Journal of
Materials Chemistry A

REVIEW

View Article Online
View Journal | View Issue

Cite this: *J. Mater. Chem. A*, 2024, **12**, 31074

AI and ML for selecting viable electrocatalysts: progress and perspectives

Anastasiia P. Dmitrieva, Anna S. Fomkina, Chantal T. Tracey, Elizaveta A. Romanenko, Ali Ayati, Pavel V. Krivoshapkin and Elena F. Krivoshapkina *

Received 18th July 2024
Accepted 10th October 2024

DOI: 10.1039/d4ta04991h

rsc.li/materials-a

The urgent need to address the current energy crisis and combat excess carbon dioxide (CO_2) in the atmosphere has emphasised the importance of transitioning to sustainable energy sources. Fossil fuel reliance has significantly contributed to global warming, underscoring the need for solutions that can mitigate CO_2 emissions effectively. One promising technology is electrolysis, which not only removes CO_2 from the atmosphere but also generates clean and renewable fuels with net-zero CO_2 emissions. By utilising a direct electric current to drive non-spontaneous oxidation and reduction reactions, electrolysis has the potential to convert CO_2 into more useful chemical species. Another electrochemical technique, water splitting focuses on producing molecular oxygen and hydrogen, offering an environmentally friendly alternative to traditional fossil fuel-dependent methods. However, electrolysis is often plagued by sluggish kinetics; thus, electrocatalysts are used to enhance reaction rates. Nanoscale electrocatalysts with complex structures have demonstrated excellent catalytic activity and selectivity for the desired products under certain reaction conditions; however, determining the most suitable shape and their corresponding reaction parameters can be an extremely tedious and time-consuming process. Recent studies have shown that artificial intelligence (AI) and machine learning (ML) can greatly simplify this process, saving both time and resources. This review highlights the role of AI and ML in optimising catalyst shape and reaction conditions for water splitting and CO_2 reduction, fostering the development of clean, efficient energy technologies for a sustainable future.

1. Introduction

Society is currently facing an energy crisis that raises legitimate concerns about whether conventional energy resources can meet human needs both today and in the foreseeable future. Historically, fossil fuels have served as the primary energy source; unfortunately, the combustion of these fuels not only generates substantial energy but also releases harmful CO_2, an important greenhouse gas that significantly contributes to climate change. Despite the alarming pace of global warming, modern society continues to rely heavily on fossil fuels for

---

Energy Lab, ITMO University, 9 Lomonosova Street, Saint Petersburg, 197101, Russian Federation. E-mail: kef@scamt-itmo.ru

Anastasiia P. Dmitrieva

Anastasiia Dmitrieva is a PhD student in ITMO University, Russia. Since 2021 she has been a junior researcher in the Laboratory of Applied Energy Materials. She explores the electrochemical transformation of carbon dioxide into sustainable fuels and chemicals, harnessing the power of nanomaterials. Her research into water splitting unveils pathways to clean hydrogen energy, merging innovation with environmental stewardship.

Anna S. Fomkina

Anna Fomkina is a Master's student in ITMO University in Chemistry. Her main area of interest is nanoparticles and semiconductor catalysts. Since 2021, she has been working on applying machine learning and artificial intelligence approaches to modelling and optimising catalytic processes and battery systems.

electricity generation, transportation, and heating due to their economic viability and reliability, continuing to increase atmospheric CO_2 levels. This situation underscores the urgent need to explore clean, renewable energy sources and to develop methods for capturing CO_2 emissions until mankind can completely transition away from fossil fuels to more sustainable energy sources. Emerging technologies, such as electrolysis, offer promising solutions to both challenges. Studies have shown that electrolysis can be used to effectively remove CO_2 from the atmosphere and generate clean, renewable fuels with net-zero emissions.[1,2] Electrolysis uses a direct electric current to speed up non-spontaneous reactions, including the conversion of harmful atmospheric CO_2 into useful chemical precursors like methane and ethanol, as well as the splitting of water into hydrogen fuel (H_2) and O_2. However, these reactions are often hindered by slow kinetics. As a result, heterogeneous electrocatalysts are often used to increase reaction rates. Nanostructured electrocatalysts have been shown to be particularly effective in increasing reaction rate; however, their performance heavily depends on their composition and structure as well as reaction conditions – the optimisation of which requires both substantial time and energy. The identification of highly-efficient, nontoxic electrocatalysts made from various materials is essential for achieving the desired efficiency of the production process. However, the number of viable catalysts that can be synthesised is enormous, far exceeding what can be investigated through laboratory experiments alone. Fortunately, over the past few years, artificial intelligence (AI), and machine learning (ML) have assisted researchers in identifying suitable catalysts and optimising their properties.

AI, which uses computers to simulate human intelligence, has been used to automate the high-throughput synthesis and characterisation of a variety of electrocatalysts in self-driving laboratories. This has facilitated the rapid generation and analysis of large datasets, accelerating the identification of effective catalysts.[3] A subset of AI, ML specifically focuses on developing algorithms that learn from the generated datasets to make predictions and decisions. The integration of ML into materials science could lead to significant changes in the theoretical frameworks and methods used in the development

Chantal T. Tracey

Chantal Tracey is a post-doc researcher at ITMO University where she specializes in the development of various materials for practical application. She received her bachelor's degree from the University of the West Indies, Mona; her master's degree from Peter the Great St. Petersburg Polytechnic University; and her PhD from ITMO University. She is particularly interested in the use of biomaterials and nanomaterials to create sustainable systems.

Pavel V. Krivoshapkin

Dr Pavel Krivoshapkin is a chemist focused on developing theories to control interactions between metal oxide nanoparticles and various materials. Since 2017, he has served as the Head of the Scientific and Educational Center for Chemical Engineering and Biotechnology at ITMO University. His work aims to enhance the understanding and application of metal oxide nanoparticles in various scientific fields.

Ali Ayati

Dr Ali Ayati, an associate professor at Quchan University of Technology in Iran, specializes in chemical engineering, focusing on nanotechnology, green chemistry, and adsorption processes. He earned his PhD from Ferdowsi University of Mashhad in 2013 and has served as a senior researcher at ITMO University in Russia. His research emphasizes innovative and sustainable approaches in materials science, particularly in wastewater treatment, catalysis, and advanced material design. Dr Ayati is involved in various collaborative projects and is dedicated to advancing interdisciplinary research that tackles global environmental challenges using cutting-edge technologies.

Elena F. Krivoshapkina

Dr Elena Krivoshapkina is a chemist with experience in various fields, including synthesis methods for metal nanoparticles and oxides, assembly of hybrid systems, and development of smart materials. She obtained her PhD in Chemistry in 2012 and earned the degree of Doctor of Chemical Sciences in 2022. In 2017, she joined ITMO University, where she currently serves as the Head of the Laboratory of Applied Materials for Energy Technologies. Her current research interests include the application of nanostructured materials in electrochemical synthesis and the study of sustainable processes for electrochemical CO_2 reduction.

of next-generation functional materials. Unlike traditional global structure search algorithms like CALYPSO,[4] Universal Structure Predictor: Evolutionary Xtallography (USPEX),[5] and Materials Data Facility (MDF),[6] ML can predict potential structures much more quickly and cost effectively, which in return leads to a vastly more efficient research process. ML relies on an iterative data-driven approach involving the construction of models that can be trained and tested on various datasets to identify optimal parameter sets that correlate multidimensional input data with desired outputs, making it indispensable in the identification of important physical and chemical principles. In terms of catalysis, these algorithms can specifically predict combinations of elements that yield catalysts with desired characteristics, such as enhanced activity, high stability, and good cost-effectiveness. For example, such algorithms have been used to optimise the various properties of high-entropy alloys (HEAs) for electrocatalytic applications and have found viable alternatives to expensive platinum-based catalysts used in electrochemical reactions.[7]

By combining computational models with experimental data, researchers can predict the catalytic activity of new materials without needing to conduct multiple experiments with slight variations. Using surrogate models to rationalise thermodynamic proxies and predict adsorption and activation energies, ML models can speculate on the behaviour of electrocatalysts based strictly on structural and compositional information, which allows researchers to focus on only the most promising candidates for further testing, significantly reducing the time and resources required for catalyst development.[3,8] AI and ML are also being used to elucidate the mechanisms of electrocatalytic reactions by analysing the kinetics and thermodynamics of electrocatalysis in order to identify patterns and correlations to inform the design of more efficient materials.[7,8] Recent advances in nanoparticle production have made it possible to achieve significant control over their shape at the nanoscale. However, selecting the optimal shape of the catalyst and determining suitable conditions for electrochemical reactions in order to achieve high yields of target products requires significant resources, such as time, reagents, and energy. In this context, a promising approach would be to utilise AI and ML techniques to determine optimal values for the factors that influence the reaction process. This would help streamline the process, reduce waste, and improve overall efficacy.

In this review, we examine the influence of ML and AI on the comprehension and advancement of heterogeneous electrocatalysts in CO_2RR and water splitting, drawing examples primarily from the latest literature. Our objective is to present an overview of the current strategies for accelerating the discovery and optimisation of electrocatalysts through close connection between computational models and experimental approaches. This review is divided into four sections. Section 1 introduces ML and AI and its applications in materials science. Section 2 discusses the process of utilising AI and ML techniques in electrochemical CO_2RR, including a brief overview of relevant ML algorithms and their advantages and limitations. Section 3 examines the literature on ML-driven development of electrocatalysts for the electrochemical water splitting reaction, including examples of their discovery, optimisation, and application on materials such as metals and oxides. Lastly, Section 4 presents conclusions on current challenges and future directions in catalysis based on ML approaches.

2. AI and ML in materials science

With the ability to learn from data and make predictions without relying on predefined rules or features, ML and AI have emerged as transformative technologies in the field of materials science (Fig. 1), paving the way for more efficient data-driven materials research and discovery, improving the efficiency of synthesis and materials characterisation using large language models, and facilitating experimental research by automating instrument control and data analysis.[9] AI has significantly transformed materials science by expanding the possibility of

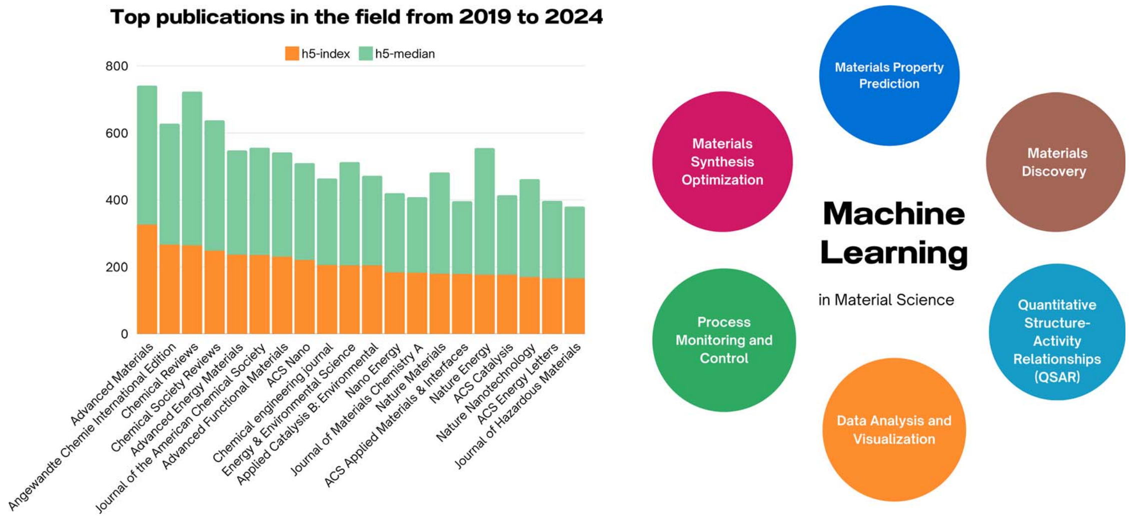

Fig. 1 An analysis of recent influential publications that utilise ML in the field of materials science.

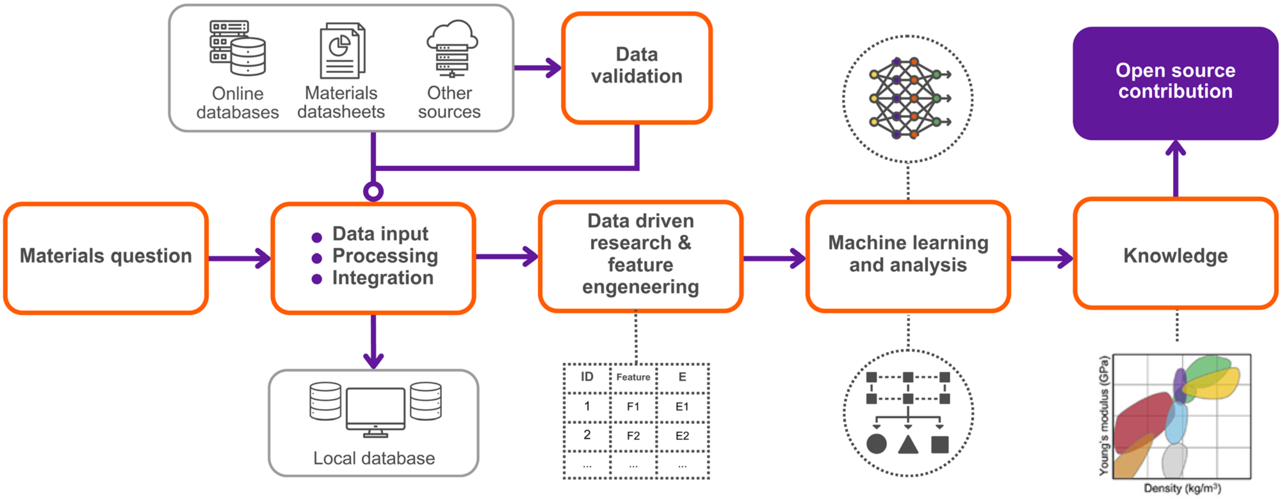

Fig. 2 A schematic of research on materials science using ML [reprinted with permission from ref. 12 Copyright 2020 Chemistry of Materials].

discovering and designing new materials using advanced computing technologies, such as deep learning. Deep learning approaches offer promising results, bypassing manual feature engineering and making accurate predictions with limited training samples, which eliminates reliance on serendipity and enhances the efficiency of materials research.[10] ML algorithms simulate material processing, providing insights into complex interactions, leading to process optimisation and new technique designs. ML has the potential to enhance productivity, achieve cost savings, and improve material properties through improved process parameters, quality control, energy efficiency, predictive maintenance, and process simulation across various sectors,[10] from the aerospace industry to soft robotics.[11]

Research has shown that ML can be divided into four main steps: data collection, feature engineering, model selection, and training and model evaluation (Fig. 2).[12,13] Data collection, an essential part of the ML process, uses various methods to gather information, including simulations, experiments, and the analysis of open-source databases. The expansion of materials science databases in recent years has facilitated easy access to large raw datasets.[14] Different databases on computed materials properties, such as Novel Materials Discovery, Open Quantum Materials Database, and Materials Project,[15,16] are available for open-access use. There are also several online databases that provide information on experimental materials structure, such as the Cambridge Structural Database and the Inorganic Crystal Structure Database.[17] There are also some valuable resource databases on nanomaterials like the Computational 2D Materials Database and Nanomaterial Registry.[18] Specialised databases may be useful for some specific applications, such as direct catalyst screening with calculated surface reactions (Table 1).[19] Despite strict quality assurance measures, databases may contain errors and inconsistencies due to human factors such as perception and measurement errors. Additionally, variability may arise when experiments are performed under different conditions across various laboratories or when properties are determined using different methods. Therefore, it is essential to examine and pre-process data sets prior to training ML models in order to avoid various errors, such as unrealistic, infinite or missing values, or mismatched formatted data.[13]

After data collection comes feature engineering, where the collected data is converted into a suitable mathematical form for training ML models. This involves describing the key features of molecules, compounds, and clusters on compound surfaces using a series of numerical values. These numerical values are attributes of the materials, which are known as descriptors, features, or fingerprints in the literature. The process of creating these mathematical representations of material objects is known as functionalisation or object creation, and it plays a crucial role in the quality and interpretability of the models constructed from them. This stage requires a significant amount of human input and expertise, but recent advancements in the field of automated molecular object characterisation using specific types of deep neural networks hold promise for paradigm shifts in this area. Effective descriptors should differentiate between objects in a given data space and capture features relevant to the simulated or predicted property. While there is a large number of possible descriptors that could be generated, they are selected depending on the scientific question being addressed. The set of descriptors should include unique information regarding the structure, composition, and physicochemical properties of materials, ensuring that each material has a distinct set. Although it is possible to generate a large number of descriptors, it is important to avoid overfitting by using a limited number of descriptors to ensure accurate model predictions. Redundant descriptors may include those that have a high degree of correlation with other features within the dataset as well as those whose values vary little across the dataset (low variability), or those that have little or no correlation to the target property being modelled. In addition to encoding structural, compositional, and physico-chemical properties of materials, information regarding the synthesis method of the catalyst and its post-treatment may also be included.

Table 1 Summary of databases on catalysis materials and their characteristics

Database	Description	Features and diversity of available data	Ref.
2D materials encyclopedia	The database contains 6351 materials (2940 – exfoliated, 3409 – chemically substituted)	Materials can be used as catalysts, catalytic supports or electron donors, demonstrating their versatility in catalytic applications. The growth conditions and physical characteristics of two-dimensional materials grown on different catalysts are described	20
AFLOW database	3 530 330 material compounds (734 308 640 calculated properties, and growing)	Crystal structures, including unique atomic decorations, compound compositions, and correspondences with existing prototypes	21
ASM alloy database	Composition, structure, properties, performance of engineering materials (metals and alloys)	ASM alloy centre – over 7500 data volumes ASM micrograph – 4500 data volumes ASM failure analysis – over 1200 data volumes ASM alloy phase diagram – over 40 300 data volumes ASM corrosion performance – 30 000 data volumes	22
Atomwork advance	Globally accessible database containing: crystal structures – 379 736, X-ray diffraction – 689 445, properties – 504 325, phase diagrams – 47 347	Data on crystal structure, X-ray diffraction, properties, and state diagrams of inorganic materials. It also includes a matrix function to assess dataset availability for binary material combos and an automatic charting function for material property variable graphs	23
Cambridge structural databases	Database containing 1 159 253 structural data	Three-dimensional structural data of molecules, containing at least carbon and hydrogen. It includes a wide range of organic, metal–organic, and organometallic molecules	24
ChemSpider	Integrated database of 114 million chemical structures from 270 sources	Nomenclature information, structural information, physicochemical properties, spectral data, safety information, synthetic methods	25
Crystallography open database	Open database containing 484 154 records of crystalline structures of organic, inorganic, and metal–organic compounds	Open-access collection of crystal structures. Inorganic, organic, metal–organic compounds, and minerals data. Exclusion of biopolymers. Data and software integration from CrystalEye	26
High-performance alloy database	Contains 171 alloys with 52 359 curves, primarily for industrial applications	Database includes composition, mechanical properties, thermal properties, and corrosion resistance data, making HPAD a valuable resource for industries requiring high-performance alloy information	27
Materials project	Offers open access to computed and predicted materials information. Contains 144 595 compounds, 530 243 porous materials, 76 240 band structures	MP offers a comprehensive platform for accessing materials data, DFT-based data, prediction tools catalyst/MOF	16
Open quantum materials database	Database containing 1 022 603 calculated thermodynamic and structural data	High-throughput DFT calculations of materials. Tools for screening prototypes and using ML for materials discovery. Availability of ML tools for materials science, such as magpi	28
Open catalyst 2022	62 331 density functional theory relaxations, which amounts to approximately 9 854 504 single point calculations	Dataset covers a wide range of oxide materials, coverages, and adsorbates, showcasing a diverse set of data for studying the oxygen evolution reaction and oxide electrocatalysts	29
NIST computational chemistry comparison and benchmark database	Experimental and computed thermochemical data for a selected set of 2186 gas-phase atoms and small molecules	Tools for comparing experimental and computational ideal-gas thermochemical properties, vibrational frequencies, rotational constants, electric dipole, electric quadrupole, polarizabilities	30

Experimentally determined descriptors may include temperature, pH, pressure, reaction time, and reagent concentration. Computational descriptors are preferred for modelling as they do not require experiments to be performed, reducing the likelihood of errors and repetition.

