Review | Received 10 November 2024; Accepted 23 December 2024; Published 15 January 2025
https://doi.org/10.55092/aimat20250003

Applications of artificial intelligence in materials research for fuel cells

Hong Liu[1], Heming Guo[1], Zibo Gao[1], Hetong Pan[1], Junhao Zheng[1], Jiajun Linghu[2,*] and Zhipeng Li[1,*]

[1] Frontiers Science Center for Flexible Electronics, Xi'an Institute of Flexible Electronics, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, China
[2] Department of Applied Physics, Chang'an University, Xi'an, China

* Correspondence authors; E-mails: linghujiajun@chd.edu.cn (L. J.); iamzpli@nwpu.edu.cn (L. Z.).

Highlights:

- AI applications, especially ML, for predicting and designing fuel cell material properties reviewed.
- Machine learning and high-throughput methods boost efficiency while cutting costs.
- General strategies and structured workflows for data-driven material discovery are proposed.
- Challenges in AI, like sparse data, are addressed, with SHAP and ensemble solutions analyzed.

Abstract: The increasing global energy demand and the growing environmental problems have intensified the pursuit of clean and sustainable energy solutions. Hydrogen, with its high energy density and clean by-products, is a promising candidate as an energy source. Fuel cells play a key role in harnessing hydrogen energy, but this technology faces challenges such as the trade-off between material stability and ion conductivity, which limits its widespread application. To address these challenges, designing material properties and adjusting system parameters are highly desirable. However, the traditional trial-and-error approach is no longer feasible when dealing with the vast array of possibilities. Fortunately, the advancement of artificial intelligence (AI) offers a new approach which can dramatically speed up the material design and parameter control. This article reviews the application of AI in fuel cells, especially its ability to accelerate material development. The review begins by outlining the mechanisms and classifications of fuel cells, as well as the property requirements for each part of the fuel cells. Subsequently, the article introduces the basic concepts of AI and its application in materials science, including the workflows of data aggregation, feature construction, model training, and experimental validation. Importantly, the applications of AI in predicting fuel cell material performance are highly emphasized and discussed. In addition, the challenges encountered in AI applications are introduced, including sparse datasets, complex feature engineering, the limitations of general models, and the weak interpretability of AI models, along with their respective development blueprints.

Keywords: artificial intelligence; fuel cells; machine learning; material properties

1. Introduction

In the last few decades, technological and industrial development has reached a new height worldwide. These developments have led to high energy consumption, prompting the search for clean, secure, affordable, and sustainable energy resources. Consequently, scientists from various science branches worldwide are collaborating to find such sources. Traditionally, fossil fuel-based sources are used to meet energy demands. Hydrogen, as an excellent clean energy source, has an energy density of 143 MJ/kg, which is three times higher than that of liquid hydrocarbon fuels [1]. At the same time, the utilization of hydrogen as a fuel generates minimal harmful emissions, with water being the sole byproduct [2]. The final key to the hydrogen energy system lies in using the fuel efficiently in internal combustion engines, conventional combustion turbines, and fuel cells. In 1839, the British physicist Sir William Grove demonstrated that hydrogen and oxygen could be electrochemically combined to produce water and electricity through devices now known as fuel cells [3]. To this day, fuel cells, with their high energy efficiency, extremely low emissions, and nearly silent operation, have become one of the most forward-looking and potential technologies for the utilization of hydrogen [4].

At present, the actual fuel utilization rate of Solid oxide fuel cells can reach up to 85% [5,6], much more than the 40% of conventional internal combustion engines [7]. However, there are still numerous difficulties and challenges in the research and development of fuel cell materials such as the lack of catalyst activity, which limits the overall performance of the fuel cell [8,9]. Meanwhile, poor durability is also a significant problem, for example, structural degradation and corrosion in long-term operation can lead to a shortened lifespan of the cell [10–12]. In addition, the high cost, especially the reliance on precious metal catalysts, also restricts the sustainable deployment of fuel cells [13]. Hence, a significant barrier persists before fuel cells can be widely adopted for large-scale commercial use.

The future development of fuel cells heavily relies on the enhancement of material properties and system control. Nevertheless, the traditional research method of trial-and-error is inefficient and makes it challenging to identify high-performance materials from a large number of candidates. As early as 1995, Xiang *et al.*, invented an experimental method known as a “high-throughput composite material chip” to enhance the efficiency of discovering new materials [14]. Drawing inspiration from integrated circuits and gene chips, this technique merges and evaluates a diverse array of thin film libraries, each with uniform composition, structure, and material properties, on a single substrate. Utilizing any element as the fundamental building block, it achieves this through sophisticated design, allowing for rapid characterization. The method features a minimal sample size of 200 microns by 200 microns, translating to a library density of 10,000 sites per square inch. In terms of artificial intelligence (AI), researchers such as Amil Merchant and Ekin Dogus Cubuk at Google DeepMind have used graph networks for materials exploration (GNoME) and deep learning techniques to greatly speed up the process of discovering new materials [15]. The graph networks trained at scale can reach unprecedented levels of generalization, improving the efficiency of materials discovery by an order of magnitude. Rational use of artificial intelligence to accelerate material research is imperative. The role of AI in accelerating research in the field of lithium-ion batteries and solar cells has been demonstrated [16–19], but there is

not much research using AI in the fuel of fuel cells. Therefore, this article provides a brief review of AI research in the field of fuel cells.

This review will be developed from the following aspects: firstly, the requirements and research status of the fuel cell materials; secondly, an introduction of artificial intelligence; thirdly, the application of artificial intelligence in the field of fuel cell material research; and finally, the summary. The review conducted in this paper can promote the application of AI in the research of fuel cell materials and accelerate the development and application of fuel cells.

2. Requirements and research status of fuel cell materials

2.1. The working principles and main types of fuel cells

A fuel cell is a device that generates electricity through a chemical reaction. It consists of two electrodes: an anode (negative electrode) and a cathode (positive electrode), separated by an electrolyte [20]. Fuel cells can operate continuously by supplying fuel to the anode and oxidant (typically air) to the cathode. On the cathode (anode) and oxygen ions (protons), the fuel and oxidants react separately and produce ions, which are conducted in the electrolyte to maintain the charge balance in the system, produce water, and heat. At the same time, the electrons flow through an external circuit, creating an electric current [21,22]. Compared with conventional power generation methods, fuel cells offer several advantages, including high efficiencies, high power densities, compact size, low emissions, low noise, and high-quality power. Fuel cells are modular and maintain their efficiency even at small scales, making them well-suited for aerospace applications and distributed power generation, thus reducing transmission and distribution losses [23].

Based on the choice of fuel, electrolyte type, operating temperature, efficiency, application, and cost, fuel cells are divided into 6 main categories: alkaline fuel cell (AFC), phosphate acid fuel cell (PAFC), solid oxide fuel cell (SOFC), molten carbonate fuel cell (MCFC), proton exchange membrane fuel cell (PEMFC), and direct methanol fuel cell (DMFC) [24]. The chemical reactions and components of these 6 types of fuel cells are shown in Table 1 and Figure 1, respectively.

Table 1. Chemical reactions of the 6 main fuel cells.

	Anode reaction	Cathode reaction	Overall reaction
AFC	$H_2+2OH^- \rightarrow 2H_2O+2e^-$	$1/2O_2+H_2O+2e^- \rightarrow 2OH^-$	$2H_2+O_2 \rightarrow 2H_2O$
PAFC	$H_2 \rightarrow 2H^++2e^-$	$1/2O_2+2H^++2e^- \rightarrow H_2O$	$H_2+1/2O_2 \rightarrow H_2O$
SOFC	$H_2 \rightarrow 2H^++2e^-$ $H_2+O^{2-} \rightarrow H_2O+2e^-$	$2H^++2e^-+1/2O_2 \rightarrow H_2O$ $1/2O_2+2e^- \rightarrow O^{2-}$	$H_2+1/2O_2 \rightarrow H_2O$
MCFC	$H_2+CO_3^{2-} \rightarrow H_2O+CO_2+2e^-$	$CO_2+1/2O_2+2e^- \rightarrow CO_3^{2-}$	$H_2+1/2O_2 \rightarrow H_2O$
PEMFC	$H_2 \rightarrow 2H^++2e^-$	$1/2O_2+2H^+ +2e^- \rightarrow H_2O$	$H_2+1/2O_2 \rightarrow H_2O$
DMFC	$CH_3OH+H_2O \rightarrow CO_2+6H^++6e^-$	$3/2O_2+6H^++6e^- \rightarrow 3H_2O$	$CH_3OH+H_2O+3/2O_2 \rightarrow CO_2+3H_2O$

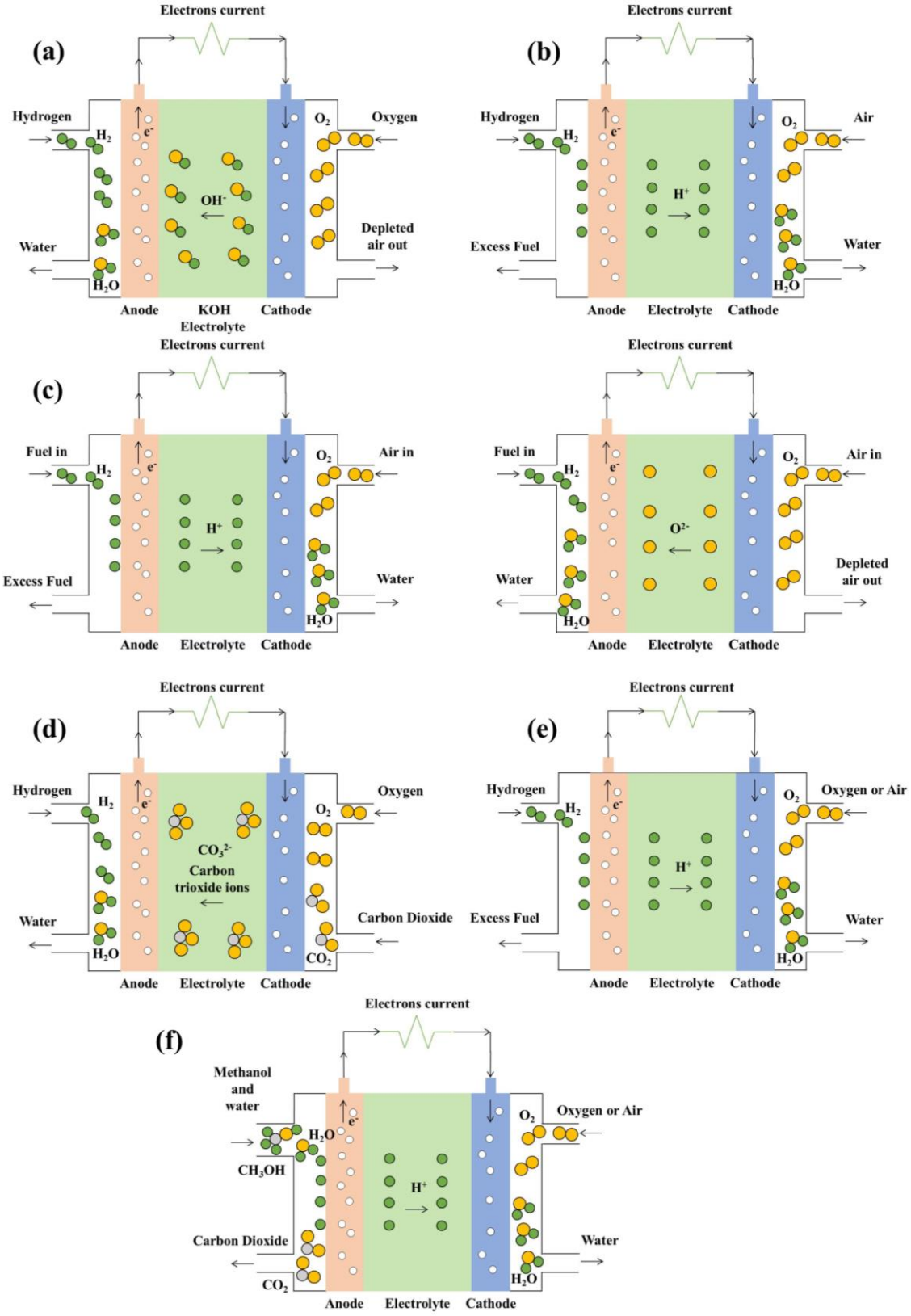

Figure 1. (a) Typical schematic of AFC working principle; **(b)** Typical schematic of PAFC working principle; **(c)** Typical schematic of SOFC working principle, which is divided into oxygen ion conduction type and proton conduction type; **(d)** Typical schematic of MCFC working principle; **(e)** Typical schematic of PEMFC working principle; **(f)** Typical schematic of DMFC working principle.

