, some transition metals, such as V and Co, are still being actively researched despite their toxicity. Besides, this color-coded table may be updated as some materials, including Sn, Nb, Mo, and W, demonstrated acceptable electrochemical performances in recent reports. Reproduced with permission from ref. 145. Copyright 2013, the American Chemical Society.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

superior electrochemical characteristics, carbon-based electrodes have been extensively explored experimentally and employed in the production of supercapacitors. To improve the performance of supercapacitors, research and development efforts are concentrating on enhancing the specific capacitance and energy density supplied by carbon electrodes, while maintaining their high-power density to improve the performance of supercapacitors. Generally, it is believed that increasing the number of micropores in the electrodes, which have larger surface areas than mesopores and macropores, should increase their capacitance, and consequently, the overall performance characteristics of supercapacitors. However, the capacitance and power density are reported to decrease when the micropore surface area increases in a few experiments.[149–152] This necessitates the development of effective approaches that help in comprehending the relationship between electrode structural features and supercapacitor performance, in addition to researching measurement parameters such as the scan rate. Therefore, recently, a few studies have demonstrated the value of using machine learning approaches to comprehend the charge storage characteristics of supercapacitors.[153–155] In the literature, it has been discussed how robust machine learning (ML) algorithms such as generalized linear regression (GLR), random forest (RF), support vector machine (SVM), and artificial neural network (ANN) can be used to investigate the relationships between the performance of supercapacitors and their various features. In this context, several studies have examined the relationship between the structural characteristics of electrodes and the performance metrics for supercapacitors using machine learning (ML) approaches. The consensus is that porous carbon materials (PCMs) with a high specific surface area (SSA), porosity, conductivity, and thermal stability are good candidates as supercapacitor materials. The performance of supercapacitors is affected nonlinearly by their structural characteristics (SSA, porosity, pore distribution, type of dopant, concentration, *etc.*) and operational parameters (type of electrolyte, concentration of electrolyte, potential window, *etc.*).[156–159] Recent years have witnessed the use of machine learning (ML) models to comprehend the relationship between the structural and operational parameters and the performance of carbon-based supercapacitors such as activated carbon (AC), activated carbon derived from biomass (BAC), and heteroatom-doped carbon materials.

Four regression models, *i.e.*, ANN, RF, SVM, and generalized linear regression (GLR), were employed for predicting the required properties of activated carbons (ACs) to achieve the highest energy and power density.[153] The capacitance and power density were the output parameters, whilst the surface area of the micro and mesopores and scan rate were employed as the input parameters for training the models. An improved connection between the predicted and observed power densities was provided by the ANN model, which also performed better at predicting the capacitance of activated carbon. Also, according to the ANN model, the highest energy density can be attained for ACs with surface areas of 920 $m^2\ g^{-1}$ for micropores and 770 $m^2\ g^{-1}$ for mesopores. In a different study, the influence of electrolyte solvent properties on the EDLC capacitance was investigated.[160] The experimental data utilized to train the ML models included a variety of diameters, viscosities, dipole moments, dielectric constants, boiling temperatures, and frequency-dependent capacitance for various solvents. For the investigation of capacitance, ML models such as SVR, MLP, M5 model tree (M5P), M5 rule (M5R), and LR were utilized. Among them, the MLP, M5R, and M5P models outperformed the SVR and LR. The outcome of the ML model showed that the solvent type and dielectric constant have a

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

substantial impact on the EDL capacitance, whereas the dipole moment, viscosity, and boiling temperature showed slight influence. The relationship between structural features and capacitance led to the development of a novel method for predicting the capacitive performance of polypyrrole-multiwalled carbon nanotube-cellophanes (PCMs).[161] For the purpose of predicting the capacitance of PCMs, 105 sets of various PCMs with 11 structural features, such as SSA, pore volume (PV), pore size (PS), mesopore (V_{meso}) and micropore (V_{micro}) volume, were acquired. In addition to supervised learning models such as ANN, SVM, and multiple linear regression (MLR), three ensemble models were used to predict the capacitance, *i.e.*, RF, gradient boosting machines (GBM), and extreme gradient boosting (XGBoost). With R, RMSE, and MAE values of 0.892, 25.50, and 19.56, respectively, the outcome showed that the ensemble XGBoost model exhibited a better prediction performance of supercapacitors (SCs). Among the features, the contribution of the structural features of S_{micro}/SSA was the most significant, followed by SSA and PS, to the capacitive performance of PCMs. An ML-based prediction model for the energy storage capabilities of biomass-derived activated carbons (BACs) made from mango seed husk was investigated.[162] For the prediction of capacitance, the synthesis conditions and structural and electrochemical features gathered through experimented measurements were employed. The model was trained using the input variables of activation temperature, SSA, I_D/I_G ratio, PV, average pore diameter, and current density. The multilayer perceptron (MLP) model outperformed the other ML models including decision tree (DT), logistic regression (LR), SVR, and MLP, with just a slight gap between the predicted and experimental data and values of R^2 (0.9868), MSE (4.1651), and MAE (1.5741).

Four machine learning algorithms, ANN, DT, RF, and XGBoost, were used to predict the electrochemical performance of biochar materials generated from biomass.[163] The biochar properties (C/O, C/N, atom, and SSA) and activation settings (activator ratio, activation temperature, and time), as well as the capacitance test conditions, were among the nine input parameters utilized to train the ML models. The strong R^2 and low error of the DT and XGBoost models indicated their good performance. The activator ratio, activation duration, and SSA, together with C/N and C/O, are key factors for predicting biochar capacitance, according to further research into the significance of the input features using the DT and XGBoost models.

Chemical doping with heteroatoms such as boron, nitrogen, sulphur, phosphorus, and other elements was further used to improve the capacitive performance of carbon-based materials.[164–167] To predict the performance of supercapacitors, while considering the influence of % doping components together with other structural and operational characteristics, efforts have also been directed toward the usage of ML models. Machine learning models such as ANN, LR, and LASSO were employed to predict the capacitance of heteroatom-doped carbon-based materials.[155] A data set consisting of 681 carbon-based supercapacitors was used, which was compiled from more than 300 articles. The ML model was trained using five input features, *i.e.*, SSA, PS, I_D/I_G, N-doping level, and voltage window. With an R^2 value of 0.91, it was shown that the ANN model performed better than the LR and LASSO models. Su *et al.*[168] used ML algorithms to analyze the influence of carbon material structural characteristics and heteroatom doping on capacitance performance. For the purpose of training the ML models, 121 sets of carbon-based supercapacitors with input parameters such as SSA, PV, PS, PW, I_D/I_G, % N-doping, and % O-doping were gathered from the literature. The MLP and RT models outperformed the other four ML models, *i.e.*, LR, SVR, MLP, and RT, in terms of capacitance prediction, with the RMSEs of 67.62 and 68.45, respectively. Using ML models, a group examined how the heteroatom doping arrangement of carbon electrodes affects the capacitance and retention rate.[154] The capacitance and power density of carbon electrodes in 6 M KOH electrolyte were used to measure the supercapacitor performance. The model was trained using eight input parameters, including the surface area of micro- and mesopores on the carbon materials, the chemical makeup of N/O co-doping, and the scan rates utilized in the cyclic voltammetry studies. The effects of the surface composition and structural properties of N/O co-doped carbon electrodes on the performance of supercapacitors were revealed using four distinct ML models, including GLR, SVM, RF, and ANN.

For performance prediction and design optimization of pseudocapacitive supercapacitors, including oxides and composite materials, ML approaches are also used. For the purpose of predicting the performance of various pseudocapacitive supercapacitors, many researchers have investigated ANN models. For example, the ANN model presented by Farasi and Gobal[169] for predicting the performance of a mixed-type supercapacitor. The crystal size, surface lattice length, exchange current density, and cell current were employed as input parameters for the network, while energy density, utilization, and power density were all examples of the outputs.

4.3 Fuel cell materials search and development by artificial intelligence

The chemical and material fields frequently utilize machine learning to identify novel material properties and develop next-generation materials for fuel cells.[170–172] The conventional techniques for determining or predicting the properties of a material are experimental measurement, characterisation, and theoretical calculation. These techniques are typically costly in terms of money, time, and computational resources. Furthermore, the complexity of the elements influencing material properties makes it more challenging to find the best possible material synthesis using only conventional techniques. Using current information, machine learning may help with the material selection and property prediction, which is helpful for accounting for unknown physics and chemistry and considerably enhancing the efficiency in the development of materials for fuel cells. Due to the immense potential of machine learning in chemistry and materials science, specialized tools and common machine learning frameworks have been

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 1 Accessible ML tools for chemistry, physics, and materials, and structures and properties databases. Adapted from ref. 198 as edited and most relevant information added

Database name	Description	URL
Machine learning tools for chemistry and material		
Amp	Package to facilitate machine learning for atomistic calculations	https://bitbucket.org/andrewpeterson/amp
ANI	Neural-network potentials for organic molecules with Python interface	https://github.com/isayev/ASE_ANI
COMBO	Python library with emphasis on scalability and efficiency	https://github.com/tsudalab/combo
DeepChem	Python library for deep learning of chemical systems	https://deepchem.io
GAP	Gaussian approximation potentials	https://libatoms.org/Home/Software
MatMiner	Python library for assisting machine learning in materials science	https://hackingmaterials.github.io/matminer
NOMAD	Collection of tools to explore correlations in materials datasets	https://analytics-toolkit.nomad-coe.eu
PROPhet	Code to integrate machine-learning techniques with quantum-chemistry approaches	https://github.com/biklooost/PROPhet
TensorMol	Neural-network chemistry package	https://github.com/jparkhill/TensorMol
Computed structure and property databases		
AFLOWLIB	Structure and property repository from high-throughput *ab initio* calculations of inorganic materials	https://aflowlib.org
Computational	Materials repository infrastructure to enable collection, storage, retrieval and analysis of data from electronic-structure codes	https://cmr.fysik.dtu.dk
GDB	Databases of hypothetical small organic molecules	https://gdb.unibe.ch/downloads
Materials project	Computed properties of known and hypothetical materials carried out using a standard calculation scheme	https://materialsproject.org
The materials project	Computed information on known and predicted materials including inorganic compounds, organic molecules, nanoporous materials	https://materialsproject.org
NOMAD	Input and output files from calculations using a wide variety of electronic structure codes	https://nomad-repository.eu
Open quantum	Computed properties of mostly hypothetical structures carried out using a standard	https://oqmd.org
Materials database	Calculation scheme	
NREL materials	Computed properties of materials for renewable-energy applications	https://materials.nrel.gov
Database		
TEDesignLab	Experimental and computed properties to aid the design of new thermoelectric materials	https://tedesignlab.org
ZINC	Commercially available organic molecules in 2D and 3D formats	https://zinc15.docking.org
Experimental structure and property databases		
ChemSpider	Royal Society of Chemistry's structure database, featuring calculated and experimental properties from a range of sources of chemical information based on chemical structures, including physical and chemical properties of compounds	https://chemspider.com
Citrination	Computed and experimental properties of materials	https://citrination.com
Crystallography	Open database structures of organic, inorganic, metal–organic compounds and minerals	https://crystallography.net
CSD	Repository for small-molecule organic and metal–organic crystal structures	https://www.ccdc.cam.ac.uk
ICSD	Inorganic crystal structure database	https://icsd.fiz-karlsruhe.de
MatNavi	Multiple databases targeting properties such as superconductivity and thermal conductance	https://mits.nims.go.jp
MatWeb	Datasheets for various engineering materials, including thermoplastics, semiconductors and fibres	https://matweb.com
NIST	Chemistry high-accuracy gas-phase thermochemistry and spectroscopic data	https://webbook.nist.gov/chemistry
WebBook		
NIST materials	Repository to upload materials data associated with specific publications	https://materialsdata.nist.gov
Data repository		
PubChem	Biological activities of small molecules	https://pubchem.ncbi.nlm.nih.gov

developed. In addition, several structure and property databases for molecules and solids are widely accessible for model training (see Table 1).

The primary electrochemical mechanism in energy conversion systems such as fuel cells is the oxygen reduction reaction (ORR).[173,174] The anode hydrogen oxidation reaction (HOR) is often significantly faster than the cathode ORR, which effectively limits the total performance of fuel cells.[175] Hence, to increase the kinetics to a practically useable level for a fuel cell, an ORR catalyst is needed. Due to the need for catalyst optimization in fuel cells, Zhu *et al.*[176] integrated the calculation of density functional theory (DFT) and machine learning to

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

effectively screen the dual-metal-site catalyst (DMSC) features that improve ORR activity. Subsequently, they streamlined the identified features, saved the most relevant features for the database, and trained the fitting equation between the ORR activity and catalyst property through combined application DFT and machine learning.

