Chinese Chemical Letters 37 (2026) 112021

Contents lists available at ScienceDirect

Chinese Chemical Letters

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

Recent advances in machine learning-driven discovery of alloy electrocatalysts for hydrogen evolution reaction

Na Qin[a,1], Wenxin Guo[b,1], Fangxiu Li[a], Houfeng Zhang[a], Hong Liu[a], Chang Zhang[a,c], Lipiao Bao[a,c], Lei Liu[d], Muneerah Alomar[e], Siqi Zhao[a,*], Jian Zhang[a,*], Xing Lu[a,c,*]

[a] School of Chemistry and Chemical Engineering, Hainan University, Haikou 570228, China
[b] Key Laboratory of Precision and Intelligent/School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China
[c] School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[d] School of HUST-Wuxi Research Institute, Wuxi 214174, China
[e] Department of Physics, College of Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

ARTICLE INFO

Article history:
Received 24 June 2025
Revised 20 October 2025
Accepted 22 October 2025
Available online 24 October 2025

Keywords:
Hydrogen evolution reaction
Machine learning
Alloy catalysts
Artificial intelligence
Deep learning

ABSTRACT

The hydrogen evolution reaction (HER) is a pivotal process for clean energy conversion, yet the development of efficient and cost-effective electrocatalysts remains a major challenge. Alloy catalysts, with their tunable electronic properties and promising catalytic performance, have shown great potential for HER. However, the design of component types and ratios, along with structural optimization, has largely relied on traditional trial-and-error approaches, which are very complex and time-consuming. The rise of machine learning (ML) provides an efficient strategy for discovering and optimizing alloy catalysts by enabling rapid analysis of extensive experimental and simulation datasets. This review highlights the recent advances in applying ML techniques for the design and optimization of alloy electrocatalysts for HER, covering binary and multinary (ternary, quaternary and high-entropy alloys). In particular, by employing supervised learning and deep learning techniques, ML has achieved remarkable success in the rapid screening of alloy catalysts and in improving prediction accuracy. It also demonstrates the merit and capability of ML in accelerating this process. In the end, we discuss current challenges and future prospects for integrating ML into advanced HER catalysis, highlighting its potential to revolutionize catalyst development and promote sustainable hydrogen energy solutions.

© 2025 Published by Elsevier B.V. on behalf of Chinese Chemical Society and Institute of Materia Medica, Chinese Academy of Medical Sciences.

1. Introduction

The growing global energy demand and the pressing need to mitigate climate change have intensified the exploration of sustainable and energy-efficient solutions [1,2]. Hydrogen, with its high energy density and zero-carbon emissions during use, plays a crucial role as an energy carrier in the transition toward a carbon-neutral future [3–5]. Among the various hydrogen production methods, electrochemical water splitting presents a highly promising approach by converting renewable energy into storable chemical energy [6,7]. However, its practical implementation is hindered by the slow reaction kinetics, therefore requiring highly efficient and cost-effective electrocatalysts to achieve the scalable production [8]. Noble metal catalysts such as Pt have long been the benchmark for the electrochemical hydrogen evolution reaction (HER) due to their exceptional catalytic activity. However, their high cost and scarcity limit large-scale applications [9]. Compared with monometallic catalysts, alloy catalysts offer a more cost-effective alternative due to their lower material cost and tunable electronic structures [10]. These features allow for the optimization of hydrogen adsorption, enhancement through synergistic effects, and increased density of active sites, ultimately improving catalytic performance [11]. Despite these advantages, discovering and optimizing efficient alloy catalysts for HER remains time-intensive with the traditional trial-and-error approaches.

Density functional theory (DFT) simulations have partially alleviated this challenge by providing insights into the electronic properties of complex alloy materials [12]. Nevertheless, the inherent complexity of alloy catalysts and the vast number of potential compositions make high-throughput computational screening a formidable challenge. The emergence of machine learning (ML) techniques in alloy research has begun to reverse this trend [13–15]. Notably, in 1998, Yoshitake *et al.* [16] developed a Bayesian

* Corresponding authors.
E-mail addresses: sqzhao@hainanu.edu.cn (S. Zhao), zhangjian7@hust.edu.cn (J. Zhang), lux@hust.edu.cn (X. Lu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cclet.2025.112021
1001-8417/© 2025 Published by Elsevier B.V. on behalf of Chinese Chemical Society and Institute of Materia Medica, Chinese Academy of Medical Sciences.

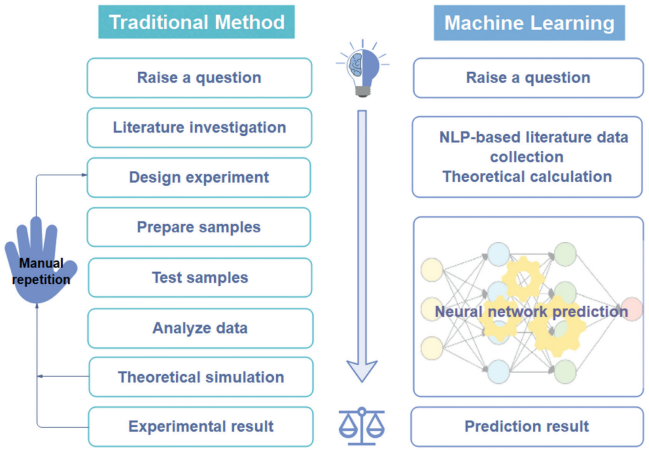


Fig. 1. Comparison diagram of traditional method and machine learning method.

framework-based neural network (NN) model to predict the lattice constants of γ and γ' phases in Ni-based superalloys. The predicted results showed strong agreement with the experimental data, demonstrating the potential of ML in alloy materials. Furthermore, the Materials Genome Initiative at United States further advanced ML in materials discovery, promoting the development of extensive materials databases that provide critical training data for ML models.

Leveraging vast amounts of experimental and simulation data, ML models have become indispensable tools for predicting catalytic properties in alloy systems with both high speed and accuracy [17,18]. Fig. 1 visually compares the traditional 'trial-and-error' approach to catalyst development with the ML-driven high-throughput screening method, highlighting the advantages of ML in accelerating catalyst discovery. It enables researchers to identify complex structure-property relationships, perform high-throughput screening, and rapidly narrow down the search space for promising alloy candidates [19–22]. Initial efforts focused primarily on binary systems. For instance, the universal ML framework developed by Chen *et al.* [23] quickly screened 43 high-performance bimetallic alloys with a computational speed approximately 100 times faster than conventional DFT calculations. With the progressive validation of ML-guided predictions, its application has been increasingly extended to more complex alloy systems. Li *et al.* [24] applied the labeled site crystal graph (DimeNet-LSCG) model to predict adsorption-related descriptors and successfully identified $FeCu_2Pt$ as a promising HER catalyst, exhibiting performance characteristics comparable to those of pure Pt (111). Further integrating ML with experimental validation, Kim *et al.* [25] successfully developed a ternary alloy, $Pt_{0.65}Ru_{0.30}Ni_{0.05}$, which exhibited an even lower overpotential than pure Pt. Recently, Huang *et al.* [26] integrated DFT calculations with three different ML algorithms to optimize the composition of a high-entropy alloy IrPdPtRhRu. Notably, their incorporation of synthetic minority over-sampling technique for regression with Gaussian noise (SMOGN) oversampling with Bayesian optimization yielded a 400% efficiency improvement over conventional non-Bayesian approaches, demonstrating the growing efficacy of ML in designing complex alloy systems. However, despite these promising advancements, a detailed understanding of how ML is influencing the rational design of HER alloy catalysts remains fragmented.

In this review, we explore the transformative role of ML in advancing HER alloys electrocatalysis. We begin by outlining the fundamental principles of ML and its integration into materials science, highlighting its advantages over traditional methods.

Specific attention is given to recent breakthroughs in applying ML to various alloy catalyst classes, including binary and other multi-alloy catalysts. Finally, we discuss the challenges and future prospects of ML in HER research, emphasizing its potential to revolutionize catalyst discovery and accelerate the transition to sustainable hydrogen energy.

2. Process of machine learning

ML technology is a sub-field of artificial intelligence, relying on statistics, algorithms, and data science to analyze large datasets, identify patterns, and establish predictive models [27]. Unlike traditional programming, which requires explicit instructions, ML enables systems to "learn" from data and improve their performance over time [28]. ML methods can be categorized into four main types: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (RL). Among them, supervised learning is one of the most widely used ML techniques. In this approach, the model is trained on labeled data, which includes input-output pairs. By learning the mapping relationship between inputs and outputs, the model can predict outcomes of new and unlabeled data.

Fig. 2 outlines the essential stages of applying ML techniques in catalyst development. The process begins with data collection, where relevant data sources are identified and gathered, followed by data representation and feature engineering, which ensure that the data is appropriately processed and refined for modeling. Algorithm selection then focuses on choosing suitable ML algorithms based on specific problem requirements and dataset characteristics, setting the foundation for effective learning. The model training and optimization phase involves iterative training, validation, and tuning to maximize predictive accuracy and generalization. Each of these stages plays a critical role in building an effective ML workflow for catalyst development, as described below.

2.1. Data collection

Data collection forms the foundation of the ML process, where the data quality directly impacts model accuracy. For a defined objective, data is sourced from experiments, computational models, simulations, and established databases such as Materials Project (MP), Crystallography Open Database (COD), and Inorganic Crystal Structure Database (ICSD) (Table 1) [29–51]. Published literature is another valuable data source, however, manual extraction is time-consuming. The integration of natural language processing (NLP) technology, powered by ML, has significantly accelerated this process, enabling efficient analysis of large volumes of text and extraction of relevant chemical information [52–59].

While integrating data from diverse sources, challenges such as inconsistent formats and incompatible data types may arise [60]. Post-collection data quality assessment is critical to address these issues, ensuring accuracy, integrity, and consistency [19,61]. For example, when developing HER electrocatalysts, maintaining uniform experimental conditions (*e.g.*, pH) is essential for reliable analysis.

2.2. Data representation and feature engineering

Effective data representation and feature engineering ensure that the model can interpret and process inputs accurately. Processing steps include removing duplicate entries, handling missing data (*e.g.*, using mean or median imputation), and correcting outliers. Afterward, data is standardized, normalized, and converted into numerical formats through methods like one-hot encoding and label encoding [62].

Next, feature selection or descriptor selection plays a crucial role in enhancing model interpretability and accuracy [63]. Statisti-

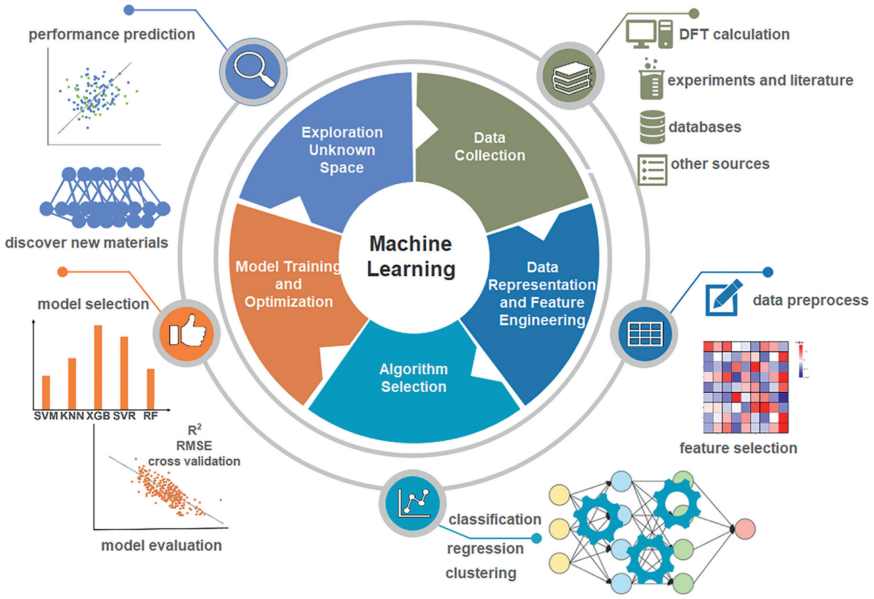


Fig. 2. Schematics of machine learning workflow.

Table 1
Summary of common materials science databases.