2.2. Requirements for materials of fuel cells

The electrolyte constitutes the most critical component of a fuel cell, serving as a basis for classifying various fuel cell technologies (for instance, in SOFC, the electrolyte is in solid form, whereas in PEMFC, a polymer film is utilized as the electrolyte). The primary function of the electrolyte is to inhibit electron diffusion within the cell body, thereby acting as an insulator between the fuel and oxidizer. Furthermore, oxide ions (or protons) need to traverse the electrolyte to engage in redox reactions, so the electrolyte must exhibit high ionic conductivity. In addition, electrolyte must be stable, which means it should have thermal, chemical, and phase stability, and it should be compatible with both electrodes.

During the fuel cell reaction process, the oxidation or combustion of the fuel occurs on the anode side. Therefore, the anode must possess high electrocatalytic activity for fuel oxidation. The porosity of the anode structure is another crucial factor, as it ensures the transport of fuel to the reaction sites. When the porosity of the anode reaches 30–40%, the single fuel cell powered by pure hydrogen fuel shows better gas diffusion and current conduction characteristics, and the corresponding output voltage is also relatively high [25]. Another important aspect is the chemical compatibility of the anode, which means it should not react with the interconnects and electrolyte, even at high temperature.

Oxygen reduction, a process which produces oxygen ions, occurs at the cathode. Therefore, the cathode must exhibit high catalytic activity for oxygen reduction. Similar to the anode, the cathode must have sufficient porosity to facilitate oxygen flow and excellent exhibit chemical compatibility with other fuel cell components. The requirement of materials in a fuel cell is summarized in Table 2.

Table 2. The requirements of materials in fuel cells.

Type	Requirement
Electrolytes	High ionic conductivity, chemical stability, temperature resistance, low electronic conductivity, low permeability
Catalysts (Unnecessary in high temperature)	High catalytic activity, high temperature resistance and corrosion resistance, durability
Anode and cathode	High electronic conductivity, certain ionic conductivity, appropriate porosity, matching thermal expansion coefficients, high stability, catalytic activity

2.3. Typical challenges among material research of fuel cells

Among these 6 types of fuel cells, the most famous two are PEMFC and SOFC, which will be taken as examples and discussed in detail.

PEMFC is classified as low-temperature fuel cells, operating within a temperature range of 60 ℃ to 100 ℃ [26]. The fundamental principle of PEMFC is proton conduction, which is typically the primary characteristic assessed for the suitability of a membrane in fuel cell applications. As the core component of the proton exchange fuel cell, the proton exchange membrane should fulfill the roles of conducting protons, isolating anode and cathode reactants and preventing electron flow. Based on the latest membrane research, the use of composite materials and material doping modification remain the primary approach for proton exchange membrane development. The future of proton exchange membranes is trending toward high-temperature operation [27].

The most critical challenges facing current fuel cell membranes include high-temperature degradation, fuel crossover, water condensation, and balancing production costs with safety issues. Therefore, the optimization of fuel cell membranes requires extensive research involving the selection and modification of materials, the utilization of various technologies and conditions, and an in-depth study of the membranes' structure and morphology [28].

Unlike PEMFC, which can only use pure H_2 as fuel, SOFC utilizes not only pure hydrogen but also hydrocarbon as fuel. However, SOFC is a high-temperature fuel cell that usually operates above 800 ℃. Although the resultant heat is often harnessed to generate additional electricity via gas turbines, raising cogeneration efficiency to 70% to 80% [29], more efforts are being made to reduce the working temperature of SOFC to the range of 450 ℃ to 650 ℃. The oxide electrolyte is the key component that determines the operating temperature of a SOFC. At present, yttrium oxide-stabilized zirconia (YSZ) is the most commonly used electrolyte in SOFC. However, YSZ has disadvantages such as a narrow operating temperature range and low ionic conductivity [30].

For the anode, nickel-based oxide is widely used due to its good catalytic ability. However, there are several practical issues with nickel-based anode materials, such as carbon deposition, sulfur poisoning, low redox cycle stability, and particle growth. Similarly, hydrocarbon fuels contain a certain amount of sulfur, which can form hydrogen sulfide (H_2S) in the reducing atmosphere of the anode. The entry of H_2S into the anode leads to a decrease in fuel cell performance [31,32].

In SOFC, perovskite-based materials, such as strontium-doped lanthanum manganite (LSM), are commonly used as cathodes. LSM possesses high electrical conductivity and a thermal expansion coefficient comparable to that of YSZ electrolyte. As a cathode, its advantages include a stable structure, good compatibility with the electrolyte, and high electronic conductivity [33]. However, due to the lack of oxygen vacancies in the structure, LSM exhibits low ionic conductivity during operation, a characteristic typical of electron-conducting cathode materials [34]. At present, to improve ionic conductivity, it is a common practice to introduce materials with high ionic conductivity [35–37].

Based on the above problems faced by the materials, how to efficiently develop electrode or electrolyte materials with excellent performance remains a focus for current scientific researchers. However, the typical experimental methods are time-consuming and laborious, especially the sintering process, which is inseparable from the fabrication of electrolytes of SOFC. This process usually requires a long processing time and high energy consumption (1200–1800 ℃). Taking $La_{0.9}Sr_{0.1}Ga_{0.8}Mg_{0.2}O_{3-\delta}$(LSGM) as a famous example of electrolyte material. LSGM powder was prepared using a solid phase reaction method with La_2O_3, MgO, Ga_2O_3, and $SrCO_3$. The preparation process requires repeated sintering and calcining (at temperatures above 1300 ℃), with a total duration of 111 hours [38]. Obviously, such complicated and time-consuming fabrication processes are not suitable for studying a large number of materials.

Theoretical research significantly enhances the efficiency of new material development. First-principles calculations, particularly those based on density functional theory (DFT), have long been the preferred approach for investigating material properties. For example, Liu *et al.* used DFT to explore defective carbon materials for catalytic nitrogen reduction reactions (NRR), assessing the impact of various doping methods, locations, and amounts [39]. However, due to the complexity of material structures and the variability in compositions, first-principles calculations remain resource-intensive and are not widely adopted.

Fortunately, emerging methods that combine machine learning (ML) with high-throughput screening have significantly enhanced the efficiency of developing new materials, making this approach a hot topic and a development trend in the field of materials. Numerous studies have demonstrated that predicting material properties through ML, followed by high-throughput screening, and then experimental research on the selected materials, can greatly reduce trial and error costs and enhance the success rate of discovering materials with superior properties [40,41]. These cases illustrate that the integration of artificial intelligence technologies has not only significantly enhanced the efficiency of new material design and performance optimization, but has also substantially improved material discovery and manufacturing processes, thereby accelerating the overall research and development of novel materials. This trend is similarly applicable in the domain of fuel cell material design. In Figure 2, we present a trend chart depicting the annual increase in the number of publications related to "ML + Fuel cell" in the Web of Science (WOS) database. This provides a clear visualization of the growing research interest and practical applications in this field.

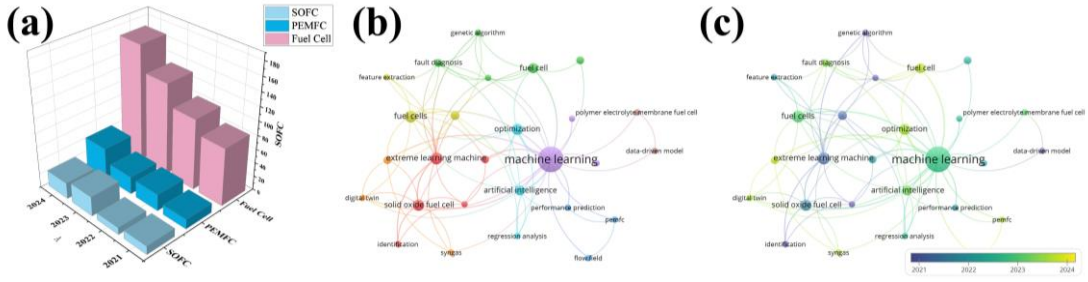

Figure 2. (a) The number of papers on the combination of ML with SOFC, PEMFC, and Fuel Cells in WOS; **(b)** Keywords network; **(c)** Keyword novelty.

3. Introduction to AI

3.1. Basic characteristics of AI

AI is a system that uses probabilistic models to learn from data and make predictions, mimicking human intelligence [42]. These data can come from a variety of sources and include various aspects, such as material conductivity recorded during tests. Given a database focused on a specific problem, an appropriate mathematical model of artificial intelligence is selected to solve the problem through a programming language. The algorithm is used to obtain the answer to the initial question and then to perform further analysis on the final result [43,44]. The emerging technologies of AI are triggering a revolution in the field of material design [45–47].

The application of AI in the fuel cell field is becoming increasingly important, and it can facilitate the development and optimization of fuel cell technology in several ways. For material discovery, AI technologies, especially ML and data mining, can accelerate the discovery of new materials. This can help researchers identify materials with potential applications and optimize the properties of materials [48]. For system optimization, AI is capable of processing complex datasets to provide accurate predictions. It has been used to predict the performance and lifetime of fuel cells, as well as for fault diagnosis, facilitating battery design and operation [49]. AI can also be used to develop intelligent control systems

for fuel cells to achieve real-time monitoring and to optimize the working status of the cell, such as the hydrogen-oxygen flow ratio and temperature [50].

AI has broad application prospects in the field of fuel cells. It can not only accelerate the development speed and efficiency of fuel cell materials but also improve the performance and efficiency of fuel cell systems. It can also help reduce costs and promote the commercialization process. With the continuous progress of AI technology, its future applications in the field of fuel cells are expected to be more in-depth.

3.2. AI in material design solutions

Driven by advancements in ML technologies and the increasing availability of high-quality data, AI has made significant progress in predicting material properties, optimizing material design, and accelerating the discovery of new materials [51]. In terms of data collection and analysis, large language models such as ChatGPT can transform traditional research paradigms by efficiently processing vast amounts of research resources and extracting insights from complex literature. This approach addresses various challenges, including the fragmentation of research literature, difficulties in identifying relevant information, and the time-consuming iterations involved in research design [52].

