Applied Physics Reviews REVIEW pubs.aip.org/aip/are

Advanced theoretical modeling methodologies for electrocatalyst design in sustainable energy conversion

Cite as: Appl. Phys. Rev. **12**, 011316 (2025); doi: 10.1063/5.0235572
Submitted: 28 August 2024 · Accepted: 30 December 2024 ·
Published Online: 6 February 2025

Tianyi Wang,[1] Qilong Wu,[2] Yun Han,[3] Zhongyuan Guo,[4] Jun Chen,[2,a)] and Chuangwei Liu[1,5,a)]

AFFILIATIONS

[1]School of Material Science and Engineering, Northeastern University, Shenyang 110819, China
[2]Intelligent Polymer Research Institute, Innovation Campus, University of Wollongong, Squires Way, North Wollongong, NSW 2500, Australia
[3]Queensland Micro- and Nanotechnology Centre, School of Engineering and Built Environment, Griffith University, Nathan Campus, QLD 4111, Australia
[4]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China
[5]State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

Note: This paper is part of the APR Special Topic on Frontiers in energy materials research: novel measurement, modelling and processing approaches.
[a)]**Author to whom correspondence should be addressed:** junc@uow.edu.au and cwliu@dicp.ac.cn

ABSTRACT

Electrochemical reactions are pivotal for energy conversion and storage to achieve a carbon-neutral and sustainable society, and optimal electrocatalysts are essential for their industrial applications. Theoretical modeling methodologies, such as density functional theory (DFT) and molecular dynamics (MD), efficiently assess electrochemical reaction mechanisms and electrocatalyst performance at atomic and molecular levels. However, its intrinsic algorithm limitations and high computational costs for large-scale systems generate gaps between experimental observations and calculation simulation, restricting the accuracy and efficiency of electrocatalyst design. Combining machine learning (ML) is a promising strategy to accelerate the development of electrocatalysts. The ML-DFT frameworks establish accurate property–structure–performance relations to predict and verify novel electrocatalysts' properties and performance, providing a deep understanding of reaction mechanisms. The ML-based methods also accelerate the solution of MD and DFT. Moreover, integrating ML and experiment characterization techniques represents a cutting-edge approach to providing insights into the structural, electronic, and chemical changes under working conditions. This review will summarize the DFT development and the current ML application status for electrocatalyst design in various electrochemical energy conversions. The underlying physical fundaments, application advancements, and challenges will be summarized. Finally, future research directions and prospects will be proposed to guide novel electrocatalyst design for the sustainable energy revolution.

TABLE OF CONTENTS

I. INTRODUCTION

Sustainable energy conversion and storage are essential for addressing global energy challenges and mitigating environmental crises by utilizing renewable energy sources instead of fossil fuels, thereby reducing greenhouse gas emissions.[1,2] Advanced electrochemical reaction technologies represent promising strategies to convert abundant small molecules into value-added products, including hydrogen evolution reaction (HER), oxygen reduction reaction (ORR), oxygen evolution reaction (OER), nitrogen reduction reaction (NRR), carbon dioxide reduction reaction (CO_2RR), nitrate reduction reaction (NO_3RR), etc.[3–6] These essential electrochemical processes are critical for achieving a carbon-neutral society and a sustainable energy future.[7,8] Currently, exploring high-performance and low-cost electrocatalysts is an open challenge for their industrial application.

The development of electrocatalysts is intrinsically constrained by the capacity to discover new materials and fully comprehend their behaviors. Traditional experimental methods, based on trial-and-error approaches, are time-consuming, costly, and resource-intensive; thus, they are inadequate for satisfying the urgent demand for catalyst development. To accelerate the exploration of electrocatalysts, it is essential to transition from traditional trial-and-error methodologies to more targeted and efficient approaches.[9] Recent advancements in computational capabilities and algorithms have facilitated the development of computer-aided electrocatalyst design. Density functional theory (DFT) has emerged as a widely applicable and efficient tool for analyzing the intrinsic characteristics of existing electrocatalysts and forecasting the performance of novel catalysts, achieving considerable success across materials science, chemical reactions, surface science, and related fields. However, the high computational cost limits DFT application to large-scale systems, and its inherent algorithmic properties contribute to discrepancies between experimental observations and DFT calculations.[10,11]

In the past two decades, significant advancements in computer science have propelled the application of data-driven methodologies across various domains, including materials and chemical sciences. Big data-driven artificial intelligence (AI) offers a robust framework for the automated and efficient exploration of high-performance materials, representing the fourth paradigm of science [Fig. 1(a)]. Machine learning (ML), a key subset of AI, constructs nonlinear mappings by modeling complex functions based on input data, uncovering underlying correlations within intricate datasets. For example, ML has accelerated innovation in battery technologies, including novel electrode discovery, accurate property prediction, capacity retention, etc.[12] Compared with DFT calculations, ML emphasizes data correlations rather than relying on a single physical model, which emphasizes structure–property–performance relationships. The ML-based frameworks generalize broadly applicable approaches across various electrochemical reaction networks. Integrating ML with DFT and experiment characterization techniques is widely adopted within the catalysis community, significantly shortening development cycles and reducing costs.[13–15] As an inductive tool, ML-based frameworks derive conclusions by fitting principles, enabling the extraction of patterns and trends from existing data. ML-based frameworks also function as deductive tools, facilitating the development of hypotheses for novel catalysts and reaction mechanisms.[16,17] The partnership between ML-based frameworks and electrochemical reactions will rapidly establish strategies to handle situations inaccessible to traditional theoretical modeling methodologies and build efficiencies within rational materials design efforts. As Fig. 1(b) shown, published papers about integrating ML and DFT in catalyst studies have increased significantly in recent years. In addition, ML-based methods could also accelerate the solution of molecular dynamics (MD). Computational simulations provide detailed predictions of electronic structure, adsorption energy, and reaction pathways, forming theoretical foundations for catalyst design. Experiment characterization techniques validate and refine these predictions, bridging the gap between theoretical models and empirical observations. For example, x-ray photoelectron spectroscopy probes the electronic states, oxidation states, and surface intermediates during electrochemical reactions, corroborating electronic and surface characteristics predication. Similarly, structural and morphological insights from x-ray absorption spectroscopy and transmission electron microscopy enable direct comparisons between calculated and synthesized materials.[18–20] This synergy is essential for advancing a comprehensive understanding of electrocatalytic processes, linking atomic-scale intermediate interactions to macroscopic catalyst performance, and guiding rational catalyst design. However, experiment characterization techniques often generate high-dimensional datasets, presenting challenges in data interpretation. ML provides a powerful and time-efficient approach to analyzing complex datasets by reducing noise and human errors.[19] Integrating ML with experimental and

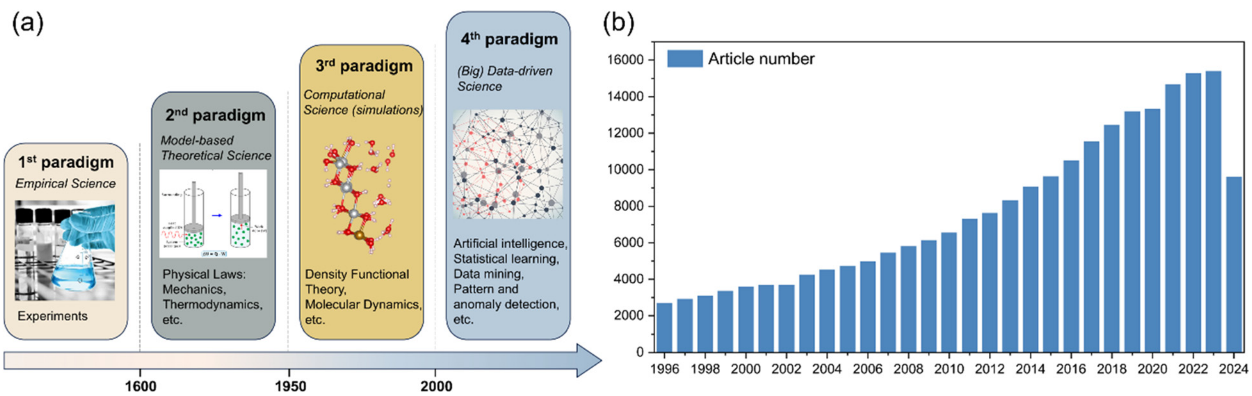

FIG. 1. (a) The four paradigms of science. (b) Recently published papers about the integration of ML and DFT in catalysts studies.

theoretical frameworks can enhance the understanding of catalyst behaviors, foster data-driven insights, and establish stronger connections between empirical observations and theoretical predictions.

Notably, small datasets and the black-box nature of algorithms significantly constrain the transferability and interpretability of ML-based frameworks in real-world applications.[21] This review will summarize the DFT development process and recent advances in ML approaches integrated with DFT, MD, and experimental characterization techniques. It will also identify the challenges associated with ML-based framework implementation and propose new directions for ML applications in electrocatalyst design and reaction mechanism exploration.

II. FUNDAMENTALS OF COMPUTATIONAL SIMULATION

A. DFT in electrocatalysts

Grounded in quantum-mechanical theory, DFT methods precisely determine the ground state energy of atoms in specified configurations, which facilitates the prediction of catalytic performance and the elucidation of reaction mechanisms through comprehensive analysis of electronic structures.[22] Recent advancements in engineering and scientific disciplines have enhanced the applicability of DFT calculations, significantly accelerating catalyst design by offering in-depth insights into the mechanisms.[1,23–25]

The DFT calculation provides a robust framework for evaluating electrocatalyst activity by quantifying the adsorption energies of key reaction intermediates, and the widely accepted Sabatier principle underscores the necessity of achieving an optimal balance in adsorption strength. Specifically, effective electrocatalysts exhibit adsorption energies that are neither too weak, which would hinder the activation of reactants, nor excessively strong, which would obstruct the conversion or desorption of intermediates and products. For HER, free adsorption energy (ΔG) for hydrogen near zero is desired; a highly negative ΔG_{H^*} restricts proton desorption, while a positive ΔG_{H^*} is unfavorable for H adsorption on catalytic sites.[26] Similarly, in NRR, the N_2 adsorption and NH_2 desorption are critical descriptors of catalytic activity. Moreover, the adsorption energies of OH and CO are commonly used as indicators for ORR and CO_2RR, respectively.[27–29] Notably, two primary parameters dictate OER activity: the difference between the adsorption energies of OOH and OH and the relative position of ΔG_{O^*}. Optimal OER performance is achieved when ΔG_{O^*} is positioned between ΔG_{OOH^*} and ΔG_{OH^*}.[30] The scaling relations play a pivotal role in understanding complex catalytic systems by correlating the adsorption energies of various intermediates. These relationships effectively reduce the degrees of freedom in multistep catalytic reactions, simplifying the analysis of adsorbate–surface interactions. Volcano plots serve as a valuable tool for visualizing activity trends, representing catalytic performance as a function of the adsorption energies of key intermediates. In such plots, weak adsorption limits reaction rates by the adsorption step, while excessively strong adsorption results in desorption as the rate-determining step.[31,32] Importantly, variations in adsorption environments can induce different responses in adsorbed species, which can uncover catalytic performance beyond conventional scaling relationships, thus offering new insights for catalyst design.