To predict the fuel cell performance and status, DNN with convolutional layers was utilized to calculate the water coverage ratio for a proton-exchange membrane fuel cell (PEMFC).[119] The DNN model was trained using a total of 32 pictures with 176 × 176 × 3 shapes. Particularly, the evolutionary algorithm was used to optimize the structure of the DNN model, which consisted of two dense layers and four convolutional hidden layers. The amount of water in the fuel cell was the DNN output (divided by six classes). The test results revealed that the trained DNN can identify the water coverage ratio with an accuracy of 94.23%.

The performance characteristics and internal states of PEMFCs were modeled using data by machine learning techniques such as artificial neural network (ANN) and support vector machine regressor (SVR).[177] Cell current, temperature, reactant pressures, and humidity were utilized as the PEMFC operating conditions, and the anticipated cell voltage, membrane resistance, and membrane hydration level for various operating circumstances were used as feature parameters in the model. In this study, it was further demonstrated that machine learning techniques incorporating the dropout technique can provide very accurate predictions with $R^2 \geq 0.99$ for all the predicted variables, indicating the ability to build precise data-based models using only the data from verified physics-based models, thus reducing the need for an extensive series of experiments.

5. Identified gap and envisaged solution

According to our survey on the applications of artificial intelligence in the performance prediction of electrochemical energy storage systems, an interesting observation was made that can be a future research direction in an attempt to achieve high performance EESSs. At present, AI-based approaches have not yet reached a mature stage because of lack of test-bed technique(s) that can allow for exhaustive utilisation of the available huge databases for the prediction of high-performance EESSs. Given that the metrics for the performance evaluation of EESSs involve diverse features or factors, AI-based technologies can be applied to more rich areas beyond the current focus. Although the current stage of research covers several aspects of the development of EESSs, there is still one area of importance that is not well addressed. This is the connection between the large datasets that represent metrics of performance evaluation to the macro-scale performance of EESSs. This has been challenging with the traditional trial and error experimental approach, whereas AI-based technology can be a powerful tool in handling and solving such daunting challenge. The cathode, anode, and electrolyte are the most important active materials of EESSs, while the energy density and power density are the two parameters that are essentially used to evaluate the performance of EESSs, which are commonly presented pictorially in a Ragone plot (Fig. 2a). Thus, given that the performance of EESSs is dependent on several features or factors, single-parameter or few-parameter feature(s) is not sufficient to describe their performance. A feature is any characteristic that has significant influence on the performance of an EESS, and thus can be represented as a descriptor. Therefore, to avoid confusion, multiple descriptor families need to be created to fully describe the performance of EESSs for the prediction and development of new and high performance systems for future needs. In this context, features or factors with similar characteristics can be grouped into a descriptor family. For example, given that EESSs with multiple features are complicated, these features are required to be made descriptor families, connected with each other and combined together to develop more robust predictive universal descriptors that are suitable to accommodate several parameters as metrics for the performance evaluation. Hence, it is expected that AI-aided approaches through the application of the ML process can quicken the utilisation of multiple descriptor families to expel doubt and ambiguity in the prediction and development of new and high-performance EESSs.

To shorten the time-frame required for the development of EESSs, it is necessary to switch from the trial and error approach of investigating useful materials to a more selective and/or extractive practice based on data-driven model predictions. Presently, the content performance of EESSs is very limited, given that there is still lack of material systems that exhibit the desired performance and longevity. These materials characteristically perform multiple functions, and thus the challenge is not only to find materials with appreciable functionalities but the ones exhibiting all the required functions efficiently. Specifically, EESSs are comprised of multiple material phases, particularly the electrode and electrolyte, where their overall functionalities strongly depend on how they interact with each other. Therefore, given that chemistry-based and physics-based analyses have become increasingly common to quantitatively describe the structure–property and property–performance relationship, ML can be helpful to facilitate the overall predictability of materials. This will definitely decrease experimental efforts and cost, as well as expel the rate-limiting processes of development of high-performance EESSs. The success of this may lead to a shift in science and engineering; however, its breakthrough depends on our ability to understand, reason, and formalise the underlying chemical and physical mechanisms. This is because the predictability of the material(s) response is essential in the rational design of high-performance EESSs.[156,178–180] Traditionally, the mapping of the electrochemical performance of EESSs is carried out experimentally, which is often achieved *via* multiple property combinations of materials, and thus not cost- and time-effective. However, when datasets are available, mapping can be generated with data-driven modelling and several properties can be unveiled.[181] Consequently, it avoids all possible multiple

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

property combinations of materials, which is prohibitively expensive and laborious. Therefore, reliable AI-based approaches that can accurately perceive multi-descriptors or features environment and quickly make decisions based on the categorisation of each unit are necessary in the development of high-performance EESSs.

It is obvious that features or descriptor families are characterized by definite set of parameters, which cannot all be used to predict all the prerequisite performance features of an electrochemical system for energy storage. In addition, the complexity of performance metrics in the EESS space requires appropriate descriptors for the successful training of ML models. Therefore, it has become necessary to devise methods for creating a subset of descriptors from multiple descriptors or descriptor families (*e.g.*, materials descriptors, electrochemical descriptors, and operating/environmental conditions descriptors), which will ultimately be more universal, predictable, and measurable in predicting new and high-performance EESSs. Universal descriptors can be achieved by integrating the descriptor families into one comprehensive descriptor that will accurately predict the finest features as metrics for the performance evaluation or create a finite set of descriptors from the original set of descriptors, as displayed in Fig. 10. It is believed that not all sets of features are key in determining or predicting the performance. However, when several descriptor families are combined, the number of features or descriptors increases, and likewise the chemical space, and therefore the hypothesised chemical space must be further reduced by discarding redundant and irrelevant features through feature selection/extraction.

The selection of the descriptor families is based on the fact that one core component of EESSs that determines their electrochemical performance is their active materials.[182–184] For materials with a given chemical composition, their physical and electrochemical characteristics are significantly influenced by their microstructure, crystal structure, and electronic structure.[185] However, although the electrochemical properties of electrodes are influenced by their intrinsic crystal and electronic structure of the selected materials, their microstructures vary greatly with the methods and conditions of their synthesis.[185] Recently, microspherical MFe_2O_4 (M = Ni, Co, and Zn) nanocomposites were prepared *via* a facile wet-chemical method.[186] According to this study, it was obvious that all the electrochemical performances of MFe_2O_4 exhibited a volcano-shaped relationship with their inversion degree, and thus suggested that it can be used as an electrochemical descriptor for high-performance Li ion batteries. Hence, materials and electrochemical descriptors are justifiably useful in predicting the performance of EESSs. Alternatively, besides the materials and electrochemical descriptors, their performance also depends on and/or limited by the operating or environmental conditions under which the performance was claimed and delivered. Accordingly, this should also be considered when specifying the performance of EESSs to match the other the materials and electrochemical descriptors. Otherwise, this can give rise to the problem of underestimation or prediction, given that materials and electrochemical descriptors alone cannot give the best indication of the high performance of EESSs without their corresponding operating or environmental conditions. Recently, for example, the electrochemical performance of an Ag_2O thin film supercapacitor was improved with an increased capacitance and reduced resistance with an increase in the operating temperature.[187] Thus, normally, the performance of redox flow batteries (FBs) varies with the materials and operating conditions.[167] However, it is important to note that the electrolyte for RBs is composed of active materials, solvent, and supporting electrolyte (which are generally all materials). Therefore, an AI-based approach is urgently needed for the robust prediction of the performances of EESSs by establishing models between materials, electrochemical and operating/environmental conditions (multiple descriptor families) and performance. This is important given that the identification of descriptors related to the performance of EESSs and their use in ML have the potential to significantly accelerate the discovery of EESSs with desired high performances.

As depicted in Fig. 10, it is envisaged that multi-descriptor feature engineering utilizes the revolutionary algorithm-based feature selection/extraction method, which is based on an artificial intelligence unified framework. The selection will be made from the number of descriptor families using feature selection/extraction algorithm(s). This is to form a subset or a new set of descriptors much less than the whole share of the combined parameters of the multiple or number of descriptor families. This process will be terminated by cross result validation for the creation of universal descriptors for accurate prediction of performance of EESSs. The datasets of all the multiple descriptor families can be sourced from both historical and online databases.

6. Feature engineering

Feature engineering is a pre-processing phase of ML, which mines features or descriptors from raw datasets. It employs relevant knowledge in data science and engineering for the creation of features or descriptors with which ML algorithms can achieve the best predictive models. In essence, after the construction of original datasets, feature engineering is an important step for the application of ML algorithms. Thus, it helps to represent underlying problem in predictive models in a better way, thereby improving the accuracy of the models for unobserved data. In short, feature engineering method chooses or selects the most practical and useful predictor variables for a proposed model, while the predictive model mainly contains predictor variables and an outcome variable. Operationally, feature engineering involves basically three processes, namely, feature creation, transformation, selection or extraction, as demonstrated in Fig. 11.

After the features or descriptors are created (which requires the intervention of electrochemical energy storage experts, particularly chemists, physicists, and materials scientists and engineers), the descriptors are transformed into computer-recognised mathematical representations. This is followed by feature selection or extraction depending on the purpose of action.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

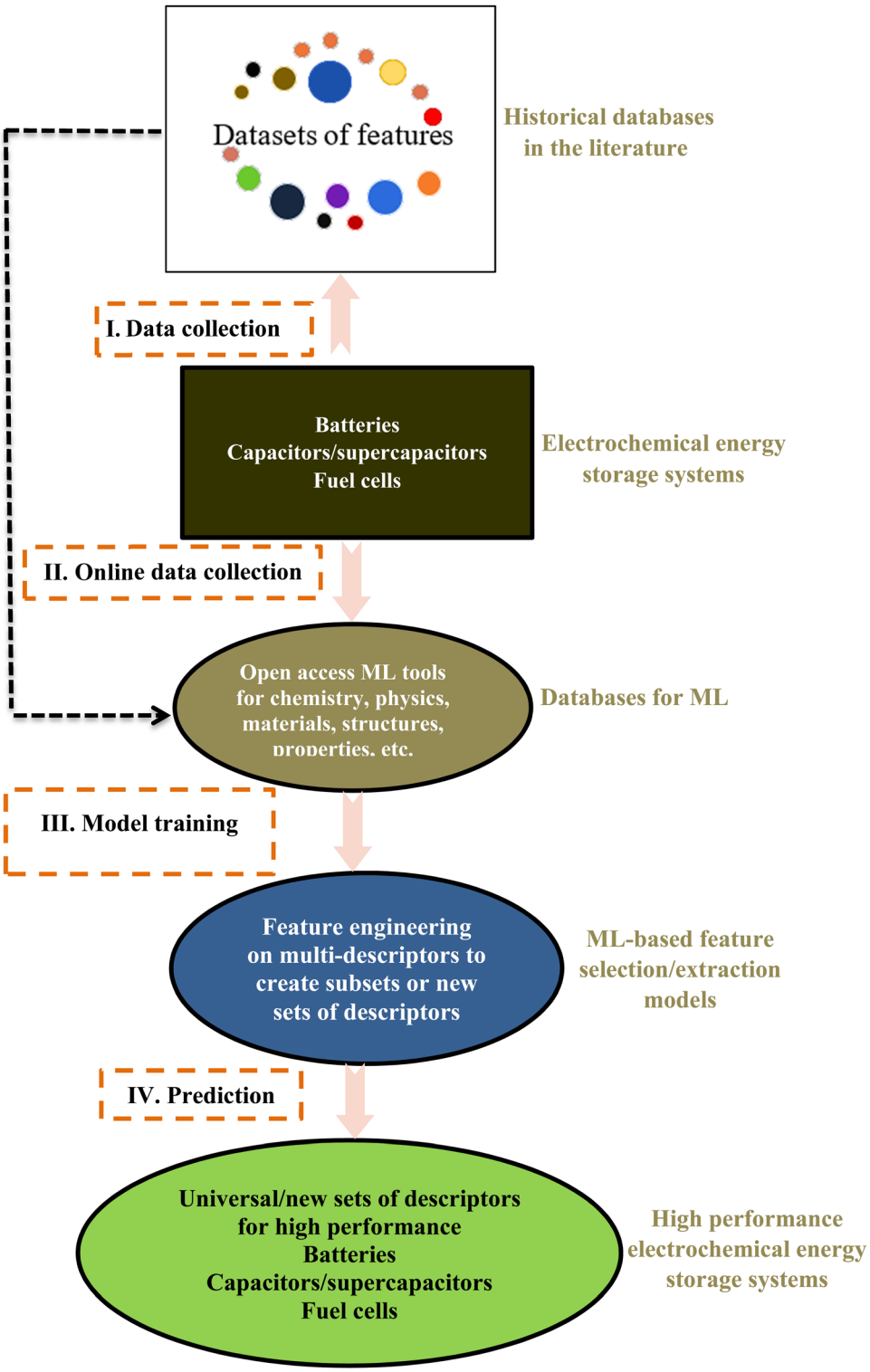

Fig. 10 Envisaged AI-based approach to predict new and high-performance EESSs based on multi-descriptor feature engineering.