In addition to training models using material features, SHapley Additive exPlanations (SHAP) can be employed to analyze the relationship between features and target performance, helping to uncover the factors influencing performance and deepening our understanding of it. Furthermore, feature heatmaps can be used to explore the interactions and correlations between different features with respect to target performance. In Figure 3, we present the two types of ML problems—classification and regression. Classification problems focus on grouping material types, while regression problems aim to predict the properties of unknown materials and display the synergistic or antagonistic relationships between features in the heatmap.

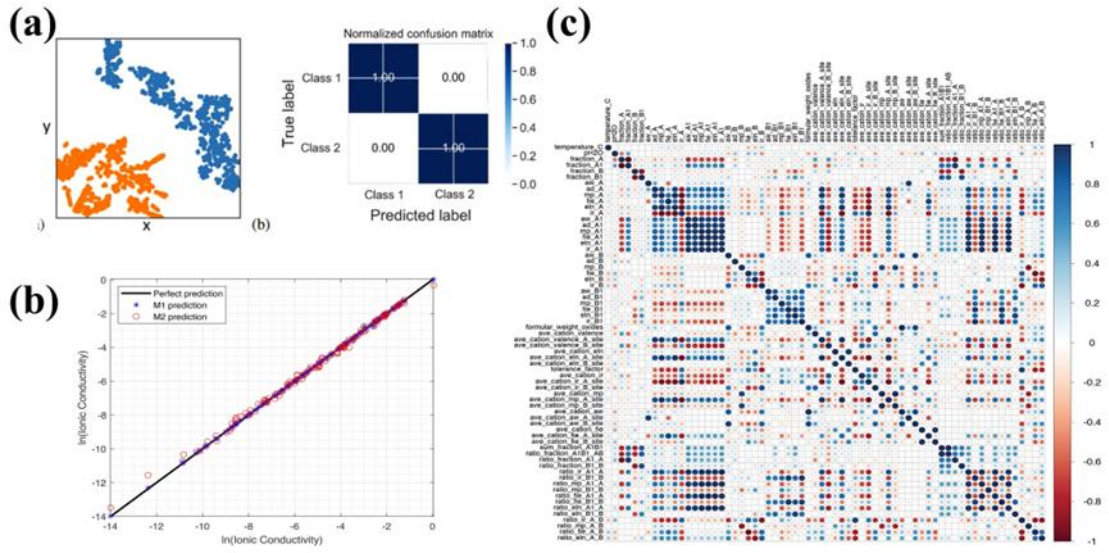

Figure 3. (a) Classification of material types [53] (Copyright 2020, Journal of Applied Physics); **(b)** Prediction of material properties [54] (Copyright 2022, Chemical Physics); **(c)** Feature relationships in the heatmap [55] (Copyright 2022, Advanced Materials).

3.3. Steps of ML to accelerate material development

The three key aspects of ML are: (1) data integration, normalization, and formatting; (2) feature engineering, including feature extraction, selection, and construction; (3) model training and evaluation. The steps of ML process are shown in Figure 4.

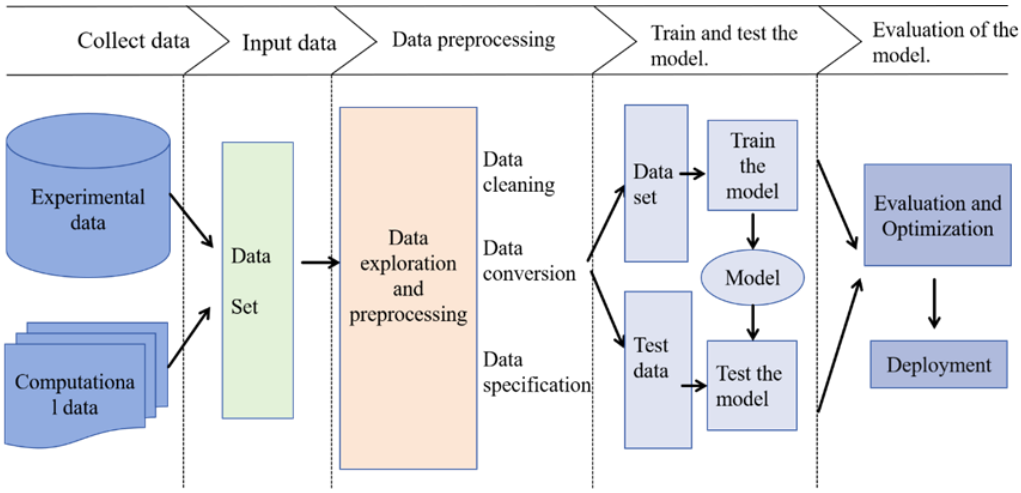

Figure 4. Machine learning flowchart.

The process of ML in accelerating material development is similar to the diagram above, but there are many details worth our attention. The following key stages are typically involved and shown in Figure 5, and Table 3 provides a list of authoritative databases developed in the field of materials.

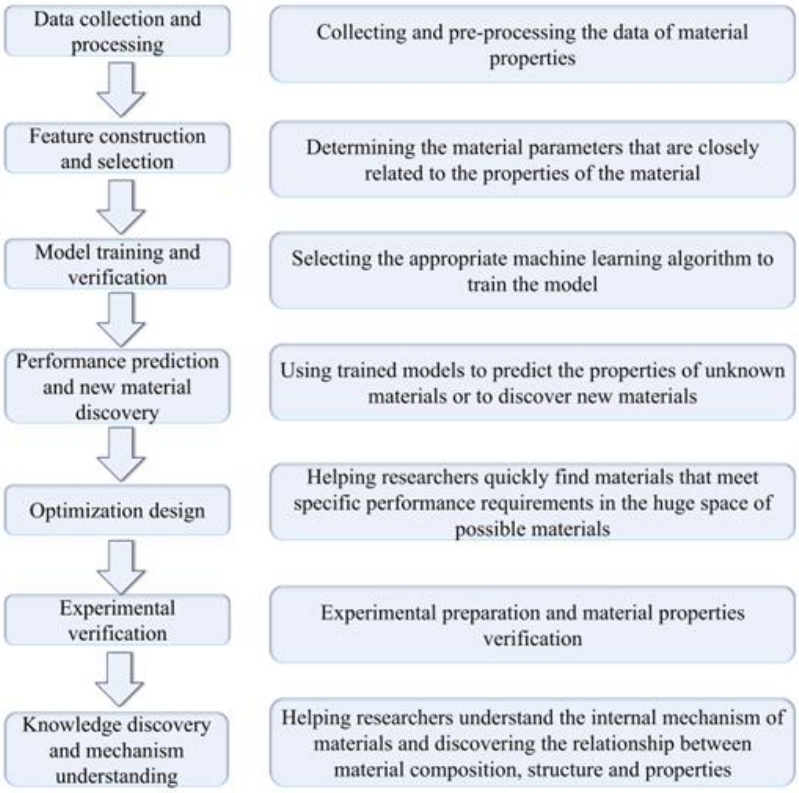

Figure 5. The process of ML in accelerating material development.

Table 3. Famous database.

Database	Brief Description	Website
Materials Project [56]	As of 2024, it contains over 150,000 materials with a focus on inorganic compounds, including both crystalline and non-crystalline structures.	http://materialsproject.org/
The Open Quantum Materials Database (OQMD) [57,58]	It is a database of DFT calculated thermodynamic and structural properties of 1,226,781 materials, created in Chris Wolverton's group at Northwestern University.	http://oqmd.org/
Computational Materials Repository (CMR) [59]	It contains data on approximately 4,000 two-dimensional (2D) materials distributed across more than 40 different crystal structures.	https://cmr.fysik.dtu.dk/
AFLOW [60]	A globally available database of 3,530,330 material compounds with over 734,308,640 calculated properties, and growing.	http://www.aflowlib.org/
Inorganic Crystal Structure Database (ICSD) [61–63]	It contains detailed information on the structure of more than 210,000 inorganic crystals published since 1913, covering structural data for pure elements, minerals, metals, and intermetallic compounds.	https://icsd.nist.gov/
Cambridge Structural Database (CSD) [64]	It contains more than 1.25 million structures, all derived from experimental results.	https://www.ccdc.cam.ac.uk/solutions/software/csd/
Crystallography Open Database (COD) [65–67]	It contains more than 520,000 entries. Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers.	https://www.crystallography.net/cod/
Springer Materials	It contains more than 290,000 materials and 3,000 properties, incorporating integrated multi-source data from all major topics in the fields of materials science, chemistry, physics, and engineering.	https://materials.springer.com/

4. ML applications in fuel cells

Fuel cells are complex systems with several components and a large number of parameters. For one component, there are numerous properties such as thermodynamic phase stability, ion conductivity, electrical conductivity, catalytic performance, *etc.* Between two components, there are issues such as matching problems, interface effects, phase transitions, *etc.* For the entire system, many factors such as temperature, atmosphere, and fuel flux have a significant effect on the overall efficiency. These variables are interrelated and complex, making the design and optimization of fuel cell very difficult.

After the application of ML has been extended into the experimental field, it offers a highly efficient method to develop, characterize, and optimize devices, saving time and effort by avoiding numerous manual experiments [68]. Especially, ML has attracted the attention of many scholars due to its rapid and precise capabilities in simulating and predicting the properties of materials and devices. The parameters involved in applying ML for device optimization have been discussed in some articles [69–72]. Therefore, this paper only summarizes and discusses the use of ML to predict the properties of materials. In Figure 6, we illustrate some actual workflows of applying ML to material design.

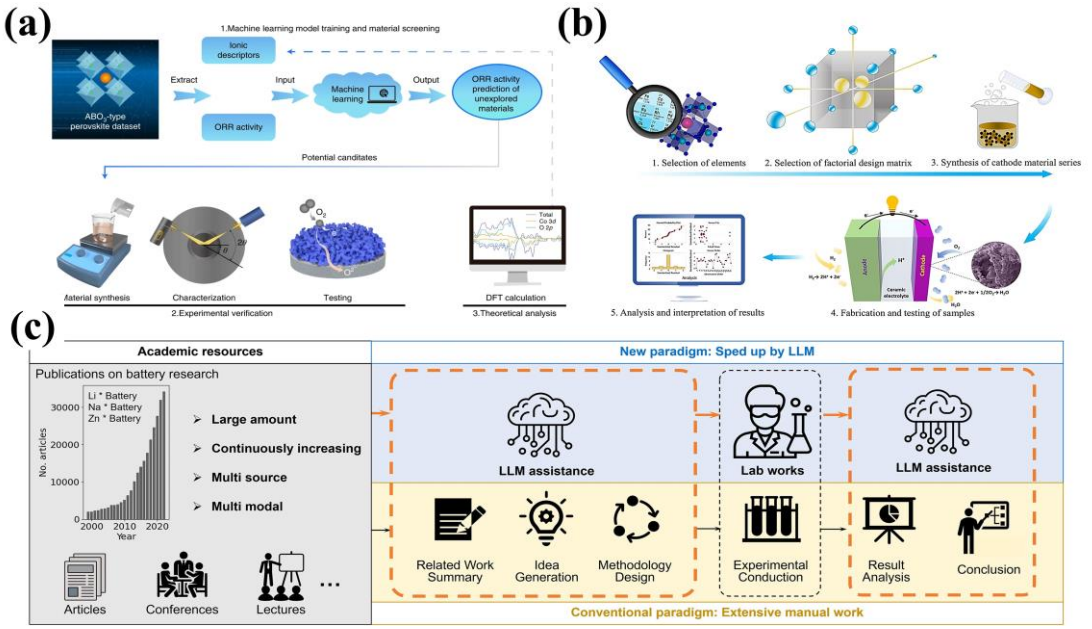

Figure 6. (a) After ML prediction and screening, the selected materials are experimentally prepared and undergo characterization tests, which are combined with using DFT calculations to verify their properties simultaneously [73] (Copyright 2022, Nature Energy); **(b)** Simplified cathode material development framework based on the experimental design paradigm (EDP). Selection of elements, selection of factorial design matrix, synthesis of cathode materials series, fabrication and testing of samples, and analysis of results [74] (Copyright 2023, Chemical Engineering Journal); **(c)** Comparative illustration of conventional and LLM-transformed research paradigms [52].