DFT also serves as a powerful tool for investigating electronic structures and elucidating electrochemical reaction mechanisms. Bader charge analysis is widely used to assess charge distribution at catalytic sites, providing insights into the bonding environment and charge population. Charge density difference calculations, which quantify the charge transfer between the electrocatalyst and reaction intermediates, are critical for characterizing bond strength. The extent of charge transfer is directly correlated with bonding interactions, influencing the adsorption and desorption of intermediates.[33,34] The density of states provides a detailed understanding of orbital interactions and electronic structure. The abundant states near the Fermi level facilitate electron evolution and transfer processes, indicating high catalytic activity. Projected density of states analysis further allows the examination of individual orbital contributions, elucidating the interactions between the molecular orbitals of adsorbates and the catalyst surface. Such analysis is crucial for understanding the electronic interactions that influence catalytic performance.[35,36] Crystal orbital Hamilton Population is an indispensable method for extracting detailed chemical bonding information from electron density calculations, which enable the characterization of bonding and antibonding states within the electronic structure, providing critical insights into the electronic interactions that govern catalyst behaviors. The integrated crystal orbital Hamilton population (ICOHP) is a quantitative descriptor of bond strength, and less negative ICOHP values correspond to stronger bonding interactions.[37,38]

As a physics and quantum chemistry-based methodology, DFT calculations enable explaining catalytic behaviors and reaction mechanisms via fundamental physical laws. However, their prohibitively high computational cost limits their applicability in catalyst design. The data-centric nature of ML has predominantly positioned its scientific applications within data-intensive areas, and its applications in the chemical domain are expanding rapidly. ML techniques use training data from experiments and DFT calculations to develop surrogate models, in which catalyst properties are characterized by specific attributes relevant to their applications. This synergy framework identifies the relationships between compositions, structures, and morphologies, facilitating the discovery of novel catalysts by incorporating the featurization step. This integration offers a transformative approach to deciphering complex catalysts across scales with higher accuracy and computational cost.[10,39,40] Integrating DFT and ML is an unavoidable trend in catalyst communities, achieving remarkable success in capturing optimal crystal structures in large search space.[41,42]

B. DFT accuracy assessment

DFT is significant among modern quantum chemical methodologies, evolving from the traditional wavefunction-based approach. The Schrödinger equation, introduced in the early 20th century, underpins quantum chemistry by providing the mathematical framework of quantum mechanics to describe electronic structures and atomic properties, thus enabling the prediction of chemical system behaviors. With the DFT development [Fig. 2(a)], the Kohn–Sham (KS) approach reformulates the problem of finding the ground-state electron density by introducing a set of noninteracting, single-electron orbitals. The KS transforms the complex many-body problem into a set of coupled single-particle equations, which can be solved iteratively using standard numerical methods.[43] The thermodynamic energy of the elementary step is then evaluated using the following equation:

$$E[\rho] = T[\rho] + V[\rho] + J[\rho] + E_{XC}[\rho],$$

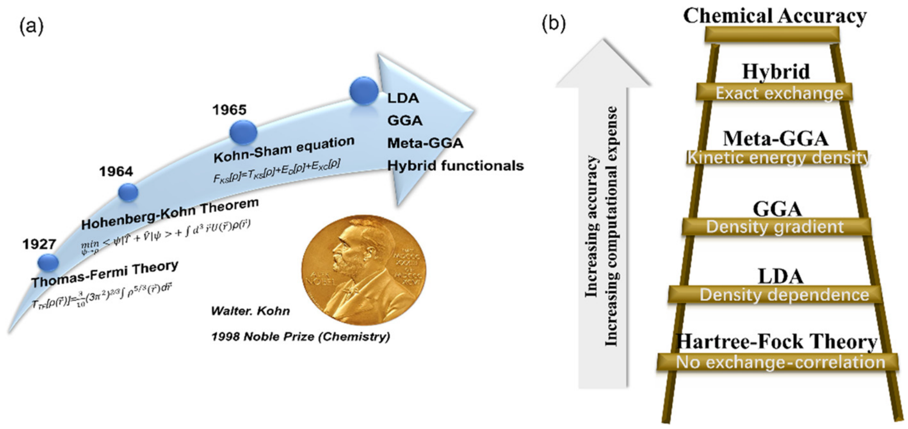

FIG. 2. (a) The summary of DFT development. (b) The schematic representation of different DFT functional approximations.

where the T is the kinetic energy of ideal noninteracting electrons. V and J are the potential energies of the classical electron–electron repulsion and the nucleus–electron interaction, respectively. The E_{XC} is the exchange–correlation energy to describe the nonclassical correction of the electron–electron interaction. The T, V, and J have explicit mathematical formulas, while the E_{XC} is an approximation function. Thus, the E_{XC} term has significant impacts on the accuracy and cost of DFT calculations.

The inherent errors in DFT functionals generate inaccuracy in describing the behavior of real physical systems. The accuracy of DFT results is highly sensitive to the choice of exchange–correlation functionals due to intrinsic errors associated with the KS equation [see Fig. 2(b)]. Each E_{XC} functional approximates the exchange–correlation energy in distinct ways, influencing the precision and reliability of computed results.[44] Local density approximation (LDA) is the simplest E_{XC} functional used in DFT, which approximates the exchange–correlation energy based on the energy per particle in a uniform electron gas.[45] Although the LDA has achieved practical success, it tends to underbind core electrons in atoms and overbind atoms in molecules or solids, resulting in errors exceeding 1 eV.[46] To address the limitations of LDA, generalized gradient approximations (GGAs) were developed to enhance the accuracy of the exchange–correlation energy calculations. Among these, the Perdew–Burke–Ernzerhof (PBE) and revised Perdew–Burke–Ernzerhof (RPBE) functionals are widely utilized in the design of electrochemical catalysts and analysis of reaction mechanisms. The PBE and RPBE functionals are based on similar construction logic and physical criteria, with the primary difference lying in the mathematical form of the exchange energy enhancement factor κ:1.245 for RPBE and 0.804 for PBE.[47] The PBE functional is extensively used in physics and surface science, while RPBE has demonstrated improvements over PBE in atomization and chemisorption energies.[48] Given the complexity of electrochemical catalysts involving multiple electron interactions, RPBE is often favored for most electrochemical reactions due to its enhanced accuracy in describing adsorption properties. Beyond GGAs, additional functionals have been developed to address the limitations of exchange–correlation functionals. Meta-generalized gradient approximations (meta-GGAs) extend the GGAs by incorporating kinetic energy density, enabling the distinction between localized and delocalized electron density regions. This feature makes meta-GGAs versatile for accurately predicting the adsorption energies of molecules on surfaces, lattice constants, and surface energies of transition metals.[49,50] In addition, hybrid functionals improve upon conventional exchange–correlation functionals by introducing a fraction of the exact Hartree–Fock exchange into the semi-local functional. This approach reduces self-interaction errors, enhancing the accuracy of thermodynamic properties, reaction energies, and electronic structures. Hybrid functionals are particularly effective in calculating more realistic band gaps in semiconductors and insulators and improving predictions of molecular properties, such as geometries, dipole moments, and vibrational frequencies.[51–56] However, the significant computational expense of meta-GGAs and hybrid functionals limits their applicability in large-scale or high-throughput calculations. Upon evaluating the strengths and limitations of various functionals, it is evident that the RPBE and PBE functionals are the most prevalently employed in exploring catalytic performance and material properties, offering a favorable trade-off between computational efficiency and accuracy. Notably, the PBE and RPBE functionals exhibit a mixed record of success and failure in metal-based catalysts, because their treatment of the exchange–correlation hole leads to a diffuse and extended large-U tail in solid-state systems. To overcome the limitations of GGAs in strongly correlated systems, DFT is frequently supplemented with a Hubbard U correction term (DFT+U). The Hubbard U term introduces an on-site Coulomb correction potential to address self-interaction errors that arise from electron delocalization, thereby enhancing the accuracy of the electronic structure description in strongly correlated systems.[54,57]

C. MD in electrocatalysts

DFT provides detailed electronic structures for small systems with tens or hundreds of atoms. However, DFT struggles to account for realistic reaction conditions, like temperature, pH, and specific cations, leading to discrepancies between theoretical predictions and experimental observations. Volcano-shaped relationships and microkinetic models simulate catalytic performance under operational conditions, but encounter significant challenges when applied to large-scale systems and long-time scales, including phase transitions, intricate reaction networks, multiple product formations, etc.[58–60] MD simulations track atomic motions over time, offering valuable insights into dynamic phenomena such as phase transitions and system evolution. Furthermore, MD simulations facilitate the exploration of fundamental interactions at the solid–liquid interface, exploring the characterization of processes such as bond formation and dissociation, intermediate states identification, and rate-limiting step determination.[61] These capabilities make MD particularly effective for elucidating the influence of electrolytes on solvation structures and catalyst surface interactions, facilitating a deeper understanding of electrocatalytic processes under realistic conditions.[62–65] As a multiscale approach, integrating MD with DFT bridges the gap between atomic-level details and macroscopic behavior, providing a comprehensive framework for studying electrochemical systems in complex environments.[66]

Force fields are fundamental to MD simulations, describing the potential energy of particles (atoms, ions, or molecules) and the forces between them. They capture key physical interactions, facilitating the simulations of molecular dynamics and thermodynamics in complex systems across timescales ranging from microseconds to milliseconds and system sizes from thousands to millions of atoms. This makes them invaluable for investigating diffusion, reaction kinetics, and conformational changes under realistic conditions. However, classical force fields rely on empirical parameters derived from experimental or quantum data, limiting their transferability and accuracy. Moreover, they lack explicit electronic structure and quantum effects, such as charge transfer and polarization, which are critical for accurately simulating systems with strong covalent or ionic interactions.[67] The ML offers a transformative approach to generating force fields with enhanced accuracy and transferability. ML-based force fields, trained on high-fidelity quantum mechanical datasets, capture intricate interactions that classical force fields often fail to represent. This advancement facilitates simulations of larger systems over extended timescales, significantly improving the reliability of MD simulations in complex electrocatalysis systems.

III. MACHINE LEARNING APPLICATION FOR ELECTROCATALYST DESIGN ACCELERATION

The systematic DFT calculations have made remarkable contributions to the design of electrocatalysts and the investigation of reaction mechanisms, enabling an understanding of electronic structures and the identification of key intermediate steps.[68,69] However, the vast number of potential catalyst structures, the diverse physicochemical properties of complex catalysts, and the intricate reaction mechanisms remain poorly understood, owing to the prohibitive computational costs. Fortunately, ML algorithms establish a nonlinear map between input computational/experimental data and targeted properties, avoiding unnecessary experiments and calculations.[70–73] The ML-guided workflow is also designed to screen potential reactions and identify mechanisms of interest, accelerating the discovery and optimization of electrochemical reactions by efficiently narrowing down the most promising candidates for further investigations.[74] ML-based frameworks significantly enhance the efficiency of developing and commercializing electrocatalysts.

A. The integration of DFT and ML

As a statistical analysis tool, the data-driven ML approach, combined with DFT-calculated data, offers an efficient framework to understand unexplored factors influencing catalytic performance. Subsequently, the established structure–property–performance relationships serve as predictive frameworks to screen potential catalyst candidates for various electrochemical reactions, significantly accelerating the selection of optimal catalysts.[60,75,76] The combination of DFT and ML techniques addresses the challenges associated with the high costs of experimental trials and conventional DFT calculations.

1. Understand the origin of catalytic activity

Adsorption energies of key intermediates on catalyst surfaces serve as crucial indicators of catalytic activity in electrochemical reactions. For example, the adsorption energy of N atom (ΔG_N) indicates catalytic activity for nitrogen fixation in NRR.[77–79] Jiang *et al.* adopted the Sure Independence Screening Sparsifying Operator (SISSO) method to screen transition metal single-atom catalysts (TM-SACs) supported on the N-doped graphene and nanographene. The single descriptor model based solely on ΔG_H* revealed that single V, Rh, and Ir atoms embedded in N-doped nanographene significantly enhance HER activity. This was because the stronger interactions between the SACs and the substrate generate hybridized electronic states at lower energy levels, leading to a downshift in the *d*-band center. Additionally, the extent of hybridization between the SACs and the carbon atoms was more pronounced in smaller nanographene structures compared to their 2D graphene counterparts, highlighting the importance of substrate size and dimensionality in modulating catalytic performance.[80] Furthermore, a two-descriptor model incorporating ΔG_H and work function exhibited improved predictive accuracy in acidic and alkaline environments because the work function addresses the overestimation of Cu activity observed in transitional volcano plots.[81] Additionally, Huang *et al.* employed a random forest regression (RFR) to predict the catalytic performance of the hybrid metal (M) and nonmetal (NM) embedded in g-CN, using seven structural parameters as key features. The Re-Se/g-CN and Fe-Te/g-CN catalysts exhibited high activity, emphasizing the importance of the hydrogen affinity of the M center and the spatial configuration of hydrogen adsorption sites. The coordination environment of the M-NM pair synergistically influenced catalytic activity. In particular, the distance between H and NM species, and the angle between the H-M and M-NM bonds were determined by the interaction strength between the M and NM components. A stronger M-NM interaction generated weakened hydrogen adsorption, as it reduced empty electronic states for back-donation from the hydrogen atom.[82] For similar carbon-based catalysts, including SACs embedded in N-doped carbon and TM/Ln metal-doped graphdiyne (GDY), the Bag-Tree approach has identified valence electron number and covalent radius as the primary determinants influencing ΔG_H.[83,84] The CatBoost regression also predicted the activity and selectivity of SACs supported on mono- or dual-type nonmetal doped g-C_3N_4 by leveraging electronic, geometric, and thermodynamic features, such as the lowest-unoccupied Kohn–Sham eigenvalue and unpaired *d*-electrons number. By extensively screening 364 catalysts, Fe/P_3, Mn/P_4, and Fe/P_4 exhibited ultra-low HER overpotentials between -0.01 and -0.03 V, outperforming commercial Pt-based catalysts. The findings indicated that ΔG_{H*} was largely influenced by the coordination environment, the distinctive charge transfer facilitates the HER activity. The position of the *d*-band center relative to the Fermi level reflected the nature of orbital hybridization. For TMs with a *d*-band center close to the Fermi level, strong hybridization occurred between the H-*s* orbital and the TM-*d* orbital, making the metal top site the most favorable for hydrogen adsorption. Conversely, for TMs with a *d*-band center located farther from the Fermi level, the primary orbital hybridization involved the H-s orbital and C-p orbitals, in which the TM-C bond edge sites became more active for H adsorption.[85]