Database name	Data type	Quantity
Materials Project [29]	Inorganic compounds, metal alloys, battery materials, *etc.*	>130,000 species
Crystallography Open Database (COD) [36]	Inorganic, organic and organometallic frameworks (MOF), crystal structure of compounds	>400,000
Inorganic Crystal Structure Database (ICSD) [37]	All known inorganic crystal structure	>200,000
ChemSpider [38]	Organic, inorganic, organometallic and biological macromolecules, composite materials and nanomaterials, *etc.*	>67 million species
AFLOWLIB [39]	All kinds of inorganic materials, metals, oxides, alloys, *etc.*	About 100,000 species
Open Quantum Materials Database (OQMD) [40]	Inorganic compounds and elements	>1 million species
Pauling File [41]	Various inorganic compounds	>1 million
Cambridge Structural Databases (CSD) [42]	The most comprehensive organic and organometallic molecular structures	>1 million species
Atomwork Advance [43]	Metal alloys, ceramics, organic materials and composite materials, *etc.*	About 1.8 million
High Throughput Experimental Materials (HTEM) [44]	Including photovoltaic materials, battery materials and other materials related to energy technology	>140,000
Nanomaterial Registry [35,45]	Nanophase materials	Thousands
Computational 2D Materials Database (C2DB) [34,46,47]	Two-dimensional materials, including single-layer graphene and transition metal sulfides	About 4000 species
2D Material Encyclopedia [48]	Two-dimensional materials, such as graphene and transition metal sulfides	>6000 species
Catalysis-hub.org [31]	Specifically for the field of catalytic science	>10 thousand
Open Catalyst 2020 (OC20) [49]	Focused on important molecules in renewable energy applications (metal oxides, sulfides, carbon materials, *etc.*), it is the largest electrocatalyst structure dataset to date.	Over 1.3 million
Open Catalyst 2022 (OC22) [50]	It is an extension and supplement to the OC20 dataset, specifically designed to support the research of oxide electrocatalysts, particularly for OER.	62,331 DFT relaxations (~9854,504 single point calculations) across a range of oxide materials
Atomly [51]	Inorganic compounds	>340,000

cal tests (*e.g.*, chi-square test and F-test) and model-based methods (*e.g.*, L_1 regularization) identify and remove irrelevant or redundant features, simplifying the dataset and improving model performance. For HER, features such as adsorption energy of hydrogen ($E_{H,ads}$), surface structure, and electronegativity are crucial in capturing complex relationships between material properties and catalytic activity [64].

2.3. Algorithm selection

Algorithm selection depends on the specific tasks and dataset characteristics. Choosing a suitable algorithm can significantly improve the performance and accuracy of the model. Common tasks include classification, regression, and clustering. For classification tasks, algorithms such as support vector machine (SVM), decision trees (DT), random forest (RF), and NN are used to classify data into predefined categories. For regression tasks, algorithms such as linear regression, ridge regression, least absolute shrinkage and selection operator (LASSO) regression, support vector regression (SVR) and regression neural networks are applied to predict continuous numerical values. For clustering tasks, algorithms such as K-means and hierarchical clustering are commonly used to group data into clusters.

Algorithm selection should consider dataset size, dimensionality, type, and distribution. For instance, K-means is effective for large datasets, while DT excels with smaller, nonlinear datasets.

Tools like Scikit-learn, TensorFlow, and PyTorch facilitate algorithm implementation and tuning. In the context of developing HER electrocatalyst, regression models such as linear regression or SVR are trained on data sets containing experimental measured and calculated $E_{H,ads}$. By studying the relationship between alloy composition, structure, and $E_{H,ads}$, these models can accurately predict the Gibbs free energy of hydrogen adsorption (ΔG_{H^*}) for novel, untested alloys. Choosing the appropriate ML algorithm not only affects the accuracy of the model, but also determines the training speed and scalability.

2.4. Model training and optimization

The purpose of model training and optimization is to achieve the best model performance. The dataset is divided into training, validation, and test sets, with the training set typically constituting at least 70%. After the initial training phase, the validation set is used to monitor the model performance and guide hyperparameter tuning. Hyperparameters, such as learning rate, regularization strength, and the number of layers or neurons in a NN, can significantly affect the model convergence and predictive accuracy [65]. Effective hyperparameter optimization ensures a balance between underfitting and overfitting, enabling the model to generalize well to unseen data.

Evaluating model performance requires the use of appropriate metrics tailored to the specific task. For regression problems, metrics such as mean-absolute-error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R^2 score are commonly employed. In classification tasks, metrics like accuracy, precision, recall, and the F1-score are more appropriate. These metrics provide a quantitative basis for assessing how well the model captures the relationships in the data and how effectively it makes predictions.

Optimization is an iterative process aimed at enhancing model performance by fine-tuning its configuration. Some key techniques include: (1) Cross-validation: Dividing the data into multiple folds to ensure the model performs consistently across different subsets of the data; (2) Early stopping: A method to prevent overfitting by halting training once the validation error stops improving for a predefined number of iterations.

After training and optimization, the independent test set is used for final performance evaluation. By following these steps: dataset preparation, training, hyperparameter tuning, and performance evaluation, a robust and reliable ML model can be developed.

Such a model not only achieves high accuracy but also maintains the ability to adapt to diverse and complex data scenarios encountered in catalyst design and beyond [27].

3. Machine learning-assisted design of alloy catalysts

Alloying is an effective strategy for the effective hydrogen production because of their highly tunable composition and structure. By combining different metal elements, the electronic structure of the active site of the catalyst can be significantly changed, thus optimizing the $E_{H,ads}$ and improving the catalytic performance of HER [66]. However, traditional trial-and-error optimization methods typically rely on extensive experimentation and testing, which is very time-consuming and labor-intensive [67]. In recent years, the rapid development of ML has provided a revolutionary approach to the study of HER alloy catalysts. It combines the advantages of DFT, molecular dynamics (MD), and Monte Carlo (MC) simulations to describe complex chemical systems efficiently and accurately at a lower computational cost [68]. Specifically, ML can use a small amount of DFT data to predict the properties of a larger system, significantly reduce the calculation time. ML can also identify the key configurations and pathways in MD and MC simulations, reduce the required simulation steps, and capture the complex interactions in alloy systems that are often ignored by traditional methods. On a larger time and space scale, ML can generate high-precision simulation data, effectively guiding the design and optimization of catalysts. By employing ML models, it is possible to learn underlying patterns from existing experimental and computational data, quickly predict the activity of catalysts, and significantly enhance research efficiency [69]. In the following sections, we explore the application of ML in HER alloy catalysts, with a focus on analyzing the research progress of binary, multi-nary, HEA catalysts.

3.1. Binary alloy catalysts

Binary alloys, composed of two different metallic elements, are a fundamental category of alloy materials. These catalysts enhance catalytic efficiency by modifying electronic structures and surface properties. This section delves into how ML aids in the design and optimization of binary alloy catalysts for HER, providing a systematic approach to understanding their performance.

3.1.1. Bimetallic alloys

Bimetallic alloys usually refer to alloys composed of two different metal atoms mixed in a random or ordered manner at the microscopic scale, with microstructures that may be uniformly mixed or have configurations such as core-shell structures. The application of these catalysts in HER relies on their ability to provide a large number of active sites and optimized electronic properties. ML plays a crucial role here, employing large-scale data analysis and pattern recognition to identify the most effective alloy compositions and structural configurations. This accelerates the catalyst design and performance evaluation process, enhancing the development of efficient catalysts for HER [70].

The performance of alloys largely depends on their microstructures. With the rapid development of nanotechnology, nanoclusters, as a special form of bimetallic alloys, have garnered extensive attention from researchers [71,72]. For instance, Jäger *et al.* [73] utilized the Kernel Ridge Regression (KRR) ML model to predict the $E_{H,ads}$ values on AuCu nanoclusters. As shown in Fig. 3, by comparing various structural descriptors such as atom-centered symmetry functions (ACSF), many-body tensor representation (MBTR), and smooth overlap of atomic positions (SOAP), they examined how different representations influence model performance. Figs. 3a and b showed that as the size of the training set increases, the RMSE of all models decreases significantly, with the SOAP descriptor performing the best on larger datasets. Figs. 3c–e further illustrate that the SOAP descriptor is particularly sensitive to the local environment, provided the most accurate predictions for larger datasets. Furthermore, increasing the size of the training set continuously improved the prediction accuracy, highlighting the benefits of larger datasets in enhancing ML model performance. This work not only improved the accuracy of predicting catalytic active site $E_{H,ads}$, but also reduced the costs of modeling. Subsequently, they expanded their research to include bimetallic alloys of various shapes and sizes. Fig. 4a showed an automated workflow that systematically progresses from the generation of nanoclusters to the submission of production jobs, and finally to the prediction of $E_{H,ads}$, achieving efficient structure screening and performance evaluation [74]. Fig. 4b demonstrated the excellent prediction accuracy, showing good agreement between the predicted values and DFT calculation results. Fig. 4c illustrated the distribution characteristics of $E_{H,ads}$ for different binary combinations: some combinations exhibit convex behavior, indicating that the mixed cluster is less favorable than the pure components, while others show

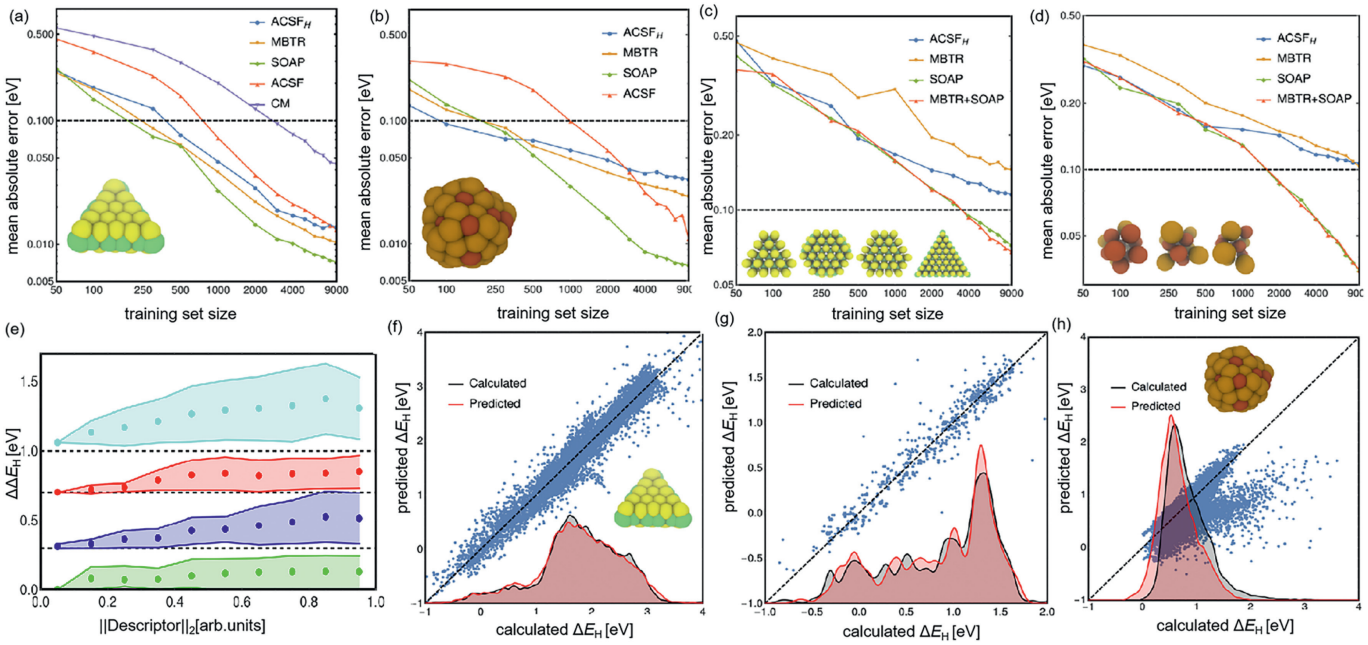


Fig. 3. Study of hydrogen adsorption on nanoclusters through machine learning using structural descriptors. (a–d) Learning curves for different datasets showing the MAE for different training set sizes. Descriptors CM, SOAP, MBTR, and ACSF were used as features in KRR to predict ΔE_H. The following datasets were used: (a) MoS_2 (single), (b) $Au_{40}Cu_{40}$ (single), (c) MoS_2 (multiple), (d) AuCu (multiple). (e) Mean of data point pairs on the axes of $\Delta(\Delta E_H)$ and (dis) similarity defined by $d = ||\text{Descriptor}||_2$ within bins of size 0.1. The colored area highlights the standard deviation in those bins. The data set MoS_2(multi) was used to compare the descriptors CM (cyan, offset 1.0 eV), SOAP (red, offset 0.7 eV), MBTR (blue, offset 0.3 eV) and ACSF (green). (f–h) Parity plots of predicted *versus* calculated ΔE_H for different training and testing sets, along with histograms of the predicted (red) and calculated (black) energy distributions. (f) The data set of multiple clusters MoS_2 (multi) was used as a training set and the data set MoS_2 (single) cluster was used as the displayed test set. (g) The data set of multiple clusters MoS_2(multi) was used as a training set and a data set of local minima on frozen clusters was used as the displayed test set. (h) The data set of multiple clusters AuCu(multi) was used as a training set and the data set $Au_{40}Cu_{40}$(single) cluster was used as the displayed test set. Reproduced with permission [73]. Copyright 2018, Springer Nature.