Model selection and ML training/evaluation are critical components of the ML workflow, each serving distinct yet complementary purposes. Model selection involves identifying the most suitable algorithm and model architecture for a specific task or dataset by evaluating various models based on performance metrics such as accuracy, precision, recall, and F_1 score, often utilizing techniques like cross-validation to ensure the model's robustness against unseen data. In contrast, ML training and evaluation focus on the actual learning process where the selected model is trained on data to recognize patterns and make predictions. Evaluation is conducted using a separate validation or test dataset to determine the model's effectiveness by selecting the optimal descriptors, and ensuring it does not overfit to the training data. Ultimately, while model selection aims to pinpoint the best candidate from a pool of options, training and evaluation are concerned with the practical application and performance assessment of that chosen model, both of which are essential for successful ML implementation. There are two main approaches to selecting descriptors: down-selection and dimensionality reduction. Down-selection, also known as sparse object selection, involves reducing a large number of potential material descriptors to a smaller set using a variety of statistical techniques. In regression models, a regularisation term known as L1 is often introduced to reduce terms that are less important to the model to zero. This process is referred to as the least absolute shrinkage and selection operator (LASSO). Tree-based algorithms, such as random forests (RF), are also frequently used to determine the significance of each descriptor following training. A good example of down-selection is the study conducted by Pande *et al.*, which involved sampling parameters for the Sacramento Soil Moisture Accounting Model (SAC-SMA) structure. A total of 500 parameter sets were sampled from specified ranges, which were crucial for estimating the model's complexity and performance.[31] The complexity of the SAC-SMA model was quantified based on the approach proposed by Arkesteijn and Pande.[31] This measure of complexity is derived from how a model simulation deviates from its expected value across multiple realisations of input forcing. Regularised RF models penalise complexity, allowing for better generalisation in predicting catalytic activity.

Another approach to feature optimisation is dimensionality reduction, which involves transforming features from a high-dimensional space into a lower-dimensional one. This process involves generating new features by linearly combining the original ones, with principal component analysis (PCA) being a commonly used method. PCA identifies the principal components, or sets of orthogonal vectors, which serve as new features with maximal data point values. This technique not only accelerates the training process of machine learning models after reducing the dimensionality, but also addresses the issue of high dimensionality, making it an effective strategy for feature optimisation. While PCA constructs linear models, the interconnections between structures and attributes are frequently non-linear in nature. Consequently, additional learning techniques have been devised for non-linear data dimension reduction in order to discern these intricate associations. The core of PCA relies on standard PCA to execute non-linear dimensional reduction by implementing a non-linear conversion to map data points into a multi-dimensional feature space, where conventional PCA can subsequently be employed. Multiple learning algorithms produce a low-dimensional projection that maintains the topological structure of the original high-dimensional dataset, enabling visualisation and modelling. The *t*-stochastic neighbour embedding algorithm, based on the *t*-distribution, is commonly used to visualise high-dimensional datasets in two or three dimensions. Nevertheless, the algorithm's high computational complexity and the variability in visualisations, which heavily depend on hyperparameter selection, restrict its practical application. The advanced Uniform Manifold Approximation and Projection (UMAP) algorithm is a powerful tool for reducing the dimensionality of data for visualisation and nonlinear dimensionality reduction. Compared to traditional methods like *t*SNE, UMAP significantly reduces the computational time while preserving the global structure of the data. This makes it particularly useful in catalyst screening and analysis of structure and properties. The development of inductive versions of UMAP, such as iGLoMAP, enables the algorithm to generalise to unseen data, which is beneficial for large-scale catalyst datasets (Fig. 3).[32]

To expedite the search for catalyst candidates in large combinatorial spaces, researchers have successfully combined ML with quantum mechanics methods.[33] ML approaches can reveal subtle details and patterns on the potential energy surface that might remain undetected in conventional visual data analysis. ML models significantly reduce the number of density functional theory calculations required to assess the activity of various catalysts. Feature engineering plays a significant role in optimising the performance of ML models.

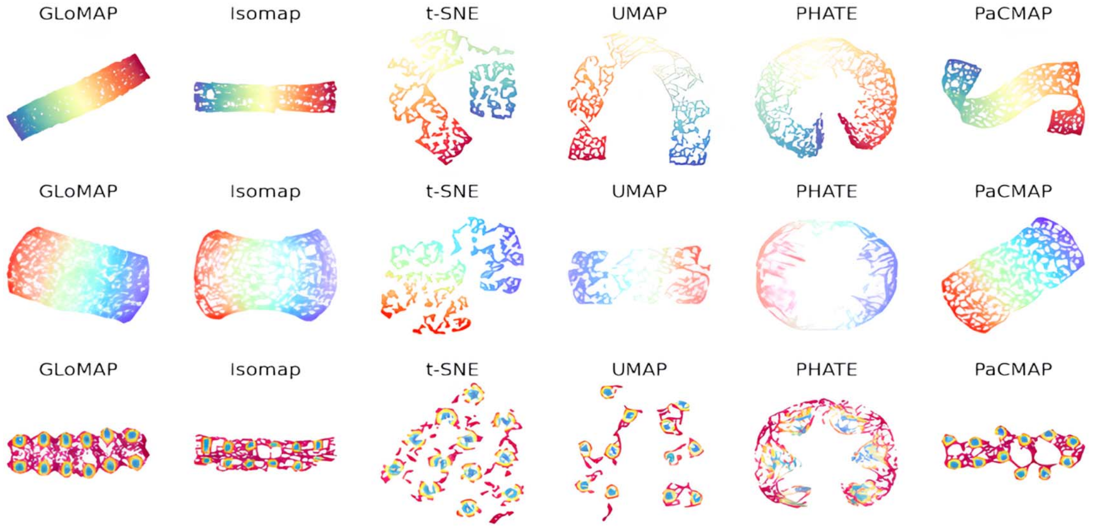

Fig. 3 The dimensional reduction for the *S*-curve, severed sphere, eggs dataset [reprinted with permission from ref. 32 Copyright 2024 *arXiv*].

Table 2 Common types of ML algorithms in materials science

Algorithm	Features	Advantages	Disadvantages	Application
k-nearest neighbors (kNN)	Does not require model training	Easy implementation, works well for small datasets	Sensitive to noise and outliers	Spectral analysis
Naive Bayes	Assumes independence of features	Effective for categorical features, quick training	Does not consider relationships between features	Determine the importance of descriptors
Gradient boosting	Sequential training of weak models	High accuracy, resistant to overfitting	Takes a lot of time to train	Materials screening, discovery, and property prediction
Neural networks	Data processing using parallel distributed learning	Ability to capture complex dependencies	Requires a lot of data for training, complex interpretation	Material design, discovery, and property prediction
K-means clustering	Divides data into a specified number of clusters	Easy to implement, good for identifying data groups	Sensitive to initial cluster centre selection	Spectral analysis
Logistic regression	Model for binary classification	Easily interpretable, relies only on linear relationships	Requires preprocessing of data	Property prediction
Multi-layer perceptron (MLP)	Consists of many neurons and layers	Handles nonlinear dependencies well	Requires tuning of optimal architecture, can overfit	Material design investigation of chemical reactions
Convolutional neural network (CNN)	Designed for image analysis	Effective in image recognition tasks	Requires large amounts of data, complex architecture tuning	Binding energies prediction Adsorption energy prediction
Support vector machine (SVM)	Effective in high dimensional spaces, works well with small data sets	Performs well in complicated domains with clear margin of separation	Not suitable for large datasets	Catalytic activity prediction and simplification of DFT calculations

High-level quantum chemical calculations provide accurate descriptors of reaction activity, but their use is limited by high computational complexity. ML methods offer an alternative approach for modelling catalyst reaction activity based on correlations between structural descriptors and reaction properties, reducing computational costs and providing rapid predictions of catalyst reaction capabilities.

To achieve high predictions and generalisation abilities, ML algorithms must be selected carefully. There are different types of ML algorithms in materials science with their own advantages and limitations (Table 2).[34] Typical algorithms used in electrocatalyst design, including classification, clustering, and regression, are well-reviewed in some articles.[35] Classification algorithms, such as descriptor-oriented machine learning and artificial neural networks, classify electrocatalysts based on their geometrical, electronic, and activity characteristics. This helps identify promising candidates for specific reactions and enables rapid screening of material properties.[36] Clustering techniques, such as data-driven clustering, group similar electrocatalysts based on performance metrics.[8] This helps to identify trends in electrocatalyst performance and optimise design strategies. Regression models, such as the Adaptive Neuro-Fuzzy Inference System (ANFIS),[37] combined with optimisation techniques such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) are used to predict adsorption energies, which are crucial for assessing the efficiency of electrocatalysts. Regression analysis is also used to establish quantitative structure–activity relationships, which helps predict electrocatalytic performance based on material properties.[38] A kernel-based algorithm is increasingly popular in material science, due to the ability to handle a variety of complex regression problems.

In one of the most promising applications in materials science, AI and ML have revolutionised the design and optimisation of catalytic systems. This approach drastically reduces the time and cost involved in catalyst discovery and enables the identification of novel catalysts that may have been overlooked using traditional methods.[39] This approach not only enables a substantial decrease in computational time when compared to traditional methods, but also allows for the analysis of complex datasets to extract meaningful insights.[40,41] Traditionally, designing electrocatalysts often relied on intuition and trial-and-error – a method that can be both costly and time-consuming. However, the integration of AI and ML has transformed this process by enabling researchers to identify meaningful descriptors or fingerprints of viable electrocatalysts. These technologies facilitate the formation of structure–activity relationship models that correlate compositional and structural information to electrocatalytic performance. Additionally, AI and ML explore uncharted areas of the chemical parameter space, rapidly suggesting elements and reaction conditions for innovative materials. This shift not only accelerates the discovery of new catalysts but also enhances the understanding of fundamental mechanisms, paving the way for more efficient energy conversion and storage solutions. For example, Zhang *et al.* used AI and ML to design efficient bimetallic catalysts by developing a comprehensive ML framework using *ab initio* datasets and descriptor-based kinetic analysis derived from quantum-chemical simulations. By establishing a catalyst database and optimising artificial neural networks, they successfully predicted complex adsorbate/metal interactions for the selective catalytic reduction of nitrogen oxides on Mn–Cr couplings. The predicted efficiency was experimentally validated, showing a significant increase in catalytic activity.[42] The potential applications of AI and ML extends beyond catalyst identification; these technologies show the potential to revolutionise the optimisation process itself. By using advanced algorithms that analyse vast datasets derived from previous research, researchers can not only predict performance outcomes, but also refine synthesis parameters in real-time, minimising resource expenditure and maximising efficiency.[43] This dynamic approach allows for a more nuanced understanding of how different variables interact in electrocatalytic systems, ultimately leading to the development of catalysts tailored for specific applications, such as CO_2RR or hydrogen production.[44] ML can also be used to inverse-design material structures, providing a cost-efficient alternative to traditional algorithms such as USPEX and CALYPSO.[45,46] As these methodologies mature, a paradigm shift may occur where the design and deployment of nanostructured electrocatalysts become increasingly more automated, further accelerating advancements in sustainable energy technologies.

3. Electrochemical conversion of carbon dioxide

Methods for the conversion of CO_2 involve a variety of advanced techniques designed to facilitate its transformation into economically valuable products or fuels, thereby aiding in the mitigation of greenhouse gas emissions and the repurposing of pollutants as resources. One notable methodology is the photoelectrochemical (PEC) conversion of CO_2, which uses solar energy to convert the harmful greenhouse gas into formic acid and carbon monoxide (Fig. 4a). Another beneficial strategy within the chemical industry for the conversion of CO_2 is the electrochemical carbon dioxide reduction reaction (CO_2RR), which has the notable advantage of converting CO_2 into valuable organic compounds that can serve as essential precursors for a wide range of applications. CO_2RR yields a much wider range of products than PEC, including valuable chemicals and fuels, such as methane, ethanol, and several hydrocarbons (Fig. 4b), making it more suitable for various applications.

The CO_2RR process is governed by a multi-step-based coordination chemistry comprising 2, 6, 8, 10 and 12 electrons for the formation of the common products CO, CH_3OH, CH_4, C_2H_4, and C_2H_5OH, respectively (Fig. 4b and Table 3). The reaction typically comprises four critical stages: (I) chemisorption of CO_2 on the surface of the electrocatalyst; (II) breakage of the C–O bond or formation of C–H bonds through the transfer of electrons and/or protons and other reduced species; (III) C–C bond generation; and (IV) reorganisation of the product configuration to allow desorption from the surface of the electrocatalyst and diffusion into the electrolyte solution. Step III is the most challenging step, in which different C_{2+} hydrocarbons and oxygenates are formed. Below, Fig. 5 depicts the formation of formic acid as an example of the stages involved in CO_2RR. Stages of CO_2RR consist of (1) chemisorption of CO_2 on the surface of the electrocatalyst, (2) breakage of the C–O bond or formation of C–H bonds through the transfer of electrons and/or protons and other reduced species, and (3) reorganisation of the product configuration to allow (4) desorption from the surface of the electrocatalyst and diffusion into the electrolyte solution. Before 3 step, the step at which C–C bond generation occurs. This is the most challenging step, to form different C_{2+} hydrocarbons and oxygenates.

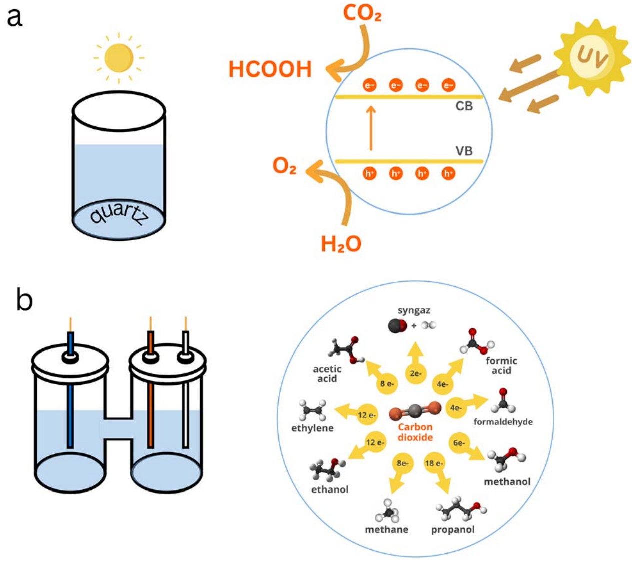

Fig. 4 (a) Photo- and (b) electroconversion of carbon dioxide into valuable chemicals.

The typical setup for CO_2 electroreduction includes a CO_2 electrolyser consisting of two chambers with distinct electrodes separated by an ion conducting membrane (Fig. 4b). As the process is nonspontaneous, an external voltage is applied to initiate the reaction. The CO_2 molecule contains fully oxidised carbon and has a non-polar structure with a high bond energy, resulting in its low reactivity and a long lifetime in the atmosphere. Due to these structural characteristics, significant technological endeavours are required to carry out electrochemical CO_2RR such as high energy (750 kJ per mole to break the C=O bonds), accompanied by the need for applying large overpotentials to obtain certain products. As a result, the precondition for the subsequent electrochemical processes and the inhibition of competing hydrogen evolution is the ability of CO_2 to chemisorb in the initial step. Thermodynamically, there is a negligible difference between the equilibrium potentials for the reduction of CO_2 to various products and the hydrogen evolution reaction (HER) (Table 3). This is valid under conditions of a pH of 7, in aqueous electrolyte solutions, at atmospheric pressure, 25 °C, and with 1 M concentration of electrolyte. Therefore, HER appears as a significant by-product

Table 3 Electrochemical reactions at thermodynamic equilibrium potentials in an aqueous electrolyte at pH 7, 1 atmosphere of pressure, and 25 °C, referenced to the RHE

Chemical reaction	$E°/V$ *vs.* RHE	Product name
$CO_2 + 2H^+ + 2e^- \rightarrow CO(g) + H_2O$	−0.11	Carbon monoxide (CO)
$CO_2 + 2H^+ + 2e^- \rightarrow HCOOH(aq)$	−0.12	Formic acid (HCOOH)
$CO_2 + 6H^+ + 6e^- \rightarrow CH_3OH(aq) + H_2O$	0.03	Methanol (CH_3OH)
$CO_2 + 8H^+ + 8e^- \rightarrow CH_4(g) + 2H_2O$	0.17	Methane (CH_4)
$2CO_2 + 10H^+ + 10e^- \rightarrow CH_3CHO(aq) + 3H_2O$	0.06	Acetaldehyde (CH_3CHO)
$2CO_2 + 12H^+ + 12e^- \rightarrow C_2H_4(aq) + 4H_2O$	0.08	Ethylene (C_2H_4)
$2CO_2 + 12H^+ + 12e^- \rightarrow C_2H_5OH(aq) + 3H_2O$	0.09	Ethanol (C_2H_5OH)
$2CO_2 + 14H^+ + 14e^- \rightarrow C_2H_6(aq) + 4H_2O$	0.14	Ethane (C_2H_6)
$2CO_2 + 16H^+ + 16e^- \rightarrow C_2H_5CHO(aq) + 5H_2O$	0.09	Propionaldehyde (C_3H_6O)
$3CO_2 + 18H^+ + 18e^- \rightarrow C_3H_7OH(aq) + 5H_2O$	0.10	Propanol (C_3H_7OH)
$2H^+ + 2e^- \rightarrow H_2$	0.00	HER
$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$	1.23	OER

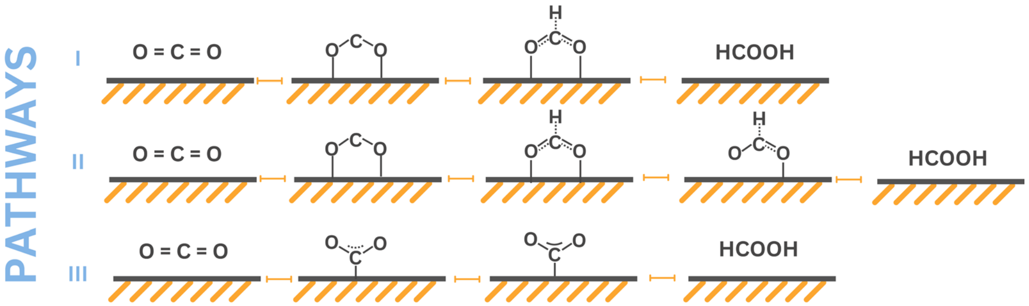

Fig. 5 Stages of CO_2RR to formic acid: 3 different pathways of reaction. In pathway I, CO_2 is adsorbed onto the catalyst and transformed into a bent intermediate, followed by hydrogenation to form formic acid directly. In pathway II, CO_2 is adsorbed and reduced to form a formate species, which is then protonated and desorbed as formic acid. Pathway III involves the adsorption of CO_2, reduction to a carboxyl intermediate, and further protonation to produce formic acid.

due to its similar thermodynamic potential compared to that of CO_2RR products.