SACs have also emerged as promising catalyst candidates in OER and ORR. The few-shot learning algorithm revealed that the intrinsic catalytic ORR activity of TM-SACs supported on defective carbon surfaces is closely related to topological, bonding, and electronic structures of the carbon support, as well as the interactions at carbon defect sites.[86] The RFR was employed to screen 1200 double-atom and nanocluster-based catalysts, and the carbon-based Fe−Ce catalyst was predicated as the best-performing for ORR. The binding energy, Bader charge transfer, and reaction kinetics were important features influencing the adsorption energies of critical intermediates.[87] Fan *et al.* utilized the ML-DFT frameworks to systematically investigate the bifunctional catalyst performance and intrinsic factors of Pt-doped dual TM Janus-MXenes-based SACs for ORR/OER. The RFR model identified the most crucial determinants of ORR catalytic activity as Pt binding energy, d-band center, and hydride formation enthalpy. For OER, the extreme gradient boosting regression (EGBR) highlighted Pt binding energy, lattice constant, and charge transfer to Pt as the most predictive descriptors. Pt-VO-MnTiCO$_2$ and Pt-VO-PdTiCO$_2$ were exceptional catalysts for the ORR and OER, with overpotential ($\eta^{ORR/OER}$) of 0.24/0.38 and 0.33/0.36 V, respectively. Electronic structure analysis reveals that Janus-MXenes enhanced the catalytic activity by regulating the charge arrangement, thereby altering their electronic properties. The enhanced catalytic performance was also attributed to the synergistic effects of dual transition metals (DTMs), which modified inherent electron configurations. Compared with Ti_2CO_2, DTM-based Janus-MXenes exhibited significant changes in electronic configuration, electronic properties, and the strength of the binding interaction between Pt and the substrate, which modulate catalytic activity collectively.[88] The EGBR was also employed to assess the feasibility of TM-SACs/defective g-C_3N_4 for bifunctional ORR and OER, exhibiting that the Rh/V_N-g-C_3N_4 demonstrated low overpotentials of 0.43 V for ORR and 0.32 V for OER. After an in-depth examination of the correlations among adsorption behavior, structural descriptors, and atomic descriptors, the first ionization energy and TM charge transfer were identified as the most critical factors influencing the adsorption capability.[89] The RFR was employed to predict TM/MnPS$_3$ in ORR/OER, in which Rh/MnPS$_3$ and Ni/MnPS$_3$ were identified as efficient bifunctional electrocatalysts. Key features included Ni-O bond length, d electrons number, the d-center, the atomic radius, and the first ionization energy (I_1) of TM atoms.[90] Additionally, as Fig. 3 shown, the gradient boosting regression (GBR) provided insights into the catalytic activity of TM-anchored AlP, identifying that the d-electrons number, TM atom radius, and TM charge transfer were primary descriptors influencing the adsorption behavior. Co- and Ni-based defective AlP systems demonstrated $\eta^{OER/ORR}$ of 0.38/0.25 and 0.23/0.39 V, respectively. The interactions between TM d-orbitals and electronic states of adsorbates formed bonding and antibonding states. Specifically, the overlap of TM d-orbitals and O p-orbitals resulted in the splitting of hybridized energy levels, with bonding states occupying energy levels below the Fermi level and antibonding states above it. The d-orbital electron configuration of the TM atom was important in determining adsorption energy. The lower d-band center correlated with greater occupancy of antibonding states, leading to a weakened adsorption strength, while a higher d-band center enhanced

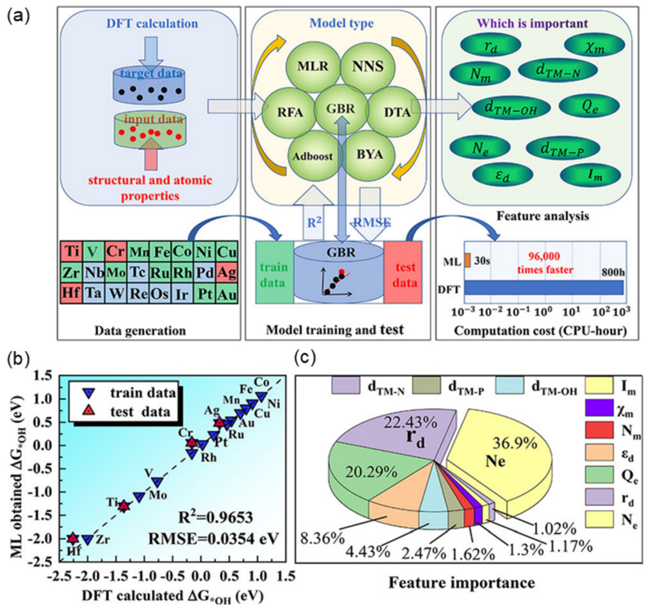

FIG. 3. (a) Brief description of the ML process for exploring TM-anchored AlP bifunctional electrocatalysts in ORR/OER, including DFT calculated data, the ML model data, and the feature importance analysis. (b) Comparison of ΔG_{OH*} value from DFT calculation and the GBR algorithm. (c) Feature importance ranking of each descriptor. Reproduced with permission from Liu *et al.*, Appl. Mater. **14**, 1249–1259 (2022). Copyright 2025 American Chemical Society.[91]

adsorption by reducing antibonding state occupancy.[91] Notably, the active-learning algorithm accelerated the identification of electrochemically stable Ir oxide polymorphs, and a novel α-IrO$_3$ with an uncharacterized FeF$_3$-type structure was discovered with thermodynamic amorphous synthesizability and global stability. The octahedral local coordination environments were preferred for nearly all low-energy structures, and subsequent Pourbaix Ir−H_2O analysis showed that α-IrO$_3$ was the globally stable solid phase under acidic OER conditions.[92] Furthermore, the crystal graph convolutional neural network (CGCNN) predicted surface coverages and activity for various active sites and configurations on O*- and HO*-covered surfaces. The low-index surfaces of α-IrO$_3$ had lower theoretical overpotential by 0.2 V than the benchmark rutile IrO$_2$(110).[93] Similarly, the adaptive Gaussian process searched 4000 perovskite structures, and ten unknown candidates were screened as potential candidates in OER. The bulk electronic structure features offered valuable molecular orbital insights into the OER activity of perovskites. Specifically, the electronic characteristics of the e_g orbitals of the metal B-site were crucial in influencing the ΔG_{O*} and ΔG_{OH*}, which was also critical for Ni-based MOF catalysts in OER. The d_{Z^2} orbital, as a component of the e_g orbitals, directly overlaps with the p orbitals of oxygen intermediates at surface catalytic sites, thereby influencing the binding energies of these intermediates.[94,95]

Importantly, the ML-DFT framework exhibits outstanding efficacy in analyzing intricate electrochemical reaction networks. Evaluating catalyst selectivity is essential for optimizing desired product yields within specific reaction systems, as different products are governed by distinct descriptors. These descriptors serve as key parameters for deciphering and quantifying complex reaction mechanisms, offering a comprehensive understanding of catalytic behavior and performance. By providing a framework for differentiating competing reaction pathways, such descriptors enable the systematic analysis and rational design of catalysts to enhance target product selectivity. For example, CO_2RR is crucial for converting CO_2 into valuable industrial feedstocks. However, its intricate reaction mechanism involves various carbon-containing products, with the HER serving as a competing pathway.[96,97] The composition and coordination number of active sites are governed by the size and geometry of catalysts, which influence the binding strengths of various reaction intermediates. Tuning the adsorption strengths of these intermediates modulates reaction pathways and enhances catalyst selectivity. Among pure metals, late transition metals, such as Pt and Rh, exhibited superior HER activity due to their strong hydrogen binding affinity.[98–100] In contrast, precious metals, like Ag and Au, favored CO^* production as their weak binding energies for OH^* and CO^* suppress the formation of $HCOO^*$ and the protonation of CO^* to COH^*.[101–103] Similarly, *p*-block elements selectively promoted formate production, as their weaker bonding with CO^* compared to OH^* facilitated the formation of $HCOO^*$ over $COOH^*$.[104] Notably, Cu was a unique catalyst capable of producing various products due to its moderate binding energies for CO^*, OH^*, and H^*. This balance in adsorption energies enabled Cu surfaces, such as Cu(111) and Cu(211), to synthesize C_{1+} product, rendering Cu a versatile catalyst for CO_2RR with diverse carbonaceous products. Consequently, understanding and controlling reaction pathways was critical to optimizing the efficiency and selectivity of CO_2RR.[105,106] The RFR screening of 465 elemental combinations identified Cu-Pd and Cu-Ga binary alloys as exhibiting intrinsic selectivity for C_{1+} products, respectively. For Cu-based alloy, a high proportion of edge and corner sites with low coordination numbers, coupled with high Cu content, favored C_{1+} product formation. Conversely, higher coordination numbers and reduced Cu content enhanced formate production, highlighting the critical role of active site structure and composition in determining product selectivity.[101,103,107] Furthermore, altering the substrate selectively tuned the adsorption energy of reaction species and influenced the CO_2RR pathways. The RFR analysis revealed that the polarized charge and magnetic moment of the metal center affected the catalytic performance of SACs supported on GDY and holey graphyne (GHY). A larger polarized charge on the central metal correlates with increased charge transfer to CO_2, suggesting that a vacant *d*-orbital facilitates the capture and activation of CO_2. However, excessive polarization generated CO_2 overactivation, resulting in overly strong binding between the central metal and the oxygen atom. This induced more negative adsorption energy for OH species, thereby increasing the overpotential for the HER. These observations suggested that modifying the supports can effectively tune the electron polarization of the central metal's *d*-orbitals, and reduced polarized charge on the metal center promotes moderate CO_2 activation. A comprehensive screening across the periodic table identified Mn/GDY, Co/GHY, Ru/GDY, and Os/GDY as promising candidates for CH_4 production with low limiting potentials ranging from -0.22 to -0.58 V. The morphologies of catalysts determine the composition and coordination number of active sites.[108] Additionally, Zhen *et al.* utilized the EGBR to predict ΔG_{CO} across 1060 metal−nonmetal co-doped graphene systems, which incorporated key features such as the Pauling electronegativity and covalent radius of metal atoms. After the rapid screening, the Sc-C_2O_2, Sc-CN_2O, and Y-CN_2O were identified as potential catalysts with lower overpotentials in CO_2RR.[109] Furthermore, the deep neural network (DNN) and light gradient boosting machine (LGBM) models were employed to identify optimal boron-doped graphene SACs for NRR, which incorporated geometrical structure and bonding characteristics as key features. Among 182 catalyst candidates, CrB_3C_1 exhibited excellent selectivity with a minimal overpotential of 0.13 V.[110] Chen *et al.* utilized the boosted regression tree ensemble method with seven intrinsic features to predict the catalytic activity of TM-SACs supported on 2D substrates. After screening 139 candidates, isolated electron numbers in *d* orbitals exhibited the strongest correlation with NRR activity. Ru/N_4-C was predicated high NRR activity, and subsequent experiments validated the potential of 0.55 V.[111]

To assess accuracy, empirical corrections to calculated energetics are often necessary to align computational predictions with experimental data, providing reliable and comparable thermodynamic results. Symbolic regression (SR) was utilized to simultaneously incorporate the octahedral factor (μ) and the tolerance factor (t) in identifying the optimal oxide perovskite catalysts for OER, enhancing its robustness and applicability.[112] The statistical learning (SL) algorithm merged experimental and DFT results to assess the activity and selectivity of the hydrodehalogenation of CH_2X_2 (for X = Br, Cl). Such a hybrid data approach had the potential to overcome the limitations of classical microkinetic methods.[61] Moreover, the SISSO approach, which incorporated primary features derived from both experimental and computational results, was employed to investigate the catalytic performance of Co nanoparticles supported on Si for CH_3OH production via CO_2RR (Fig. 4). This method incorporated features such as the Co species reducibility, the intermediate's adsorption strength, and the chemical properties of additive metals, revealing that V and Zn can enhance the CH_3OH selectivity of Co NP/Si.[113] Owing to experimental limitations, extensive hypothetical reaction mixtures were overlooked, resulting in the omission of numerous potentially groundbreaking and useful synthetic reactivities. The discovery pipeline involving "experimental testing—ML analysis—prediction and redesign" provided significant insights into catalyst optimization. Cu catalysts were synthesized using various additive combinations under controlled conditions, and their performance was evaluated based on Faradaic efficiency (FE) and current density in CO_2RR. The RIT method was then employed to identify critical features of catalyst composition that correlate with the FE of different products. From ML insights, additives containing aliphatic OH groups enhanced the FE for generating C_{1+} products, and Sn-based additives increased the yields of CO and HCOOH. Aliphatic alcohols further enhanced HCOOH production by facilitating the reduction of Cu_2O to metallic Cu. Guided by these ML insights, new catalysts were synthesized and validated by experiments. This iterative approach facilitated the ML model refinement, enhancing its predictive accuracy and reliability for guiding catalyst design.[114] The RFR also combined experimental and DFT results to deliver quantitative predictions for optimal catalysts for OER across a broad doping space, which thoroughly considered various factors including composition, morphology, phase, electrolyte pH, and working electrode type. The $Ni_{0.77}Fe_{0.13}La_{0.1}$ was predicated as a high-activity hydroxide catalyst with the experiment validated low overpotential of 0.226 V.[115]

The electrochemical reaction network is characterized by considerable complexity and uncertainty under practical conditions, as it