6.1 Feature selection

Finding the best features from datasets is the process of feature selection.[188] The most important and difficult part of machine learning is feature selection, which deals with inappropriate, pointless, or unnecessary features.[188,189] The method of feature selection (variable reduction) enhances the prediction performance, lowers the processing demands, and lessens the impact of the cause of dimensionality.[190] Finding a selection of variables from the input that may accurately define the data input, while minimizing the influence of noisy or unneeded variables

© 2023 The Author(s). Published by the Royal Society of Chemistry

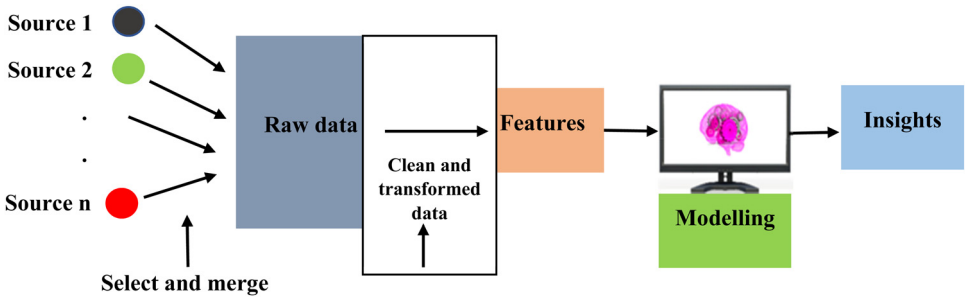

Fig. 11 Schematic illustration of feature engineering processes.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

and still providing robust predicted results is the goal of the feature selection technique.[190,191] These variables are as follows: (1) irrelevant and noisy, (2) weakly relevant and non-redundant, (3) redundant and weakly relevant.[190] A feature is irrelevant or unimportant if it is not necessary for accurate prediction. Models, search strategies, feature performance criteria, and feature evaluations are a few examples of the typical ways in the feature selection process that fall under the purview of the wrapper and filter methods support the integrity of the process. This procedure includes a sequence of selection events in the chain of activities in the feature selection, including biological issues, microarray analysis (to choose the best gene from a group of candidate genes), text mining (to find the best terms, words, or phrases), and picture analysis (to select the best visual contents, pixels, and color).[188]

The three most widely used feature selection strategies fall under the three categories of filter, wrapper, and embedding based on how they interact with the learning model.[188–190,192,193] The filtering method chooses attributes based on categorization rules or statistical measurements. It can be used for high-dimensional data because it is unaffected by the learning process and takes less time to compute than wrapper approaches.[188,191,194] Due to the multiple learning stages and cross-validation, which assess the variable subsets using predictor performance as the objective function, wrapper techniques are considerably more expensive than filter methods.[191] Consequently, this feature selection process is improved in anticipation of the implementation of the classification algorithm. Furthermore, the filter method is less precise than the wrapper strategy.[192] In a typical wrapper approach, a particular learning algorithm executes two occurrences of operations, as follows: (i) searching for a subset of features and (ii) analyzing the features found. To meet the given halting requirements, the second method is repeatedly executed.[193] Examples include genetic algorithms, sequential feature selection methods, and recursive feature removal algorithms. Three different search techniques, *i.e.*, exponential, sequential, and randomized selection strategies, are used in the wrapper selection approach. In the exponential approach, the number of analyzed features increases exponentially as a function of feature size. Accordingly, once a feature is added to or deleted from the selected subset, it cannot be changed again, leading to the local optimum in the addition or removal of characteristics in a sequential manner in the algorithm. The sequential algorithms include best first, linear forward selection, floating forward or backward selection. To widen the search area and prevent the algorithms from getting stuck in local optima, randomization selection employs randomized algorithms.[194]

The filter and wrapper methods are combined to generate embedded methods.[193] The training process is held with the classifier in embedded techniques, and feature selection is a part of it. Additionally, given that embedded methods use a learning algorithm to function, they will be categorized as wrapper approaches.[188,195]

It is also significant to note that this strategy differs significantly from traditional computational approaches such as DFT. In this method, data must first be gathered to create a training dataset, and then relevant descriptors must be generated and selected to translate the properties of the materials of interest, an appropriate algorithm must be selected to create the desired model, and finally the predictive power of the built model and its interpretability quality must be assessed. This approach is distinct from feature extraction and its details will be presented later.

6.1.1 Data collection for multi-descriptors selection. The largest issue in building an ML model for an AI-based feature selection strategy is how to gather reliable and adequate data for the materials under consideration. The datasets must contain specific input variables for the model (which can be materials, electrochemical, and operating conditions descriptors of the EESSs) as well as the properties of interest that the model is trained to predict in this case, which are related to the prediction of high-performance EESSs.[196,197] In this instance, the proposed multi-descriptor attributes can be determined from a variety of sources, including literature databases, experimental and computational measurements, and multi-descriptors. Table 1 the contains datasets that are commonly used for cheminformatics studies, while Table 2 provides some examples of critical material descriptors for high-performance EESSs. Subsequently, the most helpful descriptors are identified as features for high-performance prediction once informative variables have been synthesized and extracted from various sources.

It worth mentioning that the credibility of the published data must be practically assessed and appraised, even though

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Examples of key descriptors of materials for high-performance EESSs

EESSs	Description	Descriptors	Ref.
Batteries			
Crystalline solid batteries	Ion mobility was used as critical performance parameter	Ionic radii, oxidation states, Pauling electronegativities of the involved materials	199
Li ion battery	12 top descriptors were identified, of which 3 were the top descriptors,	Dipole polarizability, average gas phase basicity, and average heat of fusion	200
Na ion battery	While the performance of the batteries is dominated by their crystal structure, and a different descriptor was proposed	Cationic potential	201
Redox flow batteries			
Organic redox active materials	8 different material descriptors to represent 9,10-antraquinone-2,7, disulfonic acid were used organic flow batteries	Fingerprints, electronic density, symmetry function, bag of bonds, chemical environment, coulomb matrix, atom–atom radial distribution function, substructure fragmentation	202
Capacitors/supercapacitors	No descriptors mentioned but the future development is on the way of combining different morphologies of materials for creating composite materials to offer large surface area, good conductivity, high stability and more importantly high capacitance. Better choice of electrolyte was also mentioned as an important metric for high performance		203
	In addition, absence of standard performance metrics hindered the development of the energy storage system		204
Fuel cells	In energy storage system such as fuel cells, electrochemical performance is revealed by relation between electrocatalysts and charged species	d-Band center, e_g-electron number, and charge transfer capacity	205

chemical information can be mined and collected from the literature utilizing the ML-based information extraction approach. This is because there may be discrepancies in data from scattered sources due to variations in electrochemical reaction conditions from various laboratories. Therefore, developing a standard technique for reliability evaluation of published data in the literature requires collaboration between materials scientists, catalysis specialists, and computer scientists.

6.1.2 Generation and selection of relevant descriptors for features selection. The next step after gathering data is to easily translate the attributes of the materials under investigation into mathematical representations that can be used to train ML models.[206] In this instance, it entails the labeling of important electrochemical characteristics associated with the effectiveness of EESSs by a group of integers known as vectors.[206,207] These vectors are electrochemical performance characteristics that are frequently referred to in the literature as descriptors, genes, or fingerprints. The crucial technique of feature engineering in this situation is the mathematical representation of materials. This action is crucial because it calls for human intelligence and awareness. Descriptor creation or generation and descriptor selection are the two stages of the procedure.

Generation of descriptors. Although several descriptors can be produced for a given class of materials, the selection and generation of descriptors is mostly determined by the specific electrocatalytic issue that needs to be resolved.[208] There are some controlling the norms related to the generation of descriptors, although it is problem- or circumstance-dependent. According to this viewpoint, a descriptor for electrochemical features, for instance, must describe the unrivaled electrochemical qualities of a material to guarantee that they are distinct for each EESS. Additionally, the number of descriptors should not be too large to risk impairing their capacity to predict the desirable characteristics of fresh EESS data, which is typically the result of model overfitting.[208] Descriptors can typically be categorized into several classes during the generation process based on the encoded information they contain. Depending on the materials, electrochemistry, and operational circumstances of EESSs, the information can be encoded. However, it is crucial to understand that a variety of descriptors must be carefully picked to guarantee that they capture the characteristics of the EESS being simulated. The generation of descriptors in electrocatalysis should not be restricted to a single form. Specifically, to ensure that all aspects of a given EESS are accurately captured, and a robust ML model is produced, multiple families of descriptors are unavoidably needed. For instance, a dataset containing 2275 fibrous structures was built using the stochastic reconstruction method, morphological algorithm, and lattice Boltzmann method to predict the rational design of porous electrodes with a large surface area and high hydraulic permeability for redox flow batteries. For the construction of the model to estimate the precise surface area and hydraulic permeability of porous electrodes of redox flow batteries,[93] the LR, ANN, and RF algorithms were specially used. Thus, the model gave insight into the sensible design of more than 700 promising electrodes. Accordingly, descriptor generation is essential in the logical design of high-performance EESSs.

Selection of descriptors. It has become necessary to generate a subset of acceptable descriptors for the desired modelling due to the complexity involved in managing too many descriptors for ML modelling. The predictability of models produced by overfitting is compromised when ML models are built with too many descriptors, whereas optimal descriptor selection results

 © 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

in models with the highest predictivity and stress-free model interpretation.[208,209] The down selection and dimensional reduction approaches are the two main methods utilized in the selection of descriptors. When using the down selection approach, several material descriptors, for instance, are reduced using a variety of statistical techniques to a manageable amount. Regularization is used to exclude descriptors with lower relevance to a built ML model by shrinking them to zero through a procedure called least absolute shrinkage and selection operator, which is most frequently applied to regression models (LASSO).[210] Sure independence screening and sparsifying operator (SISSO), which converges pertinent characteristics to the attributes of the material of interest based on the optimal grouping, has recently been published with a demonstrated competence in handling a vast feature space.[211] Additionally, regularized RF is also frequently used, and after preparation, this method establishes the significance of each descriptor.[212] Recently, the LR and ANN algorithms were used to forecast how a membrane treated with various solvents will behave.[94] Consequently, the models determined that alcohols were the best solvent for controlling the porous structure. Although down selection is effective in reducing the number of descriptors, it occasionally compromises and discards descriptors that provide essential information about the properties of materials. This is the main drawback associated with this selection approach.