Similarly, the difficulty in achieving a specific product from CO_2RR arises from the small differences in thermodynamic potentials between different products, as shown in Table 3. In contrast, the actual potentials required to drive reactions are significantly more negative than thermodynamic equilibrium potentials (overpotentials). Density functional theory (DFT) studies have been widely used to provide atomistic insights into the underlying mechanisms of CO_2RR to C_1 and C_{2+} products.[47–50] Some of the key rate-determining steps include *CO*, *COH*, *CHO*, *COCO*, *CHCHO*, *COCOH*, *OCCO*, HOCCH* and HOCHCH* (where the superscript denotes a covalently bound intermediate on the catalyst surface). The factors that regulate their long-term functionality and the reaction kinetics that lead to their generation and modification should be elucidated in order to aid the design of rational catalysts for the selective reduction of CO_2 to C_{2+} products.

As stated earlier, electrochemical reactions, including CO_2RR, suffer from sluggish kinetics; therefore, they are sped up using electrocatalysts. The efficiency of the catalyst and comparative assessment for CO_2RR developed in a manner that is most suitable by defining a number of significant parameters that allow for an accurate assessment of catalyst design strategies. One of the most significant parameters is the Faraday efficiency (FE) coefficient, which is defined as the proportion of the total current used to produce a specific product. An increase in FE content to a predefined level indicates improved selectivity and can be calculated using the following equation:

$$\mathrm{FE} = anFq \times 100\% \tag{1}$$

where a, n, F, and q are the number of electrons transferred, number of moles of desired product, Faraday constant (96 485 C mol^{-1}), and total charge transferred, respectively.

Another important parameter for determining catalytic efficiency is current density or partial current density. This is calculated by dividing the recorded current value by the geometric surface area of the electrode. The current density has a direct relationship with the reaction rate, as it is a measure of how fast the reaction is taking place. The catalytic activity of a given product can be determined by the partial current density, which is calculated by multiplying the total current density by the fraction of the current that is contributed by that product. Both the current density and the partial current density depend on the amount of catalyst used and the specific surface area available for reaction.

An important parameter defining the electrochemical process is energy efficiency, which equals the amount of energy used to produce a product and calculated using the following formula:

$$\mathrm{EE} = E^{\circ} \times g \times \mathrm{FE} \tag{2}$$

where EE is energy efficiency, E° is the equilibrium potential, g is the overpotential, and FE is the Faraday efficiency.

The equilibrium potential is the voltage necessary to achieve a specific level of product formation, while the overpotential is the difference between this equilibrium potential and the actual applied potential. The onset potential, typically lower than the CO_2 conversion equilibrium potential, ensures substantial product formation. Overpotential is crucial for evaluating electrocatalyst performance, defined as the difference between the applied potential and the theoretical potential needed for complete reaction. Catalysts with a lower overpotential generally exhibit better performance. The Tafel plot, which graphs current density against overvoltage, is instrumental in assessing catalyst performance. A steeper Tafel slope indicates lower performance, whereas a flatter slope suggests superior efficiency.

Manual control in CO_2RR catalytic reactions limits efficiency and accuracy, leading to suboptimal outcomes. Overcoming these challenges requires solutions that advance the field. Automation and optimization are key to streamlining processes and enhancing selectivity. Harnessing technology and innovative approaches can lead to a more efficient and precise future in catalysis. Collaboration among researchers, scientists, and engineers is crucial to push boundaries, realize the full potential of catalytic reactions, and unlock new applications. Continuous exploration and development can overcome manual control limitations and revolutionise catalysis.

3.1. AI/ML in CO_2 electroconversion

The integration of ML and AI in electrocatalysis has been transforming the way researchers approach the discovery, design, and optimisation of electrocatalysts. ML has significantly advanced the development of catalysts for converting CO_2 into valuable C_{2+} products. This approach enhances the efficiency and selectivity of electrocatalysts, addressing the challenges posed by traditional methods. In this part of comprehensive overview examines on date the most efficient electrocatalysts (such as Cu) where ML and AI play significant roles, categorised by various project groups or conditions.

3.1.1. Copper-based electrocatalysts. To date, a plethora of experimental investigations elucidating the exceptional selectivity of Cu in the CO_2RR have been reported.[2,51–53] Nonetheless, these findings have not yet brought us any closer to unravelling the fundamental mechanisms underlying this phenomenon. The end products resulting from the copper electrode are a complex mixture of carbon compounds whose structural characteristics are significantly influenced by the applied potential, due to the various mechanisms of intermediate stabilisation. Initially, it was thought that only metallic copper (Cu^0) exhibited catalytic activity, regardless of the source material. However, research has shown that a certain degree of oxidation in copper particles can increase the efficiency of CO_2 conversion to C_{2+} products.[54] Oxide electrodes have demonstrated that oxygen atoms on the surface of copper oxide and Cu^+ oxide sites promote the binding of molecules and modulate the reactivity, significantly affecting the distribution of products.[55] A team of researchers led by Xu has demonstrated the photoreduction of Cu(II) atoms into Cu(I), and then further into Cu(0). This process effectively promotes the creation of methane (CH_4) from carbon dioxide (CO_2).[56] Copper-based catalysts possess a moderate binding energy for important intermediates, such as carbon monoxide (CO), which contributes to the generation of subsequent crucial intermediates and ultimately C_{2+} products. The study revealed substantial differences in the selectivity for producing more than 16 distinct polycrystalline copper compounds, 12 of which are C_2 or C_3 products.[57] This highlights the complexity of the reaction and facilitates the potential use of CO_2 as a starting material in various chemical syntheses. However, the exact mechanism of the CO_2 reduction reaction on copper surfaces remains unclear. The factors that influence the structural evolution and catalytic properties of electrocatalysts can be divided into external and internal factors. External factors include the applied potential, which can lead to changes in the composition and structure of the catalyst due to an increased likelihood of exceeding the redox potential of the active substances under real operating conditions. Internal factors are related to the specific properties of the catalyst material, such as its chemical composition and structural features. On-site characterization techniques can be used to assess the surface reconstruction and selectivity of the electrocatalyst. These techniques allow for a more accurate understanding of the behaviour of the catalyst under specific conditions, which is essential for optimising its performance in electrochemical applications. It is also worth noting that the pH of the electrolyte used can influence the redox potential of electrocatalytic components, which in turn affects the activity and selectivity of electrochemical reactions. Internal factors, such as the physicochemical properties of catalysts produced by different synthesis methods, can significantly impact the performance of electrocatalysts. However, current analysis techniques, such as XANES, can only collect data within milliseconds or seconds when dynamic structural changes occur within picoseconds or nanoseconds. Therefore, it is essential to develop a novel method for ultrafast data acquisition that can instantly detect intermediates and track structural modifications during electrocatalysis in real time.

In this context, ML and AI emerge as invaluable resources. While they cannot be considered a universal solution, when combined with other computational and prognostic instruments, they can offer profound insights into the intricate interplay of internal and external elements that impact catalytic activity. An example in the same vein is computational screening, which began with the observation that copper appears to be an essential building block of an effective catalyst for the CO_2RR process. Therefore, only intermetallic compounds containing copper were investigated, which still accounts for more than 200 candidates in the Materials Project database. The generation of probable surfaces and the enumeration of potential active sites for CO adsorption as a proxy server led to more than 200 000 active site motifs. To explore this vast space, an active learning algorithm was applied to limit the DFT calculation to only the most important area. However, about 4000 DFT calculations were performed. From the analysis of the most promising active sites, it became clear that the intermetallic compounds Al–Cu seem very promising. Therefore, appropriate validation experiments were conducted, which demonstrated a significant increase (from 35% to 60%) in the effectiveness of ethylene compared to pure copper, which confirms the correctness of the screening strategy.[58]

In this case scientists used DFT as a well-established tool for atomistic understanding of electrocatalytic systems, but it comes with a fairly high computational cost.[59] This motivates the development and use of more efficient methods, especially considering the size of the electrocatalytic interface (thousands of atoms), its dynamics (at least nanoseconds) and even the huge number of catalysts that would like to be computationally evaluated. In this context, ML is currently best viewed as a way to construct system-specific "force fields" – mathematical functions that derive the energy of a system as a function of the positions and nature of atoms. These functions are usually called machine-trained potentials (MLP), and they are many orders of magnitude faster than DFT. Other levels of theory can be used instead of DFT, if the accumulation of sufficient training data is feasible. An example is ANI-1ccx (described earlier), which achieves the accuracy of single, double and perturbative triples of almost connected clusters for organic molecules of a neural network, and for the condensed phase, the random phase approximation was used to go beyond the accuracy of DFT using MLP.[60,61] Regardless of the MLP architecture, MLP training is the most time-consuming for the user: since MLPs are system-specific, a special training set must be created for each application. The size of the training set is on the order of 10^{3-4} DFT energy estimates, and the quality of the MLP depends at least as much on the representative diversity of the training set as on the MLP architecture: geometries that deviate too much from the training set will have completely incorrect energies and forces. In practice, currently the most popular software for parameterization and use of MLP in electrocatalysis are LASP and DeepMD.[62,63] The advantage of the former is the presence of a large set of predefined MLPs, while the latter has very powerful active learning capabilities, which allows you to effectively parameterize system-specific MLPs. After learning MLP, one can utilise the standard set of atomic modelling techniques. A striking example of MLP application is the identification of realistic surface structures for catalysts that are not completely crystalline. For example, the restoration of the surface of a copper oxide-based catalyst exposes areas of metallic copper that are not completely smooth. This was proved using MLP simulations conducted using LASP for several nanoseconds, which aimed to reproduce an experimental protocol where the system undergoes stepwise recovery reactions.[64] It has been shown that these defect-rich surfaces have different active sites with contrast selectivity for CO_2 electrical recovery products. It is assumed that the active sites with square steps are responsible for alcohol products, while flat and convex-square active sites are more favourable for the production of ethylene.

To achieve a catalyst with desired selectivity and activity, researchers must be able to predict the product of CO_2 reduction on a material's surface by examining intermediate reaction pathways and the kinetics of competing hydrogen evolution and undesired methane production. Many groups have used *ab initio*-based methods to study the reaction pathways of CO_2RR catalysts, such as Nørskov and Studt *et al.*[65] have done study on Cu(100) surfaces and Sn based materials. The computational hydrogen electrode (CHE) model was applied to the CO_2RR by examining a network involving 41 different intermediate steps on the Cu (211) surface. From this pathway, numerous routes to the major products CO, H_2, HCOOH, CH_4 and C_2H_4 are possible and the lowest energy pathways were found as a function of applied potential. Results show CO_2RR over a stepped Cu(211) facet as a model catalyst with one proton–electron pair transfer on the *x*-axis. The free-energy diagrams at 0 V show an uphill process that turns into a downhill process, at larger cathodic potentials, that indicates the ease of intermediate formation at higher negative potentials. In addition, the lowest energy pathway is obtained for the C_1 product formation (mainly HCOOH and CO) rather than C_{2+} products. The carboxyl (*COOH) species forms *via* adsorption of proton–electron pairs that then reacts with another proton–electron pair to yield HCOOH while this work is insightful, it does not perfectly simulate the electrocatalyst in an experimental setting, requires a good understanding of computational chemistry methods. Authors in work[66] take a similar approach (RBPE) and compare it with BEEF-vdW functional which was specifically designed to address vdW forces reasonably well while maintaining an accurate description of chemisorption energies of molecules on surfaces. It was noted that the outcome of DFT calculations can be highly dependent on the functional used. The RPBE functional predicts that the transition state for CO hydrogenation is 0.44 eV higher in free energy compared to CO_2 hydrogenation, with the highest barriers for HCOOH and H_2CO being 1.84 and 1.90 eV, respectively, suggesting a significant difference in reaction rates of about 4–5 orders of magnitude. This finding contradicts experimental evidence indicating that CO_2 is the primary carbon source in methanol production over Cu-based catalysts. In contrast, the BEEF-vdW functional shows that intermediates in CO_2 hydrogenation interact more strongly with the Cu(211) surface, resulting in comparable highest free energy states for CO and CO_2 hydrogenation (1.53 and 1.56 eV, respectively). The presence of ZnO is expected to enhance the rate of CO_2 hydrogenation significantly, making it faster than CO hydrogenation in the Cu/ZnO system. This work is a good start to bridging the gap between experimental work and theoretical simulations, but it is limited to predictions on known catalysts. With development AI/ML methodologies, it is possible to automatically compare a targeted CO_2RR product to the known experimental data and theoretical predictions on an array of numerous materials, and to predict the probability of finding a good CO_2RR product selectivity at the early stage of catalyst development on to a material which has not been tested.

In work[67] researchers studied the conversion of CO_2 to C_2-type products in order to improve the efficiency of the process. They found that a better understanding of the reaction mechanisms and optimal catalytic surfaces was needed to achieve better results. By jointly applying density functional and ML methods, the researchers analysed the adsorption of *COCOCOH on eight different bimetallic copper catalytic alloys and identified ideal surfaces for catalytic activity, through configuration space learning. Eight different ML models considering descriptors such as element period, group, electronegativity and number of unpaired d-electrons were used for analysis. The best models successfully predicted the adsorption energy of *COCOCOH on copper-based bimetallic alloys with MAE less than 0.095 eV. The most accurate models proved to be Cu/Ag and Cu/Au based BACS with nano-islands of 2–3 atoms on the surface and high silver/gold concentration in the subsurface having the most favourable energy reaction line corresponding to the lowest *COCOCOH adsorption energy.

In another paper,[68] the researchers report on the stage process of using ML to find and optimise additives in the preparation of Cu catalysts for CO_2RR. The authors explored the reaction of electrochemical copper deposition as a technique for producing a copper catalyst using various metal salts and organic molecules as additives. They used a three-step learning process involving several algorithms, including decision trees, random forests, logistic regression, and XGBoost, to select the optimal catalyst composition. In the initial stage, a dataset comprising 12 metal salts and 200 water-soluble organic compounds was created. This allowed for the generation of 2000 potential combinations. The analysis revealed that the metal additive Sn plays a significant role in enhancing FE-CO, while the aliphatic OH functional group in organic additives is crucial for increasing FE-C_{2+}. Ligands containing nitrogen heteroaromatic rings, carboxyl groups, and aliphatics amino acids were selected to achieve high levels of FE-CO and FE-HCOOH, respectively. Additionally, it was observed that at higher local pH values most catalysts exhibit a higher ratio of FE-HCOOOH to FE-CO. This is consistent with the notion that the underlying environment is not conducive to breaking the C–O bond. This was first discovered in the Sn/Cu system. It was concluded that aliphatic hydroxyl groups contribute to Faraday efficiency in C_{2+} product formation, potentially by controlling Cu_2O formation in the catalyst precursor. The Cu_2O phase in the bimetallic Sn/Cu system is also essential for CO/HCOOH production. The authors have demonstrated a unique application of ML to accelerate the discovery of additives for highly efficient CO_2 reduction catalysts.[68]

3.1.2. Metal–organic frameworks (MOFs). In the issue of highly efficient electrochemical processing of CO_2, its sorption and storage play an important role. In this vein, there are also methods to optimise this process using ML and AI methods. Organometallic frameworks, also known as MOFs, are a type of porous crystalline material that consists of a metal assembly and organic linker. These materials have attracted significant research interest due to their unique properties, such as high porosity and a large specific surface area. The porosity and surface area of MOFs make them ideal for gas adsorption and storage applications. They have abundant adsorption sites, which can be used to capture CO_2 molecules. MOFs with a larger surface area and pore volume provide more adsorption sites for CO_2, making them more effective at absorbing the gas. However, the choice of MOF for CO_2 capture depends on the specific requirements of the application. Different types of MOFs have varying physical and chemical properties due to differences in metal assemblies, linkers, and topologies. This makes identifying the most suitable MOF for a given application a challenging task.

Elucidating the pore chemistry and molecular structure of metal–organic frameworks (MOFs) in relation to CO_2 adsorption at various pressures is crucial for effectively screening MOF molecules. This process is particularly important because the development of MOF materials is complex, lengthy, and often expensive. Therefore, efficient screening methods are essential to streamline the identification of promising candidates for CO_2 capture. The Grand Canonical Monte Carlo (GCMC) simulation has emerged as a powerful tool in this context. It primarily utilises van der Waals and Coulomb forces to simulate interatomic potentials, making it effective for evaluating the gas adsorption capabilities of MOFs. Numerous studies have successfully employed GCMC to predict the adsorption of gases such as CO_2, H_2, and CH_4, yielding promising results.[69] However, it is important to note that gas adsorption simulations often require manual intervention and substantial computing resources, which can complicate the screening process. This reliance on computational power can be a barrier to rapid development and testing of new MOF materials.

Over the past decade, many high-quality large-scale MOF datasets have appeared, such as the Computing-Ready Experimental (CoRE) database, MOF31, the Quantum MOF (QMOF) dataset and the hypothetical metal dataset-organic framework (hMOF). Most studies have considered the structural information of MOF molecules as descriptors, such as the proportion of voids and the diameter of the smallest cavity.[70–72] Some studies have added molecular descriptors to identify chemical reactions of the adsorption process, such as atomic species and functional groups. The researchers proposed a MO model to predict the properties of CO_2 adsorption in MOF without calculation and modelling. Teng and Shan implemented a plethora of computational attributes and trained them in conjunction with structural and molecular characteristics on the hMOF dataset, with the goal of enhancing the precision of CO_2 adsorption forecasts at various pressures.[73] The incorporation of design attributes boosted the accuracy of the model by 15–20% across a range of pressures, providing a deeper comprehension of the operational mechanism of the model. Visualisation of feature significance revealed that during CO_2 absorption, the structure of the MOF molecule and pore size assume a dominant role, with computational attributes such as molecular polarizability, electronegativity, and atomic mass also contributing significantly. This ML model can be interpreted through the lens of pore chemistry, molecular dynamics, and chemical interactions, enabling rapid and precise prediction of CO_2 sorption by MOFs on a large scale.