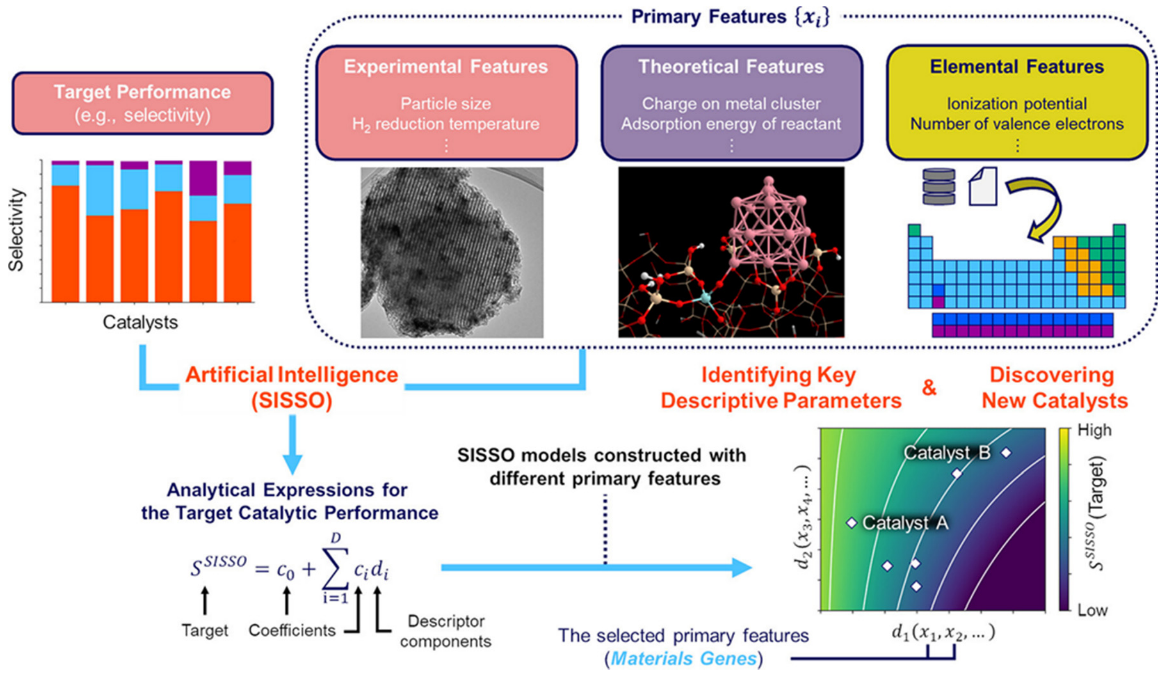

FIG. 4. Schematic outline of the ML-DFT framework with experimental data correlations of Co nanoparticles supported on silica. By integrating information from various material parameters (primary features) using the SISSO approach, identify analytical expressions and key descriptive parameters (material genes) correlated with the experimentally measured target catalytic performance. The SISSO models, which utilize primary features with low acquisition costs, further facilitate the accelerated discovery of high-performance catalysts. Reproduced with permission from Miyazaki *et al.*, J. Am. Chem. Soc. **146**, 5433–5444 (2024). Copyright 2024 Authors, licensed under a Creative Commons Attribution (CC BY) license.[113]

encompasses a multitude of potential intermediates and reaction pathways. This inherent complexity significantly complicates understanding various experimental phenomena and optimal catalyst design. DFT is combined with the ML method to intrinsically describe the catalytic nature, thereby accelerating the screening and prediction of high-performance electrochemical catalysts. It should be noted that different algorithms describe adsorption using distinct features. The identified design principles also provide a roadmap for closing the gap between current artificial photocatalysts and enzyme catalysts.[116,117] Recent ML-DFT frameworks for investigating the origin of activity in electrocatalysts are listed in Table I.

2. Explore rational active sites

The active site is central to the function and efficiency of catalysts, impacting reaction rates, substrate specificity, reaction mechanisms, etc.[69,118] Catalysts with compositional variations, such as bimetallic crystals, are systematically enumerated and cataloged, resulting in hundreds of potential active sites. Usually, active site identifications rely on experimental and theoretical analyses of existing catalysts, but many potentially important active sites remain underexplored. Employing newly developed ML potentials as surrogate models for DFT calculations is utilized to analyze feasible active sites within practical constraints, providing rational explanations of experimental observations and uncovering new and effective catalytic sites.

Norskov *et al.* introduced a surrogate model utilizing Gaussian process regression (GPR) to predict adsorption energies based on group additivity fingerprints. This model, combined with transition state information, served as a simplified classifier for identifying the rate-limiting step in the syngas reaction on Rh(111). This framework facilitated the efficient identification of the most probable reaction mechanisms under uncertain conditions, thereby significantly reducing the computational costs of high-accuracy DFT calculations.[119] Figure 5 shows the workflow of the neural network approach combined with the Monte Carlo method and experiments consistently, demonstrating that PtFeCu was highly stable and active for ORR because the atomic distribution of Cu led to beneficial modulation of surface strain and segregation.[120] Support vector regression (SVR) estimated 2500 active sites of various edge-doped graphene nanoribbons in OER, and 25 ideal active sites were selected with lower overpotential ($\eta < 0.5$ V) and validated via DFT calculations.[121] Moreover, the simple linear ML model was employed to predict the ΔG_{OH} and ΔG_O on high-entropy alloys (HEAs) of IrPdPtRhRu. Due to randomly accessible binding sites on the IrPdPtRhRu surface, the number and type of nearest-neighbor atoms emerged as critical features in the predictive model. The IrPt demonstrated substantial enhancements in ORR activity compared to pure Pt(111), highlighting the importance of HEAs composition and surface structures.[122]

Notably, the RFR discovered the origin of the volcano-shaped relationship between the catalytic activity and ΔG_O in ORR, which further incorporated element-specific activity by considering the outer electron number and oxide formation enthalpy, providing accurate predictions of ΔG_O with reduced time and computational costs.[78] The ML approaches have also proven effective in breaking scaling

TABLE I. Recent ML-DFT frameworks for investigating the origin of activity in electrocatalysts.

Reaction	Catalysts	ML algorithms	Main features	Reference
HER	TM/Ln-GDY	Bag-tree regression	Valence electron number Covalent radius	84
HER	DACs-g C_3N_4	CatBoost regression	Lowest-unoccupied Kohn–Sham eigenvalue Unpaired d-electrons number	85
ORR	SACs-carbon	Few-shot learning algorithm	d-electron count First ionization energy Electronegativity	86
ORR	Carbon-based DACs	Random forest regression	Binding energy Bader charge transfer Reaction kinetics	87
ORR	Janus-MXenes-based SACs	Random forest regression	Pt binding energy d-band center hydride formation enthalpy	69
OER		Extreme gradient boosting regression	Pt binding energy Lattice constant Charge transfer to Pt	
OER	TM-SACs/g-C_3N_4	Gradient boosting regression	First ionization energy TM charge transfer	89
ORR/OER	TM/$MnPS_3$	Random forest regression	Ni-O bond length d electrons number First ionization energy	71
ORR/OER	TM-anchored AlP	Gradient boosting regression	d-electrons number TM atom radius TM charge transfer	91
OER	perovskites	Gaussian process	e_g orbitals of the metal	94
OER	Ni-based MOF	Active learning	Electronic characteristics of the e_g	95
CO_2RR	SACs-GDY	Random forest regression	Polarized charge Magnetic moment of metal atoms	108
CO_2RR	DMCs-graphene	Extreme gradient boosting regression	Pauling electronegativity Covalent radius of metal atoms	109
NRR	SACs-B-graphene	Deep neural network Light gradient boosting machine	Electronegativity Atomic radius Atomic number of TMs	82
NRR	TM-SACs-2D materials	Boosted regression tree	Isolate electron number in *d* orbitals	83

relationships via altering active sites. Singh *et al.* developed an ML model to search optimal catalysts by evaluating 1280 adsorption sites on a HEA composed of FeCoNiCuMo. Their approach circumvented traditional scaling relations between the adsorption energies of $COOH^*$, CO^*, and CHO^* through the strategic rotation of $COOH^*$ and CHO^*, and the lowest potential of the designed active site was 0.29 V for CH_4 production.[123] A comprehensive multiple linear regression (MLR) analysis of 149 candidates also revealed that their catalytic performance was closely related to factors, such as TM-O band hybridization, *d*-band center, and charge transfer between M and OOH^*. The framework also broke the ORR scaling relations and identified that the Zn/PcN_4 possessed high activity and selectivity toward H_2O_2 production at 0.15 V.[124] Moreover, Sargent *et al.* developed an integrated framework combining DFT simulations, ML regression, and ML prioritization in a cyclical process to systematically explore surface orientations, adsorption sites, and Al–Cu ratios for identifying near-optimal ΔG_{CO}. After conducting 4000 DFT simulations, the optimized Cu–Al alloys demonstrated an 80% FE for C_2H_4 production via CO_2RR, highlighting that multimetal catalysts preferred to surpass the performance of single-component systems. The Cu-Al combination created a favorable Cu coordination environment that enhanced C–C dimerization to improve catalytic performance.[125] Similarly, Uliss's group employed active learning to screen 1499 intermetallics, identifying 131 candidate surfaces with potential high CO_2RR activity due to near-optimal ΔG_{CO} values. This model highlighted the importance of the distribution of active sites across intermetallic surfaces, and combining two weak-binding elements can enhance surface activity. The strongest-binding Ga and Au surfaces exhibited binding energies of

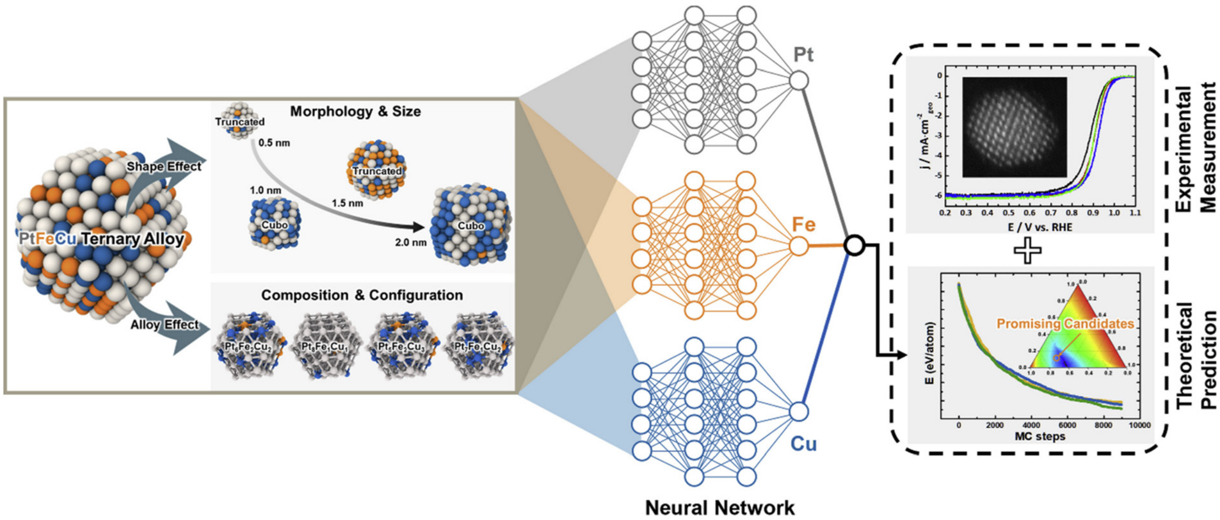

FIG. 5. Schematic representation of the search for ternary alloy configurations, including theoretical predictions and experimental validation for ORR. Reproduced with permission from Chun *et al.*, Chem. Catal. **1**, 855–869 (2021). Copyright 2025 Elsevier.[120]

approximately −0.44 and −0.53 eV, respectively. However, a Ga–Ga bridge site on the $AuGa_2(100)$ surface achieved a near-ideal binding energy of −0.57 eV, illustrating the potential for improved catalytic performance through strategic site engineering.[28] The neural network also identified novel active site motifs on Ni/Ga intermetallics for CO_2RR, incorporating seven features related to the surrounding environment and metal ratios. The isolated Ni atom surrounded by Ga atoms, a configuration not previously explored, demonstrated step-like kinetic behavior and correlates with the observed catalytic activity.[126]