concave behavior, suggesting synergistic effects at certain compositions. Fig. 4d further analyzed the impact of surface reconstruction and adsorption site drift on the prediction results, revealing that these factors have minimal effect on model accuracy. The results also show that the maximum value of the d-band Hilbert transform ε_u is closely related to the $E_{H,ads}$ at the nanocluster level, which can be used as a screening feature available at the nanocluster level. In 2021, Mao *et al.* [75] investigated clusters with core-shell structured clusters. They optimized and evaluated 7924 $Cu_{55-n}M_n$ ($M = Co$, Ni, Ru, Rh, $n \leq 22$) alloy clusters and found that the CuNi alloy clusters exhibited superior HER activity, with the active sites mainly located at the bridging and triple sites of the shell Cu atoms. Moreover, they defined a surface charge-based descriptor, ΔQ_{Cu-Cu}, which effectively assesses the HER activity of alloy clusters by calculating the average charge difference between surface atoms of the clusters. As shown in Figs. 5a and b, the ΔQ_{Cu-Cu} values on $Cu_{55-n}M_n$ alloy clusters vary with increasing dopant concentration. Each type of $Cu_{55-n}M_n$ alloy cluster exhibits a maximum ΔQ_{Cu-Cu} value, and there is an approximately linear relationship between ΔQ_{Cu-Cu} and the number of active sites on the $Cu_{55-n}M_n$ clusters. A ML NN trained on a large DFT database using this descriptor can rapidly and accurately predict ΔG_{H*} values on nanocluster surfaces. As shown in Figs. 5c and d, the model achieves MAE of 0.07 eV and RMSE of 0.11 eV on the test set. This study promotes the application of high-throughput computational method in the field of electrocatalyst design. In the latest research, Kang *et al.* [76] employed an machine learning potential (MLP) model to conduct annealing MD simulations, revealing the structure of ligand-free AuPt nanoclusters. Through these simulations, they found that Au atoms tend to segregate to the surface, forming a Pt(core)/AuPt(shell) structure. By investigating the interactions between Au and Pt, they found that when the Au content is low, Au atoms enhance the hydrogen adsorption on the adjacent Pt atoms. This enhancement is attributed to the altered electronic structure of Pt caused by the presence of Au, leading to stronger hydrogen binding.

By uncovering the complex structure-property relationships in nanoclusters, ML models enable efficient high-throughput screening over a wider compound space. Using this method, researchers can quickly determine the most promising candidate materials, providing a solid foundation for further application development and performance optimization.

For example, Chen *et al.* [23] developed a universal ML framework (Fig. 6), which utilizes crystal graph convolutional neural networks (CGCNN) and SchNet models to predict the ΔG_{H^*} for a range of bimetallic alloys (such as Pd-Cu, Cu-Zn, Pt-Cu, and Pt-Mn), achieving computational speeds approximately 100 times faster than traditional DFT methods. This study rapidly screened 2974 candidate alloys and identified 43 high-performance compositions. The efficacy and practicality of this ML framework were further confirmed through experimental validation of the selected AgPd alloy. Zhang *et al.* [77] developed a light gradient boosting (LGB) ML model by incorporating the electronic and structural characteristics of alloys, such as electronegativity and electron efficiency. They successfully screened 84 potential alloys with $|\Delta G_{H^*}|$ values <0.1 eV from 2290 candidate alloys, including ScAu, TbSb, TbCd and TiAu. This study provides new insights for the application of ML in electrocatalyst optimization.

Using ML and high-throughput screening techniques, researchers can swiftly identify candidates with high-performance potential from vast databases of bimetallic alloys. These preliminary screening results provide a reference for further exploring the performance of alloys in practical applications. For instance, Tran and Ulissi [78] utilized active learning (AL) to guide DFT calculations in predicting adsorption energies on bimetallic alloys, identifying 258 surfaces suitable for HER from 102 alloys. Simi-

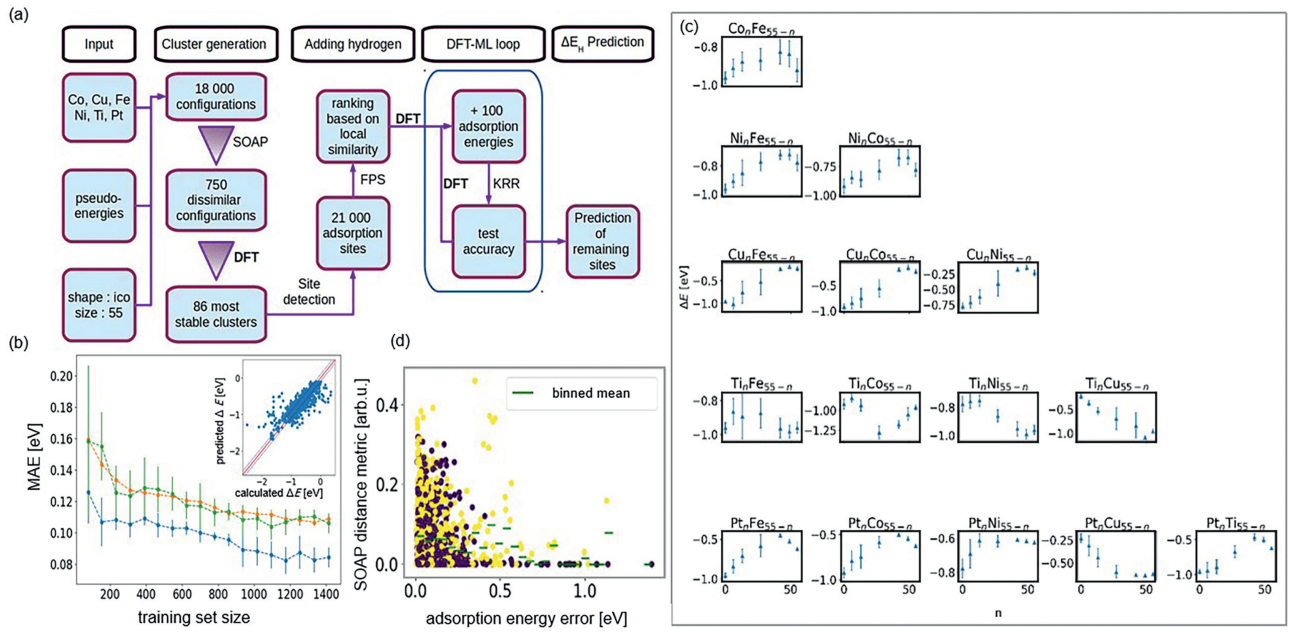


Fig. 4. Machine learning-assisted study of $E_{H,ads}$ on bimetallic nanoclusters. (a) Workflow outline showing the process from forming clusters to predicting $E_{H,ads}$ distribution. (b) Learning curve for KRR, with an inset image showing the parity between calculated and predicted $E_{H,ads}$ from 1767 DFT calculations. (c) Predicted $E_{H,ads}$ distribution. (d) Evaluation of machine learning accuracy in the presence of adsorption site drift and surface reconstruction. Reproduced with permission [74]. Copyright 2020, American Chemical Society.

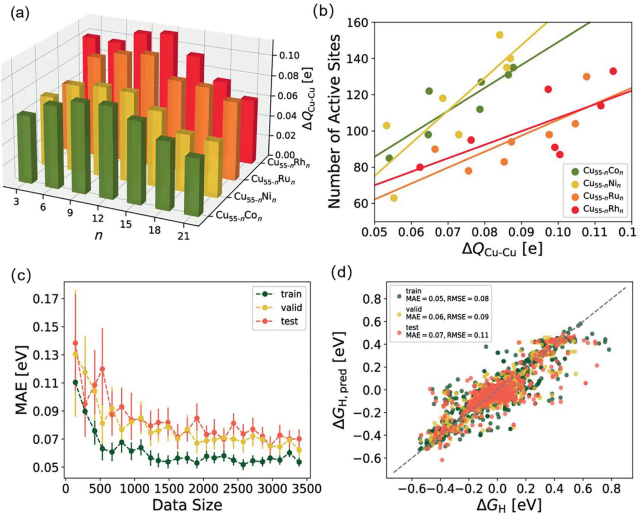


Fig. 5. Study on screening alloy nanoclusters for electrocatalytic hydrogen evolution. Descriptors used for evaluating HER performance: (a) The average charge difference $\Delta Q_{Cu\text{-}Cu}$ between two adjacent Cu atoms on the edge site and vertical site for $Cu_{55\text{-}n}M_n$ (M = Co, Ni, Ru, and Rh) alloy clusters; (b) linear relationship between the number of active sites ($|\Delta G_{H^*}| < 0.15$ eV) and the average charge difference $\Delta Q_{Cu\text{-}Cu}$. Machine learning predictions: (c) Learning curve of the neural network, (d) parity plot between predicted values and DFT calculated ΔG_{H^*} values. Reproduced with permission [75]. Copyright 2021, Springer Nature.

larly, Li *et al.* [79] employed a three-layer Artificial Neural Network (ANN) with backpropagation algorithm (BPNN) to explore the catalytic activity on (100) surfaces of bimetallic alloys for HER. They discovered promising theoretical HER activity in acidic media for $Pd_xAg_{1\text{-}x}$ and $Pd_xAu_{1\text{-}x}$ (100) surfaces. This ML-based analysis deepens the understanding of alloy properties and their catalytic mechanism. Wang *et al.* [80] used RFR combined with the sure independence screening and sparsifying operator (SISSO) model to rapidly predict adsorption energies on metallic and bimetallic alloy surfaces. Compared to linear scaling relations (LSR), this approach enables to make more accurate predictions lowering predictive RMSE by a factor of two and more general to predict adsorption energies of various adsorbates on thousands of binary alloys surfaces, thus paving the way for the discovery of novel bimetallic catalysts. Similarly, Martínez-Alonso *et al.* [81] used an RFR model to predict the surface activity of HER bimetallic catalysts, identifying 27 promising candidates. Additionally, Paz-Castany *et al.* [82] used an AL approach to optimize the performance of Ni-W alloy thin films for the HER in acidic media. Their findings indicated that the more electroactive Ni-W thin films should be deposited at slightly higher negative current densities ($-j = 7$–9 mA/cm^2) and a temperature of 62 °C. The predictions were validated experimentally, demonstrating accuracy while also saving material resources and laboratory time.

3.1.2. Single-atom alloys

Single-atom catalysts represent an emerging field in catalysis research, characterized by maximized atomic utilization and isolated active sites [83]. In 2011, the concept of single-atom catalysis was formally proposed by Professor Zhang and co-workers [84]. The structural characteristics of single-atom catalysts determine that the interaction between the dispersed metal atoms and the coordinating atoms (C, N, O, S, P) on the support is very important for the stability of the metal atoms. Beyond the typical bonding between metals and non-metal coordinating atoms, a special class of single-atom catalysts, known as single-atom alloys (SAAs), can also be prepared on metallic supports through metal-metal interactions [85,86]. In SAAs, the typical non-metal supports of single-atom catalysts are replaced by metallic supports, with the active metal atoms interacting with the metal supports *via* metal-metal bonds [87,88]. Throughout the development of single-atom catalysts, the study of metal-support interactions (MSI) has been integral to exploring their synthesis, structure, and catalytic reaction mechanisms. An early example of SAAs used in catalysis includes the activation of H_2 on Pd/Cu (111) and Pd/Au (111) surfaces [89].