Dimensional reduction algorithms, which define methods that reduce the amount of features in a dataset based on dimensionality, are another strategy for the best selection of multiple descriptors. For instance, the method unrolls a comparable Swiss-roll from 3D to 2D in a number of input characteristics, while simultaneously saving the greatest amount of data. The original descriptors are projected from high-dimensional space into a lower-dimensional space using this technique, and the new descriptors are just a linear combination of the old ones.[213] Among the dimensional reduction procedures, principle component analysis (PCA) is the one that is most widely used. Although PCA is frequently used to model the properties of materials from a large number of descriptors (features), it only creates linear projections, whereas the correlations between structure and property and between property and performance are typically nonlinear in nature.[214] Thus, additional methods have been created to perform nonlinear dimension reduction functions considering the challenge at hand. Kernel PCA, which uses standard PCA to carry out nonlinear dimension reduction, was created to manage the difficulties of the nonlinearity nature of structure–property and property–performance correlations in electrocatalysis.[215] It is interesting to note that PCA preserves the original descriptors or characteristics. Typically, this is done by creating low-dimensional representations of the original high-dimensional data, while maintaining neighborhood links.[216,217] To anticipate high-performance EESSs, it may theoretically be used to choose discriminatory features from a variety of descriptors (such as materials, electrochemical, and operating conditions descriptors). Consequently, feature selection can classify the available descriptors based on their relative importance in terms of the capacity to draw inferences, given that each of these performance characteristics contains many descriptions. It includes determining which of the many descriptors already in use provides the most information for a decision, and it does not result in the production of any new features or descriptors; rather, it creates subsets of the original set, while maintaining the original meaning of these subsets. In essence, feature selection is an algorithm that chooses a subset of the original collection of features based on the relevance of each feature. However, with little to no information loss from the complete set of descriptors, the subset(s) should capture all the information that the entire set may offer. Consequently, the procedure can handle several EESS descriptions with various performance parameters. Fig. 12 illustrates the

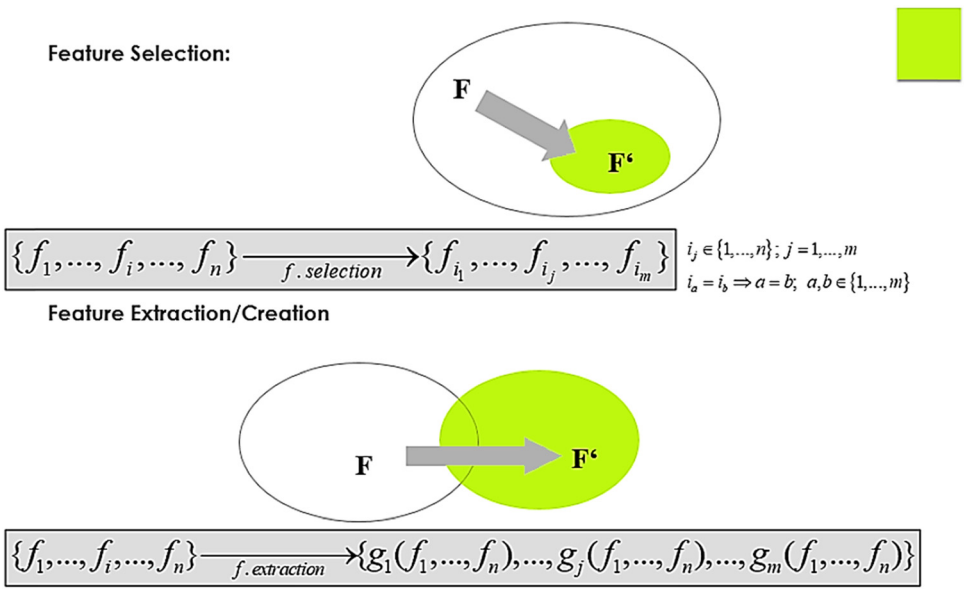

Fig. 12 Illustration of difference between feature selection and feature extraction.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

distinction between feature selection and extraction for simple understanding.

6.1.3 Selection of appropriate algorithm. Machine learning models can be trained using a wide range of algorithms as soon as the descriptors or feature subsets are formed. To create models for prediction, datasets can be divided into training sets.[204] There are numerous algorithms used in the creation of ML models, but the ones most frequently used in the design, discovery of new materials, and development of electrocatalysts are LR, SVM, RF, NN, and EL, and other algorithms, including active learning and genetic algorithms. However, genetic algorithms appear to be the most suitable for this task when considering the viewpoint presented here as well as the expected usage of the ML model for the prediction of new and high-performance EESSs. When using genetic algorithms, many EESS performance descriptors can be seen as genomes that store databases of compositional, electrochemical, and operating condition characteristics. The performance characteristics of a particular EESS can be determined by chance or past performance. To produce future generations of high-performing EESS, the fittest factors or descriptors will be subjected to genetic operators such as crossover and mutation. Electrochemical performance is a factor in the evaluation and selection of fitness levels. Each genetic optimization algorithm iteration is used to determine fitness. Until the intended descriptors are entirely optimized or no significant improvement takes place, the cycle repeats continuously. Then any descriptors that are less appropriate are discarded. Consequently, using a genetic algorithm, ML models can be trained on descriptors related to materials, electrochemistry, and operating conditions, for instance, and then used to produce universal descriptors that can be used as a substitute for a fitness function to choose performance-related factors for subsequent iterative cycles. Therefore, using ML models with genetic algorithms as fitness functions can be evolutionary and allow efforts to be focused on the EESS genomes with the best performance, thereby replacing the requirement for time-consuming and tedious amounts of tests in the construction of high-performance EESSs. Genetic algorithms have been widely used in ML modelling for discovery, property prediction, and optimization;[218–220] consequently, they are practical and can be helpful in this direction.

However, how can one choose the sort of feature selection algorithm knowing that feature selection algorithms fall under one of the three classes, *i.e.*, wrapper methods, filter methods, and embedding methods? The most important factors to consider in practice are as follows: (1) a small sample size (<1000), (2) a high dimensional space with sparse relationships between features and the target decision or variable and (3) a small sample size in a high dimensional environment. Regardless of the three aforementioned characteristics, it is advised to test various feature selection techniques and evaluate the outcomes if the dataset is large enough. However, a stability index can be used to evaluate the stability of any method under consideration.[221]

7. Envisaged workflow to address the challenge

It is thought that not every collection of performance-related criteria is essential for determining or forecasting high-performance EESSs. Additionally, when the three descriptor families that have been proposed are combined, both the number of features and the chemical space increase. Consequently, the chemical space must be further reduced by eliminating duplicate and irrelevant characteristics. It is important to note that when combined with other features, a single feature or a group of information may become superfluous or useless when used to forecast the behavior of model materials.[222]

The three descriptor families that are proposed will each have a feature chosen individually using a feature selection technique, which will finally condense the features into a much smaller subset than the total of the combined features of the three sets of multiple descriptor families. Based on the collection of several descriptors, the algorithm will be built to select the most pertinent feature. Following cross-validation of the results and selection criteria, the combined subset of multiple descriptors produced from the three descriptor families will be used to create universal descriptors for the precise prediction of high-performance EESSs, as shown in Fig. 13. Hence, to develop universal descriptors and reveal high-performance EESSs, a multi-descriptor feature selection strategy is undoubtedly crucial.

The dataset can be created using datasets of employed numerous descriptors in an effort to build a general dataset intended to forecast high-performance EESSs. The dataset for the machine learning (ML) technique will be composed of the combination of these various descriptors, as shown in Tables 1 and 2 and demonstrated in Table 3. Subsequently, the workflow technique shown in Fig. 10 can be used to operationalize the dataset for best feature selection.

By applying the feature selection technique using the objective function displayed in eqn (1), the universal descriptors is represented as the function f, as follows:

$$\text{Minimize } f(\text{MT}_i, \text{EC}_i, \text{OC}_i) = \sum_{i=1}^{n} \text{MT}_i + \sum_{i=1}^{m} \text{EC}_i + \sum_{i=1}^{p} \text{OC}_i \quad (1)$$

Subject to the following constraints

$$1 \leq \text{MT} \leq n$$

$$1 \leq \text{EC} \leq m$$

$$1 \leq \text{OC} \leq p$$

where MT, EC, and OC represent the materials, electrochemical and operating conditions descriptors, respectively. Also, n, m and p represent the number of characteristic features in the described descriptors, respectively. At any ith value of these descriptors, the most fitted descriptor(s) will be selected in the

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

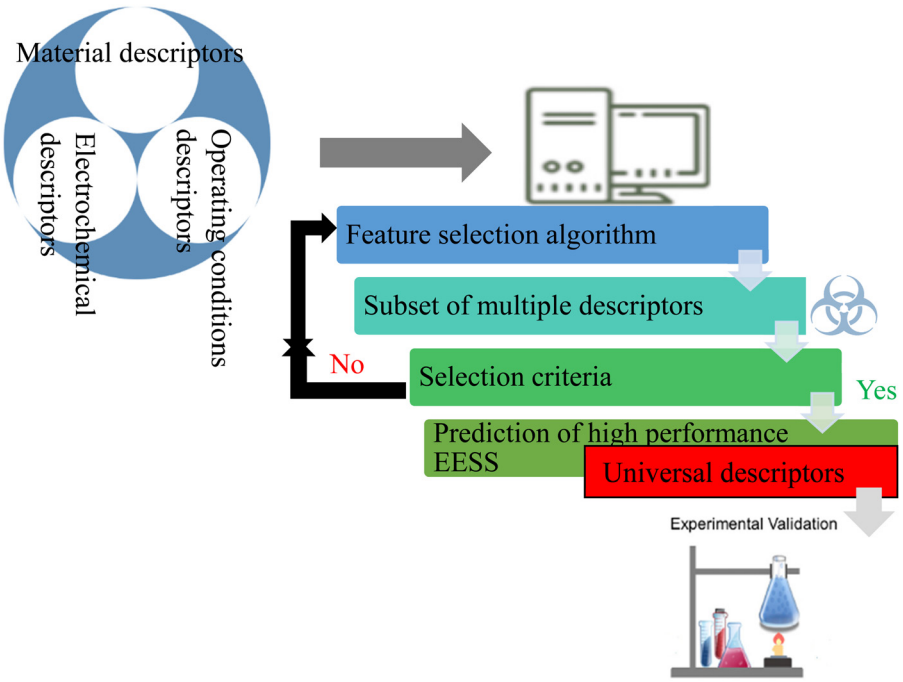

Fig. 13 Illustration of envisaged workflow process of feature selection of multiple descriptor families.

Table 3 Simple illustration of the construction of dataset for high-performance EESSs for ML approach

EESS	MT descriptors	EC descriptors	OC descriptors
$EESS_1$	$MT_1 \ldots MT_n$	$EC_1 \ldots EC_m$	$OC_1 \ldots OC_p$
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
$EESS_\infty$	∞	∞	∞

Where, $EESS_1 \ldots EESS_\infty$ represent an existing list of a chosen electrochemical energy storage system, $MT_1 \ldots MT_n$, $EC_1 \ldots EC_m$ and $OC_1 \ldots OC_p$ represent, for example, the materials, electrochemical and operating conditions descriptors, respectively.

range of one (1) and "n" in the case of the material descriptors (MT_i), one (1) and "m" in the case of the electrochemical descriptors (EC_i), and one (1) and "p" in the case of the operating conditions descriptors (OC_i). However, n, m, and p must not be the same due to the variations in the number of performance factors in the different descriptor families available for selection, as illustrated in Table 3.

8. Feature extraction

The construction or generation of new descriptors or features that are composites of the original or current dataset of descriptors is known as feature extraction and transformation (Fig. 12). It entails the development or generation of information that may not always be included in the dataset that is current or original. Consequently, feature extraction/transformation involves creating a fresh set of features. For instance, if there are n descriptors, and the decision maker does not want to remove any of them, they can apply a specific operation that yields m descriptors instead of n, where m is always $<n$ if n is the total number of the original characteristics or descriptors. Consequently, feature extraction entails transforming the entire set of "n" descriptors to obtain a new set of "m" descriptors. Then, the "m" number of descriptors or features is a transformed version of the original "n" number of descriptors. Consequently, the content of "m" is taken from the original set of descriptors, and its meaning is different. In short, the number of features obtained by feature extraction is a decision parameter given that prior to the operational transformation, a decision maker specifies the number of features. The original features may be combined in a linear or nonlinear way in the feature transformation. To produce extra, more important qualities, it implements various adjustments to the original characteristics.[223,224]

Principal component analysis (PCA), latent semantic analysis (LSA), linear discriminant analysis (LDA), partial least square (PLS), and other feature extraction techniques are only a few examples. Karl's principal component analysis is the method of feature extraction that is most well-liked and frequently employed (PCA). Convolutional neural network (CNN) performs better than any other well-known classifier, albeit in the majority of instances.[225] It can automatically extract features, learn and organize features, and uses a deep learning technique to do so from an input dataset.