Active sites not only influence catalytic activity but also affect reaction mechanisms. C_1 products are particularly favorable in CO_2RR, but different active sites can alter reaction pathways and the resulting products. Integrating GPR with DFT calculations was utilized to explore CO_2RR trends of 15 000 SACs/GDY, which revealed that acidic environments favor intermediate migration within the electroactive region by lowering energy barriers. Additionally, the framework provided comprehensive predictions for C_3 pathways in CO_2RR, highlighting critical C−C−C coupling trends and identifying strongly correlated factors.[127] Investigating key intermediate adsorption steps helps to select crucial active sites. The key C−C bonding scission reaction is a critical step in ethanol reforming, and the least-absolute shrinkage and selection operation (LASSO) method based on computed and experimental data identified that Ni−Pt−Pt(111) and Pt−Ni−Pt(111) were promising catalysts.[128] For ORR, O_2 binding activity, as the initial step, received extensive attention in reaction mechanisms investigation. The multilayer perceptron method was applied to classify active sites of graphite-conjugated catalysts concerning their O_2 binding activity. C atoms located at the ortho or para to N atoms and the edge of aromatic systems exhibit elevated catalytic activity, attributing to greater electron density transfer from the catalyst to the terminal O atom. Both electronic and structural characteristics, including spin density and variations in partial charge at the C site, significantly influenced the catalytic behavior.[129] The C−H activation reaction of alkanes is an important chemical reaction with applications in many industrial fields. However, many factors influence the rate of C−H activation with complex relationships, making it challenging to predict reaction rate constants using traditional methods. The gradient boosted regression trees (GBRT) model can quantitatively describe experimentally measured reaction rate constants in terms of multiple cluster electronic features (average electron occupancy of valence s orbitals, the minimum natural charge on the metal atom, cluster polarizability, and energy gap involved in the agnostic interaction), demonstrating that the general mechanisms governing the C−H activation of butane. After analyzing over 100 Rh-based clusters, the Rh atoms number, metal cluster shape, and the electronic structure of the metal clusters had the most significant impact on the C−H activation reaction rate constants.[130]

3. Accelerate solutions of DFT and MD

The core of DFT is the Kohn–Sham (KS) equations, ML-based methods can approximate the function of the KS equations, potentially bypassing the direct solutions, thus saving substantial computational time. Felix and colleagues applied Kernel Ridge Regression to learn density-potential and energy-density mappings, generating a machine-learned density functional for malonaldehyde within MD simulations that successfully captured the intramolecular proton transfer process.[131] As Fig. 6 shown, neural networks trained on reference DFT results have proposed paradigms for high-fidelity emulation of KS functionals, calculating the Fermi level and density of states with reduced computational expense.[132] Similarly, Bayesian linear regression using a local similarity kernel allowed interrogation of catalytic activities based on simulated local atomic configurations, as validated by studies on NO decomposition on RuAu alloy nanoparticles. Furthermore, considering kinetic analysis offered detailed information on structures of active sites, facilitating the examination of size- and composition-dependent catalytic activities.[133]

MD simulations can provide detailed information about the behavior of large systems over time scales, including the movement of atoms and molecules, interactions, structural changes, and

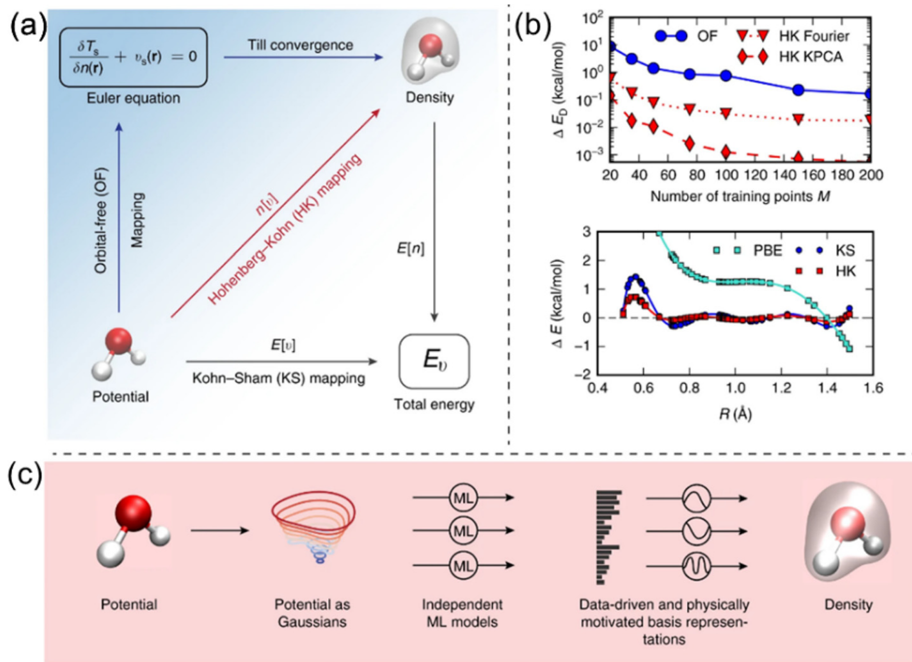

FIG. 6. (a) Mappings for bypassing the Kohn–Sham equations with ML approaches. The black arrow indicates E[*v*], which represents a conventional electronic structure calculation (KS-DFT), where the ground-state energy is obtained under the external potential. E[*n*] denotes the total energy density functional. The red arrow represents the Hohenberg–Kohn (HK) mapping n[*v*], which translates the external potential to its corresponding ground-state density. (b) Analyzing the energy error depends on *M*, the number of training points, for ML-OF and ML-HK with different basis sets for the 1D problem (top). Errors in the DFT-calculated energies and the ML maps as a function of interatomic spacing. (c) The prediction process of ML–Hohenberg–Kohn map. Gaussians represent the molecular geometry; many independent Kernel ridge regression models predict each basis coefficient of the density. Reproduced with permission from Brockherde *et al.*, Nat. Commun. **8**, 872 (2017). Copyright 2017 Authors, licensed under a Creative Commons Attribution (CC BY) license.[131]

thermodynamic properties. Integrating ML with DFT and MD simulations enhances the range and speed of simulations with high accuracy. For instance, Rice and colleagues employed a high-dimensional neural network in conjunction with long-timescale MD simulations to investigate the HER mechanisms on Pt(111) under acidic conditions. This model explicitly considered interactions such as adsorbate–adsorbate, adsorbate–solvent, and ion solvation effects, revealing that the preferred reaction mechanism is influenced by surface coverage. The study highlighted that manipulating solvent–substrate interactions was essential for advancing HER catalysts beyond Pt.[134] The solvent–substrate also played a significant role in influencing thermodynamics and kinetics performance in ORR. The RFR approach was developed for ultrafast covalency competition screening of over 300 spinel oxides, demonstrating the O-*p* band center and the relative reactivity between the O-*p* and TM-*d* band influenced their activity. Spinel oxides underwent surface reconstruction into amorphous or hydroxide structures.[135]

In MD, interatomic potentials describe interactions among atoms, allowing the accurate simulations and analysis of systems' dynamic behavior and properties. Traditional interatomic potentials are typically based on specific physical models and involve adjusting parameters to match experimental or calculated results. This approach relies on predefined functional forms, which limits its accuracy and applicability in complex systems. Recent studies have demonstrated that ML techniques can approximate extensive electronic structure data obtained from DFT calculations, leading to the development of ML-based interatomic potentials (ML-IAPs). These ML-IAPs can achieve accuracy comparable to DFT calculations while significantly reducing computational costs.[136–139] ML-IAPs have been effectively utilized to identify ground-state configurations for Re- and Cs-promoted Ag catalysts, improving the understanding of structure–selectivity relationships. The charge transfer from Re and Cs promoters to Ag catalysts was defined as a significant descriptor for ethylene oxide selectivity.[140] ML-AIPS also successfully studied the thermodynamic and mechanical properties of HEA. The moment tensor potentials were applied as a reference potential within the Langevin Dynamics method for evaluating the full vibrational free energy of chemically complex disordered VnbMoTaW. Monte Carlo simulations were further employed to investigate the phase stability, phase transitions, and chemical short-range order of the NbMoTaW alloy, which confirmed that the single-phase formation of NbMoTaW remains stable at room temperature.[141,142] Similarly, the on-lattice Metropolis Monte Carlo algorithm utilized ML-AIP to investigate the short-range ordering of CoCrFeNi HEA, indicating that Fe and Cr atoms form sublattices at specific temperatures.[143] Additionally, ML was also important in the development and optimization of force fields. The Δ-machine learning provides a flexible framework to predict redox potentials (Fe^{3+}/Fe^{2+}, Cu^{2+}/Cu^{+}, and Ag^{2+}/Ag^{+}), which were well matched with experimental estimations.[144] The consistent success of the smooth overlap of atomic-Gaussian approximation potentials

(SOAP-GAP) framework across a range of materials, molecules, and biological systems demonstrates the feasibility of bypassing explicit electronic structure and free energy calculations. This success can be attributed to the SOAP-GAP framework's general and mathematically rigorous approach to representing local chemical environments, including its ability to effectively capture assumed correlations between the contributions of different chemical elements.[145,146]

B. The integration of ML and experiment characterization techniques

Theoretical approaches for evaluating activity descriptors are based on correlating variations in catalyst performance with changes in structural and electronic properties.[147,148] However, catalytic species can exhibit multiple structural configurations and oxidation states under reaction conditions. The limited understanding of intermediate species characteristics and diverse catalyst configurations is an open challenge for identifying reliable descriptors. Fortunately, advanced experiment characterization techniques allow for the direct observation of surface structures and intermediate configurations, offering a robust reference for theoretical analysis and prediction. However, analyzing data generated from these techniques remains a complex and time-demanding task.[149] Integrating ML with characterization techniques represents a cutting-edge approach to facilitate reaction mechanisms exploration and active site identification.[150,151]

Transmission electron microscopy (TEM) is a powerful tool for elucidating the morphological and temporal evolution of catalysts through sequential snapshots of their microstructure. However, TEM images often suffer from background noise and indistinct boundaries, necessitating noise reduction during image analysis. Traditional manual analysis of TEM videos is laborious, error-prone, and time-consuming. Variability in researcher proficiency and individual preferences further contributes to inaccuracies. To address these challenges, the YOLO deep learning (DL) model has been employed to automate the analysis of TEM data for FeCrAl alloys, including size, shape, density, motion trajectory, and diffusion coefficient.[152] The DL-aided TEM data recognition realized real-time characterization and analysis of defects during *in situ* TEM operations and then guided *in situ* TEM experiments, accelerating material mechanism discovery. However, it is difficult for the DL model to recognize particle boundaries in high-resolution images and then accurately predict particle activation states. The extensive noise, intricate details, and blurry regions presented in high-resolution images pose significant challenges for feature learning in ML models. Implementing the U-Net architecture, complemented by enhancements such as batch normalization, additional convolutional layers, and Otsu's thresholding technique, can significantly enhance the feasibility of analyzing chemical images.[153,154]

In addition to direct imaging methods, ML-based approaches can extract quantitative insights from indirect techniques, such as various spectroscopic and mass spectrometry methods. The x-ray absorption spectroscopy is a valuable element-specific technique for observing the oxidation state of local atomic and electronic structure of active sites under working conditions. It provides insights into potential active moieties and reaction mechanisms, enhancing the understanding of catalytic processes.[18,155] Extended x-ray absorption fine structure (EXAFS) and x-ray absorption near-edge structure (XANES) are two distinct regions of x-ray absorption spectroscopy, each requiring different analytical approaches for data interpretation. EXAFS focuses on the higher energy region relative to the absorption edge, while XANES pertains to the area near the absorption edge. XANES is a well-established technique for the quantitative structural determination of active sites in catalysis, gaining a deeper understanding of the structure and performance of catalysts. Utilizing clustering methods streamlines data analysis by reducing the complexity of XANES spectra and facilitating the visualization of species distributions. As Fig. 7 shown, combining the EXAFS spectroscopy, molecular dynamics, and the neural network was applied to Pt and PdAu NPs, which demonstrated the finite size effects on the nearest neighbor distributions, bond dynamics, and alloying motifs in mono- and bimetallic particles.[18,156–158] Ridge regression and extra trees regression are also successfully applied to XANES spectroscopy, determining the structure of CO, CO_2, and NO adsorbed on Ni^{2+} sites within the CPO-27-Ni MOF, which was closely related to the Ni-adsorbate distance and the molecular bending angles of the adsorbed species.[159] Lee *et al.* employed a three-hidden-layer backpropagation artificial neural network (ANN) to extract hidden features of local chemical environments and structural information from EXAFS data, analyzing structural information of graphene-based Co SACs in HER (Fig. 8).[160] Convolutional neural networks (CNNs) were used to analyze operando time-resolved XANES data, revealing the complex local structure around Ni sites of Ni-MOx supported catalysts in "as-prepared" and "after reaction" states under realistic working conditions. After that, the electronegativity of metal atoms in MOx supports was identified as a crucial descriptor for evaluating CH_4 production via CO_2RR.[161] Furthermore, unsupervised and supervised ML approaches fitted simulation-based XANES and EXAFS data were employed to monitor the dynamic changes of active sites in Ni-based TMNC catalysts during CO_2RR. The XANES spectra described oxidation state and coordination environment of Ni species, while EXAFS spectra provided information on coordination numbers and bond lengths, identifying the structures of the initial catalysts, intermediate, and final states under reaction conditions.[162] The principal component analysis for XANES and ANN-EXAFS was employed to identify Co chemical state transitions and active site transitions in tetrahedral/octahedral coordination environments of $Co_xFe_{3-x}O_4$ during OER, respectively. These transitions were attributed to the conversion of disordered oxides to spinel structures, the transformation of spinels into active oxyhydroxides, and changes in the degree of spinel inversion, which helps elucidate the active species and the underlying OER mechanism.[162,163] In addition, the Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS) enables visualization of the surface chemical composition and its dynamics under an electrical field. However, interpreting multidimensional data is challenging, due to the strong correlations between chemical signatures and the requirement to simultaneously monitor multiple spectral peaks. The ML–non-negative matrix factorization (NMF) workflow addressed these challenges by providing detailed insights into the chemical nature of dynamic species, ion accumulation, and interfacial electrochemical reactions. For example, the electrochemical behavior of $(CH_3NH_3)PbBr_3$ at the Au electrode interface has been analyzed using light- and voltage-dependent ToF-SIMS imaging. Despite the inherent complexity of ToF-SIMS spectra, the ML-NMF workflow effectively identified and visualized active components involved in interfacial processes and successfully detected redox reactions occurring at the device interface under applied voltage, which enhanced the understanding of electrochemical phenomena and the dynamics of interfacial reactions.[164]