3.1.3. Noble metal-free catalysts.

Noble metal-free catalysts are crucial for CO_2RR due to their cost-effectiveness, as they rely on abundant and inexpensive materials rather than scarce noble metals like platinum and gold. This not only reduces overall costs but also aligns with sustainability goals, enabling the development of environmentally friendly technologies. Recent advancements demonstrate that these catalysts can achieve comparable or even superior activity and selectivity compared to noble metals, while often exhibiting enhanced durability and stability under operational conditions. Moreover, the versatility in design of noble metal-free catalysts allows for innovative approaches to optimize their performance for specific reaction conditions, making them a promising alternative for efficient CO_2 utilization in the fight against climate change.

The primary output of CO_2RR is carbon monoxide (CO), which is generated with maximal efficiency on catalysts based on noble metals. The research team has described a fundamentally novel mechanism for the production of CO through the process of catalytic hydrogenation of carbon dioxide *via* the reverse reaction of the RWGS, paving the way for innovative technologies in the synthesis of methane, methanol, dimethyl ether, and hydrocarbons. Copper (Cu), gold (Au), silver (Ag), and molybdenum (Mo) have been identified as particularly effective in promoting high levels of CO or methanol production. *In situ* generation of Cu–Al spinel surfaces has been observed to be highly stable and exhibit excellent catalytic activity at high volumetric flow rates. Density-functional theory (DFT) calculations and diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) have demonstrated the enhancement of larger active sites due to the presence of Cu_{2+} vacancy sites and oxygen on the catalyst surface.[74] In work,[67] a multilayer perceptron neural network demonstrated high accuracy in predicting the adsorption energy of COCOH over a wide range of values (average absolute error of 0.095 eV). ML models in this study relied heavily on the number of unpaired d-orbital electrons, particularly for Os, Ir, Ru, and Rh. The correlation of descriptors to adsorption energy (E_{ads}) highlighted the significant influence of the number of unpaired d-orbital electrons. The neural network model demonstrated outstanding accuracy and the ability to predict outliers, making it a cost-effective and efficient alternative to DFT calculations for predicting adsorption energy. This model can be applied to other systems with minimal pretraining, serving as a versatile tool in the early stages of catalyst design. While the ALIGNN model showed slightly better performance, the descriptor-based neural network captured most of the catalytic dynamics derived from the chemical properties of material systems. The inability of graph neural networks to comprehensively study complex material properties led to the exclusion of the ALIGNN model in this study.

ML has been combined with DFT to predict the presence of metal-oxo intermediates in the reaction pathway for 2100 CO_2RR catalysts to accelerate CO_2RR catalyst discovery.[75] 2D visual representation of the catalysts using principal component analysis and the *t*-distributed stochastic neighbour embedding approach were used, and both approaches revealed a strong dependence of the reaction pathways on the often-cited descriptor for CO_2RR, $\Delta G_{CO_{2-}}$.[76] The approach, intended for functional integration with other ML models, highlights the capability of ML in helping to understand a wide range of catalyst metrics and their relationship with reactivity under bifurcations in the CO_2RR reaction network. The final model had a 0.817 correlation coefficient, and a root mean square error of 0.056 eV on the training data and a 0.257 validation error on the training data and a 0.151 validation error on an external quasi-training set with 45 data points beyond the training set. This model has thus developed a QSPR machine that can predict the oxygen binding energies of Fe-based molecular catalysts at a credible validation level, which is equivalent to 5–6 kcal mol^{-1} for CO_2 reduction. No other computational model of this size has achieved such validation with the ~0.1 eV oxygen functionality compared to adsorption data. Despite the promise of investigating the reaction mechanism using conventional base metal catalysts on an atomic level, there is a lack of studies that combine ML and AI applications with experimental validation. Nevertheless, a significant portion of this type of research focuses on examining the nature of reactions occurring on single atom catalysts.

Single atom catalysts (SACS) are a promising type of non-noble metal catalysts. SACS have been created and actively developed in the last decade. The concept of efficacy of this group of catalysts is that reducing the size of catalysts increases the number of active atoms and accelerates their use, improving catalytic properties and reducing cost. A large number of experimental and theoretical studies have shown that isolated single atoms, especially metal atoms embedded or loaded into the corresponding bases, can effectively catalyse CO_2RR.[77–79] In another study, the authors propose design principles for selective carbon production of CO at high current densities in acidic electrolytes. These principles are based on the mass transfer of CO_2 and protons (H^+) and interfacial electrokinetics. Their theoretical and experimental study showed that the CO_2 reduction process in a gas-fed flow cell has carbon selectivity over hydrogen (H_2) due to the acidic proton reduction reaction. However, the water (H_2O) reduction reaction reduces the carbon selectivity, which can be effectively reduced by using catalysts with higher activity towards CO. Single-atom nickel catalysts deposited on carbon nanotubes (Ni-SAC-CNT) were used for the study, which showed a high CO selectivity of more than 90% at a current density of -100 mA cm^{-2} and pH 2.0. Weakly hydrated cations in electrolytes were also found to contribute to the generation of high carbon selectivity in acids by stimulating the CO_2 reduction reaction and suppressing the electrokinetics of HER. In addition, a systematic study of the carbon balance during CO_2 electrolysis showed that the efficient formation of CO from Ni-SAC-CNTs leads to the highest carbon selectivity of the process, reducing carbonate formation and achieving approximately 77% product selectivity factor for CO_2 in CO without further adjustment of the chemical environment.[80] However, SAC have limitations in the synthesis of multielectron products, such as methanol and methane, due to their single active site, which makes optimising the bond strength of intermediates challenging. To address this issue, double metal-site catalysts (DMSC) have been developed. These catalysts exhibit a higher metal atom loading and more functional active sites, resulting in improved interaction between reagents and catalytically active centres.

However, there are several challenges associated with DMSC research, including the difficulty in screening potential catalysts and identifying the factors that determine their catalytic activity. ML techniques have gained significant attention for their potential in catalyst discovery and design. ML can be employed to screen a vast chemical space for potential catalysts, helping researchers focus on materials that exhibit better performance. Integrated algorithms for processing training data, based on theoretical calculations, make it possible to create predictive models and significantly reduce computational costs. By introducing ML methods into DFT calculations, it is possible to effectively investigate the potential of CO_2RR catalysis using a density matrix (DMSC). The substrates used for SAC and DMSC systems for CO_2RR are varied. However, the majority of these studies have relied on carbon supports for anchoring metal atoms.

In this research, a ML study was conducted on 20 transition metal atoms supported on a graphdiyne (GDY) substrate, resulting in the identification of 210 potential dual metal single-cluster systems (TM1-TM@GDY). To enhance the efficiency of DFT calculations, a combined ML and DFT approach was utilised, employing spin-polarised DFT computations with the Vienna *ab initio* simulation package (VASP) and the generalised gradient approximation (GGA) for modelling exchange–correlation interactions. Four supervised learning algorithms were applied, using 10 physical and chemical descriptors, with the number of d electrons in transition metal atoms identified as a key predictor of adsorption energy for COOH. The study found that combining two transition metals in a heteronuclear complex significantly alters the adsorption energies of critical intermediates, which is essential for optimising electronic structures and enhancing catalytic efficiency for CO_2 reduction reactions (CO_2RR). Notably, five TM1-TM2@GDY systems demonstrated strong affinity for CO, particularly Co-Mo@GDY and Co-W@GDY, which showed promising selectivity for converting CO_2 to CH_4 with low limiting potentials. The integration of DFT and ML not only reduced computational costs but also facilitated rapid assessments of catalytic activity and selectivity, paving the way for innovative designs in CO_2RR catalysts.[81]

Lots of research is focused on the search for catalysts for the production of CO, CH_4, or C_2 CO_2RR products. ML approaches have also been applied to predict C_3 product pathways, revealing the significance of neighbouring effects in C–C coupling processes. This insight aids in identifying effective electrocatalysts for producing high-value chemicals.[82] Among the C_3 products, *n*-propanol is the most important solvent and raw material for the chemical industry. There are studies showing that nanoporous Cu makes it possible to obtain *n*-propanol with FE of 21% at −0.56 V relative to RHE.[83] Various mechanisms have been proposed for the formation of C–C–C bonds, including the connection between intermediate compounds C_1 and C_2 to form CH_3CH_2CHO. Theoretical calculations show that direct CO* binding may be a potential pathway of CO_2RR to C_3 products. The reaction mechanisms for C_3 products are still unclear, which hinders the development of effective electrocatalysts. SAC and DMSC have the potential to realise C_3 product generation due to their high electroactivity, tunable selectivity and maximum atom utilisation efficiency. For example, Ni SAC deposited on nitrided graphene has more than 97% FE CO at an overvoltage of 0.61 V. Sun and Huang applied the first principle machine learning (FPML) method to make complex predictions for the C_3 CO_2RR pathways regarding C–C–C binding processes and GDY-based reaction pathways.[82] Direct predictions of C_3 paths were implemented based on previous results of the density functional theory (DFT) of C_1 and C_2 paths, without additional DFT calculations for C_3 products. Based on the predictions of the reaction pathways, we selected six of the most promising electrocatalysts for C_3 products, including GDY-Cu, GDY-Mn, GDY-Pd, GDY-Pt, GDY-Pr and GDY-Pm. Based on the prediction of the reaction pathway, propanol exhibits a higher reactivity compared to acetone and 1,2-propylene glycol, which is confirmed by previous experiments. A close interaction has also been identified between the binding processes and various key factors necessary to understand the mechanisms of the C_3 reaction. Due to the analysis of reaction energy and RDS barriers, several new GDY-AC are considered promising candidates for the effective production of C_3 products in the CO_2RR process. This research will contribute to the development of new electrocatalysts capable of efficiently producing C_3 products from CO_2 with high electroactivity and selectivity in the future.

3.1.4. Bimetallic catalysts. Bimetallic catalysts, particularly those based on copper, have been attracting increasing interest due to their capacity to enhance the process of converting CO_2 into valuable hydrocarbons. While monometallic copper-based catalysts encounter obstacles such as elevated activation energy and limited selectivity, the introduction of bimetallic systems offers a solution.[84] By optimising intermediate adsorption and desorption steps, these bimetallic formulations enhance both activity and selectivity.[85] The superior performance of these catalysts is attributed to their ability to catalyse intricate multistep proton/electron transfer reactions, resulting in improved efficiency and selectivity in the conversion of CO_2. Recent research has underscored the critical role of atomic configuration and precise proportions of metal atoms in bimetallic systems. These elements significantly impact the selectivity and overall efficacy of the CO_2RR.[86] Advancements in ML catalyst screening methods are also accelerating the discovery of effective bimetallic catalysts, paving the way for the development of more efficient technologies for converting CO_2.

In this regard novel workflow has been developed that integrates DSTAR-based ML binding energy prediction models and CO_2RR selectivity maps to detect active and selective catalysts.[87] Based on the concept of the CO_2RR selectivity map, originally developed by Tang *et al.*,[88] authors used three binding energies to more accurately predict activity and selectivity. The potential-dependent activity and selectivity of CO_2RR for 465 binary combinations were estimated without performing any DFT calculations and modelling the surface structure. It has also been demonstrated that the method is able to provide a more detailed design strategy by analysing activity and selectivity in accordance with the composition and coordination number of active motifs. Finally, we experimentally tested the binary alloys Cu–Pd and Cu–Ga and confirmed their high selectivity for C_{1+} and formate, respectively, which is consistent with the prediction made using HTVS. The authors expect that the HTVS strategy developed in this work will accelerate the detection of active and selective CO_2RR catalysts.

In a separate study,[35] the authors demonstrated the feasibility of their method by screening a variety of alloys composed of 31 different elements. This screening covered 50% of the d-block elements and 33% of the p-block elements. As a result, the method identified 131 potential surfaces in 54 alloys suitable for CO_2RR. A qualitative analysis was then used to prioritise the best candidates for further experimental verification. Using this approach, the Cu–Al catalyst was identified as being active for CO_2RR. Experimental tests showed high efficiency with respect to ethylene production, confirming the effectiveness of the combined DFT and ML approach in catalyst discovery. However, it is important to note that in the screening studies, CO* binding energy was used as a descriptor of activity to predict catalytic activity without considering product selectivity. Due to the fact that a diverse range of CO_2RR products have been experimentally reported, it is also essential to assess the selectivity of these products. Furthermore, most screening studies depend on open databases, such as the Materials Project, which restricts the chemical research scope and makes it challenging to discover completely novel catalysts.

In this study,[89] a methodology for predicting the energetic properties of bimetallic nanoparticle complexes with reagents was described. The investigation focused on the Cu_nNi_{1-n} nanocluster in interaction with CO_2 and intermediate products of its transformation. The authors compared the effectiveness of straightforward and rapid machine learning techniques to achieve this goal. Calculations were carried out within the DFT framework, as implemented in the VASP (Vienna *Ab initio* Simulation Package) version 5.4.4. Considering all possible metal distributions, it would have been very laborious to determine the energy of all clusters. Even when simplifying assumptions were considered, the number of possible combinations exceeded 20 000; therefore, the authors developed a procedure for predicting the stability of a particular system. The Pickle procedure determines a subset of the most preferred systems at minimal computational cost. The primary objective of the study was to minimise the time required for structure optimisation. However, the focus of this particular study was on the geometric features of nanoclusters, which requires meticulous optimization of each structure to generate input data for the model. After analysing the possible algorithms, the most popular representatives of the core and ensemble methods were selected to check the accuracy of a large number of models following the free lunch theorem. Most research efforts in this field are dedicated to predicting the properties of organic molecules. Fewer studies have been conducted on systems containing transition metals. The authors propose a method for predicting the catalytic activity of copper- and nickel-containing systems based on geometric features alone, using simple and fast learning models.

The integration of AI and ML into the field of CO_2 electroconversion represents a revolutionary step forward in the development and optimisation of catalysts. Research papers demonstrate how predictive models and selectivity maps can simplify the identification of active and selective catalysts, significantly reducing the time and resources required for traditional experimental approaches. The ability of AI/ML methods to analyse complex data sets and identify relationships between the composition, structure, and performance of a catalyst has led to the discovery of new catalysts with increased efficiency in reducing CO_2 emissions. Synergy between AI/ML and high-throughput virtual screening (HTVS) has allowed comprehensive screening of various alloy combinations, identifying promising candidates that would have gone unnoticed using traditional techniques. These achievements not only contribute to the rapid development of vast chemical spaces, but also open up the way for the development of individual catalyst designs that consider the activity and selectivity of products. As the industry develops further, improvements in AI and ML algorithms, along with experimental verification, will be essential to overcome existing challenges and improve the efficiency and sustainability of carbon dioxide electroconversion processes. Ultimately, the use of AI/ML in this field opens up significant opportunities for advancing the development of efficient electrocatalysts and contributing to global efforts to reduce climate change by utilising CO_2.

4. Water splitting: hydrogen and oxygen evolution reactions catalysts

With the depletion of fossil fuel reserves, the pursuit of viable and cost-effective alternatives becomes imperative. Hydrogen, with its multifaceted applications, emerges as a pivotal player in the realm of global energy markets. Its utility extends beyond traditional uses in oil refining and chemical industries, encompassing applications as a fuel for transportation vehicles and power generation in hydroelectric facilities. Hydrogen fuel cells find application in a diverse range of transportation modes, including automobiles, maritime vessels, and aircraft. Nonetheless, the widespread adoption of H_2 energy systems faces a multitude of obstacles.

Electrocatalytic splitting of water is the most environmentally friendly process for producing hydrogen fuel, in which an electric current is passed through water to decompose it into H_2 and O_2, converting electrical energy into chemical energy by accumulating electrons in the form of stable chemical bonds.[90–92] While electricity generated from conventional sources of natural gas and coal produces greenhouse gases, electricity generated from renewable sources such as nuclear, wind, solar, geothermal, hydro- and tidal energy does not do this, which makes the electrocatalytic splitting of water using these sources of electricity cleaner and more desirable. However, its large-scale implementation is complicated by many factors, including technological issues, cost considerations for catalysts, and stakeholder resistance.

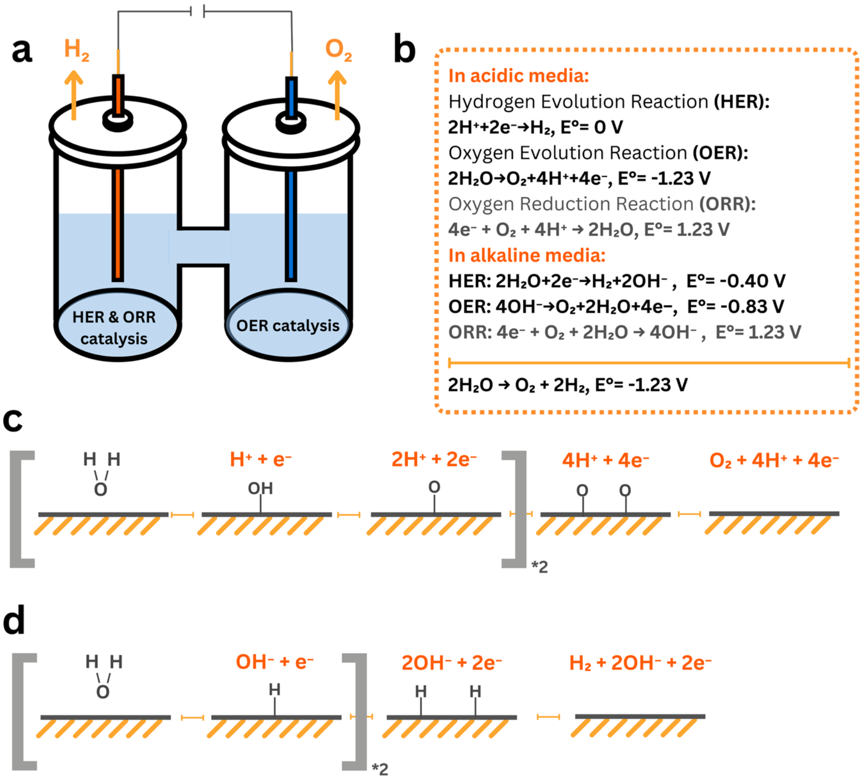

Fig. 6 Water splitting reaction: (a) scheme of the electrochemical setup, (b) half-reactions of the process in acidic and alkaline media; reaction stages of (c) OER and (d) HER, respectively.

The electrochemical decomposition of water consists of three main reactions: the oxygen evolution reaction (OER), which takes place at the anode, and the HER and oxygen reduction (ORR), which takes place at the cathode (Fig. 6).[93] These semi-reactions are three-phase reactions and require electrocatalysts to accelerate them initially sluggish kinetics.[94] Effective electrocatalysts can reduce the excess potential of oxygen and hydrogen production using these processes. In principle, this becomes possible when an alternative reaction pathway becomes available through a variety of mechanisms. Firstly, the catalyst can provide a large number of active sites for the reaction due to its advanced morphology. Secondly, electron transfer can be facilitated by modifying the electronic structure of the reagent or increasing the rate of electron transfer over a conductive surface. Another possible option is to change the reaction mechanism to a more kinetically favourable one.[95,96]

An ideal electrocatalyst should demonstrate high catalytic activity for both HER and OER, as well as long-term stability in harsh electrolysis environments. Historically, metal-based materials have been the primary electrocatalysts for these reactions. Transition metals and their oxides have been widely investigated as potential candidates, with platinum standing out due to its exceptional catalytic performance in HER. Nevertheless, the scarcity and high cost of platinum significantly limit its practical application.