Feature extraction is a vital stage in data mining given that it allows for the discovery of previously unknown information from enormous databases and the correction of inaccurate data included in the datasets.[226,227] Fig. 14 shows the proposed

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

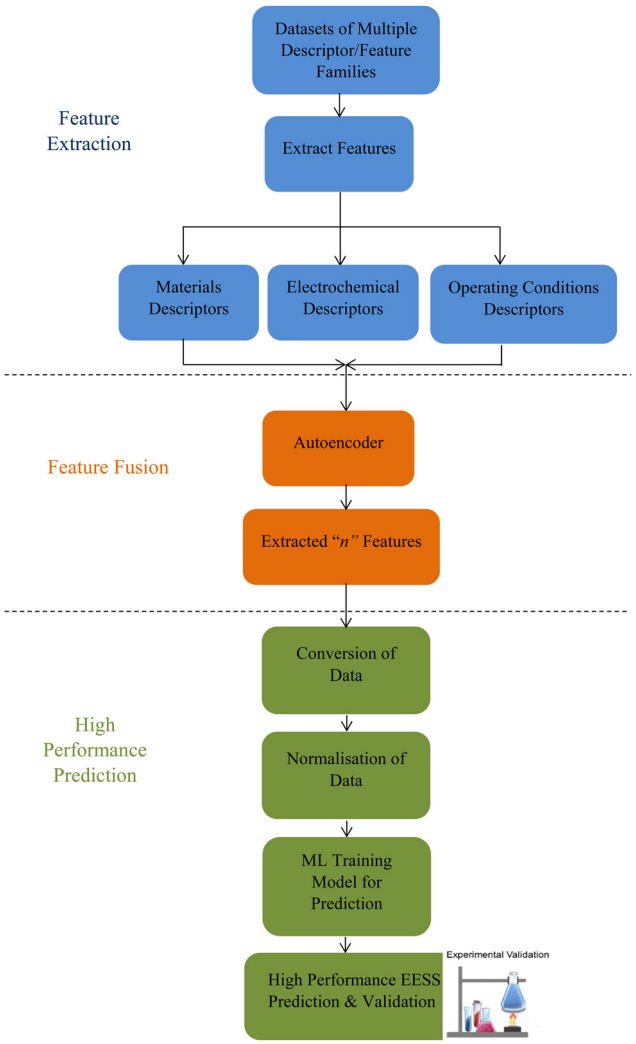

Fig. 14 Illustration of envisaged workflow process of feature extraction of multiple descriptor families.

methodology for the use of feature extraction on numerous descriptor families for high-performance EESSs. Therefore, the original data cannot be entered directly into the fusion model in an effort to extract enormous data from several descriptions of families that are not the same size. The properties of each typical descriptor family must be extracted from the original data as part of the pre-processing step. This is possibly achievable by strong and positive correlation between information representation and information dimension. Consequently, every parameter, feature, or descriptor belonging to a single family must have clear relative characteristics that are extremely sensitive to the family. Although in theory more features extracted from the source data will lead to better prediction, caution must be exercised.

The general concept of feature fusion is to reduce the feature dimension to increase the effectiveness of the prediction model. In theory, the number of highly correlated features and descriptors will directly grow as the number of extracted features and descriptors increases. In practice, this leads to redundant model information and inefficient computational processes. The most often used techniques to address this issue include autoencoder,[228] as well as PCA and ICA.[229] Autoencoder may be a good choice because there is no prior knowledge of this scenario, as shown in Fig. 15a. Consequently, Table 3 provides sets $\{MT_1 + MT_2 + MT_3 \ldots MT_n\}$; $\{EC_1 + EC_2 + EC_3 \ldots EC_n\}$; $\{OC_1 + OC_2 + OC_3 \ldots OC_n\}$ as the input data of the neutral network representing the materials, electrochemical and operating conditions descriptors, respectively, for feature fusion to train the encoder and decoder.

Following this stage, the aforementioned sets of ith data must be compressed into "n" features or descriptors that will be used as inputs in the model to forecast high-performance EESS. However, to use and train the prediction model, it is also necessary to obtain the unsupervised data generated by the autoencoder (all the outputs that match the input), which must then be transformed into supervised data. The next step is data normalization, which aims to eliminate all potential impacts that are frequently brought on by variations in value ranges. The range of all retrieved values must now be changed using the minimum–maximum normalization approach to [0, 1].[230] Deep neutral networks can be utilized to forecast high-performance EESSs because they can increase the computational capability of the training model, while also having a greater capacity for expressiveness than shallow neural networks (Fig. 15b).[231] Given that deep neural networks have the advantages of great robustness, reliability, and accuracy when handling difficult issues with several descriptors/features such to the one described here, this is very intriguing. Additionally, it features a flexible structural design that makes it easy to adapt to a variety of problems.

It is important to provide a brief history of nonlinear feature extractors when discussing feature extraction. For instance, Charles Darwin's idea of natural selection served as the foundation for the metaheuristic search method known as the genetic algorithm (GA). The main benefit and importance of GA is its capacity to manage parallelism and complex systems. To address complex issues, GA searches a large parameter space and provides a generally optimal solution (irrespective of whether the fitness function is stationary, non-stationary, linear, nonlinear, continuous, discontinuous, or with random noise).[232–234] It comes about as a result of the numerous children in the population function as independent agents. These populations explore in several directions, while simultaneously searching. The three basic GA operators are crossover, mutation, and selection/extraction.[232–237] As shown in Fig. 16, chromosomes are selected for future reproduction in selection/extraction operators (using, for example, roulette wheel selection/extraction).

When using the objective function of eqn (1), the function is initially assessed at the MT_i, EC_i, OC_i locations in P(0) of the initial population of the chromosomes of the parents.[238] Following this assessment, a new population of population denoting points P(1) is produced.[239,240] Recombination, crossover, and mutation are a succession of processes that lead to the formation of point P(1). These processes include the exchange

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

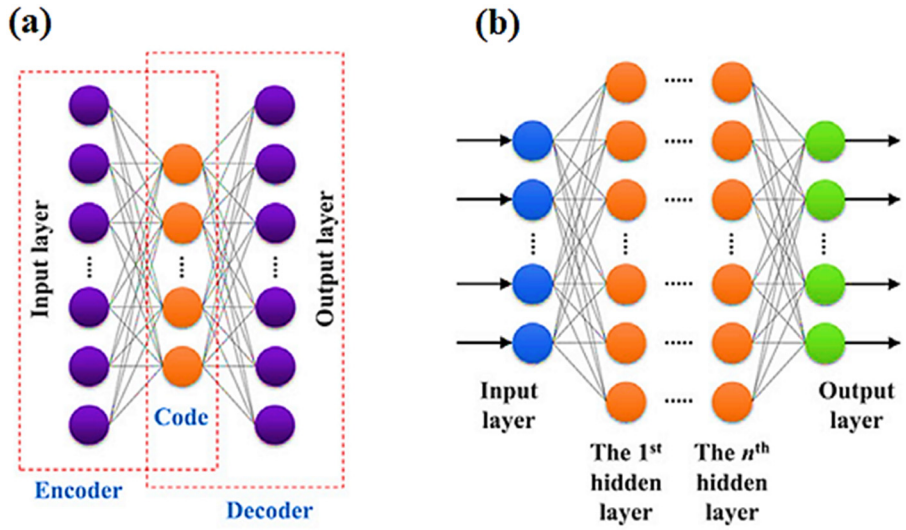

Fig. 15 (a) Envisaged autoencoder model and (b) deep neural network model.

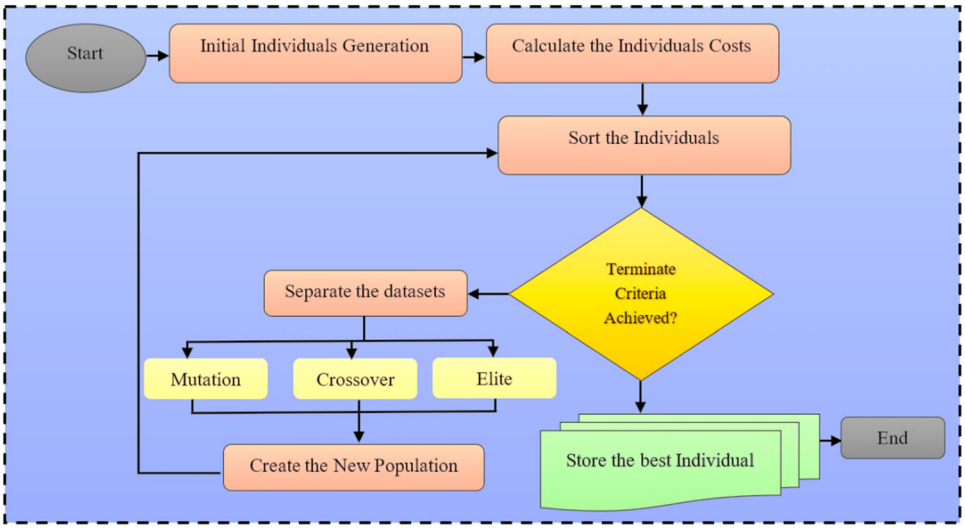

Fig. 16 Step-by step flowchart of GA.

of traits between two different parents to create offspring.[241–245] The entire procedure requires switching chromosomal substrings from both parents.[246] The crossover parent pairs are selected at random from the mating pool, and the likelihood that a chromosome will be used for operation is denoted by P_c.[247,248] In the process of mutation, a single chromosome is taken from the mating pool, and each symbol is changed at random with the probability P_m.[240,248] Consequently, probability P_m is used to represent each characteristic descriptor feature. Depending on the number of iterations specified with respect to the termination criterion, these operational procedures are repeated to generate populations P(2), P(3), P(4), *etc.*[239,249] The ideal solution is both met and offered at the end criterion.[240,250–252] Therefore, to create new sets of descriptors from the original databases or to gain the best universal descriptors, GA on the three descriptor families is done to predict the high performance of EESSs.

Nevertheless, it is still advisable to test various feature selection and/or extraction algorithms and compare outcomes when there are enough input datasets. However, a stability index can be used to evaluate the stability of any method under consideration.[253]

9. Conclusion

The development of high-performance electrochemical energy storage systems requires intense efforts of processing and preparation of cathode, anode, and electrolyte, which are the active materials targeted for high energy density and power

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

density. Owing to the importance of the properties of the active materials in EESSs, especially their impact on performance, the alignment of properties of the active materials such as crystal structure, electronic structure and their microstructural features is the key design parameter. This is traditionally achieved by the trial and error approach, which is tedious and laborious, characterised with a set of limitations such as cost- and time-consuming, and the approach is currently considered destructive. In addition, the experimental trial and error approach cannot observe all the multiple performance metrics of EESSs to facilitate and quicken the processing of the properties-performance features of the existing and emerging chemistries and physics of active materials of EESSs. In contrast, AI-based techniques can be applied to predict features/descriptors that determine or inform the performance characteristics of EESSs within hours instead of days or years. Although state estimation and prediction (*i.e.*, SOC and SOH), remaining useful life prediction (RUL), prediction of property classification, fault diagnosis and discovery, modelling, design, and optimization of EESSs are the most common problems associated with the broad application of AI-based approaches, choosing appropriate AI-based strategy(ies) to predict high-performance EESSs is also significant. However, thus far, no attempt has been made to use AI-based strategy(ies) such as feature engineering to predict high performance through multiple feature/descriptor families (materials, electrochemical, and operating/environmental conditions) of active materials of EESSs.

Author contributions

In the authorship of this paper, every person contributed in idea generation and composition, visualization, writing, and editing.

Conflicts of interest

There are no conflicting interests declared by the authors.

Acknowledgements

The authors acknowledge the funding provided by Interdisciplinary Research Center for Hydrogen and Energy Storage (IRC-HES), King Fahd University of Petroleum and Minerals (KFUPM) through project#INHE2213. H. Adamu gratefully acknowledges Abubakar Tafawa Balewa University, Bauchi, Nigeria for grating him research fellowship.