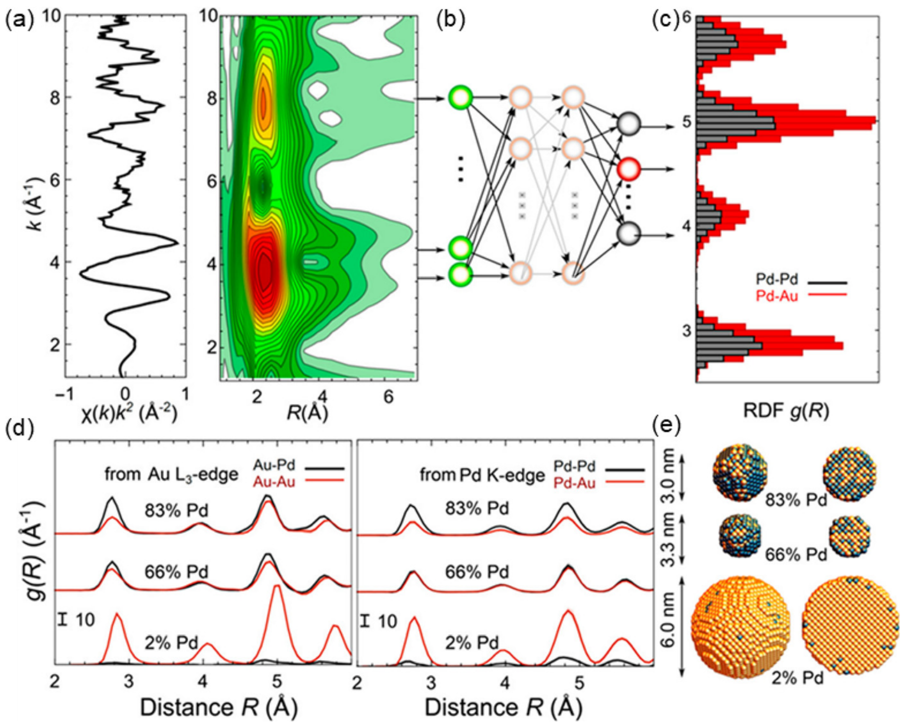

FIG. 7. The experimental Pd K-edge EXAFS data for PdAu nanoparticles (a) undergoes wavelet transform analysis and input into a neural network (NN) comprising several hidden layers (b). The NN output nodes generate a histogram of bond lengths (c), corresponding to the radial distribution functions (RDFs) of atoms, differentiating between Pd and Au atoms surrounding the absorbing Pd atoms. Partial RDFs are derived using the NN-EXAFS method for PdAu NPs of varying sizes and compositions, utilizing both experimental AuL_3-edge and Pd K-edge EXAFS data (d). The coordination numbers (CNs) obtained from the NN-EXAFS analysis are used to reconstruct 3D structural models of the NPs (e). Reproduced with permission from J. Timoshenko and A. I. Frenkel, ACS Catal. **9**, 10192–10211 (2019). Copyright 2025 American Chemical Society.[18]

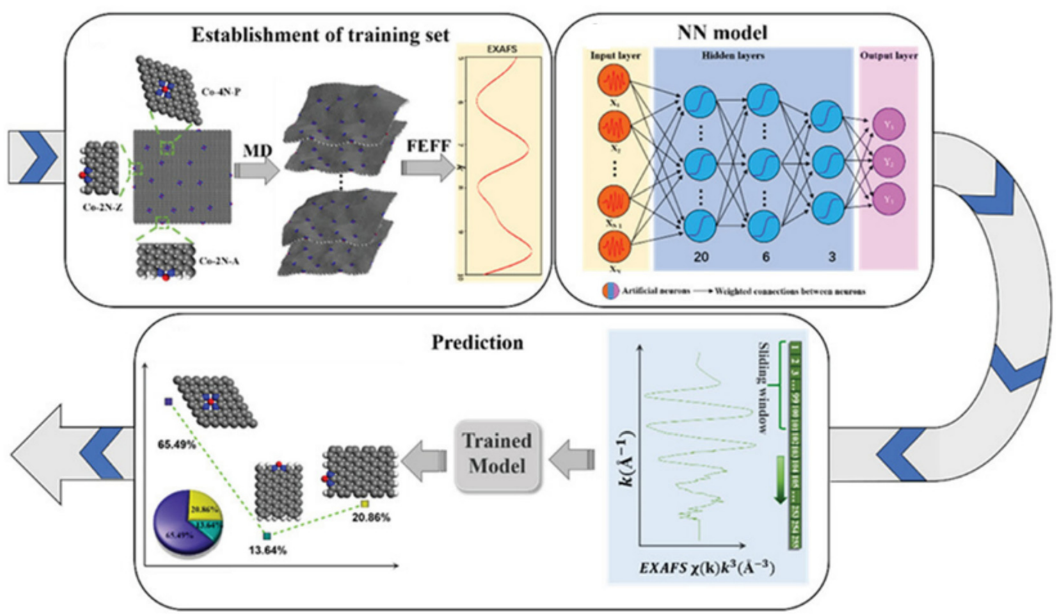

FIG. 8. Scheme of our supervised learning to interpret the EXAFS of graphene-based Co SACs in HER. First, MD-EXAFS data of Co-N doped graphene with different proportions. Then, the NN model comprises an input layer that receives the EXAFS spectrum, two hidden layers that process the data, and an output layer that produces the proportion vector. Finally, the model predicts the local structural proportion based on the experimental EXAFS measurement. Reproduced with permission from Liu *et al.*, Adv. Funct. Mater. **31**, 2100547 (2021). Copyright 2025 Elsevier.[160]

Moreover, electrochemical impedance spectroscopy (EIS) is a technique used to analyze the electrical properties of catalysts and electrochemical systems by measuring their impedance over frequencies, such as resistance, capacitance, and inductance. The distribution of relaxation times (DRT) method is widely utilized to extract temporal characteristics of electrochemical systems from EIS data, but ill-posed nature and sensitivity to experimental noise are the main obstacles. The Gaussian process has emerged as a robust solution to assume a probabilistic model and estimate the covariance of the DRT from EIS data, demonstrating notable efficacy in significant noise management, overlapping timescales, truncated data, and inductive features.[165,166] The integration of ML and experiment characterization techniques enables the gathering of structural information effectively, which is essential for the discovery and optimization of novel catalysts with tailored properties. However, ML frameworks primarily focus on limited techniques, other techniques also provide valuable insights into structural, electronic, and chemical changes under reaction conditions, such as x-ray diffraction and magnetic resonance imaging.[167,168] Expanding the application of ML to a broader range of techniques could significantly deepen our understanding of catalytic structures, thereby accelerating advancements in catalyst design and the modulation of reaction pathways.

IV. OUTLOOK

ML algorithms train experimental and simulation data to establish mapping relationships between special features and targeted properties. ML-based frameworks provide more insights into remaining problems due to effort-intensive existing approaches, which also identify promising catalysts, elucidate reaction mechanisms, and reduce the development cycle for electrochemical systems.[169,170] However, their stability and reliability are constricted by the quality and quantity of training data and customized learning algorithms. Despite these limitations, ML-based frameworks can still uncover hidden relationships between catalytic performance and features governed by underlying physical principles, as the models are designed to represent the structure within the available training data. The limited size of electrocatalyst datasets constrains the accuracy and reliability of ML models, posing significant challenges for feature selection. In small datasets, feature selection becomes highly sensitive to data quantity, where minor variations in data patterns can significantly affect model prediction. Inadequate selection of representative features may result in underfitting, where the model fails to capture essential data trends. To address this, feature combinations can generate new and informative features through mathematical combinations of original descriptors. This approach enriches the feature space, enabling better selection and improving model robustness. For example, the SISSO, based on compressed sensing, is particularly effective in transforming initial descriptors into a broad set of combined features. By efficiently identifying the most informative features, SISSO enhances the performance and reliability of models trained on limited datasets.[171–174] In addition to underfitting, the high dimensionality of small electrocatalyst datasets can lead to overfitting. Feature transformation addresses this challenge by mapping high-dimensional data onto a lower-dimensional space while preserving essential information, thus reducing model complexity. For instance, the linear discriminant analysis is a widely used dimensionality reduction technique that projects data to minimize within-class variance and maximize between-class variance, facilitating better class separation. Meanwhile, feature selection refines the feature space by selecting a subset of relevant features, enhancing predictive accuracy and generalizability. Feature selection techniques are generally classified into three categories: filter, wrapper, and embedded methods. Filter methods apply statistical measures to rank features and require predefined selection parameters, helping manage overfiltering risks. Wrapper methods use a subset search strategy, iteratively evaluating feature subsets based on model performance to identify an optimal subset. Embedded methods integrate feature selection into model training by optimizing algorithm hyperparameters and directly improving performance.[175,176]

Existing DFT models are the simplifications used in calculations and are often inadequate to match experimental conditions, lacking consideration of surface restructuring, amorphization, and solvent interactions. Experimental data were collected under controlled reaction environments, including temperature, pH, and metal load.[177] Thus, challenges persist in aggregating and comparing disparate datasets due to inherent variations and ambiguities in their composition. The statistical validity of existing literature data is uncertain because of the lack of standardized acquisition protocols. These inconsistencies make it difficult to draw reliable conclusions from aggregated data and hinder the development of robust ML models. Addressing these issues requires the establishment of standardized methodologies for data collection and reporting, which would improve the consistency and comparability of datasets across studies.[178,179] To enhance the predictive capability for novel electrocatalysts, developing reliable cross-scale models with uniform standards is crucial. However, the data acquisition process, involving time-consuming experiments and complex computations, presents a significant challenge. The high-throughput DFT explorations enable screening new catalysts, but high computational expense constrained its implementation. In addition, robot-assisted high-throughput experiments generate high-quality and consistent training data for ML models, including electrocatalyst synthesis, characterization, and testing. For example, the adaptive learning framework integrated ML algorithms and high-throughput synthesis platform searched expedited Fe–N–C catalysts in ORR, determining the critical parameters of activity and selectivity in a six-dimensional search space. The best machine learning-optimized catalyst is 33% more active than the highest-performing one in the initial datasets.[180] However, such ML-based frameworks are still in their early stages, requiring further advancements to realize their full potential.[181–183] The forecasting methods incorporated multiple electrochemical properties and relevant parameters to improve accuracy, while its complex models lead to computationally intensive computation costs and high failure rates. Thus, it is important for ML performance to establish common evaluation metrics and large datasets, facilitating qualitative comparison and understanding of different proposed approaches, and then streamlining the optimization process.

Another key limited ML framework application is the selected algorithms, which build models for predicting performance based on material parameters. The subsequent analysis of these models reveals critical relationships that are not easily identified by human inspection and traditional statistical analyses. Human interpretation of these key relationships can accelerate novel catalyst design processes. Traditional ML methodologies typically treat input parameters as relatively independent variables, lacking the ability to capture intricate structural and relational information. They are usually appreciated for large datasets, and thus, it is important to develop suitable algorithms for extracting trends from small material datasets.[184] Although emerging algorithms, like CNNs, excel at modeling complex data relationships within multi-dimensional parameter spaces, they primarily provide information on locally optimal materials without necessarily explaining the underlying materials science driving performance variations.[163,185–187] Enhancing ML model accuracy using small datasets requires strategies like crude estimation, which has proven effective in boosting predictive capability to state-of-the-art levels. Furthermore, neural network gradient analysis is employed to automate the prediction of parameter space directions that could enhance performance, thereby expanding the potential for discovery beyond existing data.[188–190] Artificial neural networks and deep neural networks loosely emulate the brain's operational mechanisms, comprising artificial neurons organized into input, output, and hidden layers. In the hidden layers, each neuron receives input signals from preceding neurons, processes these signals, and performs a straightforward computation based on the integrated results.[191] Recent emerging techniques, including meta-learning, neural turing machines, imitation learning, and the Bayesian framework, have shown promise in mitigating the challenges posed by limited datasets. Such advanced techniques are essential for molecular and materials science, fields characterized by sparse data availability and the high cost and slow pace of data acquisition.[192–195] In addition, the "black-box" nature of ML frameworks impedes the explanation of internal decision-making processes, making it challenging to understand how the model transforms input data into output prediction. This inherent lack of transparency restricts the interpretability of the models and limits the generalizability across different systems. Interpretable ML has emerged as a powerful approach for elucidating complex correlations between input physics features (descriptors) and catalytic performance. It provides analytical frameworks that establish explicit relationships between input variables and target properties, enhancing the transparency and reliability of predictive models.[196,197] Esterhuizen *et al.* demonstrated that the interpretable ML model generalized additive models (iGAM) effectively capture and elucidate the relationships between catalyst site geometry and chemisorption strength. Ensemble tree models, such as eXtreme Gradient Boosting (XGBoost) and light gradient boosting machine (LightGBM), combine high predictive accuracy with interpretability.[198] For instance, the XGBoost model identified the TM valence electron count and the N substitution level as critical features influencing charge distribution at active sites, providing valuable guidance for the rational screening TM@VB/C-Nx-BC_3 catalysts in NRR.[199] Recently, a physically meaningful feature engineering and sparsification (PFESS) method was also developed to decouple atomic (A), reactant (R), synergistic (S), and coordination (C) effects on the *d*-band shape of dual-atom catalytic sites (ARSC descriptors). Using the ARSC descriptor, this method identified Co-Co/Ir-Qv_3 as an optimal bifunctional oxygen reduction and evolution electrocatalyst from over 5000 candidates, which was further validated experimentally.[200] Thus, interpretable ML offers a pathway to explain performance modulation and addresses the limitations of "black-box" models, advancing our ability to design materials with improved catalytic performance.