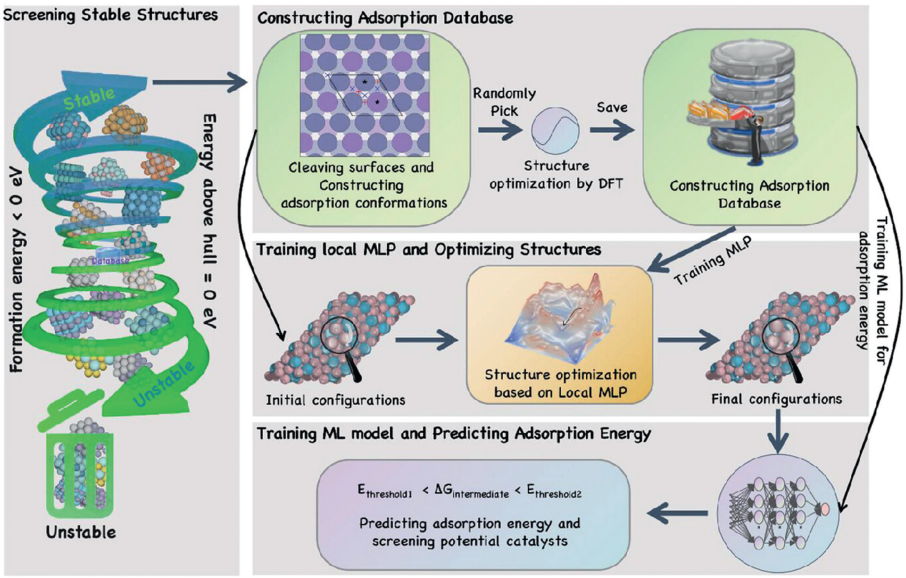


Fig. 6. ML framework used for high-throughput screening of electrocatalysts. Reproduced with permission [23]. Copyright 2022, Wiley-VCH GmbH.

In this context, ML provides a powerful tool for predicting and optimizing the performance of SAAs [90].

For example, Rao *et al.* [91] used a variety of ML models, including DT, SVM, NN, and KRR, to accurately predict the properties of 250 of the most stable and 358 near-stable SAA configurations. This approach helped form an intuitive understanding of the factors influencing the stability of SAAs. By using these diverse ML models, they could systematically analyze and deduce the underlying patterns that dictate stability, thus providing critical insights that can guide the synthesis and application of SAAs in various catalytic processes. Han *et al.* [92] employed the SISSO algorithm to identify key descriptors for SAAs. This powerful computing technology enabled them to accurately identify >200 previously unreported SAAs candidates with superior catalytic performance. Their work illustrates the critical role of data analysis in discovering optimal SAAs formulations. Subsequently, Zhou *et al.* [93] developed an automated catalyst design workflow called CATIDPy, using a genetic algorithm focused on screening and designing efficient SAAs for HER. This method successfully identified 70 candidate binary SAAs, demonstrating its application potential in theoretical predictions and experimental validations. This study not only accelerated the screening process for binary SAAs, but also promoted the efficient automation of catalyst development by reducing reliance on expert domain knowledge, which provides important guidance for the design of future catalytic materials for HER. Recently, Kayode *et al.* [94] utilized Bayesian Optimization technology to efficiently screen and optimize bimetallic catalysts for HER. Through Bayesian Optimization, the research team was able to rapidly identify the best-performing catalysts with minimal iterations, demonstrating high catalytic activity both theoretically and experimentally.

3.1.3. Dual-atom catalysts

Dual-atom catalysts (DACs) refer to catalysts featuring isolated active sites composed of two adjacent metal atoms or metal ions, which exhibit synergistic catalytic effects. Unlike single-atom catalysts containing two distinct single metal atom sites separated spatially, DACs involve direct metal–metal bonding at the active site [95]. Although this review focuses on metal alloy catalysts, DACs supported on nonmetallic substrates (such as graphitic carbon nitride, g-CN) are also discussed here because of the presence of metal–metal interactions, which align them conceptually with binary alloy systems. The catalytic properties of DACs can be finely tuned by varying the metal composition at these dual-atom sites [96].

DACs often exhibit complex atomic-level features, such as electronic structures and interatomic interactions, which require high-precision ML models to deal with these details. In this regard, the integration of DFT calculations with ML frameworks has yielded remarkable advantages, offering an innovative pathway for designing new materials [97]. Zhang *et al.* [98] conducted systematic calculations on DACs supported on g-CN substrates and employed RFR ML model to screen 26 homonuclear DACs (M_2/g-CN) and 253 heteronuclear DACs (M_IM_{II}/g-CN). They discovered that PdNi@g-CN and AgPt@g-CN exhibited exceptionally low overpotentials for HER. Subsequently, they extended this method to the domain of metal-nonmetal hybrid systems [99], demonstrating that the RFR model is capable not only of predicting known chemical systems but also of forecasting unknown chemical systems. This proves to be an effective tool for the rapid discovery and design of new electrocatalysts.

As computational capabilities improve, high-throughput computing combined with ML has increasingly become central to the research of DACs. Boonpalit *et al.* [100] innovatively applied the CGCNN model, to screen 435 dual-atom combinations on a nitrogen-doped graphene (N_6Gr) substrate, finally identifying AuCo@N_6Gr and NiNi@N_6Gr as highly efficient catalysts for HER. As shown in the workflow of Fig. 7a, the synergistic strategy combining ML and DFT significantly enhances the efficiency of discovering candidate materials. The screened AuCo and NiNi catalysts exhibited nearly ideal ΔG_{H^*} values (about 0 eV) along with exceptional stability and robust electronic performance. Subsequent DFT and electronic structure analyses highlighted that the metal synergistic interaction between Au and Co helps mitigate overly strong hydrogen binding, while the NiNi combination enhances hydrogen binding strength through dual-atom synergy. COHP analysis and charge density difference maps both confirm the significant synergistic effect between these bimetallic sites (Figs. 7b–g). These findings show the potential of ML to rapidly identify promising DAC configurations and advance the development of efficient and stable catalysts for HER. Subsequently, Wei *et al.* [101] comprehensively considered from multiple dimensions, using an ensemble learning method to construct the ML framework based on graph neural network (GNN) to identify the active sites of DACs. They employed RF model to predict free energies, used SVR deeply under-

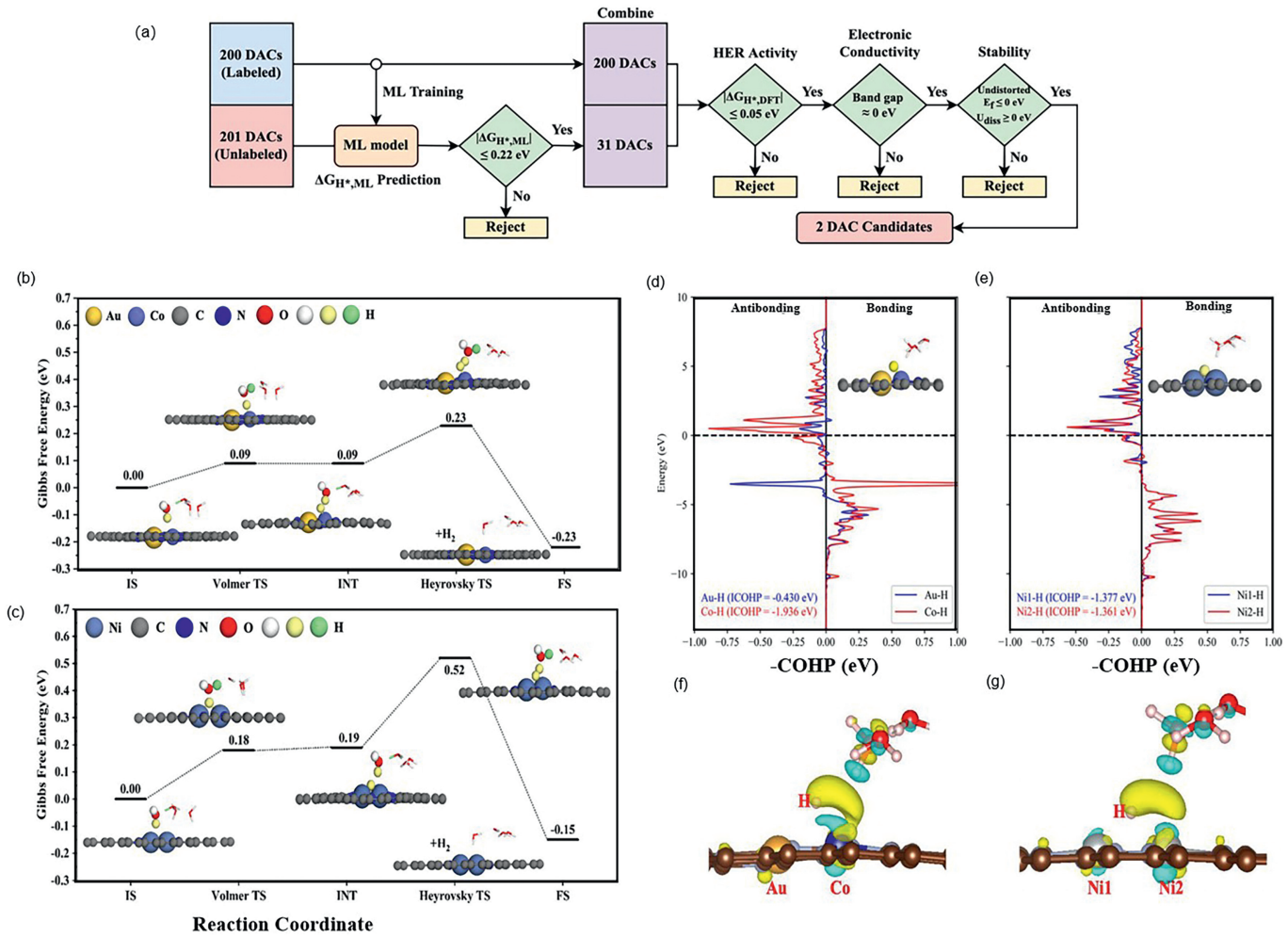


Fig. 7. Research on graphene-based HER DACs based on CGCNN and DFT. (a) Schematic representing the combined ML and DFT screening workflow. Reaction-free energy profile of HER, including explicit waters on (b) AuCo (c) NiNi, and the corresponding intermediate and transition state structures. (d, e) Projected crystal orbital Hamilton populations (pCOHP) and the corresponding integrated COHP (ICOHP) of AuCo and NiNi between metal atoms and hydrogen atoms during hydrogen binding in explicit solvent environment (right) bonding contribution and (left) antibonding contribution. (f, g) Charge density difference plot of AuCo and NiNi during hydrogen binding, (yellow) charge accumulation, and (blue) charge depletion. Reproduced with permission [100]. Copyright 2023, American Chemical Society.

stands how atoms and structures affect hydrogen bonding behavior, and employed ANN to predict binding energies. This method ensured high accuracy while reducing computational costs. Their research demonstrates the immense potential of an integrated ML-driven strategy combined with DFT computations in material design and optimizing DAC configurations. In the case of few input features, Wang *et al.* [102] employed an ANN model, based on DL algorithms, to screen from 406 transition metal (TM) DACs. They identified FeZn and VFe as two DACs with excellent acidic HER performance. This study also verified that the ANN model could accurately predict the HER activity of DACs using simple input features.

The further advancement in DACs research is reflected in the exploration of complex substrates and multi-metal combinations. For instance, Chowdhury *et al.* [103] employed DFT in conjunction with a Gradient Boosting (GB) ML model to design novel TMDACs based on α-2 graphyne (GPY). They discovered that TM-TM DACs, such as Sc-Sc, Y-Y, and Hf-Hf DACs, exhibit good thermal stability at room temperature. The GB model only predicts the stability based on the physical characteristics of components in DACs, which reduces the need for DFT calculations. Similarly, Liang *et al.* [104] employed gradient boosting regression (GBR) model to explore the synergistic effects of TM_1TM_2@BeN_4 on a BeN_4 substrate. Through shapley additive explanations (SHAP) analysis, they identified changes in the average interatomic distance and Fermi level as critical factors affecting catalytic activity. This research not only revealed the synergistic mechanisms of DACs, but also demonstrated how ML can be used to understand and predict the performance of catalysts. This provides new tools and theoretical foundations for the design and optimization of future catalysts, illustrating the significant role of machine learning in advancing catalysis research. In the latest research, Zhang *et al.* [105] developed an efficient DFT-ML framework for rapidly screening DACs on nitrogen-doped graphene (NG) for HER. Employing GBR model, which exhibited the highest R^2 values and the lowest MSE values, they successfully screened four promising HER catalysts from 1120 NG DACs. These include AgRe@N_6Gr, CrSn@N_6Gr, TiCo@N_6Gr, and RuAg@N_6Gr.