In the context of chemistry, the analysis of the above catalysts determines their ability to accelerate electrochemical reactions such as hydrogenation, water electrolysis, and other processes that are important in the fields of electrochemistry, electrolysis, and electrochemical energy conversion (Table 4). Although Pt, Pd and Au based materials are popular as anode catalysts for the electrooxidation of small organic molecules, their use as monometals faces various disadvantages as shown in the table above. To overcome these disadvantages, it is necessary to consider alloying these metals by rationally designing one material from two or three of them. Doping can help to improve the electronic structure of the catalyst and suppress the etching of the active sites caused by carbon-containing compounds. Bimetallic catalysts have a unique microstructure, exhibiting higher current densities in the forward peak.[97]

Pt, Pd and Au-based materials are commonly used as cathode catalysts for the electrochemical reduction of organic molecules and water splitting. However, their use as monometallic catalysts presents several challenges, as outlined in Table 4. To address these issues, it may be worth exploring the potential of creating alloys from two or more metals. The addition of dopants can improve the electronic structure of the catalysts and reduce the etching of active sites by carbon-based compounds. Bimetallic catalysts have a unique microstructure that results in higher current densities at the peak.[97]

The optimisation of the chemical composition, structure, and active sites of catalysts is crucial for improving the efficiency of HER. Key factors that need to be considered include mass transfer mechanisms, alloying of catalysts, thermodynamic parameters, and environmental conditions. Through careful adjustment of the redox potential on the catalyst surface, reaction rates and stability of catalysts can be significantly enhanced. Use of ML and AI tools allow us to predict reaction pathways and determine properties of catalysts effectively selecting promising materials. Focusing on HER research can lead to development of more efficient methods of hydrogen production essential for sustainable energy solutions.

4.1. AI/ML in water splitting reactions

The application of deep learning methods facilitates the analysis of intricate relationships in data, such as those between adsorption energy and catalytic activity. This correlation is significant because the catalytic properties of nanoclusters frequently depend on the structural diversity of their active sites. Through the use of sophisticated machine learning models, researchers are able to not only forecast the performance of existing nanocatalysts but also to discover novel materials with optimized properties. Utilization of deep learning within this context transforms the scope of catalyst research, enabling quicker and more efficient identification of high-performing nanocatalysts. This advancement is critical to accelerating the development of sustainable hydrogen generation technologies, which are essential for addressing global energy challenges. Considering the time-dependent and dynamic nature of the catalytic process is essential for creating accurate ML models. Exploring different chemical spaces and reducing computational costs through surrogate models and first-principles calculations are common approaches. However, the structural and energetic complexity of inorganic materials limits the application of these methods in electrocatalysis. Understanding the mechanism of a specific catalytic reaction is important for ML model development as the coding format of the initial molecules and energy descriptors directly affect the accuracy of the models and its subsequent interpretability. Depending on the environment, the HER half-reaction of the water splitting process most often proceeds through the Volmer–Geirsky or Volmer–Tafel steps.[110] In acidic and alkaline media, these reactions have their own peculiarities. In acidic media, the HER proceeds *via* kinetically simpler mechanisms. For example, in the Volmer reaction, adsorbed hydrogen (H^*) is formed from the H^+ ion, which is a rapidly occurring reaction. For the further formation of the H_2 molecule, it is important for this step to have a near-zero value of $\Delta_r G_{ads}$, which provides a high probability of molecular hydrogen formation. Whereas under alkaline conditions, associative desorption reactions have to be considered, so the optimal value of $\Delta_r G_{ads}$ would be less than zero. These differences determine the design features for acidic and alkaline environments. To date, HER in acidic media has been actively studied. The research addresses various aspects such as hydrogen adsorption behaviour, descriptor volcanic sites and computational predictions. Due to stringent industrial corrosive conditions, there are increased costs and risks related to plant safety, which in turn impose additional requirements on catalyst design for large scale applications. It is important to note that the reaction rate of HER in alkaline solutions is usually much lower than in acidic solutions. Therefore, special attention is given to the development of stable molecular catalysts for the HER process that are efficient in both acidic and alkaline environments for improved application versatility. The electrocatalyst should promote the adsorption process regulated by the hydrogen adsorption free energy ΔG_{H^*}.

An illustrative example of this trend in the general areas of science and engineering is the use of deep learning for learning representations of molecular and materials data. It can, for example, be used to predict the properties of catalysts or evaluate how a molecule would behave in several chemical reaction categories. Ultimately, principal elucidation is of great importance for the usage and development of ML models. Ensemble models, which combine multiple prediction methods, tend to perform more consistently than single-model designs. Describing the impact of design parameters on model ensemble performance towards learning tasks, for instance, does have predictive value, at least if an attempt is made to construct models based on multiple rather than a single prediction method. A comparison of this kind generally emphasizes an appealing aspect of ensemble models – they incorporate distinguished learning-method performance characteristics into predictions, and thus tend to be more regularly performing than any of the single-model designs. Special attention has to be paid to correctly rationalizing the hyperparameter and feature importance characteristics obtained from randomized and LASSO-penalized techniques. ML models tend to have a high degree of variance and small sample bias from training data and a low degree of bias and high precision while predicting test data. Overfitting and the standard error of models of weak learners can be underestimated, highlighting the need for careful consideration in real-world data applications. Deep learning has found applications in molecular and materials data analysis, particularly in predicting catalyst properties and molecular behaviour in chemical reactions. Understanding the impact of design parameters on model ensemble performance is crucial for the development and use of ML models.

Heterogeneous catalysts for HER are of particular interest due to their high configurational entropy, ability to form solid solutions, and potential for high strength and resistance to oxidation/corrosion. This chapter will structure the review of AI and ML applications in catalysis for water splitting reactions by dividing the content into sections, focusing on relatively new and highly effective heterocatalysts, categorized into eight groups.

4.1.1. Platinum-based electrocatalysts. Pt is an extremely efficient catalyst that has been used in various important clean energy applications, including biomass conversion, the HER, and the ORR. However, due to the limited availability and increasing demand for the noble metal, its long-term viability as a catalyst has become more challenging, and its cost has increased. One promising approach to addressing this issue is the alloying of Pt with base metals, such as Co, Ni or Zn. High-performance computational techniques based on ML and DFT play a crucial role in developing new catalysts by enabling the selection of a wide range of promising candidates. This contributes to the development of more efficient and cost-

Table 4 Advantages and disadvantages of various monometal-based electrocatalyst in water splitting and CO_2RR

Catalyst	Advantages	Disadvantages	Ref.
Pt-based	Outstanding effectiveness in the HER with low overpotential (86 mV at 10 mA cm^{-2}) and Tafel slopes (around 30 mV dec^{-1})	High cost and scarcity (60 USD per gram) Limited natural abundance Suffer from deactivation over time due to sintering and leaching, which compromises their long-term effectiveness Quick deactivation during extended use, especially in the OER because of slow kinetics In CO_2RR, Pt selectivity can be suboptimal, leading to lower product yields compared to other catalysts	98
Pd-based	Active in both basic and acidic conditions	Sluggish reaction kinetics *via* water splitting reaction	99
	Exceptional performance in CO_2RR, achieving a FE of 81% for CO generation with a current density of 55 mA cm^{-2} and a low overpotential of 251 mV	Suffer from structural collapse and low cycling stability Require high overpotentials to drive CO_2RR (1.0 V *vs.* RHE) Low selectivity High cost (50 USD per gram)	
Au-based	Offers increased resistance to oxygen and carbonaceous intermediates poisoning	Require higher overpotentials compared to state-of-the-art OER catalysts (0.35 V at 10 mA cm^{-2})	100
	Less expensive than Pt and Pd	Only functions effectively under basic aqueous conditions	
	Relatively low overpotential during OER (0.32 V at a current density of 10 mA cm^{-2} in alkaline media) High selectivity and FE of up to 90% for CO production High current densities of up to 200 mA cm^{-2} for CO_2RR	High cost	
Ag-based	More affordable than Pt and Pd	Performance diminishes significantly	101
	Serves as a suitable promoter for Pt, Pd, and Au catalysts	Lower catalytic activity for the OER compared to other non-noble metal catalysts (overpotential of over 500 mV)	
	Relatively low overpotential *via* HER (0.18 V at a current density of 10 mA cm^{-2} in acidic media)	Oxidation reaction halted above 1.125 V, regardless of glycerol concentration	
	Showed FE of up to 95% for CO production, one of the highest reported values	Lower current densities compared to Au (50–100 mA cm^{-2}) *via* CO_2RR Suffer from deactivation due to surface oxidation	
Ni-based	For HER in alkaline media demonstrates low overpotentials (64 mV at the same current density of 10 mA cm^{-2}) comparable to Pt, Pd, and Au	Active exclusively under basic aqueous conditions	102
	Achieve FE for CO production exceeding 80% with current densities over 100 mA cm^{-2}	Limited activity for OER with current densities about 10 mA cm^{-2}	
	Economical, durable in alkaline solutions, and resists poisoning	Limited activity for CO_2RR with current densities about 10–20 mA cm^{-2} Stability issues Selectivity challenges	
Co-based	Displays activity across a wide pH range	The overpotential for the OER relatively high (520 mV at a current density of 10 mA cm^{-2}, which increases to 890 mV at 100 mA cm^{-2})	103
	Vulnerable to poisoning by S-containing species Demonstrate significant performance in ORR, achieving a stable current density of 10 mA cm^{-2} at only 277 mV in basic media Leads to a favorable formation of intermediates like methoxy and formate, enhancing selectivity	Require high overpotentials to initiate the CO_2RR, which can lead to energy inefficiencies	
Rh-based	High catalytic activity in acidic media	The overpotential for HER relatively high, (400 mV at a current density of 10 mA cm^{-2})	104
	HER current density for Rh is significantly high (1 A cm^{-2})	Requires improvement for broader pH range application	
	Provides exceptional durability and resistance to poisoning	High cost (500 USD per gram)	
	Stability over 100 h	FE for CO production of around 70–80%, which is lower than that of other non-precious metal catalysts	
Fe-based	Demonstrates good catalytic activity in basic conditions	Challenges in performance under acidic conditions	105
	Cost-effective and abundant	Overpotentials for HER around 400–600 mV at a current density of 10 mA cm^{-2}	

Table 4 (*Contd.*)

Catalyst	Advantages	Disadvantages	Ref.
	Exhibits susceptibility to carbon deposition	FE for CO production in the range of 50–70%	
	The overpotential for the OER (237 mV at a high current density of 100 mA cm^{-2}) which is comparable to Pt, Pd, Au	The current densities for CO_2RR are typically lower around 20–30 mA cm^{-2}	
Cu-based	Effective catalyst for water splitting reactions (overpotential as low as 300 mV at a current density of 10 mA cm^{-2})	Relatively low FE for CO production (50–60%)	106
	Relatively inexpensive (3 USD per kg)	The current densities for Cu-based catalysts in CO_2RR can be limited, often around 20–50 mA cm^{-2}	
	Sustained performance over 100 h	Requires enhancement in poison tolerance	
	FE of up to 72% for the production of C_{2+} products		
	Current densities of around 50–100 mA cm^{-2}		
Ru-based	Effective in catalysing water oxidation	Expensive and less abundant compared to some counterparts (300 USD per gram)	107 and 108
	Ru achieve an overpotential for HER 30–50 mV at a current density of 10 mA cm^{-2}	Agglomeration issues during prolonged operation	
	Shows good durability and resistance to corrosion	Selectivity challenges	
	Remarkable stability during water splitting (over 1000 hours)	Overpotentials in HER exceeding 500 mV to achieve significant current densities	
	FE of up to 90% for the production of CO and other valuable products		
Mn-based	Displays considerable catalytic performance in HER	Overpotentials *via* HER around 500–800 mV at a current density of 10 mA cm^{-2}	109
	Environmentally friendly and cost-efficient	Limited stability under certain conditions	
	High FE for CO production (97% at low overpotentials around 0.49 V)	Stability issues	
		Lower FE for CO_2RR to C_2 products 60–70%	
		Selectivity challenges	
		Current density limitations for CO_2RR (10–20 mA cm^{-2})	

effective catalysts for clean energy technologies.[111] As an example the study successfully identified 12 new low-cost Pt alloy catalysts for HER using a combination of ML and DFT. The best-performing binary alloy discovered is Cu_3Pt, the (100) surface of which shows comparable performance to the Pt (111) surface, with minimal deviations in adsorption energies for H, CO, and O, indicating its potential as a viable catalyst. For ternary alloys, $FeCu_2Pt$ was highlighted as a promising alternative, demonstrating reduced costs while maintaining high catalytic activity. Authors confirmed that the d-band theory remains applicable for analysing adsorption energies in these ternary alloys, providing insights for future catalyst design. It is suggested that these low-cost alloys could serve as effective replacements for platinum in various catalytic processes, enhancing sustainability of chemical conversions.

There is a large amount of work aimed at finding the essence of the catalytic efficiency in the most active catalysts and amplifying their positive attributes *via* rational catalyst design. In particular, Cu for the CO_2RR and Pt for the HER and ORR reactions.[112] Specialized on HER study challenges the long-held belief that a thermoneutral binding energy yields maximum catalytic activity. It suggests that this criterion only maximises activity near equilibrium conditions. The experimental hydrogen binding energy of Pt was found to be +0.094 eV. This positive binding energy results in higher catalytic activity when the overpotential is increased, compared to a thermoneutral catalyst. The research indicates that the trade-off between activity near and away from equilibrium implies that traditional designs based on thermoneutrality may not be optimal for catalyst development. The authors utilised microkinetic modelling to derive a mathematical relationship between catalytic activity and binding energy, which was essential for fitting the Tafel plot. It highlighted that the fitting analysis effectively minimises R^2 function, ensuring the accuracy of the results. The lack of hysteresis in the experimental HER Tafel plot supports the validity of the steady-state approximation used in the analysis. These findings provide new insights into the design of catalysts, particularly in the context of HER.

To understand HER mechanisms, it is essential to simulate and interpret kinetics using appropriate rate expressions.[113] HER kinetics of various electrocatalysts have been extensively studied using *ab initio* simulations of reaction and activation energies.[114] A number of factors contribute to uncertainty in DFT energy calculations, including adsorbed species, hydrogen-bond networks, electric double layers, and water static layers.[115,116] In order to investigate the kinetics of HER, Zhou *et al.*[117] used an end-to-end approach, combining ML and experiments using jagged Pt nanowires as electrocatalyst (Fig. 7).

Their proposed workflow for jagged Pt nanowires simulation is illustrated in Fig. 7. Authors found that the optimum value of Gibbs free energy of adsorption for the overall rate is lower under alkaline conditions. Moreover, autobifunctional mechanism was detected in the jagged Pt nanowires, where the weaker

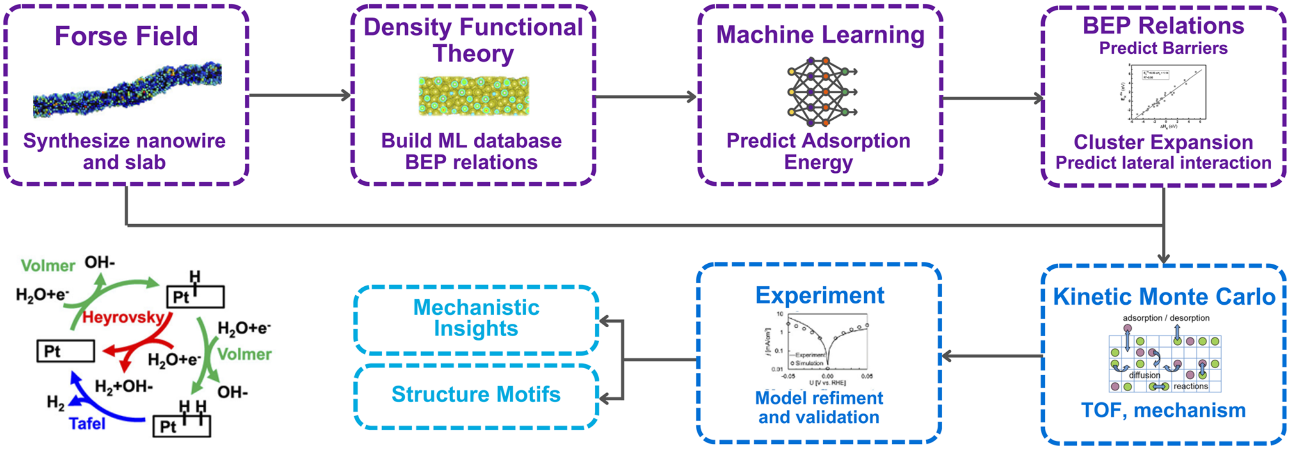

Fig. 7 End-to-end workflow for the simulation of jagged Pt nanowire, combined DFT, ML, force field, and kinetic modelling [reprinted with permission from ref. 118 Copyright 2021 Journal of the American Chemical Society].

and stronger binding sites activate hydrogen and adsorb protons respectively. For Volmer and Tafel reactions, top sites are most active, but since Tafel reactions are faster, hydrogen diffused from bridge sites is consumed by top sites. The ML model demonstrates the correlation of Gibbs free energy of adsorption with the coordination number, whereas the sites with coordination number of less than 7 exhibited higher activity.