References

1 S. S. Siwal, Q. Zhang, N. Devi and V. K. Thakur, *Polymers*, 2020, **12**, 505.
2 Q. Abbas, M. Mirzaeian, M. R. Hunt, P. Hall and R. Raza, *Energies*, 2020, **13**, 5847.
3 L. Bird, D. Lew, M. Miiligan, E. M. Carlini, A. Estanqueiro, D. Flynn, E. Gomez-Lazaro, H. Holttinen, N. Menemenlis, A. Orths and P. B. Eriksen, *Renewable Sustainable Energy Rev.*, 2016, **65**, 577–586.
4 Theme Report on Energy Transition – Towards the Achievement of SDG 7 and Net Zero Emissions''. United Nations, 2021. Accessed on the 16th September, 2022 from **https://www.un.org/sites/un2.un.org/files/2021-twg_2-062321.pdf**.
5 S. P. Badwal, S. S. Giddey, C. Munnings, A. I. Bhatt and A. F. Hollenkamp, *Front. Chem.*, 2014, **2**, 79.
6 A. N. Abdalla, M. S. Nazir, H. Tao, S. Cao, R. Ji, M. Jiang and L. Yao, *J. Energy Storage*, 2021, **40**, 102811.
7 L. Cheng, R. S. Assary, X. Qu, A. Jain, S. P. Ong, N. N. Rajput, K. Persson and L. A. Curtiss, *J. Phys. Chem. Lett.*, 2015, **6**, 283–291.
8 K. He, X. Zhang, S. Ren and J. Sun, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, 770–778.
9 J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling and S. Thrun, *2011 IEEE Intelligent Vehicles Symposium (IV)*, 2011, 163–168.
10 P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
11 Y. Sun, S. M. Ayalasomayajula, A. Deva, G. Lin and R. E. García, *Sci. Rep.*, 2022, **12**, 1–11.
12 M. Aykol, P. Herring and A. Anapolsky, *Nat. Rev. Mater.*, 2020, **5**, 725–727.
13 S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito and O. Levy, *Nat. Mater.*, 2013, **12**, 191–201.
14 A. D. Sendek, E. D. Cubuk, E. R. Antoniuk, G. Cheon, Y. Cui and E. J. Reed, *Chem. Mater.*, 2018, **31**, 342–352.
15 Y. Liu, B. Guo, X. Zou, Y. Li and S. Shi, *Energy Storage Mater.*, 2020, **31**, 434–450.
16 J. Gubernatis and T. Lookman, *Phys. Rev. Mater.*, 2018, **2**, 120301.
17 G. Pilania, C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, *Sci. Rep.*, 2013, 1–6.
18 A. Seko, H. Hayashi, K. Nakayama, A. Takahashi and I. Tanaka, *Phys. Rev. B*, 2017, **95**, 144110.
19 L. Ward, A. Agrawal, A. Choudhary and C. A. Wolverton, *npj Comput. Mater.*, 2016, **2**, 1–7.
20 A. D. Sendek, Q. Yang, E. D. Cubuk, K. A. N. Duerloo, Y. Cui and E. J. Reed, *Energy Environ. Sci.*, 2017, **10**, 306–320.
21 R. Jalem, M. Nakayama and T. Kasuga, *J. Mater. Chem. A*, 2014, **2**, 720–734.
22 Z. Ahmad, T. Xie, C. Maheshwari, J. C. Grossman and V. Viswanathan, *ACS Cent. Sci.*, 2018, **4**, 996–1006.
23 D. Morgan and R. Jacobs, *Annu. Rev. Mater.*, 2020, **50**, 71–103.
24 N. Gibbs, D. W. Pine and K. Pollack, *Artificial Intelligence: the Future of Humankind*, ed. N. Gibbs, Time Inc. Books, 2017.
25 A. Mistry, A. A. Franco, S. J. Cooper, S. A. Roberts and V. Viswanathan, *ACS Energy Lett.*, 2021, **6**, 1422–1431.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

26 R. Stevens, V. Taylor, J. Nichols, A. B. MacCabe, K. Yelick and D. Brown, AI for Science, Feb. 2020. Accessed on the 17th September, 2022 from https://www.anl.gov/ai-forscience-report.
27 B. P. MacLeod, *et al.*, *Sci. Adv.*, 2020, **6**, eaaz8867.
28 C. Mejia and Y. Kajikawa, *Appl. Energy*, 2020, **263**, 114625.
29 B. G. Pollet, I. Staffell, J. L. Shang and V. Molkov, Fuel-cell (hydrogen) electric hybrid vehicles, in *Alternative Fuels and Advanced Vehicle Technologies for Improved Environmental Performance*, ed. R. Folkson, Elsevier, 2014, pp. 685–735.
30 L. Trahey, F. R. Brushett, N. P. Balsara, G. Ceder, L. Cheng, Y. M. Chiang, N. T. Hahn, B. J. Ingram, S. D. Minteer, J. S. Moore and K. T. Mueller, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 12550–12557.
31 D. A. Howey, S. A. Roberts, V. Viswanathan, A. Mistry, M. Beuse, E. Khoo, S. C. DeCaluwe and V. Sulzer, *Electrochem. Soc. Interface*, 2020, **29**, 28–32.
32 M. Armand and J.-M. Tarascon, *Nature*, 2008, **451**, 652–657.
33 A. Franco and A. Barnard, *Batteries Supercaps*, 2022, **5**, e202200149.
34 P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y. H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang and P. K. Herring, *Nature*, 2020, **278**, 397–402.
35 L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan and M. Wei, *IEEE Access*, 2019, **7**, 162415–162438.
36 X.-S. Si, W. Wang, C.-H. Wang and D.-H. Zhou, *Eur. J. Oper. Res.*, 2011, **213**, 1–14.
37 Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo and H. E. Hoster, *Renewable Sustainable Energy Rev.*, 2019, **113**, 109254.
38 M. Murnane and A. Ghazel. A closer look at state of charge (SOC) and state of health (SOH) estimation techniques for batteries. Analog devices. Accessed on the 29th September, 2022 from https://www.analog.com/media/en/technical-documentation/technical-articles/a-closer-look-at-state-of-charge-and-state-health-estimation-techniques.pdf.
39 Y. Song, D. Liu, H. Liao and Y. Peng, *Appl. Energy*, 2020, **261**, 114408.
40 X. Tang, K. Yao, B. Liu, W. Hu and F. Gao, *Energies*, 2018, **11**, 86.
41 N. Lin, X. Xie, R. Schenkendorf and U. Krewer, *J. Electrochem. Soc.*, 2018, **165**, A1169–A1183.
42 T. Hutzenlaub, S. Thiele, N. Paust, R. Spotnitz, R. Zengerle and C. Walchshofer, *Adv. Automot. Batter. Technol. Appl. Mark. AABTAM*, 2014, **21**, 131–139.
43 M. Guo, G. H. Kim and R. E. White, *J. Power Sources*, 2013, **240**, 80–94.
44 J. Bao, V. Murugesan, C. J. Kamp, Y. Shao, L. Yan and W. Wang, *Adv. Theory Simul.*, 2020, **3**, 1900167.
45 T. Li, F. Xing, T. Liu, J. Sun, D. Shi, H. Zhang and X. Li, *Energy Environ. Sci.*, 2020, **13**, 4353–4361.
46 S. Wan, X. Liang, H. Jiang, J. Sun, N. Djilali and T. Zhao, *Appl. Energy*, 2021, **298**, 117177.
47 T. Li, W. Lu, Z. Yuan, H. Zhang and X. Li, *J. Mater. Chem. A*, 2021, **9**, 14545–14552.
48 M. S. Ziegler and J. E. Trancik, *Energy Environ. Sci.*, 2021, **14**, 1635.
49 D. Chung, E. Elgqvist and S. Santhanagopalan, *Technical Report, Clean Energy Manufacturing Analysis Center (CEMAC)*, 2015.
50 D.-W. Chung, M. Ebner, D. R. Ely, V. Wood and R. E. García, *Model. Simul. Mater. Sci. Eng.*, 2013, **21**, 074009.
51 R. García-García and R. E. García, *J. Power Sources*, 2016, **309**, 11.
52 M. Ebner, D.-W. Chung, R. E. García and V. Wood, *Adv. Energy Mater.*, 2014, **4**, 1301278.
53 D. Bruggeman, *Ann. Phys.*, 1935, **416**, 636–791.
54 B. Vijayaraghavan, D. R. Ely, Y.-M. Chiang, R. García-García and R. E. García, *J. Electrochem. Soc.*, 2012, **159**, A548.
55 M. Ebner, F. Geldmacher, F. Marone, M. Stampanoni and V. Wood, *Adv. Energy Mater.*, 2013, **3**, 845–850.
56 D. Kehrwald, P. R. Shearing, N. P. Brandon, P. K. Sinha and S. J. Harris, *J. Electrochem. Soc.*, 2011, **158**, A1393.
57 S. Korneev, H. Arunachalam, S. Onori and I. A. Battiato, *Trans. Porous Media*, 2020, **134**, 173–194.
58 P. Pietsch and V. Wood, *Annu. Rev. Mater. Res.*, 2017, **47**, 451–479.
59 O. O. Taiwo, D. P. Finegan, D. S. Eastwood, J. L. Fife, L. D. Brown, J. A. Darr, P. D. Lee, D. J. Brett and P. R. Shearing, *J. Microsc.*, 2016, **263**, 280–292.
60 F. Pouraghajan, H. Knight, M. Wray, B. Mazzeo, R. Subbaraman, J. Christensen and D. Wheeler, *J. Electrochem. Soc.*, 2018, **165**, A2644.
61 I. V. Thorat, D. E. Stephenson, N. A. Zacharias, K. Zaghib, J. N. Harb and D. R. Wheeler, *J. Power Sources*, 2009, **188**, 592–600.
62 J. Landesfeind, J. Hattendorf, A. Ehrl, W. A. Wall and H. A. Gasteiger, *J. Electrochem. Soc.*, 2016, **163**, A1373.
63 T. T. Nguyen, A. Demortière, B. Fleutot, B. Delobel, C. Delacourt and S. J. Cooper, *npj Comput. Mater.*, 2020, **6**, 1–12.
64 M. Aykol, P. Herring and A. Anapolsky, *Nat. Rev. Mater.*, 2020, **5**, 725–727.
65 X.-S. Si, W. Wang, C.-H. Hu and D.-H. Zhou, *Eur. J. Oper. Res.*, 2011, **213**, 1–14.
66 Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo and H. E. Hoster, *Renewable Sustainable Energy Rev.*, 2019, **113**, 109254.
67 Y. Zhang, R. Xiong, H. He and M. G. Pecht, *IEEE Trans. Veh. Technol.*, 2018, **67**, 5695–5705.
68 X. Hu, S. E. Li and Y. Yang, *IEEE Trans. Transp. Electrif.*, 2015, **2**, 140–149.
69 F. Xiao, C. Li, Y. Fan, G. Yang and X. Tang, *Int. J. Electron. Power Energy Syst.*, 2021, **124**, 106369.
70 C. Bian, H. He, S. Yang and T. Huang, *J. Power Sources*, 2020, **449**, 227558.
71 M. A. Hannan, D. N. How, M. H. Lipu, M. Mansor, P. J. Kar, Z. Y. Dong, K. S. Sahari, S. K. Tiong, K. M. Muttaqi, T. I. Mahlia and F. Blaabjerg, *Sci. Rep.*, 2021, **11**, 1–13.
72 M. A. Hannan, D. N. How, M. H. Lipu, P. J. Kar, Z. Y. Dong, M. Mansor and F. Blaabjerg, *IEEE Trans. Power Electron.*, 2020, **36**, 7349–7353.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