These applications demonstrate the ability of ML to understand and predict complex chemical reaction mechanisms. By integrating simulation technologies like DFT, ML can efficiently describe these complex chemical systems at a reduced computational cost, generating high-precision simulation data to guide the design and optimization of catalysts. The progression of research from single algorithms to multi-algorithm integrations, and then to material-specific NN models, demonstrates that the application of ML in the study of binary alloy catalysts is becoming increasingly refined and efficient. These studies not only accelerate the discovery and design of new binary alloy catalysts, but also provide us with new perspectives and tools to deeply understand the working mecha-

nism of binary alloy catalysts. Overall, ML exhibits great potential and promising applications in the research of HER bimetallic catalysts, effectively speeding up the discovery and performance evaluation of new catalysts.

3.2. Multi-alloy catalysts

Compared with binary alloys, multi-alloy catalysts offer more advantages in HER. These alloys cover three or more metallic elements, providing wider electronic and structural diversity, which may show higher activity, selectivity and durability in catalytic reactions. The complexity of multi-metallic alloys not only brings more opportunities for catalytic activity, but also introduces potential synergistic effects, which are rare in simple binary alloys. In this section, more complex multi-alloy catalysts will be discussed in depth, and how to explore and optimize the HER performance of multi-alloy catalysts by using ML model will be shown in detail, which will open up a new way for future energy conversion technology.

3.2.1. Ternary/quaternary alloy catalysts

In the exploration of multi-alloy catalysts, it is crucial to understand the complex kinetic pathways and reaction mechanisms under actual working conditions. ML exhibits tremendous potential in decoding complex chemical reactions by simulating and predicting dynamic changes on the alloy surfaces during catalytic processes. Schmidt *et al.* [106] selected common structures, tI10-$CeAl_2Ga_2$ and tP10-$FeMo_2B_2$, and used the Extremely Randomized Trees (ERT) model to identify potential stable and unstable phases. The results were highly consistent with experimental outcomes and conventional DFT calculations, yet the overall computational cost was reduced by approximately 75 %. This reduction significantly decreases both the cost and time required for researching and developing new materials, demonstrating the efficiency of ML in materials science research. Yoon *et al.* [107] developed a system called CatGym based on deep reinforcement learning (DRL) to predict the dynamic pathways of surface reconstruction under HER conditions for a ternary $Ni_3Pd_3Au_2$ (111) alloy catalyst. Through learning and iteration, DRL generated pathways that achieve locally minimal energy states, involving changes in surface composition and adjustments of surface atomic positions. The results show that DRL can not only explore more diverse surface components than the traditional minimum jump method, but also generate dynamic surface reconstruction paths, showing good consistency with the classical minimum energy path method, such as nudged elastic band (NEB) [108]. This study provides a novel approach for understanding and designing catalysts with specific surface kinetic properties, offering a new perspective for predicting and understanding complex catalyst surface reconstructions. The amorphous alloy $Pd_{40}Ni_{10}Cu_{30}P_{20}$ acts as a HER catalyst with exceptional electrocatalytic activity and high durability in practical experiments. Gao *et al.* [109] built the SOAP descriptor to construct a Gaussian Process Regression (GPR) ML model, accurately predicting the relationship between the local atomic environment of the active sites on the amorphous alloy surface and its catalytic performance. The model also revealed the physical source of the long-term durability of Ni dealloying, which highly aligns with experimental results. These studies demonstrate the substantial potential of ML technology in understanding complex dynamics and investigating mechanisms in catalysis research.

ML models and computational simulations can use this mechanistic information to predict the performance of a large number of potential materials, thereby reducing the number of candidate materials that require experimental validation. By integrating methods such as NN and AL, researchers can efficiently design and screen promising low-cost multi-alloy catalysts, accelerating their development and deployment in practical applications. Li *et al.* [24] combined GNN and crystal graph (CG) algorithms with AL to iteratively design and discover low-cost binary and ternary Pt alloy catalysts for HER. At last, they identified 12 new low-cost binary/ternary Pt alloys with catalytic activities similar to that of the Pt (111) surface. In particular, Cu_3Pt and $FeCuPt_2$ show catalytic performances close to that of Pt (111), with Cu_3Pt (100) exhibiting a maximum deviation of only 0.04 eV and $FeCuPt_2$ (100) having a maximum deviation of 0.16 eV in H, CO, and O adsorption energies. This research greatly reduced the cost of catalysts by incorporating more affordable metals like Fe and Cu, while maintaining high catalytic activity, potentially transforming the way catalysts are designed and discovered in the future. In addition, the analysis of the correlation between adsorption energies and the typical d-band theory descriptors indicates that d-band theory is still applicable to ternary alloys. Similarly, Kim *et al.* [25] optimized the composition and proportions of multi-metal alloy catalysts by combining AL iterative cycles with experimental methods. They successfully found a ternary alloy catalyst, $Pt_{0.65}Ru_{0.30}Ni_{0.05}$, which exhibited excellent catalytic performance with an overpotential of 54.2 mV in HER, lower than that of pure Pt catalysts. This demonstrates the potential of data-driven strategies to accelerate material discovery and optimization. Subsequently, using the Pareto AL framework combined with a GPR ML model, they discovered an optimal quaternary alloy catalyst, $Pt_{0.15}Pd_{0.30}Ru_{0.30}Cu_{0.25}$, which exhibited water splitting behavior at a voltage of 1.56 V under a current density of 10 mA/cm^2, with the cell voltage remaining below 1.6 V [110]. This method significantly enhanced the efficiency of the search process and the accuracy of catalyst performance predictions. Moreover, it can be expanded to different catalytic reaction domains and even other multifunctional applications, demonstrating a scalable and versatile approach to materials science research.

For other ternary alloys, Pandit *et al.* [111] simulated various compositions of NiCoCu alloys and used an eXtreme Gradient Boosting Regression (XGBR) ML model to predict and identify the best HER catalysts as alternatives to Pt electrodes. Figs. 8a–d show the ML model's predictions of adsorption energies for different adsorption sites (hcp, fcc, hcp-fcc, and on-top), which are in high agreement with DFT-calculated values, demonstrated the reliability of the model. Fig. 8e further indicated that the optimized and dataset-integrated XGBR model significantly reduces prediction errors. Subsequently, based on this model, they performed high-throughput screening of 5400 candidate alloy structures, ultimately identifying 27 cost-effective alloys as potential substitutes for Pt (Fig. 8f). This ML-DFT strategy not only provides a novel approach for developing low-cost HER catalysts, but also highlights the potential of ML in discovering new and efficient catalysts. In the recent research, Liu *et al.* [112] used gradient boosting trees (GBT) ML model combined with SHAP method to analyze and predict the impact of different nanoparticle sizes on the catalytic HER performance of nanoalloys. By examining the interactions between metal and support and the exposure ratio of active sites, they determined that the optimal catalyst particle size ranges from 1.5 nm to 3.0 nm. Specifically, the 2.0 nm PtCoNi catalyst achieved excellent catalytic performance with a lower noble metal content than commercial Pt/C catalysts. This study successfully guided the design and optimization of the PtCoNi alloy catalysts through a data-driven approach, validated the significant impact of particle size on catalytic performance, and corroborated the data analysis results with experimental methods.

For amorphous alloys, Zhang *et al.* [113] employed an SVR ML model to accurately predict the ΔG_{H^*} of 46,000 adsorption sites on the surface of the amorphous Pt@PdNiCuP alloy. They developed a distance contribution descriptor (DCD), used for feature engineering in the ML model, which accelerated the prediction of catalytic performance. The contribution analysis of DCD

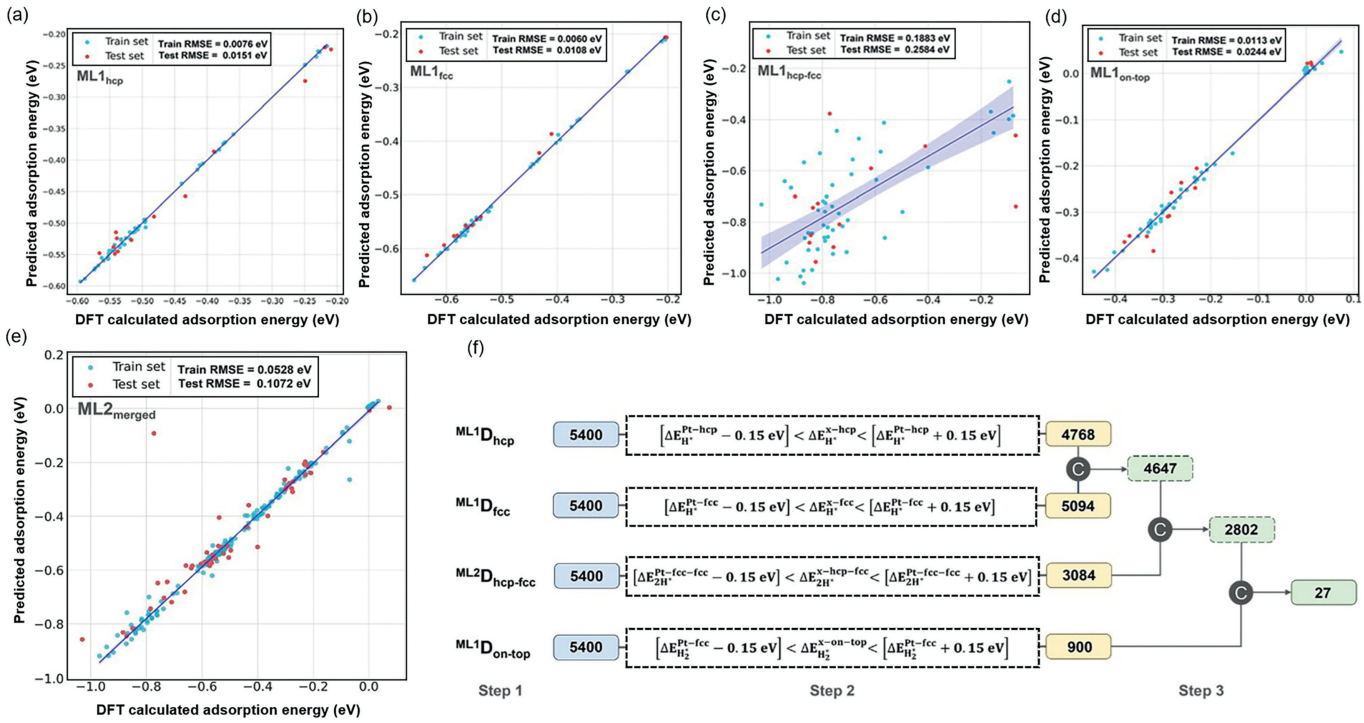


Fig. 8. Optimizing NiCoCu-based catalysts using DFT and supervised ML techniques. Plots of DFT calculated adsorption energies (ΔE_{calc}) *versus* predicted adsorption energies (ΔE_{pred}) with their respective indicated Test and Train RMSE values for (a) one H* atom adsorption at hollow hcp site with optimized XGBR model, (b) one H* atom adsorption at hollow fcc site with optimized XGBR model, (c) two H* atoms adsorptions at both, hollow hcp-fcc site with optimized SVR model, and (d) one H_2* molecule adsorption at on-top site with optimized XGBR model. ML1 represents method 1 by considering individual data sets. (e) Plot of DFT calculated adsorption energies (ΔE_{calc}) *versus* predicted adsorption energies (ΔE_{pred}) with its indicated test and train RMSE values for merged four data sets with optimized XGBR model. ML2 represents method 2 by considering merged data sets. (f) Flowchart of the performed screening and selection procedure. D_{hcp}, D_{fcc}, and $D_{on\text{-}top}$ data sets of 5400 were predicted by ML1 methods, and the $D_{hcp\text{-}fcc}$ data set was predicted by the ML2 method. Reproduced with permission [111]. Copyright 2022, American Chemical Society.

indicated that the synergistic interaction among Pt, Pd, and Ni atoms was crucial for determining the catalytic behavior of the amorphous alloy, aligning with experimental results. By calculating the percentage of atoms at different energy intervals, the optimal atomic ratio for best catalytic performance was identified as Pt:Pd:Ni:Cu:P = 0.33:0.17:0.155:0.16:0.185. It shows the potential of ML in predicting the performance of amorphous alloy catalysts.