Various theoretical models have been proposed to understand the relationship between structural features and functionality, particularly the adsorption energy of chemical intermediates with multiple material activities. However, most models are specific to certain substrates, making it challenging to determine the structure–property relationship for catalysts in advanced materials. Guo *et al.* used a machine learning model (random forest) to correlate descriptors of 149 monoatomic catalysts. Predicting specific features that influence hydrogen evolution activity without prior classification is difficult when dealing with large datasets.[119] In a similar attempt, Parker *et al.* investigated class-specific structure–property correlations of HER indicators by analysing 1300 platinum ensembles and classifying them into ordered and disordered structures to provide an accurate forecast of the structure–property relationship.[139] It was noted that for efficient hydrogen evolution, small particles of disordered materials and a larger (110) surface in ordered materials are essential. These studies focus on the significant contribution of molecular aspects and atomic plane arrangements influencing catalytic activity. Therefore, creating a map linking structure to performance promises overall savings in materials and experimental time, and provides better control over structures and surface dopants. Furthermore, activity of catalysts depends on their nature, structure and composition, directing synthetic approaches during their design and preparation, *i.e.* catalyst synthesis and/or pretreatment. Here, quantitative structure–property relationships (QSPR)[120] do not work, leading to the development of quantitative synthesis-structure–performance relationships (QS2PR) based on active learning using the kriging method proposed by Ebikade *et al.* Synthesis temperature and N-doped catalysts with high crystallinity were found to be important properties according to the research results, maximising HER performance.[120]

Recent studies also describe innovative applications of AI and ML in enhancing the understanding of OER mechanisms, predicting catalyst performance, and guiding the design of advanced materials. As an example, Omidvar *et al.* utilised the TinNet interpretable framework in conjunction with an active learning approach to investigate a broad design space of around 17 000 candidates for Pt monolayer core–shell alloys, derived from approximately 1500 stable bulk structures.[121] Through the application of explainable AI techniques, the researchers pinpointed optimal catalytic sites located near the peak of the activity volcano, uncovering electronic factors that affect reactivity trends. The findings underscored the significance of adsorbate resonance energies, arising from d-band interactions, in comprehending chemical bonding at metal surfaces, which paves the way for enhancing catalytic performance beyond merely targeting active sites. The study demonstrated the potential of adjusting active sites *via* core–shell alloying and introducing buffer layers of metals such as Ir, Rh, and Ru, aiming to boost surface reactivity while minimising the reliance on precious Pt in the ORR. It revealed that the adsorption energy of adsorbate–substrate pairs is influenced by various factors, not just the position of the band centre, indicating a more intricate interaction in catalysis.[121]

Xie *et al.* highlights the potential of single atom catalysts (SACS) to improve ORR efficiency.[122] Utilising DFT and ML techniques, the authors analysed 144 SACS to address challenges in the rational design of high-performance catalysts. They identified 25 materials that outperformed the benchmark Fe–N_4 catalyst, with Fe–Pd, Fe–Pt, and Fe–Ag demonstrating notable overpotentials of 0.980, 0.983, and 1.008 V, respectively. The research emphasizes the importance of comprehending the relationships between structure and properties, as well as optimizing the binding energies of reaction intermediates, in order to improve catalytic efficacy. Findings support the notion that ML models can substantially reduce the time and cost associated with DFT computations and experimental processes, facilitating the development of efficient and cost-effective ORR catalysts. This study lays the groundwork for prospective inquiries in the design of SAC, thereby advancing sustainable energy alternatives.[122]

Another promising catalytic material of the OER reaction is high-entropy Pt-based alloys (HEAs). Kadam *et al.* successfully synthesised HEAs such as Al–Cu–Pd–Pt, which exhibited promising catalytic activity for the ORR. Utilising ML techniques, specifically the Gradient Boosting Regressor (GBR) model, the researchers predicted O_2 adsorption energies across various surface sites of the HEAs, achieving high accuracy and generalizability. The findings revealed that the adsorption energy on HEA surfaces is influenced by the contributions of nearby metal atoms, a conclusion validated through extensive data analysis. There are indications that the current density of ORR catalytic processes is expected to vary at low frequencies, which has significant implications for optimising compositions. The GBR model effectively described the ORR processes at sensitive sites on HEA surfaces, demonstrating its potential for future research in this field. The principal outcome of the manuscript, as posited by the author, is a conceptual framework for the methodical design and synthesis of HEA catalysts, underscoring their extensive configurational diversity and the opportunity for superior catalytic efficacy.[123]

ML can help elucidate the interactions between the structural features of Pt-based catalysts and their activity in OER. By integrating computational models with experimental data, it is possible to derive insights into how modifications in the catalyst structure (*e.g.*, particle size, shape, and composition) affect performance. Li *et al.* determine the phase diagram of the bulk PtNiMo alloy using a ML approach known as Stochastic Surface Walking-Neural Network (SSW-NN), which aids in understanding the alloy's structural properties. Then researchers explore the surface structures of PtNiMo at various compositions, ranging from $Pt_{66.7}Ni_{30.6}Mo_{2.7}$ to $Pt_{83.5}Ni_{13.8}Mo_{2.7}$, identifying the most stable surface structures through Grand Canonical Monte Carlo (GCMC)/SSW-NN methods. The dynamic evolution of these surfaces is examined under varying O atom concentrations, which fluctuate according to the electrode potential, revealing how the surface structures adapt during the ORR. The work resolves the ORR activities of the stable PtNiMo surfaces under specific reaction conditions, demonstrating the influence of structural changes on catalytic performance. This investigation highlights that even low-Mo-doped Pt–Ni alloys undergo considerable structural alterations under ORR conditions, with active sites identified as $Mo_{1-2}O_x$ clusters on a Pt (111) surface. These findings contribute to a deeper understanding of the catalytic mechanisms and structural dynamics of PtNiMo alloys in electrochemical applications.[124]

4.1.2. Perovskite-based catalysts. All perovskite materials inherently exhibit structural deficiencies that undermine their stability, particularly when subjected to polar solvents such as water, ultraviolet radiation, and elevated temperatures.[125,126] Perovskite nanocrystals (PNCs) are even more vulnerable to deterioration owing to their increased surface area, which amplifies environmental interactions. This instability presents a considerable obstacle for their commercial utilization. Although methodologies such as surface coating or the application of enhanced binding agents can improve the stability of PNCs, these strategies frequently impede their charge transfer efficiency, thereby diminishing their efficacy as photocatalysts.[127] Nevertheless, research conducted by Xiao and associates has illustrated the feasibility of achieving an equilibrium between stability and charge transfer efficiency in PNC-based catalysts through the modulation of ligand quantity employed.[128]

The effective passivation of PNCs is paramount for enduring photocatalytic applications, particularly since these processes often occur under conditions that may destabilize perovskites. Xiao and associates demonstrated that an increased density of passivating ligands on the surface of $CsPbBr_3$ PNCs is vital for maintaining their photocatalytic efficacy over time, especially in the presence of water vapor for hydrogen production.[128] Their results indicated that PNCs with a higher ligand count preserved stability for 160 h, in stark contrast to only 23 h for those with a reduced number of ligands. Notably, although the ligand-rich PNCs initially exhibited no hydrogen production due to steric hindrance, photocatalytic activity initiated following the desorption of certain ligands during extended irradiation.

ML techniques have been employed to discover and optimize perovskite compositions, focusing on their electronic descriptors to enhance catalytic performance. As an illustration the research of Farghaly *et al.* employed ML methodologies to facilitate the identification and optimisation of perovskite alkaline electrolyte OER electrocatalysts, with the objective of discovering economically viable substitutes for noble metal catalysts such as iridium and ruthenium.[129] A comprehensive simulation of over 8000 perovskite materials was conducted utilizing DFT, which demonstrated that established perovskites generally exhibit low energy levels above the thermodynamic convex hull, thereby signifying their stability and potential for enhanced activity. Promising perovskite formulations, discerned through machine learning algorithms, were earmarked for high-throughput automated synthesis at Argonne National Laboratory, thereby facilitating expedited testing and validation of novel materials. Moreover, the investigation revealed a significant correlation between phase purity, annealing temperature, and OER activity, yielding essential insights for the optimization of synthesis parameters to improve electrocatalyst efficacy.[129] The integration of ML with DFT allows for the simulation of thousands of perovskite structures, identifying those with low energy configurations conducive to high HER activity. For instance, the substitution of calcium in $KMgO_3$ perovskite has been shown to lower the bandgap and enhance photocatalytic efficiency, making it a promising candidate for HER applications. The study investigates the performance of calcium-incorporated KMg catalysts for the HER and OER, revealing that the addition of calcium significantly enhances catalytic activity for both reactions. By identifying the optimal Ca content, the research demonstrated improved electrochemical performance, which can be attributed to enhanced charge transfer and an increase in active sites. Mechanistic insights showed that calcium incorporation alters the electronic structure of the catalyst, enabling more efficient reaction pathways. The authors show stability of the calcium-incorporated KMg catalysts under operational conditions, suggesting their potential for practical applications in energy conversion technologies. These findings indicate that incorporating calcium is a promising strategy to enhance the catalytic performance of KMg-based materials for HER and OER applications, thereby contributing to a better understanding of how modifying catalyst composition can lead to significant improvements in energy-related reactions.[130] Experimental results indicate that specific perovskite compositions exhibit superior electrocatalytic activity for HER, with optimal ratios leading to enhanced current densities. By way of example the authors developed a new Surface Center-Environment (SCE) feature model that combines 38 elementary properties of substitution elements with surface structure information. This model is crucial for the ML applications in predicting catalytic behaviours. Using the SCE model, a random forest ML model was constructed based on a dataset of 172 perovskite oxides. This dataset included 72 structures calculated in the study and 100 from existing literature. The model successfully predicted the adsorption free energies and overpotentials for 610 new perovskite oxides. The predictions indicated that perovskite oxides exhibit low adsorption free energies for intermediate species (HO*, O*, and HOO*) when the B site elements are Nb, Mo, Ta, W, and Os. This suggests these elements are favourable for catalytic activity. The study found that the overpotentials of perovskite oxides are low when B site elements are Mn, Fe, Co, Ru, Rh, and Ir. These elements could be utilized to synthesize new perovskite structures with enhanced catalytic performance for OER. This work identifies specific perovskite oxides with the lowest overpotentials. For instance, $YRhO_3$ (0.26 V), $BaRuO_3$ (0.27 V), and $BaRhO_3$ (0.30 V) were highlighted as top candidates based on their low overpotential values. The calculated overpotentials for $LaFeO_3$ (0.49 V) and $CaMnO_3$ (0.58 V) were found to be in good agreement with experimental results (0.51 V and 0.54 V, respectively), validating the predictive capability of the SCE model. These work is demonstration of the fact that SCE feature model and ML methods can effectively describe surface chemical reactions and accelerate the screening of perovskite catalysts with desirable properties, paving the way for future research in this area.[131]

4.1.3. Transition metals and oxides. It is a well-known fact that the optimal catalytic materials are noble metals based on Pt, the scarcity and high cost of which seriously hinder their widespread use. Potential HER electrocatalysts that could be compared with Pt-based materials are mainly based on adapter metals (TM), such as TM compounds, monatomic TM catalysts. ML models are being used for HER catalyst prediction, enabling accelerated materials design and discovery of design rules. They navigate transition-metal chemical space and guide the discovery of electrocatalysts for various reactions, including hydrogen evolution and CO.[132] ML has made significant inroads in the quest for predicting new catalysts. It facilitates a broader range of applications, from data-analytics tasks, such as feature selection, to the development of sophisticated models for reaction mechanism and materials design. For software to be modular, there is not only one code that shows what structure is needed to achieve a defined functionality. Instead, it corresponds to a set of codes that offer parallel functionality, and models that are able to mimic final codes are being developed. The generated insights are useful in multiple respects—a model for energies can be useful in both evaluating new catalyst candidate materials and in unravelling trends and activity descriptors in an existing database.[133]

One of the best examples where the impact of AI-based models can be shown is the catalysis of the Volmer reaction on metal surfaces, as it is widely accepted that it is the rate determining step for HER over a wide range of catalysts. In recent study presented 27 different models that show that the kinetic Volmer model is quite a representative concept.[134] It would be essential to be able to predict this fundamental value, however, modelling in detail this multi-step reaction occurring on the hydrophobic surface requires both on the hand quantum mechanics methods to account for electronic structure effects and on the other hand ML or kinetic Monte-Carlo methods to account for the solvation effect and dynamics of the system.[132] Pure computational methods will never reach the throughput required for this problem. Therefore, applied multivariate response statistical analysis methods on the calculated adsorption energies to build linear models of A and Volmer barrier.

The predictions of theoretical models are typically based on the calculated electronic structure and often ignore the remaining reasons for the limiting performance of the catalysts. The use of diagnostic or performance maps, on the other hand, allow for the direct comparison of real catalysts performance and give significantly more insights into structure–property–performance relationships.[75] However, due to the hundreds of theoretically and experimentally studied catalysts and factors that affect the catalyst's performance, it is not possible to fully incorporate all of them in the development of performance maps by simple statistics.

High catalytic activity of nanoclusters makes them promising candidates for HER,[135,136] but the computational screening methods face difficulties owing to numerous relevant atomic sites in nanoclusters.[137] For example, Wexler *et al.*[138] studied the stability of $Cu_{55}–nM_n$ ($M = Co$, Ni, Ru, and Rh, and $n \leq 22$) nanoclusters by screening 7924 candidates using ML and DFT computations. Based on training an artificial neural network using DFT database, they presented an automated workflow for rapid and accurate prediction of adsorption energies on nanocluster surfaces, as shown in Fig. 8a, which led to discovering highly potential HER electrocatalysts. As an example, the selected clusters of $Cu_{13}Co_{42}$ composition are shown in Fig. 8b.

On the other hand, electronic structures and reactions of solids and molecules have been studied using computational techniques, especially density functional theory.[140] Calculations based on DFT provide information difficult to obtain experimentally about heterogeneous catalysts.[139] Molecular transition states and patterns of reactivity at solid surfaces are explained

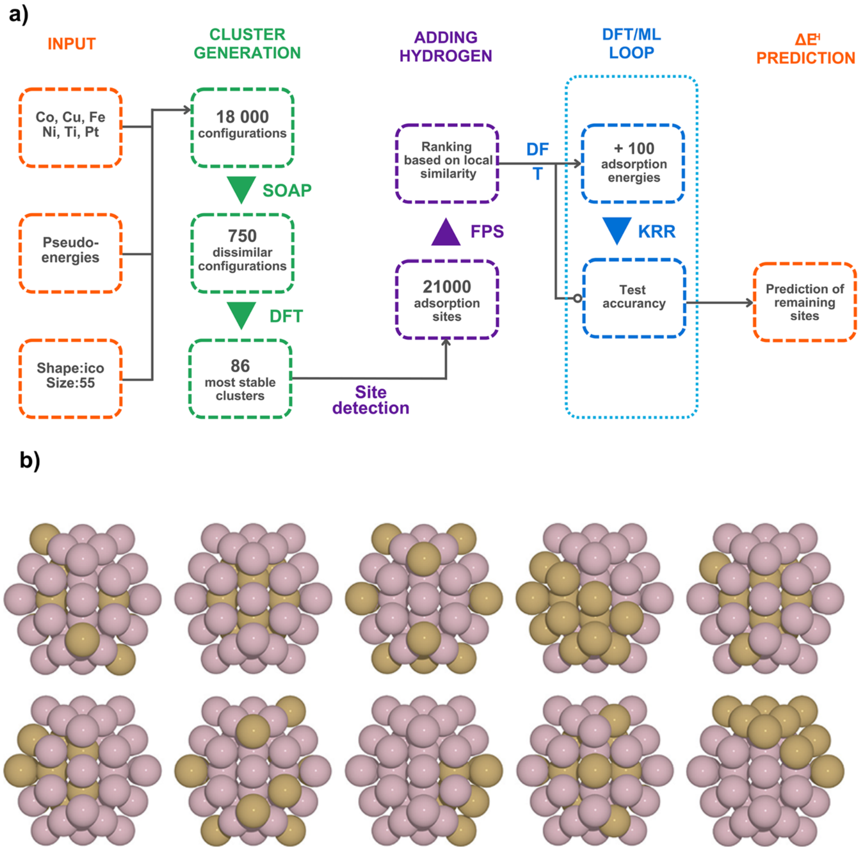

Fig. 8 (a) ML-assisted workflow proposed by Parker *et al.* to predict the adsorption energy distributions of Cu-based nanocluster, (b) generated $Cu_{13}Co_{42}$ clusters generated by Monte Carlo with various combinations of interaction and segregation energies [reprinted with permission from ref. 139 Copyright 2022 J. Am. Chem. Soc. Combinatorial Science].

by the calculations. A DFT calculation can be used to characterize the working state of a catalyst during high-pressure and high-temperature processes. DFT has also been utilized to provide the catalyst design and the optimal reaction route in HER.[141] A strong correlation was found between the adsorption energies and the maximum Hilbert-transform for the d-band. Mao *et al.*[135] also reported the identification of an alloy nanocluster electrocatalyst for HER using a ML-based workflow using DFT among 7924 candidates for the same Cu-based clusters. Their results revealed that nanocluster HER performance can be notably boosted by transition metal doping. With regard to electrocatalyst activity and structural stability, CuNi core–shell clusters were proposed as a superior catalyst for HER. In another study, they used advanced structural descriptors, such as Atom-Centered Symmetry Functions and Many-Body Tensor Representation to analyse 24 Cu–Au nanoclusters and 91 MoS_2 nanoclusters to predict hydrogen adsorption's Gibbs free energy on the nanoclusters by ML.[137] When ML was applied to different clusters simultaneously rather than separately, the mean absolute error was reduced.

The ML algorithms used in the digital design of catalysts can be divided into unsupervised and supervised learning. The most widely used algorithm under supervised learning is the regression algorithm, which can be used to find the relationship between the properties of catalysts and products. The reason why the regression algorithm is so popular is that the properties of the catalyst, the coverage of hydrogen, the coverage of intermediate, the coverage of hydrogen and intermediate, the rate of the first step, the rate of the next step, and the rate of the overall reaction are all continuous variables. Therefore, it is appropriate to use regression algorithms to find the relationship between them. In Wang's work,[142] multiple linear regression (MLR) analysis has been used to study the hydrogen evolution on transition metal dichalcogenides and to find the possible mechanism of hydrogen evolution. Unsupervised learning is another main branch of ML, and the algorithms like genetic algorithm and particle swarm optimization have been used in the optimal design of catalysts. However, it is believed that supervised learning will have a greater impact in the ML development of catalysts. This is because the direct relationship between the catalyst properties and the desired product could be found by using supervised learning, which will greatly speed up the rational design process. In 2019, Greeley and Nørskov[143] have shown that using the supervised learning algorithms in conjunction with DFT calculations can closely predict the result of the high-throughput screening experiments, which is a revolutionary work. However, up to now, the studies using ML algorithms are still limited in the real human-assisted catalyst design, partly due to the vast number of possible catalysts and their properties and the uncertainties in the output of the algorithms. A typical procedure using ML algorithms consists of the same steps that had been shown on Fig. 5. The first step is to collect the data of the known catalysts, which usually take a long period to finish the experiments. Then a high-throughput screening experiment will be applied to accelerate the data collection process.