The application of ML in the field of fuel cell material performance is summarized in Table 4 and Table 5, and some important examples are discussed below.

Table 4. Summary of ML in material development for PEMFC.

Research content	Model classification	Method	Method selection basis	Dataset	Performance indicator
Development SPARK to predict the performance and durability of PEMFC [75]	Regression	Extreme Gradient Boosting, Artificial Neural Networks (ANN), Light Gradient Boosting, Machine Gradient Boosting Decision Tree	ANN is chosen for its strong nonlinear fitting capability and CatB for demonstrating the best performance in durability prediction	The polarization curve with 405 records and a dataset for OCV degradation with 1585 records	$r > 0.965$, RMSE < 5.00×10^{-3} and $R^2 > 0.845$
Streamlining the development process for platinum-based intermetallic nanoparticle catalysts [76]	Regression	Gaussian Process Regression (GPR)	Its non-parametric nature and good performance on small datasets help avoid overfitting, while allowing updates with new data—critical in materials research where data is costly	The formation energies of 300 ordered and 300 random, and 3800 structures with varying ordering degrees	Two elements, Ni and Cu, were selected to be incorporated into the Pt_2CoM system
Integrating DFT with ML to investigate the chemisorption behavior of oxygen atoms [77]	Regression	Multivariate, Linear Regression	It can effectively predict the chemisorption energy of oxygen atoms	\	$0.847 < R^2 < 0.969$, 0.085 MAEs < 0.167 eV
The structural stability of Pt-Ni alloy nanoclusters [78]	Regression	GPR	The model provides uncertainty estimates, captures data correlations via its kernel function, and handles small sample datasets	500 DFT-calculated configurations were constructed, including 84 compositions	MSE < 0.13
Exploring the phenomenon of gold segregation and its consequential effects on the electrocatalytic activity in the Oxygen Reduction Reaction (ORR) [79]	Regression	Neural Networks (NN), Genetic Algorithm	The model, based on physical principles, predicts the structural stability of alloy nanoclusters	A total of 42,348 structures computed by DFT	RMSE < 8.7 meV per atom

Table 4. *Cont.*

Research content	Model classification	Method	Method selection basis	Dataset	Performance indicator
The limit potential of graphene-supported single-atom catalysts in ORR、OER、HER [80]	Regression	Random Forest (RF)	It enhances model accuracy and robustness, while providing feature importance to reveal the relationship between physical properties and catalyst performance	104 records	MSE=0.027 V
Establishing the ideal synthesis parameters and material characteristics necessary for facilitating four-electron transfers in the ORR process [81]	Regression	19 ML models	It chooses a variety of ML models and selects the model with the best prediction results	\	RMSE = 0.31449, R^2 = 0.67
Structural analysis of metal nanoparticle catalysts [53]	Classification and Regression	Iterative Label Spreading, Extra Trees Classifier, Extra Trees Regressor	It can analyze the relationship between multiple structures and performance, based on the average performance across many structures	The 179 structural features, including atomic, crystallographic, and topological descriptors, are reduced and standardized to 121 dimensions	Both Category 1 (disordered) and Category 2 (ordered) nanoparticles scored highly
ORR activity of complex solid solution Electrocatalyst [82]	Classification and Regression	Three optimized models based on Sequential Least Squares	These models predict ORR activity by considering binding and site interactions, and are compared with experimental data to validate their accuracy	3317 DFT calculated *OH and *O binding energies	\

Table 4. *Cont.*

Research content	Model classification	Method	Method selection basis	Dataset	Performance indicator
Efficient fuel cell electrocatalysts [83]	Regression	Slab-graph Convolutional Neural Network (SGCNN), CGCNN, CNN	They can predict the adsorption energies of various adsorbates, including H, CO, O, OH, and OOH	surface database of 9267 adsorption energies of five adsorbates	SGCNN under 5-fold cross-validation MAE = 0.16 eV
Oxygen reduction reaction Fe-N-C electrocatalysts [84]	Regression	GPR, ANN, RF	The screened catalyst was 33% more active than the best in the initial database, with ORR activity 7 times higher than similar materials	The initial 36 samples were combined with 14 new samples synthesized by adaptive learning strategies	Sample #20 from the adaptive learning framework showed the best ORR1 activity, 16.3 ± 0.4 mA/mg
Development of High-Entropy Alloy Catalysts [85]	Regression	NN	The catalytic properties of high entropy alloys can be accurately predicted, and their structural sensitivity is quantified for the first time	over 1,000 data points	MAE = 0.094eV, RMSE = 0.117eV

Table 5. Summary of ML in material development for SOFC.

Research content	Model classification	Method	Method selection basis	Dataset	Performance indicator
Stability of perovskite oxides [86]	Classification and Regression	1. Logistic Regression, Support Vector Machine (SVM), Decision Tree, NN, Extra Trees Regression (ETR) 2. Linear Regression, Kernel Ridge Regression (KRR), Decision Tree Regressor, ETR, ANN	1. Further validated the model by predicting the stability of compounds outside the training set 2. Tested three feature selection methods to remove redundant or irrelevant features	More than 1900 DFT-calculated perovskite oxide energies	The best F1 score achieve for classification is 0.881 (± 0.032) and the best RMSE value for regression of Ehull is 28.5 (± 7.5) meV/atom

Table 5. *Cont.*

Research content	Model classification	Method	Method selection basis	Dataset	Performance indicator
Oxygen ionic conductivities [54]	Regression	SVM, Partial Least Squares, ANN, and Quadratic Approximation, GPR	Use the same 117 samples to rebuild the GRP model following the fashion of constructing M2	128 samples from References include tolerance factor, charge of the A/B side, Electronegativity of the A/B side, and A complex descriptor	The final model, M2, has a CC, RMSE, and MAE of 99.87%, 0.1577, and 0.1048
Perovskites with high conductivity [87]	Classification and Regression	XG-Boost, Lasso, Ridge, Elastic Net, 7-layer ANN (100 nodes), SVM, RF	The model identifies the key predictors of conductivity and carrier type	7230 perovskite cases (585 compositions)	The test set RMSE is 0.24, R2 is 0.987. The cross-validation set RMSE is 0.25 and R2 is 0.986
Oxygen reduction reaction activity of perovskite oxides [73]	Regression	Four linear regression methods and four nonlinear regression methods	BISA (Lewis acid) shows the greatest importance in the model	95 sets of data	The log_10 area specific resistance (ASR) MSE of the best model ANN is 0.012
Predict the proton absorption amount of Co/Fe based perovskite oxides [88]	Regression	RF, XG-Boost	The importance ranking of features in PAA prediction of XG-Boost model is given as the interpretation result of model features	792 samples involve 29 features as the input variables and one output PAA as the target variable	PMSE=0.021±0.04 MAE=0.01±0.002 R^2 =0.901±0.045
ML models were developed to predict perovskite catalysis [89]	Regression	RF, NN, GPR, Linear Regression	Follow the approach of Palmer *et al.* to develop calibrated uncertainty estimates	749 data points spanning 299 unique perovskite compositions	ASR RMSE of 0.367 ± 0.049 Ohm-cm^2
An efficient ML navigation of reaction space [90]	Classification and Regression	SVM	It has significant advantages in processing complex, non-linear and high-dimensional data	74 samples	RMSE-CV of 0.7556 layers, a coefficient of determination of 0.895
Combining deep learning with FIB-SEM analysis to improve image quality and eliminate artifacts in electron microscopy [91]	Deep learning	DeepLabV3 + CNN	The DeepLabV3+ algorithm is used for automated semantic segmentation to extract the microstructure parameters	\	The relative errors range from 5.9 to 8.5% and the triple phase boundaries are calculated to be 3.2 counts/μm2 with a relative error of 16.8%
Predicted the hydrated proton concentration (HPC) and proton conductivity of 3200 oxides, accelerating the development of efficient hybrid proton-electron conductive oxides for air electrodes [55]	Regression	RF, SVM, KNN, GPR, Gradient Boosting Machine (GBM)	The HPC predicted by the ML model is in good agreement with the experimental measurements, which verifies the accuracy of the model	Contained 795 oxides under varying experimental conditions, total 66 combined features	The RMSE of the best model RF is 0.0236

Table 5. *Cont.*

Research content	Model classification	Method	Method selection basis	Dataset	Performance indicator
Develop a selective sampling procedure to efficiently identify low-energy regions, especially for atomic diffusion and ion conduction processes [92]	Regression	GPR	The proposed program can effectively identify low-energy regions characterizing proton conduction in the host crystal lattice	From the Database and DFT calculations, a set of 1768 grid points related to potential energy values is obtained	In GP4, the average number of sampling steps required for identifying all the points in the low-PE region was only 394.1
Study on auxiliary prediction of degradation of chrome-poisoned SOFC cathodes [93]	Regression	deep-DRT	It can combine experimental parameters (e.g. temperature, oxygen partial pressure) as inputs to make predictions	Up to 96 hours of data were collected by electrochemical impedance spectroscopy (EIS)	The predicted 156-hour data is in good agreement with the experimental results
The microstructure characteristics of SOFC three-phase composite cathode materials were quantified by automated semantic segmentation [94]	Deep learning	DeepLabV3 + CNN combine with an atrous convolution and an atrous spatial pyramid pooling (ASPP) module	The images inferred by deep learning closely match those of the human experts in terms of area (or volume) scores and interconnections	\	A total of 640 512 x 512-pixel electron micrographs is employed
The thermal expansion coefficient (TEC) of perovskite oxide was modeled and the microstructure of SOFC was optimized to improve the battery performance [95]	Classification and Regression	ANFIS and ANN	ML models identify intrinsic connections between material properties, microstructure, chemical composition, and external conditions in SOFC cathode studies	\	\
A generative adversarial network (GAN) framework is used to learn and generate three-dimensional (3D) microstructure of SOFC cathodes [96]	Deep learning	GAN	The microstructure generated by AN closely resembles the original in terms of visual, statistical, and topological features, and matches the property distribution of the original SOFC material in electrochemical performance simulations	3D volume image data obtained using FIB-SEM, as well as multiphase microstructure generated using DREAM.3D code package, were generated by ellipsoidal stacking scheme	\

4.1. ML used for PEMFC

For PEMFC, the electrocatalyst is crucial due to the fact that it fundamentally determines the rate of electrochemical reactions at the microscopic level. Therefore, accelerating the selection process of electrocatalysts is the main target that researchers aim to achieve using ML.

4.1.1. Pt-based electrocatalysts

ML must be highly accurate to provide reliable guidance for subsequent work. To enhance the accuracy of model predictions, one approach is to incorporate multiple descriptors into the model. Rocabado and colleagues conducted multiple regression analyses to characterize the oxygen (O) binding energy on Pt surfaces and nanoparticles (NPs), which is regarded as a crucial descriptor for the rate-determining step [77]. The ML model they proposed uses five descriptors, derived from the structural and electronic properties of Pt NPs, to predict the O binding energy with high precision. Consequently, the multivariate-descriptor model they developed has significantly outperformed traditional single-descriptor models.