Overall, integrating ML into the research of multi-metallic catalysts not only accelerates the discovery process but also provides deeper insights into interaction mechanisms, facilitating the development of catalysts for next-generation sustainable energy solutions.

3.2.2. High-entropy alloy catalysts

In 2004, Cantor [114] and Yeh [115] introduced the concept of HEA, which typically consist of 5 or more principal elements, and the mole percentage of each element is equal or nearly equal, generally with each component's atomic fraction greater than 5% (usually ranging from 5% to 35%). This design concept differs from traditional alloys, which often have 1 or 2 primary elements with additional elements added to improve performance. The key characteristic of HEA is their extremely high configurational entropy, which refers to the degree of random distribution of atoms within the material [116]. The high configurational entropy of HEA contributes to enhanced phase stability, meaning they maintain good stability at high temperatures. The application of HEA as catalysts for HER is a novel research area that has gained increasing attention in recent years. Compared to other alloy catalysts, HEAs, due to their unique properties such as high chemical stability, excellent mechanical performance, and distinctive electronic structure, show potential advantages in this field [117]. Studies have shown that some HEA-based catalysts have demonstrated HER activity comparable to or even superior to traditional noble metal catalysts [118]. For instance, some HEA containing Pt group elements exhibit very low overpotentials and high stability in HER [119]. Research is also being conducted on non-noble metal HEA, which typically include elements such as Fe, Co, Ni, Cu. The appropriate combination of these elements can lead to cost-effective and efficient catalysts for HER [120].

Although it shows many advantages, it also faces a series of challenges [121,122]. Firstly, the design of HEA is very complicated, as it requires precise control over the proportions and microstructures of multiple elements, presenting both experimental and theoretical challenges [123]. Secondly, the complex interactions among elements within HEA make predicting and optimizing catalytic activity difficult. The efficiency of HER catalysts depends on effectively reducing the overpotential for HER, which is often related to the material's electronic structure. In HEA, the electronic properties of different metal elements may influence each other, resulting in unpredictable catalytic performance [124]. Maintaining the uniformity of surface-active sites is also challenging. An ideal catalyst should have uniformly distributed active sites to provide stable and reproducible catalytic performance. However, the multi-element composition of HEA may lead to spatial heterogeneity of active sites, affecting catalytic efficiency. It is also crucial to maintain the phase stability of HEA under various operating conditions, such as different pH values, temperatures, or potentials. HEA may undergo phase separation or form unexpected intermetallic compounds under certain conditions, which can lead to a decrease in catalytic activity [123]. These challenges render traditional materials research and development methods inefficient and costly. In particular, the multi-elemental characteristics of HEAs leads to the problem of the 'curse of dimensionality'. As the number and proportion of constituent elements

increase, the combinatorial space expands exponentially, making high-throughput screening and optimization extremely challenging. To address this issue, ML techniques provide effective solutions by learning complex, non-linear relationships between composition, structure, and catalytic properties in high-dimensional datasets. By constructing surrogate models based on these learned relationships, ML can rapidly predict alloy performance and significantly narrow down the vast compositional search space. In addition, dimensionality reduction methods, such as principal component analysis, further simplify high-dimensional features while preserving essential physicochemical information, thereby reducing computational complexity. Moreover, structure-aware models like GNN can effectively capture atomic structures and inter-element interactions within alloys, thus accelerating the optimization process. Strategies such as AL and Bayesian optimization further enhance the efficiency of ML models by iteratively selecting informative data samples to improve the alloy screening process. Through efficient data processing and model construction, ML techniques not only significantly enhance the efficiency of the research and development process but also help researchers more effectively address the complex environment of HEAs, optimize catalyst design, and accelerate the discovery of new materials [125]. Firstly, ML accelerates material design through predictive modeling, allowing potential combinations to be screened more quickly in a wide alloy composition space [126]. For example, regression analysis can be used to predict properties like hardness and corrosion resistance under different alloy compositions, thereby reducing the number of experiments and associated costs. Secondly, ML models, such as classification algorithms, can predict the phase stability of alloys under specific heat treatment conditions. This helps optimize processing techniques and prevent the formation of undesired phases [127]. In performance evaluation, ML can automate the handling of large volumes of experimental data. By conducting in-depth analysis, it can identify key factors that influence performance, providing a scientific basis for alloy design.

The HEA containing noble metals show superior performance in HER catalysis due to their high intrinsic activity and stability. Kitagawa and co-workers [119,128] synthesized IrPdPtRhRu HEA nanoparticles (NPs) using a simple one-pot polyol method and conducted the first observation of the electronic structure of HEA NPs. It is confirmed that the proportion of 5 constituent elements of HEA is still stable under the extreme imbalance of 6:1:1:1:1, and it is indicated that HEA does not follow the traditional d-band theory. This finding spurred extensive subsequent research in the field. Huang *et al.* [26] combined DFT calculations with three different ML models: NN, GPR, and GNN, to optimize the composition of IrPdPtRhRu alloy. Among these, the NN served as a baseline comparison model for conventional $E_{H,ads}$ predictions, achieving the MAE of 0.079 eV, but exhibiting limited predictive capabilities in rare data regions (ΔG_{H^*} between −0.050 eV and 0.200 eV). By applying the SMOGN, the GPR model improved its predictive accuracy in low representation ΔG_{H^*} areas, obtaining a MAE of 0.084 eV. To achieve the highest modeling accuracy, the GNN model was used to capture a broader range of spatial and angular information, resulting in a MAE of 0.025 eV, which demonstrated higher precision compared to the NN model. The results showed that the Bayesian models and GNN show higher efficiency and deeper understanding in identifying ideal catalyst surfaces compared to traditional methods. The optimized IrPdPtRhRu alloy catalyst demonstrated about 40 % lower overpotential in HER than standard Pt catalysts, while the synthesis cost was about 15 % lower than that of an equiatomic alloy. This study was the first to investigate the benefits of combining SMOGN over-sampling with Bayesian learning for catalyst composition optimization. Compared with the traditional non-Bayesian method, the efficiency of Bayesian learning method in alloy composition exploration is improved by 400 %. This research not only shows the potential applications of HEA in HER, but also highlights the practicality of Bayesian learning in materials science, especially in optimizing complex alloy systems. It provides new perspectives and tools for future materials research and development.

In an effort to further reduce costs, Saidi *et al.* [129] trained 4 different ML models using 137 features to design a cost-effective, non-noble metal catalyst. The models included a convolutional neural network (CNN) and three tree-based models: DT, RF, and gradient boosting decision tree (GBDT). These models were employed to develop a multi-metallic alloy catalyst composed of Co, Mo, Fe, Ni, and Cu, named CoMoFeNiCu HEA. This catalyst as an alternative to noble metal-based catalysts, potentially lowering the costs of industrial hydrogen production. Shan *et al.* [120] used the ensemble machine learning (EML) model to predict the catalytic activity of various combinations of HEA based on high-throughput experimental data (Figs. 9a and b). The results demonstrated that among 56 screened equiatomic HEA combinations, the FeCoNiPtPd combination was the most promising. Subsequent preparations and screenings of the Fe-Co-Ni-Pt-Pd system led to the discovery of 2 promising HEA combinations, namely $Fe_{0.15}Co_{0.40}Ni_{0.05}Pt_{0.32}Pd_{0.08}$ and $Fe_{0.15}Co_{0.40}Ni_{0.05}Pt_{0.28}Pd_{0.12}$, whose catalytic performance was fully confirmed in subsequent performance tests. As shown in Figs. 9c–f, these combinations not only exhibited near-ideal ΔG_{H^*} but also achieve a good balance between cost and activity. This study provides a promising solution to the challenges posed by the diversity of HEA combinations and compositions. It accelerates the discovery process of HEA electrocatalysts, offering strong support for the rapid development and application of efficient catalysts. In the latest research, Yang *et al.* [130] proposed a "Differentiated Feature" method to predict $Fe_aCo_bNi_cCu_dMo_e$ HEAs ($0.18 < a, b, c, d, e < 0.23$, $a+b+c+d+e=1$). They successfully identified$Fe_{0.222}Co_{0.185}Ni_{0.185}Cu_{0.203}Mo_{0.203}$ HEAs with the best HER performance, exhibiting an average ΔG_{H^*} is 0.436 eV, which is close to the theoretical optimal value. Subsequently, the excellent properties of this HEA was validated by practical experiments and electronic structure analysis, indicating the synergistic effect between Fe and Ni plays a key role in enhancing HER properties. This study provides new ideas and methods for the design of efficient HEAs in the future.

From simple binary alloys to complex HEA, the integration of ML has changed the traditional trial-and-error approach, making the development of alloy catalysts more data-driven. By combining with high-throughput computational techniques, ML provides powerful support for the design and optimization of alloy catalysts. This integration enhances the efficiency of transitioning from theoretical models to experimental validation.

Overall, using ML and computational methods to develop alloy catalysts can not only accelerate the research, but also allows for more precise control over material properties, leading to the creation of more effective and economically viable catalytic solutions. This normal form shift towards a more analytical and less empirically dependent approach is setting a new standard in the field of catalysis and materials science.

4. Challenges and perspectives

In summary, this review has provided an overview of how ML is revolutionizing the design and optimization of HER alloy catalysts. We have critically evaluated recent advancements, demonstrating how ML-driven approaches accelerate catalyst discovery by efficiently processing large datasets and bridging the gap between theoretical predictions and experimental validation. By overcoming the limitations of traditional methods, ML not only reduces development time and cost but also paves the way for more innovative and high-performance catalyst solutions. Nevertheless, several key challenges remain.