Nguyen *et al.* addresses the limitations of noble metal-based catalysts for the HER, emphasizing the need for efficient and stable TM catalysts. Research proposes a ML approach combined with DFT to predict optimal TM dopants and experimental conditions for developing effective M@CQD HER catalysts. A Bayesian genetic algorithm (BGA) is employed to optimize experimental conditions for HER, categorizing input variables to minimize the number of experiments needed. The optimization process focuses on achieving the best overpotential, Tafel slope, and exchange current density for the water splitting experiment. DFT simulations are conducted to analyse the interaction of TMs with CQDs, focusing on the hydrogen adsorption energy and the stability of different doping configurations. The results indicate that Ni@CQD exhibits the most favourable electronic structure and hydrogen adsorption characteristics, supporting its superior performance in HER. DFT simulations supported the experimental findings by rationalizing the stability and electronic structure of the Ni@CQD/3D rGO hybrid, confirming that this configuration is optimal for HER. The Ni@CQD catalyst demonstrated the lowest overpotential values of 135 mV and 189 mV at current densities of 1 mA cm^{-2} and 10 mA cm^{-2}, respectively. This indicates its superior efficiency in the HER compared to other M@CQD catalysts. The Tafel slope for Ni@CQD was measured at 52 mV dec^{-1}, which, while higher than that of Pt/C (36 mV dec^{-1}), is significantly lower than the slopes for Co@CQD (64 mV dec^{-1}), Mn@CQD (60 mV dec^{-1}), and pristine CQD (78 mV dec^{-1}). This suggests that the Ni dopant enhances the kinetics of the HER. Furthermore, the stability tests revealed that Ni@CQD maintained its performance over extended periods, with minimal degradation observed after 100 hours of continuous operation, underscoring its potential for practical applications in electrochemical energy conversion. The study employs DFT to calculate the formation energy and hydrogen adsorption free energy of M@CQD, providing a theoretical basis for the observed electrocatalytic performance. This methodology ensures accurate modelling of the catalysts' behaviour under operational conditions, enhancing the reliability of the predictions made by the ML model.[144]

Ma *et al.* successfully identified several SACS based on Pt-doped dual transition metal (DTM) Janus-MXenes, showcasing their potential for the ORR and OER with ultralow overpotentials. Notable catalysts include Pt–VO–MnTiCO$_2$, which exhibits overpotentials of 0.24 V for ORR and 0.38 V for OER, and Pt–VO–PdTiCO$_2$, with overpotentials of 0.33 V for ORR and 0.36 V for OER. In this work authors used ML models, specifically RFR and XGBoost Regression (XGBR), to investigate the intrinsic factors influencing catalytic activity, which assisted in screening important descriptors related to performance. Key descriptors identified include the binding energy (E_b), the d-band centre of the Pt atom, and the adsorption energy of hydroxyl (E_{OH}), all of which significantly impact catalytic activity. Authors emphasised that the electronic properties of DTM MXenes markedly enhanced the catalytic performance of the SACS, resulting in improved conductivity and stability. Overall, the findings provide a comprehensive understanding of the catalytic mechanisms and serve as a guide for developing efficient electrocatalysts for future applications.[145]

4.1.4. Bimetallic catalysts. Bimetallic catalysts have garnered significant attention in this domain due to their unique electronic and geometric properties, which can enhance catalytic activity and stability compared to their monometallic counterparts. These catalysts leverage the synergistic effects of two different metals, optimizing the adsorption and activation of reactants involved in the water-splitting process. By harnessing vast datasets and advanced algorithms, ML can predict the performance of bimetallic catalysts, identify optimal compositions, and streamline the discovery of new materials. This review will explore the advancements in bimetallic catalysts for water splitting, emphasizing their efficiency and the transformative role of machine learning in enhancing the effectiveness of these catalytic systems.[146–149]

Through rigorous analysis and detailed experimentation, the article presents results that underscore MXene-based double-atom catalysts (DACS) incorporating non-noble metals like Fe, Co, and Ni exhibit superior bifunctional activity for ORR/OER compared to traditional catalysts like Pt and IrO$_2$, with tunable properties through strain engineering. The research highlights that MXene-based DACS specifically those modified with non-noble metals like Fe, Co, and Ni, exhibit ultrahigh catalytic activities for both ORR and OER. The performance of these DACS surpasses that of traditional uni-functional catalysts, such as Pt (111) for ORR and IrO$_2$ (110) for OER, which are known for their low overpotentials. This reveals that strain can significantly influence the catalytic activities of MXene-based DACS. It shows that novel approach to tuning the catalytic behaviour of flexible MXenes, allowing for enhanced performance in practical applications. An electrochemical model was developed using DFT and theoretical polarization curves to explain the underlying mechanisms of the observed catalytic activities. This model aligns well with experimental results, providing a deeper understanding of how these catalysts function. The research indicates that the excellent catalytic performance of the MXene-based DACS is primarily due to their electron-capturing capability. The synergistic interactions between the metal adsorbents (Fe, Co, Ni) and the MXene substrate play a crucial role in enhancing the overall catalytic activity.[148]

The data presented in the article clearly demonstrates the bifunctional catalytic activity of YBaCo$_4$O$_{7.3}$ oxide for both OER and ORR. The YBaCo$_4$O$_{7.3}$ oxide exhibits an OER activity of 1.18 mA cm^{-2} at an overpotential of 400 mV. This performance is noteworthy as it surpasses the benchmark catalyst, LaNiO$_3$. For the ORR, the oxide shows an activity of 2.76 mA cm^{-2} at an overpotential of −537 mV, indicating its effectiveness in this reaction as well. The study rationalizes the electrochemical activity of YBaCo$_4$O$_{7.3}$ using the Zaanen–Sawatzky–Allen model. This theoretical framework helps in understanding the electronic properties and behaviour of the catalyst. The presence of overstoichiometric oxygen in the material leads to the formation of octahedral Co positions, which serve as active sites for the OER. This indicates that the oxygen content significantly influences the catalytic properties of the oxide. Showed that the large Coulomb on-site repulsion energy for Co in tetrahedral coordination is crucial for its activity in the ORR. This suggests that the coordination environment of cobalt ions plays a vital role in determining the catalytic efficiency. These results collectively demonstrate that $YBaCo_4O_{7.3}$ is a promising bifunctional catalyst for both OER and ORR, with its performance being influenced by its structural and electronic properties.[146]

Takigawa *et al.* presents significant findings regarding the use of ML techniques to predict the d-band centres of metals and bimetals, which are crucial for understanding catalytic activity in heterogeneous catalysis. The study demonstrates that state-of-the-art ML methods can predict the d-band centres of metals and bimetals with minimal central processing unit time compared to traditional first-principles DFT calculations. This efficiency is crucial for accelerating the design of catalysts without extensive trial-and-error experiments. The authors utilized e_d − E_F data for 11 metals (Fe, Co, Ni, Cu, Ru, Rh, Pd, Ag, Ir, Pt, and Au) and their associated pairwise bimetallic alloys (110 pairs). The GBR method was described as particularly effective, achieving reasonable predictions for both monometallic and bimetallic surfaces. The study found that using 6 readily available descriptors yielded superior prediction accuracy compared to using fewer descriptors. Developed a strong correlation between the d-band centre values and catalytic activity, supporting the notion that the d-band centre can serve as a reliable descriptor for predicting reactivity trends in heterogeneous catalysis. Authors emphasizes that the d-band model, in conjunction with linear energy relations, is vital for understanding activity trends in transition metal-catalysed reactions. The GBR model provided insights into the importance of various descriptors used in the predictions. The feature importance scores indicated which descriptors contributed most significantly to the predictive performance, allowing for a more informed selection of parameters in future studies. The findings suggest that ML techniques can significantly enhance the discovery and design of catalytic materials, potentially reducing the reliance on traditional experimental methods.[147]

Many traditional computational approaches aim to find global minima of potential energy surfaces. However, global optimization remains a major challenge, limiting the application of these methods in exploring new materials due to the vast parameter search space. Traditional approaches heavily rely on human experiences and intuitions to determine which calculated geometries, reaction states, or materials properties warrant further exploration. Analysis and interpretation of results often require domain experts in solid-state chemistry or electrocatalysis for rational analysis and mechanistic insights. Comparing the performance of different machine learning algorithms, descriptor choices, and training data quality is an important and active research area. Optimizing the implementation of specific AI or ML approaches to achieve efficient global minimum search is a well-studied topic, highlighting the advantage of AI or ML over traditional methods. However, benchmarking and comparing AI or ML algorithms with traditional methods is challenging due to the growing number of proposed algorithms and methodologies, biased testing examples, and the subjective nature of material selection. Despite advancements, well-designed catalysts for many materials systems still rely on trial-and-error methodologies. Continual improvement of theoretical insights, algorithms, and ML regression and optimization is necessary to enhance their application in catalyst development for hydrogen evolution and other crucial reactions.

5. Prospects and perspectives

The application of AI and ML to the design of electrocatalysts for electrochemical water splitting and electrocatalytic CO_2 reduction allows accurate prediction of catalyst performance without the need for expensive theoretical calculations and extensive experiments. There are many ML models such as linear regression, random forest, deep learning and others, each with its own features and outlines to effectively analyse data and predict electrocatalyst parameters. The main challenge in using ML for HER and CO_2RR electrocatalysts is to select descriptors, *i.e.* key characteristics of the catalyst that affect its activity, selectivity, and stability. Structure, morphology, composition and surface area are the main descriptors that are widely used in the design of electrocatalysts. It is also important to be able to build models efficiently, considering different types of data and their interrelationships to ensure optimal results.

Despite the practical successes of ML, its application in electrocatalysis is still at an early stage of development, and many research opportunities remain unexplored. There is still a lot of research space for improvement of electrocatalysts with usage of ML and AI. First, the multi-step learning process should be improved in order to increase the efficiency and accuracy of ML in the search for promising electrocatalysts. Such studies exist; however, they are still quite few. It is worth noting that since the beginning of 2024, this approach is beginning to gain momentum, for example, the new multi-stage ML platform has combined data mining, active learning and domain adaptation, effectively narrowing the field of electrocatalyst research.[150] This method has led to the successful identification of advanced catalysts such as Ru–Mn–Ca peroxide, demonstrating the potential for systematic research compared to traditional methods. In general, research using multi-stage ML learning highlights the importance of the characteristics of the active centre in alloy catalysts, where an effective ML model makes it possible to predict bifunctional oxygen electrocatalysts based on their inherent characteristics, achieving remarkable prediction accuracy. The efficiency of this model is 150 000 times higher than that of traditional methods, which highlights the potential of discovering a catalyst to accelerate processes over many years.[151]

Secondly, combining different types of information together is an effective way to achieve high-precision forecasting. Recent studies have shown that the use of random forest algorithms makes it possible to achieve an average relative error of 9.3% when predicting the activity of oxygen release by TM oxides. In addition, collective learning methods have improved the prediction accuracy for transition metal phosphides, reaching an R^2 value of more than 0.85.[152] The inclusion of structural and electronic characteristics in the ML models led to high-performance forecasts, and algorithms for increasing the illumination gradient give an R^2 equal to 0.921.[153] Third, more ML models should be developed that do not depend on functional engineering, such as graph networks, which are less sensitive to the choice of atomic descriptors and support both relational reasoning and combinatorial generalization.

The development of ML models that minimize dependence on functional engineering, especially through the use of graph networks, provides significant advantages in materials science and other fields. Graph neural networks (GNNs) have become powerful tools supporting relational reasoning and combinatorial generalization, which makes them less sensitive to the choice of atomic descriptors. GNNs can achieve performance comparable to traditional models without careful function design, as demonstrated by Cong and Fung, who used simple graph structures with atoms as nodes and distances as edges.[154] Korolev *et al.* emphasized that graph convolutional neural networks (GCNNS) can predict properties in various chemical domains using a minimal set of descriptors, which increases their applicability and interpretability.[155] Also Sinha *et al.* we investigated the possibilities of logical generalization of GNNS, showing that these models can effectively adapt to new tasks using inductive distortions of relationships, which is crucial for complex reasoning problems.[156]

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Author contributions

The manuscript was written through the contribution of all authors. Anastasiia P. Dmitrieva – conceptualisation, writing – original draft. Ali Ayati – writing – original draft. Elizaveta A. Romanenko – writing – original draft. Chantal T. Tracey – writing – original draft, writing – review & editing. Anna S. Fomkina – writing – original draft. Pavel V. Krivoshapkin – conceptualization, writing – review & editing, supervision. Elena F. Krivoshapkina – conceptualisation, writing – review & editing, funding acquisition, supervision.

Conflicts of interest

All authors declare no competing financial or non-financial interests.

Acknowledgements

This work was supported by the Russian Ministry of Science and Higher Education “Science and Universities” National Project state task No. FSER-2022-0002. E. F. K. also acknowledges the Priority 2030 Federal Academic Leadership Program for infrastructural support.

References

1 E. R. Cofell, U. O. Nwabara, S. S. Bhargava, D. E. Henckel and P. J. A. Kenis, *ACS Appl. Mater. Interfaces*, 2021, **13**, 15132–15142.

2 L. Ge, H. Rabiee, M. Li, S. Subramanian, Y. Zheng, J. H. Lee, T. Burdyny and H. Wang, *Chem*, 2022, **8**, 663–692.

3 S. N. Steinmann, Q. Wang and Z. W. Seh, *Mater. Horiz.*, 2023, **10**, 393–406.

4 Calypso Database, **http://db.calypso.cn/**, accessed 2024-26-09.

5 Universal Structure Predictor: Evolutionary Xthallography, **https://uspex-team.org/en/uspex/overview**, accessed 2024-26-09.

6 B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke and I. Foster, *JOM*, 2016, **68**, 2045–2052.

7 L. Chen, X. Zhang, A. Chen, S. Yao, X. Hu and Z. Zhou, *Chin. J. Catal.*, 2022, **43**, 11–32.

8 H. Mai, T. C. Le, D. Chen, D. A. Winkler and R. A. Caruso, *Chem. Rev.*, 2022, **122**, 13478–13515.

9 D. Fébba, K. Egbo, W. Callahan and A. Zakutayev, *arXiv*, 2024, preprint, arXiv:2406.16224, DOI: **10.48550/arXiv.2406.16224**.

10 D. Jha, L. Ward, A. Paul, W. Liao, A. Choudhary, C. Wolverton and A. Agrawal, *Sci. Rep.*, 2018, **8**, 17593.

11 M. N. O. Sadiku, G. K. Suman and S. M. Musa, *Int. J. Adv. Sci. Res. Eng.*, 2021, **7**, 77–81.

12 A. Y.-T. Wang, R. J. Murdock, S. K. Kauwe, A. O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson and T. D. Sparks, *Chem. Mater.*, 2020, **32**, 4954–4965.

13 J. Kim, D. Kang, S. Kim and H. W. Jang, *J. Am. Chem. Soc. Mater. Lett.*, 2021, **3**, 1151–1171.

14 A. Bender, N. Schneider, M. Segler, W. Patrick Walters, O. Engkvist and T. Rodrigues, *Nat. Rev. Chem*, 2022, **6**, 428–442.

15 C. Draxl and M. Scheffler, *MRS Bull.*, 2018, **43**, 676–682.

16 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 11002.

17 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.

18 F. A. Rasmussen and K. S. Thygesen, *J. Phys. Chem. C*, 2015, **119**, 13169–13183.

19 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, *Sci. Data*, 2019, **6**, 75.

20 J. Zhou, L. Shen, M. D. Costa, K. A. Persson, S. P. Ong, P. Huck, Y. Lu, X. Ma, Y. Chen, H. Tang and Y. P. Feng, *Sci. Data*, 2019, **6**, 86.

21 S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo and O. Levy, *Comput. Mater. Sci.*, 2012, **58**, 227–235.

22 I. Karakaya, *ASM Handbook: Alloy Phase Diagrams*, ASM International, 2001, vol. 3.

23 Y. Xu, M. Yamazaki and P. Villars, *Jpn. J. Appl. Phys.*, 2011, **50**, 11RH02.

24 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.

25 R. Guha and E. Willighagen, *J. Cheminf.*, 2020, **12**, 4.

26 S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs and A. Le Bail, *Nucleic Acids Res.*, 2012, **40**, D420–D427.

27 *High Performance Alloys Database (HPAD)*, CINDAS LLC, **https://cindasdata.com**.

28 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, *npj Comput. Mater.*, 2015, **1**, 15010.

29 L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick and Z. Ulissi, *J. Am. Chem. Soc. Catal.*, 2021, **11**, 6059–6072.

30 Computational Chemistry Comparison and Benchmark Database, **https://cccbdb.nist.gov/**, accessed 2024-26-09.

31 S. Pande, L. Arkesteijn and L. Bastidas, Complexity regularized hydrological model selection, in *Proceedings of the 7th International Congress on Environmental Modelling and Software (iEMSs)*, San Diego, CA, 2014, pp. 15–19.

32 J. Kim and X. Wang, Inductive Global and Local Manifold Approximation and Projection, *arXiv*, 2024, preprint, arXiv:2406.08097, DOI: **10.48550/arXiv.2406.08097**.

33 H. Ma, Y. Jiao, W. Guo, X. Liu, Y. Li and X.-D. Wen, *J. Phys. Chem. C*, 2020, **124**, 15920–15927.

34 Y. Hu, J. Chen, Z. Wei, Q. He and Y. Zhao, *J. Mater. Inf.*, 2023, **3**, 18.

35 K. Tran and Z. W. Ulissi, *Nat. Catal.*, 2018, **1**, 696–703.

36 J. Liu, W. Luo, L. Wang, J. Zhang, X. Fu and J. Luo, *Adv. Funct. Mater.*, 2022, **32**, 2110748.

37 M. Gheytanzadeh, A. Baghban, S. Habibzadeh, K. Jabbour, A. Esmaeili, A. H. Mashhadzadeh and A. Mohaddespour, *Sci. Rep.*, 2022, **12**, 20859.

38 J. Li, N. Wu, J. Zhang, H.-H. Wu, K. Pan, Y. Wang, G. Liu, X. Liu, Z. Yao and Q. Zhang, *Nano-Micro Lett.*, 2023, **15**, 227.

39 Z. Li, S. Wang, W. S. Chin, L. E. Achenie and H. Xin, *J. Mater. Chem. A*, 2017, **5**, 24131–24138.

40 S. Ma and Z.-P. Liu, *J. Am. Chem. Soc.*, 2020, **10**, 13213–13226.

41 Z. Tian, C. Priest and L. Chen, *Adv. Theory Simul.*, 2018, **1**, 1800004.

42 Q. Zhang, Y. Wang, J. Zhang, Y. Yue and G. Qian, *Appl. Catal., A*, 2024, **683**, 119825.

43 R. Ding, X. Wang, A. Tan, J. Li and J. Liu, *J. Am. Chem. Soc. Catal.*, 2023, **13**, 13267–13281.

44 A. Malek, M. J. Eslamibidgoli, M. Mokhtari, Q. Wang, M. H. Eikerling and K. Malek, *ChemPhysChem*, 2019, **20**, 2946–2955.

45 C. W. Glass, A. R. Oganov and N. Hansen, *Comput. Phys. Commun.*, 2006, **175**, 713–720.

46 Y. Wang, J. Lv, L. Zhu, S. Lu, K. Yin, Q. Li, H. Wang, L. Zhang and Y. Ma, *J. Phys.: Condens. Matter*, 2015, **27**, 203203.

47 P. D. Pedersen, M. M. Melander, T. Bligaard, T. Vegge, K. Honkala and H. A. Hansen, *J. Phys. Chem. C*, 2023, **127**, 18855–18864.

48 N. K. Singh, P. Kumar, A. Yadav and V. C. Srivastava, *J. Colloid Interface Sci.*, 2024, **654**, 895–905.

49 Y. Zhang, Z. Gu, H. Yang, J. Gao, F. Peng and H. Yang, *J. Colloid Interface Sci.*, 2023, **652**, 250–257.

50 Y. Meng, L. Ying, Y. Tao, L. Ma, B. Li, Y. Xing, X. Liu, Y. Ma and X. Wen, *Langmuir*, 2024, **40**, 10663–10675.

51 T. Sun, J. Wu, X. Lu and X. Tang, *J. Mol. Catal.*, 2024, **559**, 114047.

52 F. Xue, X. Lai and Y. Xu, *ChemCatChem*, 2024, e202400590.

53 S. Soodi, J.-J. Zhang, J. Zhang, Y. Liu, M. Lashgari, S. Zafeiratos, A. Züttel, K. Zhao and W. Luo, *Chem. Synth.*, 2024, **4**, 44.