ML algorithms are highly effective at quickly generating accurate predictions about material properties. This capability made them popular and promising for future developments in material science. Zhen and colleagues created a physical niche genetic (PNG) ML program designed to identify the relationship between structural configuration and stability within the Pt-Ni alloy nanocluster framework [78]. To construct the ML model, a dataset comprising 500 optimized samples, obtained through DFT calculations, was used for training. Consequently, the ML model proved to be an accurate surrogate, capable of making predictions with a margin of error below 0.13 eV. Notably, the computational time required for the ML model, known as post-training, is virtually insignificant. Taking advantage of this benefit, the researchers employed the ML model, which requires substantially fewer computational resources, to screen 2.5×10^5 candidate structures. Ultimately, they determined the most stable configuration to be the $Pt_{43}Ni_{42}$.

Chen *et al.* trained a neural network potential (NNP) combined with a genetic algorithm (GA) to investigate a similar system of the 55-atom $Au_{13}Pt_{42}$ cluster [79]. A total of 42348 structures computed by DFT were used as training sets. After training, the ML model could precisely reproduce the DFT results with a RMSE of less than 8.7 meV per atom. Hence, searching a complex global configurational space for the most stable structures became practical, avoiding the immense computational expense of direct enumeration via DFT simulation at an astronomical level.

Lin *et al.* trained a RF as a surrogate model for the first principal simulation to compute the limiting potentials of graphene-supported SACs with different coordination structures ($M@C_3$, $M@C_4$, $M@pyridine\text{-}N_4$, and $M@pyrrole\text{-}N_4$) as electrocatalysts for the oxygen reduction reaction (ORR), oxygen evolution reaction (OER), and hydrogen evolution reaction (HER), respectively [80].

4.1.2. Nonprecious electrocatalysts

SACs represent a new research frontier due to their high active site coverage and maximum metal utilization. Lin *et al.* constructed an ML model based on the relationship between physical properties and adsorption strength of reaction intermediates to rapidly screen catalysts with specific properties [97]. The ML model successfully predicted the catalytic performance of 260 other graphene-supported SACs, and the prediction error was very small, comparable to the accuracy of the DFT calculations. Two OER catalysts, $Ir@pyridine\text{-}N_3C_1$ and $Ir@pyridine\text{-}N_2C_2$, were identified to perform better than the precious metal oxides RuO_2 and IrO_2, and one HER catalyst, $Ni@pyridine\text{-}N_3C_1$, was identified to perform better than the commonly used precious metal Pt. This study provides a new paradigm for predicting catalytic performance directly from the physical properties of catalyst candidates, significantly accelerating the design process of catalysts.

It should be noted that for nonprecious electrocatalysts, the M-N-C material system is more complex than that of Pt-based alloy NPs. For instance, the pore structure significantly influences the mass transfer process, which finally reflects the ORR performance [98]. Therefore, DFT simulations offer only a very partial description of nonprecious electrocatalysts, and more information needs to be obtained from experimental data.

Dan *et al.* used a ML algorithm to construct a predictive model to determine the number of electrons transferred during an ORR [81]. The model utilizes experimental conditions, the results from different characterization techniques, and the material's response to ORR as input variables. The authors trained 19 different types of ML regression algorithms to predict the number of electrons transferred during ORR. Through training, the Stepwise Linear Regression model was identified as the most effective.

The authors applied a reverse engineering process to determine the optimal synthesis conditions or material properties. The reverse prediction function is applied to predicted values and identifies the most important variables to guide the investigation of structure-ORR response relationships. This approach also helps to understand how to modify the material for an improved ORR response.

The authors used a ML model to predict the optimal synthesis conditions required for the transfer of four electrons during the ORR process. The predicted parameters serve as a guide for enhancing material properties. This work demonstrates the potential of ML techniques in optimizing the synthesis conditions and designing nitrogen-doped graphene materials, as well as their application value in predicting ORR properties.

4.2. ML used for SOFC

SOFC is mainly divided into two types, proton conduction types and oxygen ion conduction types. For oxygen ion conductive, air-grade materials, the primary concern is the material's oxygen ion conductivity and ORR activity. Zhang *et al.* developed a Gaussian Process Regression model based on atomic properties derived from electron negativity and ion radius to predict the oxygen ion conductivity of ABO_3 perovskites [54]. They constructed descriptors using fundamental physicochemical parameters, which were validated by both experimental and simulation, such as ionic radius and electron negativity, as these parameters are closely related to oxygen ion conductivity. A dataset comprising 128 samples, including experimental values of oxygen ion conductivity for different perovskite oxides was investigated. The characteristics analyzed included the radii of the A-site and B-site ions, the difference in electron negativity, the charge, and other factors related to the radius of the oxygen ion.

After hyperparameter adjustment and model training, the final test model (M_2) demonstrated high accuracy and stability. The correlation coefficient (R^2) reached 99.87%, the RMSR was 0.1577, and the MAE was 0.1048. Compared with four other models (SVM, partial least squares, ANN, and quadratic approximation), the performance of Gaussian process regression model is better.

The authors also investigated the correlation between features and calculated the correlation coefficients between features to identify and select those features that significantly affect the model's predictive performance. They constructed multiple models, each with different combinations of features removed, and assessed their performance. Additionally, they attempted to build a model by eliminating highly correlated features to avoid potential multicollinearity issues. But the final results indicated that the model incorporating all six features provided better performance.

The GPR model developed by the authors is a fast, robust, and low-cost tool for estimating oxygen ion conductivity. This model is applicable to all kinds of perovskites, including ordered, disordered, and non-stoichiometric varieties. It correlates readily available alloying element properties with oxygen ion conductivity, which aids in designing doped perovskites with desired ranges of oxygen ion transport performance.

In terms of ORR, by introducing Ionic Lewis Acid Strength (ISA) as an effective physical descriptor of oxygen reduction reactivity, Shuo Zhai *et al.* developed an efficient ML driven method to accelerate the screening of high-efficiency cathode materials from a large number of perovskites compositions [73].

The authors collected oxygen reduction reaction (ORR) activity data for different perovskite oxides as an initial dataset and discarded materials with very large area specific resistance (ASR) values (>5 $\Omega \cdot cm^2$) to prevent potential negative impacts on model stability. The resulting dataset comprised 85 different perovskite materials with a total of 162 different ASR values at 700 ℃ and 650 ℃.

By selecting nine ion descriptors including ISA values of A-position and B-position, ion electronegativity, ion radius, ionization energy, and tolerance factor, the dataset was fitted using different regression methods including linear and nonlinear models. Finally, an ANN model was obtained having the best performance. The study found that the accuracy of the model with A-bit and B-bit ISA values was much higher than that of the model without this feature.

Then, the ANN model was used to predict the activity of unexplored materials, and those with low predicted values were selected for synthesis and characterization. Finally, four kinds of perovskite oxides were successfully synthesized and confirmed to have a superior activity index. In particular, $Sr_{0.9}Cs_{0.1}Co_{0.9}Nb_{0.1}O_3$ (SCCN) exhibited an extremely low ASR.

Through DFT calculations, the relationship between the polarization distribution of ISA at the A and B positions and the transfer of electron pairs was investigated. It was found that the enhanced activity is mainly attributed to the polarization distribution of ISA at the A and B positions, which greatly reduces the formation energy and migration barrier of oxygen vacancies. This provides a mechanism explanation for the design of oxygen reduction electrodes.

In terms of material design, the work by Shuo Zhai *et al.* offers a valuable template. This involves selecting a specific type of material with excellent properties using ML, investigating the influence of particular characteristics on these properties, experimental preparation and verifying the material's properties, and finally analyzing the calculation results through theoretical methods such as DFT to comprehend the physical mechanisms behind the material's properties.

For proton conduction fuel cells, proton conductivity is the primary concern for researchers. To address the performance bottleneck of air electrode materials in proton conductive solid oxide fuel cells (P-SOCs), Ning Wang *et al.* used ML methods to quickly and accurately predict the proton absorption capacity (PAA) of Co/Fe-based perovskite oxides, which is a key indicator for evaluating their performance as P-SOCs air electrodes [88]. They constructed a ML model based on the Extreme Gradient Lift (XG-Boost) algorithm to predict the PAA of 27 different elements doped with Co/Fe based perovskite oxides at the B site. By combining the ML model and DFT calculations, La $(Co_{0.9}Ni_{0.1})$ O_3 (LCN_{91}) was selected as a potential air electrode material. In particular, they also prepared the designed LCN_{91} using laboratory synthesis methods and successfully applied it to P-SOCs. When used as an air electrode, it exhibits excellent electrochemical performance in both fuel cell and electrolytic modes. In particular, at 600 ℃, the current density in electrolytic mode reaches 843 mA cm^{-2} and the peak power density in fuel cell mode is 591 mW cm^{-2}.

For fuel cell electrolytes, knowing the total conductivity and carrier type of material is very important for designing electrolyte materials with high conductivity. Priya *et al.* used ML tools to design and identify ABO_3 type perovskite oxides with high electrical conductivity, suitable for various energy applications [87]. Utilizing over 7000 data points from the literature, more than 100 characteristics to determine the key predictors that influence perovskite conductivity and charge carrier type.

ML models, such as XG-Boost, lasso, ridge, elastic net, a 7-layer ANN with 100 nodes, SVM, and RF, operate by selecting feature subsets at random and constructing decision trees from limited data. The predictions from these models are then averaged to generate the final prediction. A robust learning framework was constructed and the total electrical conductivity of perovskites was classified based on temperature and environmental conditions.

It is found that the mean ionic radius, minimum electronegativity, minimum atomic mass, and minimum oxide formation energy of B-site ions are the key characteristics that determine the conductivity and charge carrier type.

In the future, perovskite materials with high electrical conductivity can be designed according to these key characteristics.

Catalysts are an integral part of SOFC, and those with high catalytic performance can greatly enhance the power generation efficiency of fuel cells. Jacobs *et al.* used ML models to conduct the following three types of studies [89]:

(1) The ML model for predicting perovskite catalytic performance was developed and proved to be superior to the prediction of the O-*p* band center descriptor method. The research showed that the ML model based on element characteristics can provide higher average accuracy than the O-*p* band center correlation in predicting key properties. It can be used to make predictions faster than O-*p* band center correlation, because ML does not require DFT computations to develop and evaluate ML models for ASR prediction.

(2) An ML model for ASR prediction was developed and evaluated. Our ASR model can achieve low error by combining elemental characteristics, single thermal coding of electrolyte type, and a novel characterization scheme for predicting the ASR activation energy barrier with a separate ML model. This characterization is then used as a feature in the ASR prediction model. The addition of ASR energy barrier makes temperature-dependent ASR prediction possible. In addition, our ASR model shows the ability to predict future promising materials using time cross-validation and provides calibrated estimates of uncertainty.

(3) Data on approximately 19 million perovskite components were searched using the ASR ML model and new promising, inexpensive, abundant perovskite materials with high stability and catalytic activity were identified. Materials containing less studied but promising elements (such as K, Bi, Y, Ni, Cu) were found and are worthy of further study.