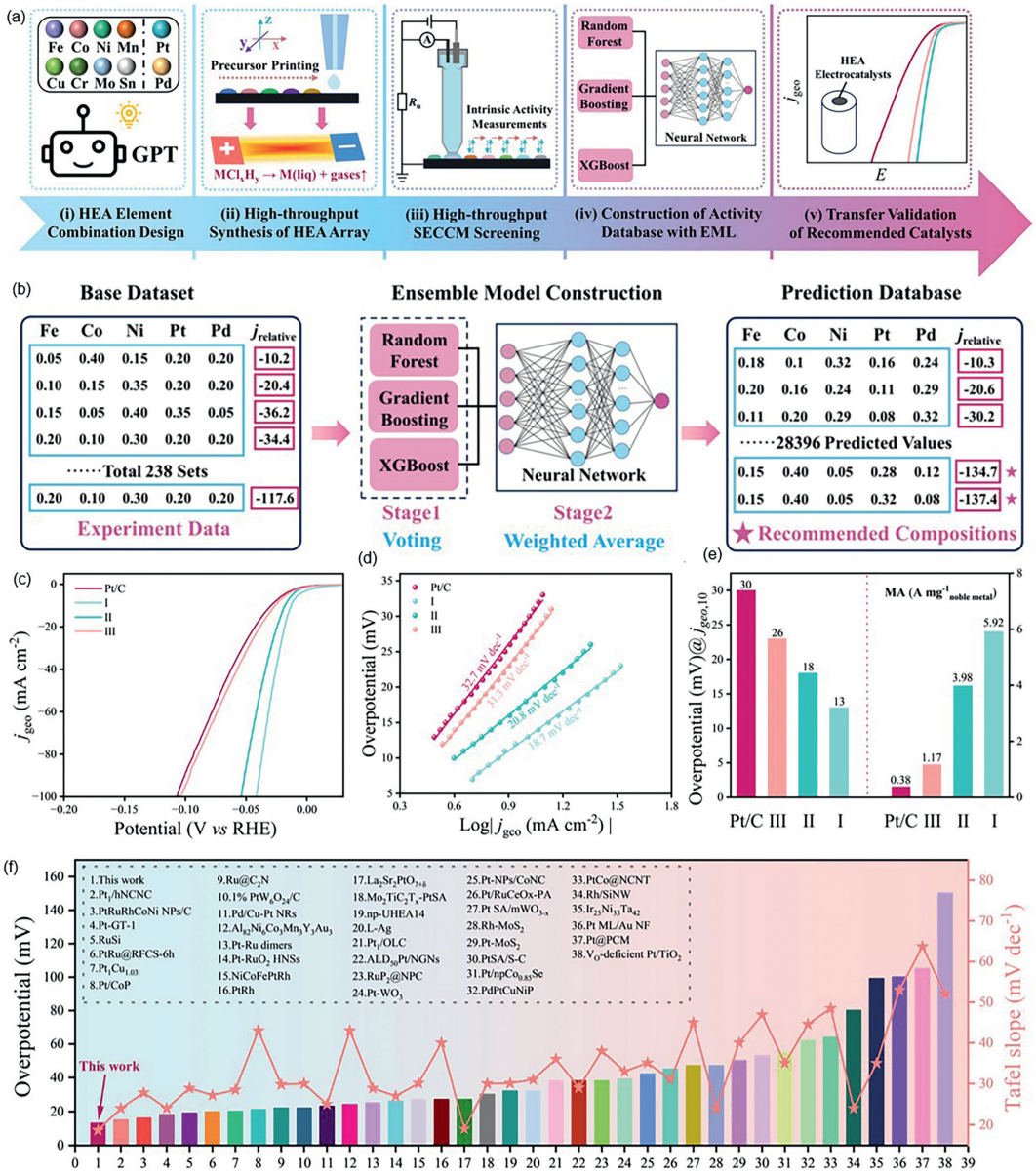


Fig. 9. Accelerating the discovery of efficient high-entropy alloy electrocatalysts through high-throughput experimentation and data-driven strategies. (a) Workflow of combining high-throughput experimentation with data-driven strategies. (b) Workflow for expanding the HER activity database of the Fe-Co-Ni-Pt-Pd system using EML model prediction. (c) HER polarization curves normalized by geometric area for catalysts I, II, and III and commercial Pt/C. (d) Tafel slope curves of samples I, II, and III and commercial Pt/C. (e) Overpotential at 10 mA/cm^2 geometric current density (left) and mass activity at −0.05 V *vs.* RHE for samples I, II, III, and commercial Pt/C. (f) Comparison of overpotentials and Tafel slopes at 10 mA/cm^2 for advanced noble-metal-based HER electrocatalysts in 0.5 mol/L H_2SO_4 electrolyte. Reproduced with permission [120]. Copyright 2024, American Chemical Society.

(1) High-quality training data. Accurate ML predictions for HER catalyst performance depend on abundant, high-quality training data. Although published experimental data provide a valuable resource, HER activity is influenced by more than the commonly reported parameters (*e.g.*, applied potential, electrolyte concentration, pH). Critical factors such as reactor type and size, flow rate, and other operational conditions are often underreported, making data from different studies difficult to compare. Automated experimental platforms, which generate comprehensive datasets under standardized conditions, could capture both primary and secondary parameters. In addition, incorporating uncertainty quantification into ML models would help assess prediction reliability and prioritize candidates for further validation. Automated experimental platforms offer a promising solution by providing diverse, high-quality datasets under controlled and standardized conditions that more accurately mimic real-world HER environments. These platforms can generate comprehensive data sets that not only encompass the primary experimental parameters but also account for secondary factors that influence catalytic behavior. In addition, integrating uncertainty quantification methods into ML models to assess the reliability of predictions can guide further experimental validation and help prioritize catalyst candidates with higher accuracy.

(2) Reliable descriptors for alloy catalysts. Developing reliable descriptors for alloy catalysts in HER remains a significant challenge. Current descriptors, such as the d-band center and local geometric parameters (*e.g.*, coordination numbers), offer valuable low-cost models to correlate complex catalytic metrics (*e.g.*, adsorption energies of key intermediates). However,

these descriptors often fall short in capturing the full complexity of alloy systems, where complex interactions among constituent elements and their coupling with adsorbates play a critical role. Moreover, conventional approaches typically overlook long-range electronic and geometric effects, as well as their gradients, which are essential for accurately reflecting the heterogeneous nature of HER catalysis. Such inaccuracy in descriptor formulation further limits the predictive accuracy of ML models when applied to complex, high-dimensional high-entropy alloy systems. A data-driven feature selection method, combining experimental data with high-throughput computational simulations, is needed to identify the most influential descriptors for catalytic performance and to enhance model reliability.

(3) Limitations of current ML algorithms. The current ML algorithms still face many challenges when dealing with complex alloy systems. At present, the algorithms widely used by ML in the design of HER alloy catalysts include SVR, RF and NN. These algorithms have achieved remarkable results in dealing with binary alloys and some multicomponent alloy systems. However, their generalization and prediction accuracy diminish when applied to highly complex multi-element systems like HEAs. Developing new algorithms capable of handling high-dimensional and nonlinear relationships is therefore crucial. DL approaches, which excel at processing complex structural data (*e.g.*, nanoclusters and polymetallic catalysts), can better capture atomic interactions and electronic structure effects. At the same time, the application of RL in the design of HER alloy catalysts is still in the exploratory stage. It optimizes the characteristics of decision-making process through interaction with the environment, which makes it have unique advantages in dynamic optimization and adaptive design. By introducing the reward mechanism, the RL model can quickly screen out the alloy combinations with the best properties and reduce unnecessary experimental verification. In the future, the combination of RL and existing high-throughput computational technology is expected to further accelerate the design and optimization of complex alloy catalysts. In addition, with the continuous development of ML technology, the automation and intelligence of algorithms will become an important trend in the future. Combined with AL technology, the model can automatically select the most valuable data for labeling and training, thus further improving data utilization efficiency and model performance.

(4) Lack of stability and durability prediction. Accurately predicting the stability and durability of alloy catalysts using ML remains a significant challenge. Experimental assessments of stability and durability often require long-term testing, making data generation both time-consuming and costly. As a result, available datasets are typically limited, fragmented, and lack consistency. A further obstacle is the absence of standardized benchmarks for evaluating catalyst stability. Unlike activity metrics such as overpotential or ΔG_{H^*}, there is no widely accepted protocol specifying, for example, which potential to apply or how long to run a durability test. This inconsistency makes it difficult to compare results across different studies or to compile reliable databases for model training. In addition, models must be capable of generalizing across a wide range of reaction environments, placing high demands on their robustness and flexibility. Recent advances in computational methods offer promising solutions to these limitations. Generative models, such as diffusion models, can expand limited datasets by producing high-fidelity simulated data that reflects realistic catalyst behavior. These models also capture atomic-level dynamics on catalyst surfaces, providing insights into degradation mechanisms that are otherwise difficult to observe. By integrating such approaches, researchers can improve both data quality and model performance, laying the groundwork for more accurate and generalizable predictions of catalyst stability and durability.

(5) The gap between predicted results and experimental validation. There is still a significant gap between ML prediction results and experimental validation, which presents an important challenge. ML models often rely on idealized structures or computational data for training, but in reality, the synthesis of alloy catalysts is influenced by various complex factors, such as inter-element interactions, compositional inhomogeneity, and the impact of different synthesis processes. These factors can lead to discrepancies between experimental results and model predictions. Especially for HEAs, whose complex composition and inter-element interactions can result in challenges such as difficult synthesis and poor phase stability during experimental validation. Additionally, catalytic performance itself is the result of multiple factors working together, including the formation of active sites, interfacial reconstruction, changes in electronic structure, and the influence of reaction conditions (*e.g.*, pH, potential, temperature), and these dynamic factors are difficult to fully represent in experimental validation. Current experimental methods also have certain limitations in accurately characterizing the transient reaction processes of catalysts, further limiting the direct validation of ML predictions. Therefore, future efforts should focus on enhancing the deep collaboration between ML and experiments, developing multiscale simulations, data-driven experimental design, and incorporating uncertainty quantification strategies, to improve the interpretability and adaptability of models, and thus increase the experimental feasibility of prediction results.

By solving these challenges and exploring new research directions, ML technology is expected to continue to promote the development of hydrogen energy technology in the future and provide strong scientific and technological support for the transformation of clean energy.

Declaration of competing interest

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

CRediT authorship contribution statement

Na Qin: Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Wenxin Guo:** Writing – review & editing, Visualization. **Fangxiu Li:** Visualization, Investigation, Conceptualization. **Houfeng Zhang:** Visualization, Investigation, Conceptualization. **Hong Liu:** Visualization, Investigation, Conceptualization. **Chang Zhang:** Visualization, Investigation, Conceptualization. **Lipiao Bao:** Project administration, Funding acquisition. **Lei Liu:** Resources, Methodology, Formal analysis. **Muneerah Alomar:** Project administration, Funding acquisition. **Siqi Zhao:** Writing – review & editing, Project administration, Funding acquisition. **Jian Zhang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis. **Xing Lu:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 22575072, 22405066, 22375067, 21925104 and 22431005), the National Key Research and Development Program of China (Nos. 2022YFA1504703 and 2022YFB4002204), the

Innovational Fund for Scientific and Technological Personnel of Hainan Province (No. KJRC2023C10), the Hubei Provincial Optics Valley Corridor Regional Collaborative Innovation Technology Project (No. 2023EGA013), and the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (No. PNURSP2025R398), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