73 S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito and O. Levy, *Nat. Mater.*, 2013, **12**, 191–201.
74 D. P. Finegan and S. J. Cooper, *Joule*, 2019, **3**, 2599–2601.
75 Y. Liu, B. Guo, X. Zou, Y. Li and S. Shi, *Energy Storage Mater.*, 2020, **31**, 434–450.
76 J. Gubernatis and T. Lookman, *Phys. Rev. Mater.*, 2018, **2**, 120301.
77 G. Pilania, C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, *Sci. Rep.*, 2013, **3**, 1–6.
78 A. Seko, H. Hayashi, K. Nakayama, A. Takahashi and I. Tanaka, *Phys. Rev.*, 2017, **B95**, 144110.
79 L. Ward, A. Agrawal, A. Choudhary and C. A. Wolverton, *npj Comput. Mater.*, 2016, **2**, 1–7.
80 D. H. Barrett and A. Haruna, *Curr. Opin. Electrochem.*, 2020, **21**, 160–166.
81 R. Jalem, M. Nakayama and T. Kasuga, *J. Mater. Chem. A*, 2014, **2**, 720–734.
82 Z. Ahmad, T. Xie, C. Maheshwari, J. C. Grossman and V. Viswanathan, *ACS Cent. Sci.*, 2018, **4**, 996–1006.
83 E. Chemali, P. J. Kollmeyer, M. Preindl and A. Emadi, *J. Power Sources*, 2018, **400**, 242–255.
84 S. Tong, J. H. Lacap and J. W. Park, *J. Energy Storage*, 2016, **7**, 236–243.
85 X. Song, F. Yang, D. Wang and K. L. Tsui, *IEEE Access*, 2019, **7**, 88894–88902.
86 S. Shen, M. Sadoughi, M. Li, Z. Wang and C. Hu, *Appl. Energy*, 2020, **260**, 114296.
87 G. Ma, Y. Zhang, C. Cheng, B. Zhou, P. Hu and Y. Yuan, *Appl. Energy*, 2019, **253**, 113626.
88 A. Veeraraghavan, V. Adithya, A. Bhave and S. Akella, *IEEE Trans. Electrif. Conf.*, 2018, 1–4.
89 Y. Wang, Y. Chen, X. Liao and L. Dong, *Proc. 2019 IEEE Int. Conf. Mechatronics Autom. ICMA*, 2019, 1067–1072.
90 J. Liu, Q. Li, W. Chen, Y. Yan, Y. Qiu and T. Cao, *Int. J. Hydrogen Energy*, 2019, **4**, 5470–5480.
91 J. Qu, F. Liu, Y. Ma and J. Fan, *IEEE Access*, 2019, **7**, 87178–87191.
92 K. Park, Y. Choi, W. J. Choi, H. Y. Ryu and H. Kim, *IEEE Access*, 2020, **8**, 20786–20798.
93 M. Lucu, E. Martinez-Laserna, I. Gandiaga, K. Liu, H. Camblong, W. D. Widanage and J. Marco, *J. Energy Storage*, 2020, **30**, 101409.
94 X. Li, T. Zhang and Y. Liu, *Sensor*, 2019, **19**, 4702.
95 L. Dong, J. Wesseloo, Y. Potvin and X. Li, *Rock Mech. Rock Eng.*, 2016, 183–211.
96 S. Kim, S. Lee and K. Cho, *J. Semicond. Technol. Sci.*, 2012, **12**, 162–167.
97 K. Song, X. Wang, J. Li, B. Zhang, R. Yang, P. Liu and J. Wang, *Electrochim. Acta*, 2020, **340**, 135892.
98 A. Berrueta, A. Ursúa, I. San Martín, A. Eftekhari and P. Sanchis, *IEEE Access*, 2019, **7**, 50869.
99 Y. Wang, Y. Song and Y. Xia, *Chem. Soc. Rev.*, 2016, **45**, 5925–5950.
100 Q. Wang, J. Yan and Z. Fan, *Energy Environ. Sci.*, 2016, **9**, 729–762.
101 C. Zhang, W. Lv, Y. Tao and Q.-H. Yang, *Energy Environ. Sci.*, 2015, **8**, 1390–1403.
102 P. Simon and Y. Gogotsi, *Acc. Chem. Res.*, 2013, **46**, 1094–1103.
103 M. Zhou, A. Gallegos, K. Liu, S. Dai and J. Wu, *Carbon N. Y.*, 2020, **157**, 147–152.
104 H. Su, C. Lian, J. Liu and H. Liu, *Chem. Eng. Sci.*, 2019, **202**, 186–193.
105 S. Parwaiz, O. A. Malik, D. Pradhan and M. M. Khan, *J. Chem. Inf. Model.*, 2018, **58**, 2517–2527.
106 S. O. T. Ogaji, R. Singh, P. Pilidis and M. Diacakis, *J. Power Sources*, 2006, **154**, 192–197.
107 M. Mehrpooya, B. Ghorbani, B. Jafari, M. Aghbashlo and M. Pouriman, *Thermal Sci. Eng. Prog.*, 2018, **7**, 8–19.
108 I. S. Han and C. B. Chung, *Int. J. Hydrogen Energy*, 2016, **41**, 10202–10211.
109 A. Kheirandish, F. Motlagh, N. Shafiabady and M. Dahari, *Int. J. Hydrogen Energy*, 2016, **41**, 9585–9594.
110 A. Kheirandish, N. Shafiabady, M. Dahari, M. S. Kazemi and D. Isa, *Int. J. Hydrogen Energy*, 2016, **41**, 11351–11358.
111 W. Huo, W. Li, Z. Zhang, C. Sun, F. Zhou and G. Gong, *Energy Convers. Manage.*, 2021, **243**, 114367.
112 V. Subotić, M. Eibl and C. Hochenauer, *Energy Convers. Manage.*, 2021, **230**, 113764.
113 B. Wang, B. Xie, J. Xuan and K. Jiao, *Energy Convers. Manage.*, 2020, **205**, 112460.
114 B. Wang, G. Zhang, H. Wang, J. Xuan and K. Jiao, *Energy AI*, 2020, **1**, 100004.
115 A. U. Chávez-Ramírez, R. Munoz-Guerrero, S. M. Durón-Torres, M. Ferraro, G. Brunaccini, F. Sergi, V. Antonucci and L. G. Arriaga, *Int. J. Hydrogen Energy*, 2010, **35**, 12125–12133.
116 Y. Bicer, I. Dincer and M. U. R. A. T. Aydin, *Energy*, 2016, **116**, 1205–1217.
117 G. Zhang, H. Yuan, Y. Wang and K. Jiao, *Appl. Energy*, 2019, **255**, 113865.
118 N. Khajeh-Hosseini-Dalasm, S. Ahadian, K. Fushinobu, K. Okazaki and Y. Kawazoe, *J. Power Sources*, 2011, **196**, 3750–3756.
119 H. Mehnatkesh, A. Alasty, M. Boroushaki, M. H. Khodsiani, M. R. Hasheminasab and M. J. Kermani, *IEEE Sens. J.*, 2020, 1.
120 W. Cai, K. L. Lesnik, M. J. Wade, E. S. Heidrich, Y. Wang and H. Liu, *Biosens. Bioelectron.*, 2019, **133**, 64–71.
121 K. L. Lesnik and H. Liu, *Environ. Sci. Technol.*, 2017, **51**, 10881–10892.
122 X. Zhu, J. Yan, M. Gu, T. Liu, Y. Dai, Y. Gu and Y. Li, *J. Phys. Chem. Lett.*, 2019, **10**, 7760–7766.
123 Z. H. Shen, H. X. Liu, Y. Shen, J. M. Hu, L. Q. Chen and C. W. Nan, *Interdiscip. Mater.*, 2022, 1–21.
124 C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng and S. P. Ong, *Adv. Energy Mater.*, 2020, **10**, 1903242.
125 Y. Liu, Q. Zhou and G. Cui, *Small Methods*, 2021, **5**, e2100442.
126 P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
127 Y. Liu, B. Guo, X. Zou, Y. Li and S. Shi, *Energy Storage Mater.*, 2020, **31**, 434–450.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

128 C. Lv, X. Zhou, L. Zhong, C. Yan, M. Srinivasan, Z. W. Seh, C. Liu, H. Pan, S. Li, Y. Wen and Q. Yan, *Adv. Mater.*, 2021, e2101474.
129 V. Etacheri, R. Marom, R. Elazari, G. Salitra and D. Aurbach, *Energy Environ. Sci.*, 2011, **4**, 3243–3262.
130 M. Hu, X. Pang and Z. Zhou, *J. Power Sources*, 2013, **237**, 229–242.
131 J.-M. Tarascon and M. Armand, *Nature*, 2001, **414**, 359–367.
132 I. A. Moses, R. P. Joshi, B. Ozdemir, N. Kumar, J. Eickholt and V. Barone, *ACS Appl. Mater. Interfaces*, 2021, **13**, 53355–53362.
133 R. P. Joshi, J. Eickholt, L. Li, M. Fornari, V. Barone and J. E. Peralta, *ACS Appl. Mater. Interfaces*, 2019, **11**, 18494–18503.
134 C. Yan, R. Xu, Y. Xiao, J. F. Ding, L. Xu, B. Q. Li and J. Q. Huang, *Adv. Funct. Mater.*, 2020, **30**, 1909887.
135 X. Chen, X. Liu, X. Shen and Q. Zhang, *Angew. Chem., Int. Ed.*, 2021, **60**, 24354–24366.
136 Y. Xiao, Y. Wang, S.-H. Bo, J. C. Kim, L. J. Miara and G. Ceder, *Nat. Rev. Mater.*, 2019, **5**, 105–126.
137 H. Che, S. Chen, Y. Xie, H. Wang, K. Amine, X.-Z. Liao and Z.-F. Ma, *Energy Environ. Sci.*, 2017, **10**, 1075–1101.
138 T. Famprikis, P. Canepa, J. A. Dawson, M. S. Islam and C. Masquelier, *Nat. Mater.*, 2019, **18**, 1278–1291.
139 E. D. Cubuk, A. D. Sendek and E. J. Reed, *J. Chem. Phys.*, 2019, **150**, 214701.
140 K. Fujimura, A. Seko, Y. Koyama, A. Kuwabara, I. Kishida, K. Shitara, C. A. J. Fisher, H. Moriwake and I. Tanaka, *Adv. Energy Mater.*, 2013, **3**, 980–985.
141 Q. Zhao, M. Avdeev, L. Chen and S. Shi, *Sci. Bull.*, 2021, **66**, 1401–1408.
142 Q. Zhao, L. Zhang, B. He, A. Ye, M. Avdeev, L. Chen and S. Shi, *Energy Storage Mater.*, 2021, **40**, 386–393.
143 A. Wang, Z. Zou, D. Wang, Y. Liu, Y. Li, J. Wu, M. Avdeev and S. Shi, *Energy Storage Mater.*, 2021, **35**, 595–601.
144 A. Ishikawa, K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama and M. Okada, *Phys. Chem. Chem. Phys.*, 2019, **21**, 26399–26405.
145 B. C. Melot and J. M. Tarascon, *Acc. Chem. Res.*, 2013, **46**, 1226.
146 V. Augustyn, J. Come, M. A. Lowe, J. W. Kim, P. L. Taberna, S. H. Tolbert, H. D. Abruña, P. Simon and B. Dunn, *Nat. Mater.*, 2013, **12**, 518.
147 M. Zhang, T. Wang and G. Cao, *Int. Mater. Rev.*, 2015, **60**, 330.
148 B. Wang, B. Luo, X. Li and L. Zhi, *Mater. Today*, 2012, **15**, 544.
149 L. Eliad, G. Salitra, A. Soffer and D. Aurbach, *J. Phys. Chem. B*, 2001, **105**, 6880–6887.
150 A. Ghosh and Y. H. Lee, *ChemSusChem*, 2012, **5**, 480–499.
151 G. Hasegawa, K. Kanamori, K. Nakanishi and T. Abe, *J. Phys. Chem. C*, 2012, **116**, 26197–26203.
152 G. J. Lee and S. I. Pyun, *Langmuir*, 2006, **22**, 10659–10665.
153 M. Zhou, A. Gallegos, K. Liu, S. Dai and J. Wu, *Carbon*, 2020, **157**, 147–152.
154 M. Zhou, A. Vassallo and J. Wu, *ACS Appl. Energy Mater.*, 2020, **3**, 5993–6000.
155 S. Zhu, J. Li, L. Ma, C. He, E. Liu, F. He, C. Shi and N. Zhao, *Mater. Lett.*, 2018, **233**, 294–297.
156 C. H. Wang, W. C. Wen, H. C. Hsu and B. Y. Yao, *Adv. Powder Technol.*, 2016, **27**, 1387–1395.
157 Q. Gao, *J. Energy Chem.*, 2019, **38**, 219–224.
158 J. Zhao and A. F. Burke, *J. Energy Chem.*, 2021, **59**, 276–291.
159 Z. Ling, Z. Wang, M. Zhang, C. Yu, G. Wang, Y. Dong, S. Liu, Y. Wang and J. Qiu, *Adv. Funct. Mater.*, 2016, **26**, 111–119.
160 H. Su, C. Lian, J. Liu and H. Liu, *Chem. Eng. Sci.*, 2019, **202**, 186–193.
161 P. Liu, Y. Wen, L. Huang, X. Zhu, R. Wu, S. Ai, T. Xue and Y. Ge, *J. Electroanal. Chem.*, 2021, **899**, 115684.
162 W. A. M. K. P. Wickramaarachchi, M. Minakshi, X. Gao, R. Dabare and K. W. Wong, *Chem. Eng. J. Adv.*, 2021, **8**, 100158.
163 X. Yang, C. Yuana, S. He, D. Jiang, B. Cao and S. Wang, *Fuel*, 2023, **331**, 125718.
164 L. Tang, Y. Zhou, X. Zhou, Y. Chai, Q. Zheng and D. Lin, *J. Mater. Sci.: Mater. Electron.*, 2019, **30**, 2600–2609.
165 Z. Shang, X. An, H. Zhang, M. Shen, F. Baker, Y. Liu, L. Liu, J. Yang, H. Cao, Q. Xu, H. Liu and Y. Ni, *Carbon*, 2020, **161**, 62–70.
166 J. Zhou, M. Wang and X. Li, *Appl. Surf. Sci.*, 2018, **462**, 444–452.
167 J. Feng, W. Song, L. Sun and L. Xu, *RSC Adv.*, 2016, **6**, 110337–110343.
168 H. Su, S. Lin, S. Deng, C. Lian, Y. Shang and H. Liu, *Nanoscale Adv.*, 2019, **1**, 2162–2166.
169 H. Farsi and F. Gobal, *Comput. Mater. Sci.*, 2007, **39**, 678–683.
170 G. S. Fanourgakis, K. Gkagkas, E. Tylianakis and G. Froudakis, *J. Phys. Chem. C*, 2020, **124**, 19639–19648.
171 D. Krishnamurthy, H. Weiland, A. Barati Farimani, E. Antono, J. Green and V. Viswanathan, *ACS Energy Lett.*, 2018, **4**, 187–191.
172 G. H. Gu, J. Noh, I. Kim and Y. Jung, *J. Mater. Chem. A*, 2019, **7**, 17096–17117.
173 B. C. Steele and A. Heinzel, *Nature*, 2001, **414**, 345–352.
174 E. D. Wachsman and K. T. Lee, *Science*, 2011, **334**, 935–939.
175 M. K. Debe, *Nature*, 2012, **486**, 43–51.
176 X. Zhu, J. Yan, M. Gu, T. Liu, Y. Dai, Y. Gu and Y. Li, *J. Phys. Chem. Lett.*, 2019, **10**, 7760–7766.
177 A. Legala, J. Zhao and X. Li, *Energy AI*, 2022, **10**, 100183.
178 A. Mistry, F. L. E. Usseglio-Viretta, A. Colclasure, K. Smith and P. P. Mukherjee, *J. Electrochem. Soc.*, 2020, **167**, 090542.
179 B. P. MacLeod, *et al.*, *Sci. Adv.*, 2020, **6**, eaaz8867.
180 A. A. Franco, A. Rucci, D. Brandell, C. Frayret, M. Gaberscek, P. Jankowski and P. Johansson, *Chem. Rev.*, 2019, **119**, 4569–4627.
181 A. Mistry, S. Trask, A. Dunlop, G. Jeka, B. Polzin, P. P. Mukherjee and V. Srinivasan, *J. Electrochem. Soc.*, 2021, **168**, 070536.
182 Y. Zhen and Y. Li, Redox flow battery, in *Studies in Surface Science and Catalysis*, Elsevier, 2020, 179, pp. 385–413.