The stability of the material is crucial for the proper operation of the fuel cell within its operating environment. Li *et al.* utilized more than 1900 datasets of perovskite oxide energies calculated by DFT to determine phase stability using Ehull analysis. They considered the magnitude of Ehull value as a direct measure of stability [86]. The researchers applied a feature generation method to generate 791 features from element attribute data, and they discovered that the first 70 features were sufficient to train accurate models without significant overfitting, using three feature selection methods: stability selection, recursive feature elimination, and univariate feature selection based on mutual information.

In terms of classification models, the Extra Trees algorithm achieved a prediction accuracy of 0.93 (±0.02) and an F1 score of 0.88 (±0.03) in the classification task. In terms of Regression models, Kernel Ridge Regression achieved a minimum root-mean-square error (RMSE) of 28.5 (±7.5) meV/atom and a mean absolute error (MAE) of 16.7 (±2.3) meV/atom in a 20% cross-validation test.

To further verify the accuracy of the model, they constructed five test sets to evaluate the model's performance in different groups of molecular Spaces. This was done by removing specific subsets of perovskite materials from the complete dataset. These test sets were constructed based on the frequency of element occurrence in the training dataset and the type of elements (for example, alkaline earth and rare earth metals in position A). For each test set, the authors trained the model on the remaining dataset, after excluding the test set, and then applied the optimal model to the excluded data to classify the stability of each material and predict the Ehull value. Fifteen new perovskite compounds were manually generated and the stability of these new compounds was predicted using the developed model. After the prediction was completed, the Ehull values of these new compounds were calculated using DFT and compared to the model's predicted values. By comparing the prediction results with the DFT calculated values, the prediction accuracy of the model in different molecular Spaces was validated. The paper demonstrates the potential of ML applications in materials science, particularly in predicting the thermodynamic stability of materials, thereby significantly reducing the time required for DFT calculations.

5. Challenge and future of ML application in the field of fuel cells

5.1. The problem of lack of datasets

The performance of ML models highly depends on the quantity and quality of the training data. The lack of a comprehensive and high-quality database is currently the main factor limiting the widespread application of ML in the field of fuel cells. Some researchers have established databases for specific problems. For example, databases such as The Materials Project, AFLOW, and others mentioned above are widely used in ML training [99,100]. Min Zou *et al.* have established models for predicting the phase structure of cobalt-based superalloys based on the experimental data. These models can predict the precipitation of harmful phases and the volume fraction of γ' phase, thereby enhancing the design of new alloys [101]. In some literature, the dataset is incomplete but the accuracy of the final ML model's calculation results is very high, exceeding 90%. The reason may be that the established models are simple, and some are linear models, which makes them easy to fit. It is also possible that the established models have some limitations and prerequisites for their use, and only have high accuracy in certain intervals. Therefore, the establishment of the sufficiently comprehensive dataset is a crucial factor influencing the model's generalization capability.

If insufficient data can be collected, there are several methods to augment the dataset. For small datasets, it can be manually expanded by applying transformations such as rotation, scaling, and cropping to generate sufficient training data [102–104]. For instance, Su *et al.* propose a transfer learning approach combined with data augmentation strategies to improve the performance of classification tasks [105]. Deep learning models require large amounts of data for training, and in real-world scenarios, there is often insufficient data in the target domain. This paper proposes a method that combines transfer learning and data augmentation. This method uses data augmentation to expand the amount of data in the target domain and applies a transfer learning model to reduce

the difference between the source domain and the target domain, thereby addressing the issue of data scarcity. A significant number of experiments have been conducted on multiple datasets to validate the effectiveness of the proposed method. The results show that data augmentation and transfer learning can serve as effective strategies to solve the problem of data scarcity in the target domain and enhance the model's generalization capability by integrating these strategies.

5.2. The problem of constructing features

Feature is another key factor that determines the performance of ML models. Nevertheless, for different targets or properties, the related features are different and often unknown, which makes it difficult to find or construct suitable features for certain properties. Blindly increasing the number of features may have the opposite effect, leading to overfitting of the model, and potentially reducing the model's accuracy [106]. Therefore, identifying the most useful features among a large number of candidates remains a challenge to be addressed. Some well-known features are highly representative, such as the tolerance factor "t" proposed by Goldschmidt a century ago, which still plays a significant role in the field of materials development today [107]. Some features are particularly useful for certain properties, such as the Lewis acid strength, which is relevant to the oxygen reduction reaction activity of perovskite oxides [73].

To identify appropriate features, constructing complex features is a common approach. Bare *et al.* developed a new tolerance factor "τ" that can reliably predict whether a composition will form a stable perovskite structure. This reduces the number of candidate materials that require expensive DFT calculations, thereby guiding high-throughput researchers in exploring a wide multi-component space [108]. This approach accelerates the process of discovery and design of new materials. Additionally, new methods have been developed to filter out the most influential features from a vast array of feature types, such as feature selection algorithms like recursive feature elimination (RFE) and model-based feature selection. More significantly, the incorporation of physical understanding into the feature construction can dramatically enhance the quality of the features.

5.3. The problem of model generalization

Model generalization refers to the ability of a ML model to make predictions about new data that has not been seen before [109]. A model with good generalization ability can learn universal patterns from the training data and successfully apply them to new data. If a model generalizes poorly, it may perform well on the training data (overfitting) but poorly on new data. Problems such as data imbalance, difficulties in feature selection and dynamically changing data distributions may affect the generalization ability of the model [110]. In order to address the generalization issues of ML models, a series of strategies can be adopted to enhance the model's predictive performance on new data. This includes using multi-task learning to share features between related tasks, increasing the diversity of training samples through data augmentation techniques such as rotation and flipping [111], adjusting data distribution to avoid model bias against specific types of data, and using larger batches of data to improve the stability of model learning. In addition, hyperparameter tuning [112], regularization techniques [113], the early stopping method [114], and ensemble learning [115] can effectively prevent overfitting and improve the generalization performance of the model. Transfer learning [116] and meta-learning [117]

enable models to quickly adapt to new tasks using knowledge gained from other tasks, especially when data is scarce. Feature importance analysis [118] and model simplification [119] are also effective methods to improve generalization ability. The combination of these methods, tailored to specific problems and data characteristics, can help the model maintain high accuracy on unseen data, thereby playing a more significant role in practical applications.

5.4. The problem of poor interpretability

In the field of fuel cell material prediction, although ML models, especially DL models, have made significant progress in terms of prediction accuracy, their "black box" nature leads to a lack of transparency and interpretability in the decision-making process. This poses a significant challenge for material design as researchers often need to comprehend the underlying basis and internal logic of the model's predictions. To verify the reliability of the model, it is essential to ensure that the predicted results are consistent with physical phenomena and chemical principles and to conduct material design and process optimization based on these predictions. Therefore, enhancing and improving the interpretability of models, developing new interpretative ML algorithms, or integrating model predictions with existing theoretical knowledge has become one of the key challenges to be addressed in this field. There are already some effective methods to improve the interpretability of models. Such as LIME [120] (Locally Interpretable Model-agnostic Explanations) and SHAP [121] are currently available, which can provide explanations for individual predictions. PDP (Partial Dependence Plot) and ICE (Individual Conditional Expectation) are visualization techniques used to show the marginal effect of one or two features on the predicted outcome of a ML model. They can help researchers determine how the model's predictions change when these features are adjusted.

6. Summary

This review comprehensively explored the current status, challenges, and potential solutions of AI applications in the field of fuel cells. This paper first introduces the working principle, main types, and basic requirements of fuel cell materials, and it especially highlights the existing problems in the current research of fuel cell materials, such as electrolyte stability and catalyst activity. Following this, the paper outlines the basic concepts of AI, including its application steps in materials science, such as data collection, feature construction, model training, and validation.

The review focuses on the specific applications of AI in fuel cell materials research, including advancements in predicting catalyst performance, electrolyte conductivity, and material stability. At the same time, the paper highlights the challenges in AI applications, such as insufficient datasets, complexity of feature construction, limitations in model generalization, and problems with model interpretability. To address these issues, the paper introduces a range of possible strategies, including expanding datasets, enhancing feature selection methods, improving model generalization, and developing novel interpretable ML algorithms.

Finally, the review summarizes the potential of AI in the development of fuel cell materials and looks ahead to future research directions, highlighting the importance of improving model generalization and interpretability, optimizing algorithms to accommodate more complex data patterns, and enhancing

the cross-integration of ML and materials science. Through these contents, the paper offers a comprehensive perspective on the application of AI in material research in the field of fuel cells.

Acknowledgments

This work was financially supported by special funds for Double-First Class construction (Grant Nos. 0515024GH0201201; 0515024SH0201201), the Fundamental Research Funds for the Central Universities (Grant No. D5000230358), and the Key Research and Development Program of Shaanxi Province (Grant No. 2024GX-YBXM-456). The authors gratefully acknowledged the kindly financial support provided by the nation to Zhi-Peng Li.

Authors' contributions

Conceptualization, J.L., Z.L., and H.L. ; investigation, H.L. and H.G.; data curation, H.L., H.G and Z.G.; writing—original draft preparation, H.L. and H.G.; writing—review and editing, Z.G. and H.P.; visualization, H.G. and J.Z.; All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

Reference

[1] Ahluwalia RK, Peng JK. Automotive hydrogen storage system using cryo-adsorption on activated carbon. *Int. J. Hydrogen Energy* 2009, 34(13):5476–5487.
[2] Olah G, Goeppert A, Prakash G. *Hydrogen as a future energy carrier*. New Jersey: Wiley. 2008.
[3] Ortiz-Rivera EI, Reyes-Hernandez AL, Febo RA. Understanding the history of fuel cells. In *2007 IEEE Conference on the History of Electric Power*, New Jersey, USA, August 3–5, 2007, pp. 117–122.
[4] Cook G. *Realizing a hydrogen future: Hydrogen technical advisory panel recommendations (Brochure)*, Golden: National Renewable Energy Lab. (NREL), 1999.
[5] Nanji HXYLJXJ. Progress in intermediate and high temperature SOFC/MGT combined power generation technology. *Integr. Intell. Energy*. 2019, 41(8): 1.
[6] Mueller F, Gaynor R, Auld AE, Brouwer J, Jabbari F, *et al.* Synergistic integration of a gas turbine and solid oxide fuel cell for improved transient capability. *J. Power Sources*. 2008, 176(1): 229–239.
[7] Daw CS, Graves RL, Caton JA, Wagner RM. Summary report on the transportation combustion engine efficiency colloquium held at USCAR, March 3 and 4, 2010. Available: https://info.ornl.gov/sites/publications/files/Pub26829.pdf (accessed on 09 November 2024).
[8] Nguyen NT, Chan SH. Micromachined polymer electrolyte membrane and direct methanol fuel cells—A review. *J. Micromechanics Microengineering*. 2006, 16(4): R1.
[9] Lu J, Zhang Z, Tang Z. Review on the development of solid oxide fuel cells. *Rare Met. Mater. Eng.* 2005, 34(8): 1177–1180.
[10] Büchi FN, Inaba M, Schmidt TJ. *Polymer electrolyte fuel cell durability*, 1st ed. New York: Springer, 2009.
[11] Wu J, Yuan XZ, Martin JJ, Wang H, Zhang J, *et al.* A review of PEM fuel cell durability: Degradation mechanisms and mitigation strategies. *J. Power Sources*. 2008, 184(1): 104–119.
[12] Wang J. System integration, durability and reliability of fuel cells: Challenges and solutions. *Appl. Energy*. 2017, 189: 460–479.
[13] Othman R, Dicks AL, Zhu Z. Non precious metal catalysts for the PEM fuel cell cathode. *Int. J. Hydrog. Energy*. 2012, 37(1): 357–372.