V. CONCLUSION

Catalytic activity is closely related to geometry and electronic structures, with different reaction mechanisms demanding varying structural sensitivity. Traditional statistical correlations often fail to establish robust cause–effect relationships. Integrating ML with traditional theoretical modeling methodologies has significantly advanced property–structure–performance relationships and provided deeper insights into reaction mechanisms through diverse descriptors, circumventing the need for computationally expensive quantum mechanical calculations. Moreover, ML-based methods could accelerate the solution of DFT and MD. Furthermore, the combination of ML with experiment characterization techniques advances the understanding of fundamental reaction mechanisms and structural transformations under various electrolytic conditions, providing more precise insights into the dynamic behavior of electrocatalysts.

Despite high predictive accuracy and efficiency, the accuracy of ML models heavily relies on the quality and quantity of training datasets and employed algorithms. Currently, small materials science datasets result in suboptimal training outcomes, and emerging robot-assisted high-throughput calculations and experiments have the potential to generate reliable datasets. Moreover, establishing algorithms suited for extracting trends from small datasets can enhance the accuracy and adaptability of ML models. Furthermore, current ML frameworks predominantly focus on electrochemical reactions with simple networks, such as the HER, ORR, and OER. Reactions involving multiple steps and complex reaction pathways, such as the NRR, NO_3RR, and CO_2RR for C_3/C_2 product formation, require more attention. Moreover, ML models should be expanded to incorporate more experimental characterization techniques, thereby providing more support for the design of electrocatalysts and the optimization of reaction mechanisms.

The evolving landscape of ML methodologies, integrated with traditional theoretical modeling methodologies and experimental techniques, offers significant opportunities for elucidating reaction mechanisms and advancing catalyst design, ultimately accelerating the discovery and optimization of next-generation electrocatalysts to drive sustainable energy conversion and achieve a carbon-neutral society.

ACKNOWLEDGMENTS

J.C. and C.L. acknowledge financial support from Australian Research Council (Grant No. DP220101290), the National Natural Science Foundation of China (Grant No. 22473108), and the National Key Research and Development Program of China (Grant No. 2023YFB3810601).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Tianyi Wang: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Qilong Wu:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Yun Han:** Data curation (equal); Formal analysis (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Zhongyuan Guo:** Data curation (equal); Formal analysis (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jun Chen:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Chuangwei Liu:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1]X. Liao, R. Lu, L. Xia, Q. Liu, H. Wang, K. Zhao, Z. Wang, and Y. Zhao, "Density functional theory for electrocatalysis," Energy Environ. Mater. **5**, 157–185 (2022).
[2]Y. Han, X. Yan, Q. Wu, H. Xu, Q. Li, A. Du, and X. Yao, "Defect-derived catalysis mechanism of electrochemical reactions in two-dimensional carbon materials," Small Struct. **4**, 2300036 (2023).
[3]Q. Fang, H. Yin, X. Mao, Y. Han, C. Yan, A. P. O'Mullane, and A. Du, "Theoretical evaluation of highly efficient nitrate reduction to ammonia on InBi," J. Phys. Chem. Lett. **14**, 2410–2415 (2023).
[4]Y. Han, X. Mao, X. Yan, Q. Wu, H. Xu, Q. Fang, Y. Jia, X. Yao, Q. Li, and A. Du, "Carbon nanotubes encapsulated transition metals for efficient hydrogen evolution reaction: Coupling effect of 3d orbital and π-bond," Mater. Today Chem. **30**, 101573 (2023).
[5]Y. Han, H. Xu, Q. Li, A. Du, and X. Yan, "DFT-assisted low-dimensional carbon-based electrocatalysts design and mechanism study: A review," Front. Chem. **11**, 1286257 (2023).
[6]Q. Wu, Y. Jia, Q. Liu, X. Mao, Q. Guo, X. Yan, J. Zhao, F. Liu, A. Du, and X. Yao, "Ultra-dense carbon defects as highly active sites for oxygen reduction catalysis," Chem **8**, 2715–2733 (2022).
[7]C. e Li, V. Sage, T. Wang, L. Tang, Y. Yang, W. J. Lee, R. Lippi, F. Wang, K. Kozielski, and J. Patel, "Evaluation and outlook for Australian renewable energy export via circular liquid hydrogen carriers," Int. J. Hydrogen **49**, 1509–1527 (2024).
[8]X. Yang, S. Mukherjee, T. O'Carroll, Y. Hou, M. R. Singh, J. A. Gauthier, and G. Wu, "Achievements, challenges, and perspectives on nitrogen electrochemistry for carbon-neutral energy technologies," Angew. Chem. Int. Ed. **62**, 202215938 (2023).
[9]Q. Wu, F. Zhu, G. Wallace, X. Yao, and J. Chen, "Electrocatalysis of nitrogen pollution: Transforming nitrogen waste into high-value chemicals," Chem. Soc. Rev. **53**, 557–565 (2024).
[10]A. Agrawal and A. Choudhary, "Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science," APL Mater. **4**, 053208 (2016).
[11]T.-Y. Dai, C.-C. Yang, and Q. Jiang, "Recent progress on catalyst design of nitrogen reduction reaction by density functional theory," Sci. China Mater. **67**, 1101–1123 (2024).
[12]Z. Wei, Q. He, and Y. Zhao, "Machine learning for battery research," J. Power Sources **549**, 232125 (2022).
[13]J. Wei, X. Chu, X. Y. Sun, K. Xu, H. X. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science," InfoMat **1**, 338–358 (2019).
[14]K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," Nature **559**, 547–555 (2018).
[15]G. H. Gu, C. Choi, Y. Lee, A. B. Situmorang, J. Noh, Y. H. Kim, and Y. Jung, "Progress in computational and machine-learning methods for heterogeneous small-molecule activation," Adv. Mater. **32**, 1907865 (2020).
[16]R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: Recent applications and prospects," npj Comput. Mater. **3**, 54 (2017).
[17]G. L. Hart, T. Mueller, C. Toher, and S. Curtarolo, "Machine learning for alloys," Nat. Rev. Mater. **6**, 730–755 (2021).
[18]J. Timoshenko and A. I. Frenkel, "Inverting x-ray absorption spectra of catalysts by machine learning in search for activity descriptors," ACS Catal. **9**, 10192–10211 (2019).
[19]J. E. Salgado, S. Lerman, Z. Du, C. Xu, and N. Abdolrahim, "Automated classification of big x-ray diffraction data using deep learning models," npj Comput. Mater. **9**, 214 (2023).
[20]L. Zhang, W. Shi, and B. Zhang, "A review of electrocatalyst characterization by transmission electron microscopy," J. Energy Chem. **26**, 1117–1135 (2017).
[21]K. Chu, J. Wang, Y-p Liu, Q-q Li, and Y-l Guo, "Mo-doped SnS_2 with enriched S-vacancies for highly efficient electrocatalytic N_2 reduction: The critical role of the Mo–Sn–Sn trimer," J. Mater. Chem. A **8**, 7117–7124 (2020).
[22]Z. Wang, "Activity and stability of catalysts for electrocatalytic water splitting in acidic media," Sci. China Mater. **67**, 1124–1128 (2024).
[23]Q. He, B. Yu, Z. Li, and Y. Zhao, "Density functional theory for battery materials," Energy Environ. Mater. **2**, 264–279 (2019).
[24]J. Wang, M. Zhou, R. Fu, J. Ge, W. Yang, X. Hong, C. Sun, X. Liao, Y. Zhao, and Z. Wang, "Iron molybdenum sulfide-supported ultrafine Ru nanoclusters for robust sulfion degradation-assisted hydrogen production," Adv. Funct. Mater. **34**, 2315326 (2024).
[25]Q. Sun, C. Zhang, Y. Mao, R. Lu, Q. He, and Y. Zhao, "Transforming main-group elements into highly active single-atom electrocatalysts for oxygen reduction reactions," Electrochim. Acta **469**, 143223 (2023).
[26]J. K. Nørskov, T. Bligaard, A. Logadottir, J. Kitchin, J. G. Chen, S. Pandelov, and U. Stimming, "Trends in the exchange current for hydrogen evolution," J. Electrochem. Soc. **152**, J23 (2005).
[27]S. Zhang, S. Lu, P. Zhang, J. Tian, L. Shi, C. Ling, Q. Zhou, and J. Wang, "Accelerated discovery of single-atom catalysts for nitrogen fixation via machine learning," Energy Environ. Mater. **6**, e12304 (2023).
[28]K. Tran and Z. W. Ulissi, "Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution," Nat. Catal. **1**, 696–703 (2018).
[29]A. A. Peterson and J. K. Nørskov, "Activity descriptors for CO_2 electroreduction to methane on transition-metal catalysts," J. Phys. Chem. Lett. **3**, 251–258 (2012).
[30]W. T. Hong, R. E. Welsch, and Y. Shao-Horn, "Descriptors of oxygen-evolution activity for oxides: A statistical evaluation," J. Phys. Chem. C **120**, 78–86 (2016).
[31]M. Busch, N. B. Halck, U. I. Kramm, S. Siahrostami, P. Krtil, and J. Rossmeisl, "Beyond the top of the volcano?—A unified approach to electrocatalytic oxygen reduction and oxygen evolution," Nano Energy **29**, 126–135 (2016).
[32]Á. Szécsényi, E. Khramenkova, I. Y. Chernyshov, G. Li, J. Gascon, and E. A. Pidko, "Breaking linear scaling relationships with secondary interactions in confined space: A case study of methane oxidation by Fe/ZSM-5 zeolite," ACS Catal. **9**, 9276–9284 (2019).
[33]Q. Liang, L. Zhong, C. Du, Y. Luo, Y. Zheng, S. Li, and Q. Yan, "Achieving highly efficient electrocatalytic oxygen evolution with ultrathin 2D Fe-doped nickel thiophosphate nanosheets," Nano Energy **47**, 257–265 (2018).
[34]B. Hammer and J. K. Nørskov, "Electronic factors determining the reactivity of metal surfaces," Surf. Sci. **343**, 211–220 (1995).
[35]J. Yin, J. Jin, M. Lu, B. Huang, H. Zhang, Y. Peng, P. Xi, and C.-H. Yan, "Iridium single atoms coupling with oxygen vacancies boosts oxygen evolution reaction in acid media," J. Am. Chem. Soc. **142**, 18378–18386 (2020).
[36]Y. Dong, P. Zhang, Y. Kou, Z. Yang, Y. Li, and X. Sun, "A first-principles study of oxygen formation over NiFe-layered double hydroxides surface," Catal. Lett. **145**, 1541–1548 (2015).
[37]R. Dronskowski and P. E. Blöchl, "Crystal orbital Hamilton populations (COHP): energy-resolved visualization of chemical bonding in solids based on density-functional calculations," J. Phys. Chem. **97**, 8617–8624 (1993).
[38]H. Xu, S. Xi, J. Li, S. Liu, P. Lyu, W. Yu, T. Sun, D.-C. Qi, Q. He, and H. Xiao, "Chemical design and synthesis of superior single-atom electrocatalysts via *in situ* polymerization," J. Mater. Chem. A **8**, 17683–17690 (2020).
[39]M. Suvarna and J. Pérez-Ramírez, "Embracing data science in catalysis research," Nat. Catal. **7**, 624 (2024).
[40]E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, and E. Olivetti, "Materials synthesis insights from scientific literature via text extraction and machine learning," Chem. Mater. **29**, 9436–9444 (2017).