[1] J. Li, Y. Sun, Z. Zhang, SmartMat 4 (2023) e1171.
[2] T. Li, Y. Deng, X. Rong, et al., SmartMat 4 (2023) e1142.
[3] S. Chu, A. Majumdar, Nature 488 (2012) 294–303.
[4] B. Lu, D. Wang, C. Zhao, et al., Mater. Chem. Front. 6 (2022) 1681–1689.
[5] M.S. Dresselhaus, I.L. Thomas, Nature 414 (2001) 332–337.
[6] A. Kudo, Y. Miseki, Chem. Soc. Rev. 38 (2009) 253–278.
[7] Y. Lei, Y. Wang, Y. Liu, et al., Angew. Chem. Int. Ed. 59 (2020) 20794–20812.
[8] F. Sun, Q. Tang, D. Jiang, ACS Catal. 12 (2022) 8404–8433.
[9] S. Trasatti, J. Electroanal. Chem. 39 (1972) 163–184.
[10] P. Zhang, S. Hong, N. Song, et al., Chin. Chem. Lett. 35 (2024) 109073.
[11] T. Yang, SmartMat 5 (2024) e1204.
[12] J. Liu, Z. Wang, L. Kou, Y. Gu, ACS Sustain. Sustain, Chem. Eng. 11 (2023) 467–481.
[13] C. Zhou, R. Yuan, W. Liao, et al., Rare Met. 43 (2024) 2884–2890.
[14] W. Chen, Y. Jiang, F. Tan, et al., Rare Met. 44 (2025) 5824–5843.
[15] Q. Tao, K. Bao, J. Cao, et al., Rare Met. 44 (2025) 7859–7875.
[16] S. Yoshitake, V. Narayan, H. Harada, et al., ISIJ Int. 38 (1998) 495–502.
[17] J. Cai, X. Chu, K. Xu, H. Li, J. Wei, Nanoscale Adv. 2 (2020) 3115–3130.
[18] S. Wei, S. Baek, H. Yue, et al., J. Electrochem. Soc. 168 (2021) 126523.
[19] M. Meuwly, Chem. Rev. 121 (2021) 10218–10239.
[20] Y. Shi, Z. Yang, S. Ma, et al., Engineering 27 (2023) 70–83.
[21] H. Huo, Z. Rong, O. Kononova, et al., npj Comput. Mater. 5 (2019) 62.
[22] H. Liu, A. Yang, C. Zhong, et al., ACS Energy Lett. 10 (2025) 1884–1891.
[23] L. Chen, Y. Tian, X. Hu, et al., Adv. Funct. Mater. 32 (2022) 2208418.
[24] X. Li, R. Chiong, Z. Hu, A.J. Page, J. Phys. Chem. Lett. 12 (2021) 7305–7311.
[25] M. Kim, M.Y. Ha, W. Jung, et al., Adv. Mater. 34 (2022) 2108900.
[26] L. Huang, Z. Gariepy, E. Halpren, et al., Small Methods 5 (2024) 2401224.
[27] P. Domingos, Commun. ACM 55 (2012) 78–87.
[28] C. Xiouras, F. Cameli, G.L. Quilló, et al., Chem. Rev. 122 (2022) 13006–13042.
[29] A. Jain, S.P. Ong, G. Hautier, et al., APL Mater. 1 (2013) 011002.
[30] N. Artrith, K.T. Butler, F.X. Coudert, et al., Nat. Chem. 13 (2021) 505–508.
[31] K.T. Winther, M.J. Hoffmann, J.R. Boes, et al., Sci. Data 6 (2019) 75.
[32] J. Deng, J. Pan, Y. Zhang, S. Du, Nano Lett. 23 (2023) 4634–4641.
[33] Y. Iwasaki, R. Sawada, E. Saitoh, M. Ishida, Commun. Mater. 2 (2021) 31.
[34] F.A. Rasmussen, K.S. Thygesen, J. Phys. Chem. C 119 (2015) 13169–13183.
[35] A. Tropsha, K.C. Mills, A.J. Hickey, Nat. Nanotechnol. 12 (2017) 1111–1114.
[36] M. Quirós, S. Gražulis, S. Girdzijauskaitė, A. Merkys, A. Vaitkus, J. Cheminform. 10 (2018) 23.
[37] A. Belsky, M. Hellenbrandt, V.L. Karen, P. Luksch, Acta Cryst. B: Struct. Sci. 58 (2002) 364–369.
[38] R. Guha, E. Willighagen, J. Cheminform. 12 (2020) 4.
[39] S. Curtarolo, W. Setyawan, S. Wang, et al., Comput. Mater. Sci. 58 (2012) 227–235.
[40] S. Kirklin, J.E. Saal, B. Meredig, et al., npj Comput. Mater. 1 (2015) 15010.
[41] P. Villars, M. Berndt, K. Brandenburg, et al., J. Alloy. Compd. 367 (2004) 293–297.
[42] C.R. Groom, F.H. Allen, Angew. Chem. Int. Ed. 53 (2014) 662–671.
[43] Y. Xu, M. Yamazaki, P. Villars, Jpn. J. Appl. Phys. 50 (2011) 11RH02.
[44] A. Zakutayev, N. Wunder, M. Schwarting, et al., Sci. Data 5 (2018) 180053.
[45] A.M. Khan, F.A. Loan, U.Y. Parray, S. Rashid, Inf. Discov. Deliv. 52 (2024) 53–61.
[46] M.N. Gjerding, A. Taghizadeh, A. Rasmussen, et al., 2D Mater. 8 (2021) 044002.
[47] S. Haastrup, M. Strange, M. Pandey, et al., 2D Mater. 5 (2018) 042002.
[48] J. Zhou, L. Shen, M.D. Costa, et al., Sci. Data 6 (2019) 86.
[49] L. Chanussot, A. Das, S. Goyal, et al., ACS Catal. 11 (2021) 6059–6072.
[50] R. Tran, J. Lan, M. Shuaibi, et al., ACS Catal. 13 (2023) 3066–3084.
[51] M. Liu, S. Meng, Sci. Sin. Chim. 53 (2023) 19–25.
[52] A.C. Vaucher, F. Zipoli, J. Geluykens, et al., Nat. Commun. 11 (2020) 3601.
[53] T. He, W. Sun, H. Huo, et al., Chem. Mater. 32 (2020) 7861–7873.
[54] P. Shetty, A.C. Rajan, C. Kuenneth, et al., npj Comput. Mater. 9 (2023) 52.
[55] P. Shetty, R. Ramprasad, iScience 24 (2021) 101922.
[56] P. Shetty, R. Ramprasad, J. Chem. Inf. Model. 61 (2021) 5377–5385.
[57] T. Gupta, M. Zaki, N.M.A. Krishnan, Mausam npj Comput. Mater. 8 (2022) 102.
[58] Z. Wang, O. Kononova, K. Cruse, et al., Sci. Data 9 (2022) 231.
[59] N. Qin, L. Liu, L. Wei, H. Chen, J. Alloys Compd. Commun. 6 (2025) 100067.
[60] J. Li, N. Wu, J. Zhang, et al., Nano-Micro Lett. 15 (2023) 227.
[61] Y. Ding, L. Fan, X. Liu, Energy Build. 249 (2021) 111208.
[62] M. Karthikeyan, D.M. Mahapatra, A.S.A. Razak, et al., Catal. Rev. Sci. Eng. 66 (2024) 997–1027.
[63] B. Wang, F. Zhang, Angew. Chem. Int. Ed. 61 (2022) e202111026.
[64] C. Wang, X. Wang, T. Zhang, et al., J. Mater. Chem. A 10 (2022) 18195–18205.
[65] P. Riley, Nature 572 (2019) 27–29.
[66] Y. Xu, X. Zhang, Y. Liu, et al., Environ. Sci. Pollut. Res. 30 (2022) 11302–11320.
[67] A.R. Oganov, C.J. Pickard, Q. Zhu, R.J. Needs, Nat. Rev. Mater. 4 (2019) 331–348.
[68] E.A. Olivetti, J.M. Cole, E. Kim, et al., Appl. Phys. Rev. 7 (2020) 041317.
[69] G.L.W. Hart, T. Mueller, C. Toher, S. Curtarolo, Nat. Rev. Mater. 6 (2021) 730–755.
[70] M. Hu, Q. Tan, R. Knibbe, et al., Mater. Sci. Eng. R.: Rep. 155 (2023) 100746.
[71] J. Liu, C. Lee, Y. Hu, et al., SmartMat 4 (2023) e1210.
[72] Y. Jeong, T. Begildayeva, J. Theerthagiri, et al., SmartMat 5 (2024) e1281.
[73] M.O.J. Jäger, E.V. Morooka, F. Federici Canova, L. Himanen, A.S. Foster, npj Comput. Mater. 4 (2018) 37.
[74] M.O.J. Jäger, Y.S. Ranawat, F.F. Canova, E.V. Morooka, A.S. Foster, ACS Comb. Sci. 22 (2020) 768–781.
[75] X. Mao, L. Wang, Y. Xu, et al., npj Comput. Mater. 7 (2021) 46.
[76] J. Kang, J. Kloppenburg, J. Sheng, et al., ACS Catal. 15 (2025) 9928–9939.
[77] J. Zhang, Y. Wang, X. Zhou, et al., Nanoscale 15 (2023) 11072–11082.
[78] K. Tran, Z.W. Ulissi, Nat. Catal. 1 (2018) 696–703.
[79] H. Li, S. Xu, M. Wang, et al., J. Mater. Chem. A 8 (2020) 17987–17997.
[80] T. Wang, J. Li, W. Shu, et al., Chin. J. Chem. Phys. 33 (2020) 703–711.
[81] C. MartínezAlonso, V. VassilevGalindo, B.M. Comer, et al., Catal. Sci. Technol. 14 (2024) 3784–3799.
[82] R. De PazCastany, K. Eiler, A. Nicolenco, et al., ChemSusChem 18 (2025) e202400444.
[83] X. Sun, Y. Song, G. Jiang, X. Lan, C. Xu, Sci. China Mater. 67 (2024) 1–17.
[84] B. Qiao, A. Wang, X. Yang, et al., Nat. Chem. 3 (2011) 634–641.
[85] R.T. Hannagan, G. Giannakakis, M. Flytzani Stephanopoulos, E.C.H. Sykes, Chem. Rev. 120 (2020) 12044–12088.
[86] Y. Da, R. Jiang, Z. Tian, et al., SmartMat 4 (2023) e1136.
[87] M.A. Deshmukh, A. Bakandritsos, R. Zbořil, Nano-Micro Lett. 17 (2025) 1.
[88] B. Geng, X. Chu, L. Liu, et al., Chin. Chem. Lett. 35 (2024) 108924.
[89] G. Kyriakou, M.B. Boucher, A.D. Jewell, et al., Science 335 (2012) 1209–1212.
[90] J. Liu, J. Zhang, H. Deng, et al., Chin. Chem. Lett. 36 (2025) 110656.
[91] K.K. Rao, Q.K. Do, K. Pham, D. Maiti, L.C. Grabow, Top. Catal. 63 (2020) 728–741.
[92] Z. Han, D. Sarker, R. Ouyang, et al., Nat. Commun. 12 (2021) 1833.
[93] C. Zhou, J.Y. Zhao, P.F. Liu, et al., Chem. Sci. 12 (2021) 10634–10642.
[94] G.O. Kayode, A.F. Hill, M.M. Montemore, J. Mater. Chem. A 11 (2023) 19128–19137.
[95] Y. Wang, F. Ma, G. Zhang, et al., Nano Res. 17 (2024) 9397–9427.
[96] X. Liang, N. Fu, S. Yao, Z. Li, Y. Li, J. Am. Chem. Soc. 144 (2022) 18155–18174.
[97] L. Wu, T. Li, Mater. Chem. Front. 7 (2023) 4445–4459.
[98] L. Zhang, X. Guo, S. Zhang, S. Huang, J. Mater. Chem. A 10 (2022) 11600–11612.
[99] L. Zhang, X. Guo, S. Zhang, T. Frauenheim, S. Huang, Adv. Energy Mater. 14 (2024) 2302754.
[100] K. Boonpalit, Y. Wongnongwa, C. Prommin, S. Nutanong, S. Namuangruk, ACS Appl. Mater. Interfaces 15 (2023) 12936–12945.
[101] C. Wei, D. Shi, Z. Yang, et al., J. Mater. Chem. A 11 (2023) 18168–18178.
[102] J. Wang, W. Xu, J. He, et al., Appl. Surf. Sci. 659 (2024) 159942.
[103] C. Chowdhury, E. Karthikraja, V. Subramanian, Phys. Chem. Chem. Phys. 26 (2024) 25143–25155.
[104] W. Liang, A. Aierken, X. Rong, et al., Appl. Surf. Sci. 682 (2025) 161745.
[105] H. Zhang, Q. Wei, S. Wei, et al., Mol. Catal. 570 (2025) 114649.
[106] J. Schmidt, L. Chen, S. Botti, M.A.L. Marques, J. Chem. Phys. 148 (2018) 241728.
[107] J. Yoon, Z. Cao, R.K. Raju, et al., Mach. Learn.: Sci. Technol. 2 (2021) 045018.
[108] G. Henkelman, B.P. Uberuaga, H. Jónsson, J. Chem. Phys. 113 (2000) 9901–9904.
[109] S. Gao, H. Zhen, B. Wen, J. Ma, X. Zhang, Nanoscale 14 (2022) 2660–2667.
[110] M. Kim, Y. Kim, M.Y. Ha, et al., Adv. Mater. 35 (2023) 2211497.
[111] N.K. Pandit, D. Roy, S.C. Mandal, B. Pathak, J. Phys. Chem. Lett. 13 (2022) 7583–7593.
[112] C. Liu, S. Zhu, F. Sun, et al., Chem. Eng. J. 476 (2023) 146918.
[113] X. Zhang, K. Li, B. Wen, J. Ma, D. Diao, Chin. Chem. Lett. 34 (2023) 107833.
[114] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Mater. Sci. Eng. R.: Rep. 375–377 (2004) 213–218.
[115] J.W. Yeh, S.K. Chen, S.J. Lin, et al., Adv. Eng. Mater. 6 (2004) 299–303.
[116] Y. Zhou, X. Shen, M. Wang, et al., Sci. China Mater. 66 (2023) 2527–2544.
[117] M. Tsai, J. Yeh, Mater. Res. Lett. 2 (2014) 107–123.
[118] J. Kwon, S. Sun, S. Choi, et al., Adv. Mater. 35 (2023) 2300091.
[119] D. Wu, K. Kusada, T. Yamamoto, et al., Chem. Sci. 11 (2020) 12731–12736.
[120] X. Shan, Y. Pan, F. Cai, et al., Nano Lett. 24 (2024) 11632–11640.
[121] L. Li, Z. Zhang, SmartMat 5 (2024) e1239.
[122] D. Wang, Z. Chen, Y. Wu, et al., SmartMat 4 (2023) e1117.
[123] L. Sun, K. Wen, G. Li, et al., ACS Mater. Au 4 (2024) 547–556.
[124] M. Sun, Y. Li, S. Wang, et al., Nanoscale 15 (2023) 13515–13531.
[125] X. Wan, Z. Li, W. Yu, et al., Adv. Mater. 37 (2023) e2305192.
[126] Y.V. Krishna, U.K. Jaiswal, R.M. R, Scr. Mater. 197 (2021) 113804.
[127] N. Islam, W. Huang, H.L. Zhuang, Comput. Mater. Sci. 150 (2018) 230–235.
[128] Y. Maruta, K. Kusada, D. Wu, et al., Chem. Commun. 58 (2022) 6421–6424.
[129] W.A. Saidi, T. Nandi, T. Yang, Electrochem. Sci. Adv. 3 (2023) e2100224.
[130] Z. Yang, Q. Yin, Z. Yin, et al., Prog. Nat. Sci. Mater. Int. 3 (2025) 631–637.