[14] Xiang X, Sun X, Briceño G, Lou Y, Wang K, *et al.* A combinatorial approach to materials discovery. *Science*. 1995, 268(5218): 1738–1740.
[15] Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, *et al.* Scaling deep learning for materials discovery. *Nature*. 2023, 624(7990): 80–85.
[16] Houchins G, Viswanathan V. An Accurate Machine-Learning calculator for optimization of Li-Ion battery cathodes. *J. Chem. Phys.* 2020, 153(5):054124.
[17] Jiang Z, Li J, Yang Y, Mu L, Wei C, *et al.* Machine-Learning-Revealed statistics of the Particle-Carbon/Binder detachment in Lithium-Ion battery cathodes. *Nat. Commun.* 2020, 11(1): 2310.
[18] Joshi RP, Eickholt J, Li L, Fornari M, Barone V, *et al.* Machine learning the voltage of electrode materials in Metal-Ion batteries. *ACS Appl. Mater. Interfaces*. 2019, 11(20): 18494–18503.
[19] Wang H, Feng J, Dong Z, Jin L, Li M, *et al.* Efficient screening framework for organic solar cells with deep learning and ensemble learning. *Npj Comput. Mater.* 2023, 9(1): 200.
[20] Mekhilef S, Saidur R, Safari A. Comparative study of different fuel cell technologies. *Renew. Sustain. Energy Rev.* 2012, 16(1): 981–989.
[21] Giorgi L, Leccese F. Fuel Cells: Technologies and applications. *Open Fuel Cells J.* 2013, 6(1):1–20.
[22] Vielstich W. *Handbook of fuel cells: Fundamentals, technology and applications*. New Jersey: Wiley. 2003.
[23] Giddey S, Badwal S, Kulkarni A, Munnings C. A comprehensive review of direct carbon fuel cell technology. *Prog. Energy Combust. Sci.* 2012, 38(3): 360–399.
[24] Kirubakaran A, Jain S, Nema R. A review on fuel cell technologies and power electronic interface. *Renew. Sustain. Energy Rev.* 2009, 13(9): 2430–2440.
[25] Yu J, Wang Y, Hui Y, Weng S. Numerical analysis of the performance of planar electrode supported solid oxide fuel cell with different porosities. *Guisuanyan Xuebao Journal Chin. Ceram. Soc.* 2010, 38(2): 246–251.
[26] Shao Y, Yin G, Wang Z, Gao Y. Proton exchange membrane fuel cell from low temperature to high temperature: Material challenges. *J. Power Sources*. 2007, 167(2): 235–242.
[27] Tong Xin, Xiong Zhe, Gao Xinyu, Hou Junwei, Liu Yiying, *et al.* Research status and development of proton exchange membrane fuel cell. *Bull. Chin. Ceram. Soc.* 2022, 41(9): 3243–3258.
[28] Athanasaki G, Jayakumar A, Kannan A. Gas diffusion layers for PEM fuel cells: Materials, properties and manufacturing—A review. *Int. J. Hydrog. Energy*. 2023, 48(6): 2294–2313.
[29] Tan LJ, Yang C, Zhou N. Performance of the solid oxide fuel cell (SOFC)/Proton-Exchange membrane fuel cell (PEMFC) hybrid system. *Chem. Eng. Technol.* 2016, 39(4): 689–698.
[30] Radenahmad N, Afif A, Petra PI, Rahman SM, Eriksson S-G, *et al.* Proton-Conducting electrolytes for direct methanol and direct urea fuel cells—A State-of-the-Art review. *Renew. Sustain. Energy Rev.* 2016, 57: 1347–1358.
[31] Wang J, Liu M. Computational study of Sulfur–Nickel interactions: A new S-Ni phase diagram. *Electrochem. Commun.* 2007, 9(9): 2212–2217.
[32] Gong M, Liu X, Trembly J, Johnson C. Sulfur-Tolerant anode materials for solid oxide fuel cell application. *J. Power Sources*. 2007, 168(2): 289–298.
[33] Mizusaki J, Yonemura Y, Kamata H, Ohyama K, Mori N, *et al.* Electronic conductivity, seebeck coefficient, defect and electronic structure of nonstoichiometric La1- xSrxMnO3. *Solid State Ion.* 2000, 132(3–4): 167–180.
[34] Ullmann H, Trofimenko N, Tietz F, Stöver D, Ahmad-Khanlou A. Correlation between thermal expansion and oxide ion transport in mixed conducting Perovskite-Type oxides for SOFC cathodes. *Solid State Ion.* 2000, 138(1–2): 79–90.
[35] Zheng Y, Shan G, Xueying Z, Hong C, Huang S, *et al.* LSM particle size effect on the overall performance of IT-SOFC. *J. Rare Earths*. 2012, 30(12): 1240–1244.
[36] Hosokawa K, Kondo A, Okumiya M, Abe H, Naito M. One-Step mechanical processing to prepare LSM/ScSZ composite particles for SOFC cathode. *Adv. Powder Technol.* 2014, 25(5): 1430–1434.
[37] Song HS, Kim WH, Hyun SH, Moon J. Influences of starting particulate materials on microstructural evolution and electrochemical activity of LSM-YSZ composite cathode for SOFC. *J. Electroceramics*. 2006, 17: 759–764.

[38] Zhu X, Sun K, Zhang N, Chen X, Wu L, *et al.* Improved electrochemical performance of SrCo0.8Fe0.2O3-δ-La0.45Ce0.55O2-δ composite cathodes for IT-SOFC. *Electrochem. Commun.* 2007, 9(3): 431–435.
[39] Liu A, Yang Y, Kong D, Ren X, Gao M, *et al.* DFT Study of the defective carbon materials with vacancy and heteroatom as catalyst for NRR. *Appl. Surf. Sci.* 2021, 536: 147851.
[40] Wang Y, Li Y, Tang Z, Li H, Yuan Z, *et al.* Universal materials model of Deep-Learning density functional theory hamiltonian. *Sci. Bull.* 2024,69(16): 2514–2521.
[41] Han Y, Xu H, Li Q, Du A, Yan X. DFT-Assisted Low-Dimensional Carbon-Based electrocatalysts design and mechanism study: A review. *Front. Chem.* 2023, 11: 1286257.
[42] Ghahramani Z. Probabilistic machine learning and artificial intelligence. *nature* 2015, 521(7553): 452–459.
[43] Coulter N. ACM'S computing classification system reflects changing times. *Commun. ACM*. 1997, 40(12): 111–112.
[44] Baştanlar Y, Özuysal M. Introduction to machine learning. *MiRNomics MicroRNA Biol. Comput. Anal.* 2014, 105–128.
[45] Tao Q, Xu P, Li M, Lu W. Machine learning for perovskite materials design and discovery. *Npj Comput. Mater.* 2021, 7(1): 23.
[46] Ma J, Cao B, Dong S, Tian Y, Wang M, *et al.* MLMD: A Programming-Free AI platform to predict and design materials. *Npj Comput. Mater.* 2024, 10(1): 59.
[47] Reyes KG, Maruyama B. The machine learning revolution in materials? *MRS Bull.* 2019, 44(7): 530–537.
[48] Wei J, Chu X, Sun X, Xu K, Deng H, *et al.* Machine learning in materials science. *InfoMat*. 2019, 1(3): 338–358.
[49] Arriagada J, Olausson P, Selimovic A. Artificial neural network simulator for SOFC performance prediction. *J. Power Sources*. 2002, 112(1): 54–60.
[50] Xu H, Ma J, Tan P, Chen B, Wu Z, *et al.* Towards online optimisation of solid oxide fuel cell performance: Combining deep learning with Multi-Physics simulation. *Energy AI*. 2020, 1: 100003.
[51] Papadimitriou I, Gialampoukidis I, Vrochidis S, Kompatsiaris I. AI methods in materials design, discovery and manufacturing: A review. *Comput. Mater. Sci.* 2024, 235: 112793.
[52] Zhao S, Chen S, Zhou J, Li C, Tang T, *et al.* Potential to transform words to watts with large language models in battery research. *Cell Rep. Phys. Sci.* 2024, 5(3): 101844.
[53] Parker AJ, Opletal G, Barnard AS. Classification of platinum nanoparticle catalysts using machine learning. *J. Appl. Phys.* 2020, 128(1): 014301.
[54] Zhang Y, Xu X. Modeling oxygen ionic conductivities of ABO3 perovskites through machine learning. *Chem. Phys.* 2022, 558: 111511.
[55] Wang N, Yuan B, Tang C, Du L, Zhu R, *et al.* Machine-Learning-Accelerated development of efficient mixed Protonic—Electronic conducting oxides as the air electrodes for protonic ceramic cells. *Adv. Mater.* 2022, 34(51): 2203446.
[56] Jain A, Ong SP, Hautier G, Chen W, Richards WD, *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* 2013, 1(1): 011002–011002.
[57] Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *Jom*. 2013, 65: 1501–1509.
[58] Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, *et al.* The open quantum materials database (OQMD): Assessing the accuracy of DFT formation energies. *Npj Comput. Mater.* 2015, 1(1): 1–15.
[59] Landis DD, Hummelshøj JS, Nestorov S, Greeley J, Dułak M, *et al.* The computational materials repository. *Comput. Sci. Eng.* 2012, 14(6): 51–57.
[60] Curtarolo S, Setyawan W, Hart GL, Jahnatek M, Chepulskii RV, *et al.* AFLOW: An automatic framework for High-Throughput materials discovery. *Comput. Mater. Sci.* 2012, 58: 218–226.
[61] Bergerhoff G, Hundt R, Sievers R, Brown I. The inorganic crystal structure data base. *J. Chem. Inf. Comput. Sci.* 1983, 23(2): 66–69.