54 S. Nitopi, E. Bertheussen, S. B. Scott, X. Liu, A. K. Engstfeld, S. Horch, B. Seger, I. E. L. Stephens, K. Chan, C. Hahn, J. K. Nørskov, T. F. Jaramillo and I. Chorkendorff, *Chem. Rev.*, 2019, **119**, 7610–7672.

55 R. M. Arán-Ais, F. Scholten, S. Kunze, R. Rizo and B. Roldan Cuenya, *Nat. Energy*, 2020, **5**, 317–325.

56 L. Yuan, S.-F. Hung, Z.-R. Tang, H. M. Chen, Y. Xiong and Y.-J. Xu, *J. Am. Chem. Soc. Catal.*, 2019, **9**, 4824–4833.

57 K. P. Kuhl, E. R. Cave, D. N. Abram and T. F. Jaramillo, *Energy Environ. Sci.*, 2012, **5**, 7050–7059.

58 M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.-S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S.-C. Lo, A. Ip, Z. Ulissi and E. H. Sargent, *Nature*, 2020, **581**, 178–183.

59 Z. W. Seh, J. Kibsgaard, C. F. Dickens, I. Chorkendorff, J. K. Nørskov and T. F. Jaramillo, *Science*, 2017, **355**, eaad4998.

60 P. Liu, C. Verdi, F. Karsai and G. Kresse, *Phys. Rev. B*, 2022, **105**, L060102.

61 J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. E. Roitberg, *Nat. Commun.*, 2019, **10**, 2903.

62 L. Zhang, D.-Y. Lin, H. Wang, R. Car and W. E, *Phys. Rev. Mater.*, 2019, **3**, 023804.

63 S. Huang, C. Shang, P. Kang, X. Zhang and Z. Liu, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2019, **9**(6), e1415.

64 D. Cheng, Z.-J. Zhao, G. Zhang, P. Yang, L. Li, H. Gao, S. Liu, X. Chang, S. Chen, T. Wang, G. A. Ozin, Z. Liu and J. Gong, *Nat. Commun.*, 2021, **12**, 395.

65 A. A. Peterson, F. Abild-Pedersen, F. Studt, J. Rossmeisl and J. K. Nørskov, *Energy Environ. Sci.*, 2010, **3**, 1311.

66 F. Studt, F. Abild-Pedersen, J. B. Varley and J. K. Nørskov, *Catal. Lett.*, 2013, **143**, 71–73.

67 Z. Gariepy, G. Chen, A. Xu, Z. Lu, Z. W. Chen and C. V. Singh, *Energy Adv.*, 2023, **2**, 410–419.

68 Y. Guo, X. He, Y. Su, Y. Dai, M. Xie, S. Yang, J. Chen, K. Wang, D. Zhou and C. Wang, *J. Am. Chem. Soc.*, 2021, **143**, 5755–5762.

69 J. Abdi, F. Hadavimoghaddam, M. Hadipoor and A. Hemmati-Sarapardeh, *Sci. Rep.*, 2021, **11**, 24468.

70 K. Gubsch, R. Bence, L. T. Glasby and P. Z. Moghadam, *DigiMOF: A Database of MOF Synthesis Information Generated via Text Mining*, 2022.

71 J. Abdi, M. Hadipoor, F. Hadavimoghaddam and A. Hemmati-Sarapardeh, *Chemosphere*, 2022, **287**, 132135.

72 F. Sha, H. Xie, K. O. Kirlikovali, W. Gong, Y. Chen and O. K. Farha, *ACS Mater. Lett.*, 2024, **6**, 1396–1403.

73 Y. Teng and G. Shan, *APL Mater.*, 2024, **12**, 081115.

74 J. A. Rodriguez and F. Illas, *Phys. Chem. Chem. Phys.*, 2012, **14**, 427–438.

75 D. R. Harper, A. Nandy, N. Arunachalam, C. Duan, J. P. Janet and H. J. Kulik, *J. Chem. Phys.*, 2022, **21**, 074101.

76 C. Duan, A. Nandy, H. Adamji, Y. Roman-Leshkov and H. J. Kulik, *J. Chem. Theory Comput.*, 2022, **18**, 4282–4292.

77 X. Wang, W. Ju, L. Liang, M. Riyaz, A. Bagger, M. Filippi, J. Rossmeisl and P. Strasser, *Angew. Chem., Int. Ed.*, 2024, **63**, e202401821.

78 Z. Liu, A. Ma, Z. Wang, Z. Ding, Y. Pang, G. Fan and H. Xu, *J. Mol. Catal.*, 2024, **559**, 114040.

79 I. Street, L. Lu, S. Mosali, Z. Chen and W. E. Mustain, *El. Chem. Soc.*, 2024, MA2024-01, p. 83.

80 G. Hwa Jeong, Y. Chuan Tan, J. Tae Song, G.-Y. Lee, H. Jin Lee, J. Lim, H. Young Jeong, S. Won, J. Oh and S. Ouk Kim, *Chem. Eng. J.*, 2021, **426**, 131063.

81 Z. Wu, J. Liu, B. Mu, X. Xu, W. Sheng, W. Tao and Z. Li, *Appl. Surf. Sci.*, 2024, **648**, 159027.

82 M. Sun and B. Huang, *Adv. Energy Mater.*, 2024, **14**, 2400152.

83 T.-T. Zhuang, Y. Pang, Z.-Q. Liang, Z. Wang, Y. Li, C.-S. Tan, J. Li, C. T. Dinh, P. De Luna, P.-L. Hsieh, T. Burdyny, H.-H. Li, M. Liu, Y. Wang, F. Li, A. Proppe, A. Johnston, D.-H. Nam, Z.-Y. Wu, Y.-R. Zheng, A. H. Ip, H. Tan, L.-J. Chen, S.-H. Yu, S. O. Kelley, D. Sinton and E. H. Sargent, *Nat. Catal.*, 2018, **1**, 946–951.

84 J. T. Billy and A. C. Co, *Appl. Catal., B*, 2018, **237**, 911–918.

85 X.-Q. Wang, Q. Chen, Y.-J. Zhou, H.-M. Li, J.-W. Fu and M. Liu, *ASEM*, 2022, vol. 1, p. 100023.

86 L. Xie, J. Liang, C. Priest, T. Wang, D. Ding, G. Wu and Q. Li, *Chem. Commun.*, 2021, **57**, 1839–1854.

87 D. H. Mok, H. Li, G. Zhang, C. Lee, K. Jiang and S. Back, *Nat. Commun.*, 2023, **14**, 7303.

88 M. T. Tang, H. Peng, P. S. Lamoureux, M. Bajdich and F. Abild-Pedersen, *Appl. Catal., B*, 2020, **279**, 119384.

89 R. Stottko, E. Dziadyk-Stopyra and B. M. Szyja, *Catalysts*, 2023, **13**, 1470.

90 R. Sen, S. Das, A. Nath, P. Maharana, P. Kar, F. Verpoort, P. Liang and S. Roy, *Front. Chem.*, 2022, **10**, 861604.

91 Z. Yan, J. L. Hitt, J. A. Turner and T. E. Mallouk, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 12558–12563.

92 X. Xiao, L. Yang, W. Sun, Y. Chen, H. Yu, K. Li, B. Jia, L. Zhang and T. Ma, *Small*, 2022, **18**, 2105830.

93 A. Wu, Y. Xie, H. Ma, C. Tian, Y. Gu, H. Yan, X. Zhang, G. Yang and H. Fu, *Nano Energy*, 2018, **44**, 353–363.

94 S. S. Siwal, W. Yang and Q. Zhang, *J. Energy Chem.*, 2020, **51**, 113–133.

95 H. Ding, D. Zha, S. Han and N. Jiang, *J. Taiwan Inst. Chem. Eng.*, 2023, **151**, 105135.

96 A. Lasia, *Int. J. Hydrogen Energy*, 2019, **44**, 19484–19518.

97 W. Dai and Z. Zhang, *Chem Catal.*, 2021, **1**, 1349–1351.

98 F. Saad, J. D. Comparot, R. Brahmi, M. Bensitel and L. Pirault-Roy, *Appl. Catal., A*, 2017, **544**, 1–9.

99 T. Shen, S. Chen, R. Zeng, M. Gong, T. Zhao, Y. Lu, X. Liu, D. Xiao, Y. Yang, J. Hu, D. Wang, H. L. Xin and H. D. Abruña, *J. Am. Chem. Soc. Catal.*, 2020, **10**, 9977–9985.

100 V. Pifferi, C. Chan-Thaw, S. Campisi, A. Testolin, A. Villa, L. Falciola and L. Prati, *Molecules*, 2016, **21**, 261.

101 X. Yang, Z. Liang, S. Chen, M. Ma, Q. Wang, X. Tong, Q. Zhang, J. Ye, L. Gu and N. Yang, *Small*, 2020, **16**, 2004727.

102 C. Zou, X. Guo, F. Wang, R. Tian, Y. Hou, F. Liu, H. Yang and J. Solgel, *Sci. Technol.*, 2022, **103**, 505–514.

103 A. Prašnikar and B. Likozar, *React. Chem. Eng.*, 2022, **7**, 1073–1082.

104 N. Cheng, L. Zhang, K. Doyle-Davis and X. Sun, *Environ. Eng. Res.*, 2019, **2**, 539–573.

105 L. Zhou, Z. Yang, D. Wei, H. Zhang and W. Lu, *Front. Chem.*, 2022, **10**, 882787.

106 X. Kong, C. Wang, H. Zheng, Z. Geng, J. Bao and J. Zeng, *Sci. China: Chem.*, 2021, **64**, 1096–1102.

107 S. Haschke, D. Pankin, V. Mikhailovskii, M. K. S. Barr, A. Both-Engel, A. Manshina and J. Bachmann, *Beilstein J. Nanotechnol.*, 2019, **10**, 157–167.

108 J. Li, J. Zeng, F. Zhao, X. Sun, S. Wang and X. F. Lu, *Energy Fuels*, 2024, **38**, 11521–11540.

109 B. Siritanaratkul, C. Eagle and A. J. Cowan, *Acc. Chem. Res.*, 2022, **55**, 955–965.

110 J. Wang, W. Zang, X. Liu, J. Sun, S. Xi, W. Liu, Z. Kou, L. Shen and J. Wang, *Small*, 2024, **20**, 2309427.

111 X. Li, R. Chiong, Z. Hu and A. J. Page, *J. Phys. Chem. Lett.*, 2021, **12**, 7305–7311.

112 H. Ooka, M. E. Wintzer and R. Nakamura, *J. Am. Chem. Soc. Catal.*, 2021, **11**, 6298–6303.

113 Y. Ding, M. Zeng, Q. Zheng, J. Zhang, D. Xu, W. Chen, C. Wang, S. Chen, Y. Xie, Y. Ding, S. Zheng, J. Zhao, P. Gao and L. Fu, *Nat. Commun.*, 2021, **12**, 5886.

114 H. S. Kim, *J. Mater. Chem. A*, 2021, **9**, 11553–11562.

115 B. Ryu, L. Wang, H. Pu, M. K. Y. Chan and J. Chen, *Chem. Soc. Rev.*, 2022, **51**, 1899–1925.

116 Y. Cui, X. Guo, J. Zhang, X. Li, X. Zhu and W. Huang, *Nano Res.*, 2022, **15**, 677–684.

117 F. Zhou, Y. Zhou, G.-G. Liu, C.-T. Wang and J. Wang, *Rare Met.*, 2021, **40**, 3375–3405.

118 J. Zhang, Y. Zhao, X. Guo, C. Chen, C.-L. Dong, R.-S. Liu, C.-P. Han, Y. Li, Y. Gogotsi and G. Wang, *Nat. Catal.*, 2018, **1**, 985–992.

119 X. Wan, Z. Zhang, W. Yu and Y. Guo, *Mater. Rep.: Energy*, 2021, **1**, 100046.

120 E. O. Ebikade, Y. Wang, N. Samulewicz, B. Hasa and D. Vlachos, *React. Chem. Eng.*, 2020, **5**, 2134–2147.

121 N. Omidvar, S. Wang, Y. Huang, H. S. Pillai, A. Athawale, S. Wang, L. E. K. Achenie and H. Xin, *Electrochem. Sci. Adv.*, 2024, e202300028.

122 E. Xie and X. Wang, Computational and AI-Assisted Design of Dual-Metal Single-Atom Catalyst for Oxygen Reduction Reaction, *ChemRxiv*, preprint, 2023, DOI: **10.26434/chemrxiv-2023-cxfm5-v2**.

123 E. Jagannath Jijaba Kadam, *Int. J. Recent Innov. Trends*, 2023, **11**, 2123–2135.

124 J.-L. Li, Y.-F. Li and Z.-P. Liu, *JACS Au*, 2023, **3**, 1162–1175.

125 C. Zhang, T. Shen, D. Guo, L. Tang, K. Yang and H. Deng, *InfoMat*, 2020, **2**, 1034–1056.

126 E. Tenuta, C. Zheng and O. Rubel, *Sci. Rep.*, 2016, **6**, 37654.

127 X. Jin, K. Ma, J. Chakkamalayath, J. Morsby and H. Gao, *ACS Energy Lett.*, 2022, **7**, 610–616.

128 M. Xiao, M. Hao, M. Lyu, E. G. Moore, C. Zhang, B. Luo, J. Hou, J. Lipton-Duffin and L. Wang, *Adv. Funct. Mater.*, 2019, **23**, 1905683.

129 A. A. Farghaly, M. Ferrandon, D. Schwalbe-Koda, J. Damewood, J. Karaguesian, R. Gómez-Bombarelli and D. J. Myers, *ECS Meet. Abst.*, 2022, MA2022-01, p. 1359.

130 M. A. Ullah, M. Rizwan and K. N. Riaz, *J. Phys. Chem. Solids*, 2023, **180**, 111457.

131 X. Wang, B. Xiao, Y. Li, Y. Tang, F. Liu, J. Chen and Y. Liu, *Appl. Surf. Sci.*, 2020, **531**, 147323.

132 M. Crosta, M. G. Lattanzi, C. Le Poncin-Lafitte, M. Gai, Q. Zhaoxiang and A. Vecchiato, On the principle of Astrometric Gravitational Wave Antenna, *arXiv*, 2022, preprint, arXiv:2203.12760, DOI: **10.48550/arXiv.2203.12760**.

133 D. Lach, U. Zhdan, A. Smolinski and J. Polanski, *Int. J. Mol. Sci.*, 2021, **22**, 5176.

134 Z. Yao, Y. Lum, A. Johnston, L. M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E. H. Sargent and Z. W. Seh, *Nat. Rev. Mater.*, 2022, **8**, 202–215.

135 X. Mao, L. Wang, Y. Xu, P. Wang, Y. Li and J. Zhao, *Comput. Mater.*, 2021, **7**, 46.

136 A. Baghban, S. Habibzadeh and F. Zokaee Ashtiani, *Sci. Rep.*, 2021, **11**, 21911.

137 S. Wei, S. Baek, H. Yue, M. Liu, S. J. Yun, S. Park, Y. H. Lee, J. Zhao, H. Li, K. Reyes and F. Yao, *J. Electrochem. Soc.*, 2021, **168**, 126523.

138 R. B. Wexler, J. M. P. Martirez and A. M. Rappe, *J. Am. Chem. Soc.*, 2018, **140**, 4678–4683.

139 A. J. Parker, G. Opletal and A. S. Barnard, *J. Appl. Phys.*, 2020, **128**, 14301.

140 N. K. Pandit, D. Roy, S. C. Mandal and B. Pathak, *J. Phys. Chem. Lett.*, 2022, **13**, 7583–7593.

141 Y. Jia, X. Hou, Z. Wang and X. Hu, *ACS Sustainable Chem. Eng.*, 2021, **9**, 6130–6147.

142 M. Wang and H. Zhu, *ACS Catal.*, 2021, **11**, 3930–3937.

143 J. Greeley and J. K. Nørskov, *Surf. Sci.*, 2007, **601**, 1590–1598.

144 D. Nguyen Nguyen, M.-C. Kim, U. Baeck, J. Lim, N. Shin, J. Kim, H. Choi, H. Seok Park, U. Sim, J. Kyu Kim and C. Jung Kyu Kim, *arXiv*, 2022, preprint, arXiv:2210.14701, DOI: **10.48550/arXiv.2210.14701**.

145 N. Ma, Y. Zhang, Y. Wang, C. Huang, J. Zhao, B. Liang and J. Fan, *Appl. Surf. Sci.*, 2023, **628**, 157225.

146 M. A. Kirsanova, V. D. Okatenko, D. A. Aksyonov, R. P. Forslund, J. T. Mefford, K. J. Stevenson and A. M. Abakumov, *J. Mater. Chem. A*, 2019, **7**, 330–341.

147 I. Takigawa, K. Shimizu, K. Tsuda and S. Takakusagi, *Nano inf.*, 2018, 45–64.

148 B. Wei, Z. Fu, D. Legut, T. C. Germann, S. Du, H. Zhang, J. S. Francisco and R. Zhang, *Adv. Mater.*, 2021, **33**, 2102595.

149 C. Wei, D. Shi, Z. Yang, Z. Xue, S. Liu, R. Li and T. Mu, *J. Mater. Chem. A*, 2023, **11**, 18168–18178.

150 R. Ding, J. Liu, K. Hua, X. Wang, X. Zhang, M. Shao, Y. Chen and J. Chen, Leveraging Data Mining, Active Learning, and Domain Adaptation in a Multi-Stage, Machine Learning-Driven Approach for the Efficient Discovery of Advanced Acidic Oxygen Evolution Electrocatalysts, *arXiv*, 2024, preprint, arXiv:2407.04877, DOI: **10.48550/arXiv.2407.04877**.

151 C. Wang, B. Wang, C. Wang, Z. Chang, M. Yang and R. Wang, *ACS Appl. Mater. Interfaces*, 2024, **16**, 16050–16061.

152 L. Zhang, J. Hou, H. Ji, D. Meng, J. Qi and X. San, *Catalysts*, 2024, **14**, 513.

153 J. Zhang, Y. Wang, X. Zhou, C. Zhong, K. Zhang, J. Liu, K. Hu and X. Lin, *Nanoscale*, 2023, **15**, 11072–11082.

154 G. Cong and V. Fung, Machine Learn, *Sci. Technol.*, 2023, **4**, 035030.

155 V. Korolev, A. Mitrofanov, A. Korotcov and V. Tkachenko, *J. Chem. Inf. Model.*, 2020, **60**, 22–28.

156 K. Sinha, S. Sodhani, J. Pineau and W. L. Hamilton, Evaluating logical generalization in graph neural networks, *arXiv*, 2020, preprint, arXiv:2003.06560, DOI: **10.48550/arXiv.2003.06560**.