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

183 K. Kakaei, M. D. Esrafili and A. Ehsani, Graphene-based electrochemical supercapacitors, in *Interface science and technology*. Elsevier, 2019, 27, pp. 339–386.
184 K. Yu, X. Pan, G. Zhang, X. Liao, X. Zhou, M. Yan, L. Xu and L. Mai, *Adv. Energy Mater.*, 2018, **8**, 1802369.
185 C. Liu, Z. G. Neale and G. Cao, *Mater. Today*, 2016, **19**, 109–123.
186 D. Wei, F. Xu, J. Xu, J. Fang, G. Wang, S. W. Koh and Z. Sun, *Ceram. Int*, 2019, **45**, 24538–24544.
187 A. I. Oje, A. A. Ogwu and A. M. Oje, *J. Electroanal. Chem.*, 2021, **882**, 115015.
188 P. Agrawal, H. F. Abutarboush, T. Ganesh and A. W. Mohamed, *IEEE Access*, 2021, **9**, 26766–26791.
189 M. Li, Q. Ma, W. Zi, X. Liu, X. Zhu and S. F. Liu, *Sci. Adv.*, 2015, **1**, e1400268.
190 B. Venkatesh and J. Anuradha, *Cybernetic. Infor. Technol.*, 2019, **19**, 3–26.
191 G. Chandrashekar and F. Sahin, *Comput. Electric. Eng.*, 2014, **40**, 16–28.
192 A. G. Karegowda, A. S. Manjunath and M. A. Jayaram, *Int J. Inf. Technol. Knowl. Manage.*, 2010, **2**, 271–277.
193 S. García, J. Luengo and F. Herrera, Feature selection, In: *Data pre-processing in data mining*, Springer, Cham, 2015, 163–193.
194 M. A. Hall, *Correlation-based feature selection for machine learning*, Citeseer, 1999.
195 S. Velliangiri and S. Alagumuthukrishnan, *Proc. Comput. Sci.*, 2019, **165**, 104–111.
196 B. Huang and O. A. von Lilienfeld, *Chem. Rev.*, 2021, **121**, 10001–10036.
197 T. Le, V. C. Epa, F. R. Burden and D. A. Winkler, *Chem. Rev.*, 2012, **112**, 2889–2919.
198 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
199 M. Sotoudeh and A. Groß, *J. Am. Chem. Soc.*, 2022, **2**, 463–471.
200 S. K. Kauwe, T. D. Rhone and T. D. Sparks, *Crystals*, 2019, **9**, 54.
201 X. Zhao, L. Z. Fan and Z. Zhou, *Green Energy Environ.*, 2021, **6**, 455–457.
202 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360–365.
203 S. Verma, S. Arya, V. Gupta, S. Mahajan, H. Furukawa and A. Khosla, *J. Mater. Res. Technol.*, 2021, **11**, 564–599.
204 J. Zhao and A. F. Burke, *Adv. Energy Mater.*, 2021, **11**, 2002192.
205 Y. Wang, W. Qiu, E. Song, F. Gu, Z. Zheng, X. Zhao, Y. Zhao, J. Liu and W. Zhang, *Nat. Sci. Rev.*, 2018, **5**, 327–341.
206 F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, *Chem. Rev.*, 2021, **121**, 9759–9815.
207 M. F. Langer, A. Goeßmann and M. Rupp, *npj Comput. Mater.*, 2022, **8**, 41.
208 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
209 Y. Pi, Q. Shao, P. Wang, F. Lv, S. Guo, J. Guo and X. Huang, *Angew. Chem., Int. Ed.*, 2017, **56**, 4502–4506.
210 N. J. O'Connor, A. S. M. Jonayat, M. J. Janik and T. P. Senftle, *Nat. Catal.*, 2018, **1**, 531–539.
211 R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler and L. M. Ghiringhelli, *Phys. Rev. Mater.*, 2018, **2**, 083802.
212 R. B. Wexler, J. M. Martirez and A. M. Rappe, *J. Am. Chem. Soc.*, 2018, **140**, 4678–4683.
213 F. Song, Z. Guo and D. Mei, *Feature Selection Using Principal Component Analysis*. In 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, Yichang, China, November 12-14, 2010, ICSEM, 2010, 27-30.
214 Z. Yu and W. Huang, *Electroanal*, 2021, **33**, 599–607.
215 B. Scholkopf, A. Smola and K.-R. Muller, *Kernel Principal Component Analysis. Advanced Kernel Method*, Support Vector Learning, 1999, pp. 327–333.
216 S. T. Roweis and L. K. Saul, *Science*, 2000, **290**, 2323–2326.
217 J. B. Tenenbaum, V. de Silva and J. C. Langford, *Science*, 2000, **290**, 2319–2323.
218 Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp and O. K. Farha, *Sci. Adv.*, 2016, **2**, e1600909.
219 S. Talwar, A. K. Verma and V. K. Sangal, *J. Environ. Manage.*, 2019, **250**, 109428.
220 Z. Zhang and Y.-G. Wang, *J. Phys. Chem. C*, 2021, **125**, 13836–13849.
221 L. I. Kuncheva, *A stability index for feature selection. In Artificial intelligence and applications*, 2007, pp. 421–427.
222 I. A. Gheyas and L. S. Smith, *Pattern Recognit.*, 2010, **43**, 5–13.
223 J. Cai, J. Luo, S. Wang and S. Yang, *Neurocomputing*, 2018, **300**, 70–79.
224 S. Khalid, T. Khalil and S. Nasreen. A survey of feature selection and feature extraction techniques in machine learning. In 2014 science and information conference, IEEE, 2014, 372–378.
225 S. Velliangiri and S. Alagumuthukrishnan, *Proc. Comput. Sci.*, 2019, **165**, 104–111.
226 L. V. Rajani Kumari and Y. Padma Sai, Classification of arrhythmia beats using optimized K-nearest neighbor classifier, in *Intelligent Systems*, Springer, Singapore, 2021, pp. 349–359.
227 M. Alweshah, S. Alkhalaileh, D. Albashish, M. Mafarja, Q. Bsoul and O. Dorgham, *Soft Comput.*, 2021, **25**, 517–534.
228 Y. Wu, C. DuBois, A. X. Zheng and M. Ester. Collaborative denoising auto-encoders for top-n recommender systems. In Proceedings of the ninth ACM international conference on web search and data mining, 2016, 153-162.
229 R. Md Shamim and J. Ma, ICA and PCA integrated feature extraction for classification, *IEEE, Int. Conf. Signal Process. IEEE*, 2017, 1083–1088.
230 G. K. Mislick and D. A. Nussbaum, *Data Normalization. Cost Estimation: Methods and Tools*, John Wiley & Sons, Inc, 2015, pp. 78–104.
231 G. B. Goh, N. O. Hodas and A. Vishnu, *J. Comput. Chem.*, 2017, **38**, 1291–1307.
232 A. Danandeh Mehr, V. Nourani, E. Kahya, B. Hrnjica, A. M. A. Sattar and Z. M. Yaseen, *J. Hydrology*, 2018, 566, 643–667.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 13 2023. Downloaded on 22.04.26 19:12:15.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

233 C. Sammut, *Encyclopedia of Machine Learning and Data Mining*, 2017, 566–567.
234 D. C. McKinney and M.-D. Lin, *Water Res. Res.*, 1994, **30**, 1897–1906.
235 S. M. R. Kazemi, B. M. Bidgoli, S. Shamshirband, S. M. Karimi, M. A. Ghorbani, K. W. Chau and R. K. Pour, *Eng. Appl. Comput. Fluid Mech.*, 2018, **12**, 506–516.
236 S. Liu, H. Tai, Q. Ding, D. Li, L. Xu and Y. Wei, *Math. Comput. Modelling*, 2013, **58**, 458–465.
237 J. R. Koza, *Stat. Comput.*, 1994, **4**, 87–112.
238 E. Pashaei and E. Pashaei, *Neural Comp. Appl.*, 2022, **34**, 6427.
239 M. Tubishat, S. Ja'afar, M. Alswaitti, S. Mirjalili, N. Idris, M. A. Ismail and M. S. Omar, *Expert Syst. Appl.*, 2021, **164**, 113873.
240 K. M. Hamdia, X. Zhuang and T. Rabczuk, *Neural Comput. Appl.*, 2021, **33**, 1923–1933.
241 I. A. Gheyas and L. S. Smith, *Pattern Recognit.*, 2010, **43**, 5–13.
242 H. Zhang, J. Thompson, M. Gu, X. D. Jiang, H. Cai, P. Y. Liu, Y. Shi, Y. Zhang, M. F. Karim, G. Q. Lo and X. Luo, *ACS Photonics*, 2021, **8**, 1662–1672.
243 J. X. Han, M. Y. Ma and K. Wang, *Neural Comput. Appl.*, 2021, **33**, 4111–4117.
244 J. Jiao, S. M. Ghoreishi, Z. Moradi and K. Oslub, *Eng. Compend.*, 2021, 1–15.
245 S. Han and L. Xiao, An improved adaptive genetic algorithm, in *SHS Web Conf.*, 2022, 140, p. 01044.
246 A. F. Gad, PyGAD: An Intuitive Genetic Algorithm Python Library, Accessed on the 30th September, 2022 from **http://arxiv.org/abs/2106.06158**.
247 A. A. Ewees, M. A. Al-qaness, L. Abualigah, D. Oliva, Z. Y. Algamal, A. M. Anter, I. R. Ali, R. M. Ghoniem and M. Abd Elaziz, *Mathematics*, 2021, **9**, 2321.
248 M. Zivkovic, N. Bacanin, A. Djordjevic, M. Antonijevic, I. Strumberger and T. A. Rashid, *Proc. Int'l Conf. Sust. Expert Systems*, Springer, Singapore, 2021, pp. 169–184.
249 K. S. Garud, S. Jayaraj and M. Y. Lee, *Int. J. Energy Res.*, 2021, **45**, 6–35.
250 S. Thawkar, *J. Ambient Intell. Humaniz. Comput.*, 2021, **12**, 8793–8808.
251 A. Jamal, A. Aldulaimi, L. Abualigah and D. J. Akram, *Cluster Comput.*, 2021, **24**, 2161–2176.
252 S. Katoch, S. S. Chauhan and V. Kumar, *Multimed. Tools Appl.*, 2021, **80**, 8091–8126.
253 L. I. Kuncheva, *A stability index for feature selection. In Artificial intelligence and applications*, 2007, 421–427.