[62] Stalick JK, Mighell AD. Crystal data: Version 1.0 database specifications. *Final Rep. Natl. Bur. Stand.* 1986.
[63] Belsky A, Hellenbrandt M, Karen VL, Luksch P. New developments in the inorganic crystal structure database (ICSD): Accessibility in support of materials research and design. *Acta Crystallogr. B*. 2002, 58(3): 364–369.
[64] Groom CR, Bruno IJ, Lightfoot MP, Ward SC. The cambridge structural database. *Struct. Sci.* 2016, 72(2): 171–179.
[65] Vaitkus A, Merkys A, Sander T, Quirós M, Thiessen PA, *et al.* A workflow for deriving chemical entities from crystallographic data and its application to the crystallography open dat abase. *J. Cheminformatics*. 2023, 15(1):123.
[66] Merkys A, Vaitkus A, Grybauskas A, Konovalovas A, Quirós M, *et al.* Graph Isomorphism-Based algorithm for Cross-Checking chemical and crystallographic descriptions. *J. Cheminformatics*. 2023, 15(1): 25.
[67] Vaitkus A, Merkys A, Gražulis S. Validation of the crystallography open database using the crystallographic information framework. *J. Appl. Crystallogr.* 2021, 54(2): 661–672.
[68] Srivastava M, Howard JM, Gong T, Rebello Sousa Dias M, Leite MS. Machine learning roadmap for perovskite photovoltaics. *J. Phys. Chem. Lett.* 2021, 12(32): 7866–7877.
[69] Xu Jianbing, Li Hanshi, Tan Jimin, Han Minfang, Chen Di. Research progress on data science of solid oxide fuel cells, lithium batteries, CO_2 electroreduction catalysts. *J. Chin. Ceram. Soc.* 2022, 50(11): 3021–3039.
[70] Guo N, Chen S, Tao J, Liu Y, Wan J, *et al.* Semi-Supervised learning for explainable Few-Shot battery lifetime prediction. *Joule*. 2024,8(6): 1820–1836.
[71] Liu Y, Chen S, Li P, Wan J, Li X. Status, Challenges, and promises of Data-Driven battery lifetime prediction under Cyber-Physical system context. *IET Cyber-Phys. Syst. Theory Appl.* 2024,9(3): 207–217.
[72] Lu T, Zhai X, Chen S, Liu Y, Wan J, *et al.* Robust battery lifetime prediction with noisy measurements via Total-Least-Squares regression. *Integration*. 2024, 96: 102136.
[73] Zhai S, Xie H, Cui P, Guan D, Wang J, *et al.* A combined ionic lewis acid descriptor and Machine-Learning approach to prediction of efficient oxygen reduction electrodes for ceramic fuel cells. *Nat. Energy* 2022, 7(9): 866–875.
[74] Bello IT. Revolutionizing material design for protonic ceramic fuel cells: Bridging the limitations of conventional experimental screening and machine learning methods. *Chem. Eng. J.* 2023,477: 147098.
[75] Rui Z, Ding R, Hua K, Duan X, Li X, *et al.* Design of proton exchange membranes with high durability for fuel cells: From the perspective of machine learning. *J. Membr. Sci.* 2023, 683: 121831.
[76] Yin P, Niu X, Li S, Chen K, Zhang X, *et al.* Machine-Learning-Accelerated design of High-Performance platinum intermetallic nanoparticle fuel cell catalysts. *Nat. Commun.* 2024, 15(1): 415.
[77] Rivera Rocabado DS, Nanba Y, Koyama M. Density functional theory and machine learning description and prediction of oxygen atom chemisorption on platinum surfaces and nanoparticles. *ACS Omega*. 2021, 6(27): 17424–17432.
[78] Zhen H, Liu L, Lin Z, Gao S, Li X, *et al.* Physically compatible machine learning study on the Pt-Ni nanoclusters. *J. Phys. Chem. Lett.* 2021, 12(5): 1573–1580.
[79] Chen D, Lai Z, Zhang J, Chen J, Hu P, *et al.* Gold segregation improves electrocatalytic activity of icosahedron Au@Pt nanocluster: Insights from machine learning. *Chin. J. Chem.* 2021, 39(11): 3029–3036.
[80] Lin S, Xu H, Wang Y, Zeng XC, Chen Z. Directly predicting limiting potentials from easily obtainable physical properties of Graphene-Supported Single-Atom electrocatalysts by machine learning. *J. Mater. Chem. A*. 2020, 8(11): 5663–5670.
[81] Dan M, Vulcu A, Porav SA, Leostean C, Borodi G, *et al.* Eco-Friendly Nitrogen-Doped graphene preparation and design for the oxygen reduction reaction. *Molecules* 2021, 26(13): 3858.
[82] Batchelor TAA, Löffler T, Xiao B, Krysiak OA, Strotkötter V, *et al.* Complex-Solid-Solution electrocatalyst discovery by computational prediction and High-Throughput experimentation. *Angew. Chem. Int. Ed.* 2021, 60(13): 6932–6937.

[83] Park Y, Hwang C, Bang K, Hong D, Nam H, *et al.* Machine learning filters out efficient electrocatalysts in the massive ternary alloy space for fuel cells. *Appl. Catal. B Environ.* 2023, 339: 123128.
[84] Kort-Kamp WJM, Ferrandon M, Wang X, Park JH, Malla RK, *et al.* Adaptive Learning-Driven High-Throughput synthesis of oxygen reduction reaction Fe-N-C electrocatalysts. *J. Power Sources*. 2023, 559: 232583.
[85] Lu Z, Chen ZW, Singh CV. Neural Network-Assisted development of High-Entropy alloy catalysts: Decoupling ligand and coordination effects. *Matter*. 2020, 3(4): 1318–1333.
[86] Li W, Jacobs R, Morgan D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* 2018, 150: 454–463.
[87] Priya P, Aluru NR. Accelerated design and discovery of perovskites with high conductivity for energy applications through machine learning. *Npj Comput. Mater.* 2021, 7(1): 90.
[88] Wang N, Yuan B, Zheng F, Mo S, Zhang X, *et al.* Machine-Learning assisted screening proton conducting Co/Fe based oxide for the air electrode of protonic solid oxide cell. *Adv. Funct. Mater.* 2024, 34(12): 2309855.
[89] Jacobs R, Liu J, Abernathy H, Morgan D. Machine learning design of perovskite catalytic properties. *Adv. Energy Mater.* 2024, 14(12): 2303684.
[90] Braham EJ, Cho J, Forlano KM, Watson DF, Arròyave R, *et al.* Machine Learning-Directed navigation of synthetic design space: A statistical learning approach to controlling the synthesis of perovskite halide nanoplatelets in the Quantum-Confined regime. *Chem. Mater.* 2019, 31(9): 3281–3292.
[91] Hwang H, Ahn J, Lee H, Oh J, Kim J, *et al.* Deep Learning-Assisted microstructural analysis of Ni/YSZ Anode composites for solid oxide fuel cells. *Mater. Charact.* 2021, 172: 110906.
[92] Toyoura K, Hirano D, Seko A, Shiga M, Kuwabara A, *et al.* Machine-Learning-Based selective sampling procedure for identifying the Low-Energy region in a potential energy surface: A case study on proton conduction in oxides. *Phys. Rev. B*. 2016, 93(5): 054112.
[93] Yang K, Liu J, Wang Y, Shi X, Wang J, *et al.* Machine-Learning-Assisted prediction of Long-Term performance degradation on solid oxide fuel cell cathodes induced by chromium poisoning. *J. Mater. Chem. A*. 2022, 10(44): 23683–23690.
[94] Hwang H, Choi SM, Oh J, Bae SM, Lee JH, *et al.* Integrated application of semantic Segmentation-Assisted deep learning to quantitative Multi-Phased microstructural analysis in composite materials: Case study of cathode composite materials of solid oxide fuel cells. *J. Power Sources*. 2020, 471: 228458.
[95] Zheng Y, Zhao C, Wu T, Li Y, Zhang W, *et al.* Enhanced oxygen reduction kinetics by a porous heterostructured cathode for intermediate temperature solid oxide fuel cells. *Energy AI*. 2020, 2: 100027.
[96] Hsu T, Epting WK, Kim H, Abernathy HW, Hackett GA, *et al.* Microstructure generation via generative adversarial network for heterogeneous, topologically complex 3D materials. *JOM*. 2021, 73(1): 90–102.
[97] Lin S, Xu H, Wang Y, Zeng XC, Chen Z. Directly predicting limiting potentials from easily obtainable physical properties of Graphene-Supported Single-Atom electrocatalysts by machine learning. *J. Mater. Chem. A*. 2020, 8(11): 5663–5670.
[98] Li J, Zhu X, Wang J, Rui Z, Zhang S, *et al.* Iron-Containing porphyrins Self-Assembled on ZnO nanoparticles as electrocatalytic materials for oxygen reduction. *ACS Appl. Nano Mater.* 2020, 3(1): 742–751.
[99] Jain A. Machine learning in materials research: Developments over the last decade and challenges for the future. 2024,33: 101189.
[100] Gossett E, Toher C, Oses C, Isayev O, Legrain F, *et al.* Aflow-ML: A RESTful API for Machine-Learning predictions of materials properties. *Comput. Mater. Sci.* 2018, 152: 134–145.
[101] Zou M, Li W, Li L, Zhao J, Feng Q. Machine learning assisted design approach for developing Γ′-Strengthened Co-Ni-Base superalloys. In *Superalloys 2020: Proceedings of the 14th International Symposium on Superalloys*, Switzerland: Springer, 2020, pp 937–947.
[102] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, *et al.* A comprehensive survey on transfer learning. *Proc.*

IEEE. 2020, 109(1): 43–76.
[103] Lin T, Chen S, Harris SJ, Zhao T, Liu Y, *et al.* Investigating explainable transfer learning for battery lifetime prediction under state transitions. *eScience* 2024, 4(5): 100280.
[104] Briceno-Mena LA, Arges CG, Romagnoli JA. Machine Learning-Based surrogate models and transfer learning for derivative free optimization of HT-PEM fuel cells. *Comput. Chem. Eng.* 2023, 171: 108159.
[105] Su J, Yu X, Wang X, Wang Z, Chao G. Enhanced transfer learning with data augmentation. *Eng. Appl. Artif. Intell.* 2024, 129: 107602.
[106] Ying X. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, IOP publishing, 2019, pp. 022022.
[107] Goldschmidt VM. Die gesetze der krystallochemie. *Naturwissenschaften*. 1926, 14(21): 477–485.
[108] Bare ZJL, Morelock RJ, Musgrave CB. Dataset of theoretical multinary perovskite oxides. *Sci. Data*. 2023, 10(1): 244.
[109] Wang J, Lan C, Liu C, Ouyang Y, Qin T, *et al.* Generalizing to unseen domains: A survey on domain generalization. *IEEE Trans. Knowl. Data Eng.* 2022, 35(8): 8052–8072.
[110] Liu J, Jiang G, Bai Y, Chen T, Wang H. Understanding why neural networks generalize well through gsnr of parameters. *ArXiv* 2020, ArXiv: 200107384.
[111] Czarnowski I, Jędrzejowicz P. Stacking and Rotation-Based technique for machine learning classification with data reduction. In *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, Gdynia, Poland, July 3–5, 2017, pp. 55–60.
[112] Koutsoukas A, Monaghan KJ, Li X, Huan J. Deep-Learning: Investigating deep neural networks Hyper-Parameters and comparison of performance to shallow methods for modeling bioactivity data. *J. Cheminformatics*. 2017, 9: 1–13.
[113] Evgeniou T, Pontil M, Poggio T. Statistical learning theory: A primer. *Int. J. Comput. Vis.* 2000, 38: 9–13.
[114] Jang G-S, Chang D-I, Oh D-G. A new syndrome check based early stopping method for DVB-S2 LDPC decoding algorithm. *J. Satell. Inf. Commun.* 2011, 6(2): 78–83.
[115] Alqahtani FH, Alsulaiman FA. Is Image-Based CAPTCHA secure against attacks based on machine learning? An experimental study. *Comput. Secur.* 2020, 88: 101635.
[116] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J. Big Data* 2016, 3: 1–40.
[117] Tian Y, Zhao X, Huang W. Meta-Learning approaches for Learning-to-Learn in deep learning: A survey. *Neurocomputing* 2022, 494: 203–223.
[118] König G, Molnar C, Bischl B, Grosse-Wentrup M. Relative feature importance. In *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, January 10–15, 2021, pp 9318–9325.
[119] Sun X, Ren X, Ma S, Wei B, Li W, *et al.* Training simplification and model simplification for deep learning: A minimal effort back propagation method. *IEEE Trans. Knowl. Data Eng.* 2018, 32(2): 374–387.
[120] Messalas A, Aridas C, Kanellopoulos Y. Evaluating MASHAP as a faster alternative to LIME for Model-Agnostic machine learning interpretability. In *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, December 10–13, 2020, pp. 5777–5779.
[121] Pezoa R, Salinas L, Torres C. Explainability of high energy physics events classification using SHAP. In *Journal of Physics: Conference Series*, IOP Publishing, 2023, pp 012082.