[41]T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," Phys. Rev. Lett. **120**, 145301 (2018).
[42]K. Kim, L. Ward, J. He, A. Krishna, A. Agrawal, and C. Wolverton, "Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary Heusler compounds," Phys. Rev. Mater. **2**, 123801 (2018).
[43]W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133 (1965).
[44]I. Barlocco, L. A. Cipriano, G. Di Liberto, and G. Pacchioni, "Modeling hydrogen and oxygen evolution reactions on single atom catalysts with density functional theory: Role of the functional," Adv. Theory Simul. **6**, 2200513 (2023).
[45]M. Segall, P. J. Lindan, M. a Probert, C. J. Pickard, P. J. Hasnip, S. Clark, and M. Payne, "First-principles simulation: Ideas, illustrations and the CASTEP code," J. Phys: Condens. Matter **14**, 2717 (2002).
[46]J. P. Perdew and W. Yue, "Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation," Phys. Rev. B **33**, 8800 (1986).
[47]B. Hammer, L. B. Hansen, and J. K. Nørskov, "Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals," Phys. Rev. B **59**, 7413 (1999).
[48]K. Yang, J. Zheng, Y. Zhao, and D. G. Truhlar, "Tests of the RPBE, revPBE, τ-HCTHhyb, ωB97X-D, and MOHLYP density functional approximations and 29 others against representative databases for diverse bond energies and barrier heights in catalysis," J. Chem. Phys. **132**, 164117 (2010).
[49]T. J. Giese and D. M. York, "Density-functional expansion methods: Evaluation of LDA, GGA, and meta-GGA functionals and different integral approximations," J. Chem. Phys. **133**, 244107 (2010).
[50]P. Hao, J. Sun, B. Xiao, A. Ruzsinszky, G. I. Csonka, J. Tao, S. Glindmeyer, and J. P. Perdew, "Performance of meta-GGA functionals on general main group thermochemistry, kinetics, and noncovalent interactions," J. Chem. Theory Comput. **9**, 355–363 (2013).
[51]F. Tran and P. Blaha, "Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential," Phys. Rev. Lett. **102**, 226401 (2009).
[52]T. M. Maier, A. V. Arbuznikov, and M. Kaupp, "Local hybrid functionals: Theory, implementation, and performance of an emerging new tool in quantum chemistry and beyond," WIREs Comput. Mol. Sci. **9**, e1378 (2019).
[53]S. Grimme, "Density functional theory with London dispersion corrections," WIREs Comput. Mol. Sci. **1**, 211–228 (2011).
[54]J. P. Perdew, K. Burke, and Y. Wang, "Generalized gradient approximation for the exchange-correlation hole of a many-electron system," Phys. Rev. B **54**, 16533 (1996).
[55]Z. Wu and R. E. Cohen, "More accurate generalized gradient approximation for solids," Phys. Rev. B **73**, 235116 (2006).
[56]J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids," Phys. Rev. Lett. **91**, 146401 (2003).
[57]J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," Phys. Rev. B **46**, 6671 (1992).
[58]J. Neugebauer and T. Hickel, "Density functional theory in materials science," WIREs Rev. Comput. Mol. Sci **3**, 438–448 (2013).
[59]A. A. Peterson, F. Abild-Pedersen, F. Studt, J. Rossmeisl, and J. K. Nørskov, "How copper catalyzes the electroreduction of carbon dioxide into hydrocarbon fuels," Energy Environ. Sci. **3**, 1311–1315 (2010).
[60]X. Ma, Z. Li, L. E. Achenie, and H. Xin, "Machine-learning-augmented chemisorption model for CO_2 electroreduction catalyst screening," J. Phys. Chem. Lett. **6**, 3528–3533 (2015).
[61]S. Pablo-Garcia, A. Sabadell-Rendón, A. J. Saadun, S. Morandi, J. Pérez-Ramírez, and N. Lopez, "Generalizing performance equations in heterogeneous catalysis from hybrid data and statistical learning," ACS Catal. **12**, 1581–1594 (2022).
[62]A. C. Garcia, T. Touzalin, C. Nieuwland, N. Perini, and M. T. Koper, "Enhancement of oxygen evolution activity of nickel oxyhydroxide by electrolyte alkali cations," Angew Chem. Int. Ed. **58**, 12999–13003 (2019).
[63]C. Yang, O. Fontaine, J. M. Tarascon, and A. Grimaud, "Chemical recognition of active oxygen species on the surface of oxygen evolution reaction electrocatalysts," Angew Chem. Int. Ed. **129**, 8778–8782 (2017).
[64]Y. Zhou, S. Sun, J. Song, S. Xi, B. Chen, Y. Du, A. C. Fisher, F. Cheng, X. Wang, and H. Zhang, "Enlarged Co-O covalency in octahedral sites leading to highly efficient spinel oxides for oxygen evolution reaction," Adv. Mater **30**, 1802912 (2018).
[65]A. Grimaud, O. Diaz-Morales, B. Han, W. T. Hong, Y.-L. Lee, L. Giordano, K. A. Stoerzinger, M. T. Koper, and Y. Shao-Horn, "Activating lattice oxygen redox reactions in metal oxides to catalyse oxygen evolution," Nat. Chem. **9**, 457–465 (2017).
[66]N. Ran, E. Song, Y. Wang, Y. Zhou, and J. Liu, "Dynamic coordination transformation of active sites in single-atom MoS_2 catalysts for boosted oxygen evolution catalysis," Energy Environ. Sci. **15**, 2071–2083 (2022).
[67]J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. De Fabritiis, F. Noé, and C. Clementi, "Machine learning of coarse-grained molecular dynamics force fields," ACS Cent. Sci. **5**, 755–767 (2019).
[68]W. Zhang, T. Wang, C. Liu, C. Duan, W. Xiong, and H. Li, "Hydrothermal self-assembly of gold nanoparticles embed on carbon felt for effective nitrogen reduction," Adv. Energy Sustainability Res. **4**, 2300056 (2023).
[69]C. Liu, Q. Li, C. Wu, J. Zhang, Y. Jin, D. R. MacFarlane, and C. Sun, "Single-boron catalysts for nitrogen reduction reaction," J. Am. Chem. Soc. **141**, 2884–2888 (2019).
[70]K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K.-R. Müller, and E. K. Gross, "How to represent crystal structures for machine learning: Towards fast prediction of electronic properties," Phys. Rev. B **89**, 205118 (2014).
[71]J. Westermayr, M. Gastegger, K. T. Schütt, and R. J. Maurer, "Perspective on integrating machine learning into computational chemistry and materials science," J. Chem. Phys. **154**, 230903 (2021).
[72]M. Wang and H. Zhu, "Machine learning for transition-metal-based hydrogen generation electrocatalysts," ACS Catal. **11**, 3930–3937 (2021).
[73]N. R. Knøsgaard and K. S. Thygesen, "Representing individual electronic states for machine learning GW band structures of 2D materials," Nat. Commun. **13**, 468 (2022).
[74]A. F. Zahrt, Y. Mo, K. Y. Nandiwale, R. Shprints, E. Heid, and K. F. Jensen, "Machine-learning-guided discovery of electrochemical reactions," J. Am. Chem. Soc. **144**, 22599–22610 (2022).
[75]S. Lin, H. Xu, Y. Wang, X. C. Zeng, and Z. Chen, "Directly predicting limiting potentials from easily obtainable physical properties of graphene-supported single-atom electrocatalysts by machine learning," J. Mater. Chem. A **8**, 5663–5670 (2020).
[76]P. G. Ghanekar, S. Deshpande, and J. Greeley, "Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis," Nat. Commun. **13**, 5788 (2022).
[77]T. Wang, B. Li, H. Liu, X. Zhang, R. K. Hocking, and C. Sun, "First principles study of single Fe atom supported on TiO_2(001) for nitrogen reduction to ammonia," Appl. Surf. Sci. **572**, 151417 (2022).
[78]Y. Ying, K. Fan, X. Luo, J. Qiao, and H. Huang, "Unravelling the origin of bifunctional OER/ORR activity for single-atom catalysts supported on C_2N by DFT and machine learning," J. Mater. Chem. A **9**, 16860–16867 (2021).
[79]Q. Wu, H. Zou, X. Mao, J. He, Y. Shi, S. Chen, X. Yan, L. Wu, C. Lang, and B. Zhang, "Unveiling the dynamic active site of defective carbon-based electrocatalysts for hydrogen peroxide production," Nat. Commun. **14**, 6275 (2023).
[80]V. Fung, G. Hu, Z. Wu, and D-e Jiang, "Descriptors for hydrogen evolution on single atom catalysts in nitrogen-doped graphene," J. Phys. Chem. C **124**, 19571–19578 (2020).
[81]F. C. Østergaard, A. Bagger, and J. Rossmeisl, "Predicting catalytic activity in hydrogen evolution reaction," Curr. Opin. Electrochem. **35**, 101037 (2022).
[82]L. Zhang, X. Guo, S. Zhang, T. Frauenheim, and S. Huang, "Hybrid double atom catalysts for hydrogen evolution reaction: A sweet marriage of metal and nonmetal," Adv. Energy Mater. **14**, 2302754 (2024).
[83]A. Baghban, S. Habibzadeh, and F. Zokaee Ashtiani, "On the evaluation of hydrogen evolution reaction performance of metal-nitrogen-doped carbon electrocatalysts using machine learning technique," Sci. Rep. **11**, 21911 (2021).
[84]M. Sun, A. W. Dougherty, B. Huang, Y. Li, and C. H. Yan, "Accelerating atomic catalyst discovery by theoretical calculations-machine learning strategy," Adv. Energy Mater. **10**, 1903949 (2020).

[85]M. Umer, S. Umer, M. Zafari, M. Ha, R. Anand, A. Hajibabaei, A. Abbas, G. Lee, and K. S. Kim, "Machine learning assisted high-throughput screening of transition metal single atom based superb hydrogen evolution electrocatalysts," J. Mater. Chem. A **10**, 6679–6689 (2022).
[86]L. Wu, T. Guo, and T. Li, "Rational design of transition metal single-atom electrocatalysts: A simulation-based, machine learning-accelerated study," J. Mater. Chem. A **8**, 19290–19299 (2020).
[87]Y. Luo, X. Du, L. Wu, Y. Wang, J. Li, and L. Ricardez-Sandoval, "Machine-learning-accelerated screening of double-atom/cluster electrocatalysts for the oxygen reduction reaction," J. Phys. Chem. C **127**, 20372–20384 (2023).
[88]N. Ma, Y. Zhang, Y. Wang, C. Huang, J. Zhao, B. Liang, and J. Fan, "Machine learning-assisted exploration of the intrinsic factors affecting the catalytic activity of ORR/OER bifunctional catalysts," Appl. Surf. Sci. **628**, 157225 (2023).
[89]H. Niu, X. Wan, X. Wang, C. Shao, J. Robertson, Z. Zhang, and Y. Guo, "Single-atom rhodium on defective g-C3N4: A promising bifunctional oxygen electrocatalyst," ACS Sustainable Chem. Eng. **9**, 3590–3599 (2021).
[90]X. Li, S. Lin, T. Yan, Z. Wang, Q. Cai, and J. Zhao, "Machine-learning-accelerated screening of single metal atoms anchored on MnPS 3 monolayers as promising bifunctional oxygen electrocatalysts," Nanoscale **15**, 11616–11624 (2023).
[91]X. Liu, Y. Zhang, W. Wang, Y. Chen, W. Xiao, T. Liu, Z. Zhong, Z. Luo, Z. Ding, and Z. Zhang, "Transition metal and N doping on AlP monolayers for bifunctional oxygen electrocatalysts: Density functional theory study assisted by machine learning description," ACS Appl. Mater. Interfaces **14**, 1249–1259 (2022).
[92]R. A. Flores, C. Paolucci, K. T. Winther, A. Jain, J. A. G. Torres, M. Aykol, J. Montoya, J. K. Nørskov, M. Bajdich, and T. Bligaard, "Active learning accelerated discovery of stable iridium oxide polymorphs for the oxygen evolution reaction," Chem. Mater. **32**, 5854–5863 (2020).
[93]S. Back, K. Tran, and Z. W. Ulissi, "Toward a design of active oxygen evolution catalysts: Insights from automated density functional theory calculations and machine learning," ACS Catal. **9**, 7651–7659 (2019).
[94]Z. Li, L. E. Achenie, and H. Xin, "An adaptive machine learning strategy for accelerating discovery of perovskite electrocatalysts," ACS Catal. **10**, 4377–4384 (2020).
[95]J. Zhou, Z. Han, X. Wang, H. Gai, Z. Chen, T. Guo, X. Hou, L. Xu, X. Hu, and M. Huang, "Discovery of quantitative electronic structure-OER activity relationship in metal-organic framework electrocatalysts using an integrated theoretical-experimental approach," Adv. Func. Mater. **31**, 2102066 (2021).
[96]Q. Wu, C. Liu, X. Su, Q. Yang, X. Wu, H. Zou, B. Long, X. Fan, Y. Liao, and L. Duan, "Defect-engineered Cu-based nanomaterials for efficient CO_2 reduction over ultrawide potential window," ACS Nano **17**, 402–410 (2023).
[97]Q. Wu, J. Gao, J. Feng, Q. Liu, Y. Zhou, S. Zhang, M. Nie, Y. Liu, J. Zhao, and F. Liu, "A CO_2 adsorption dominated carbon defect-based electrocatalyst for efficient carbon dioxide reduction," J. Mater. Chem. A **8**, 1205–1211 (2020).
[98]T. F. Jaramillo, K. P. Jørgensen, J. Bonde, J. H. Nielsen, S. Horch, and I. Chorkendorff, "Identification of active edge sites for electrochemical H_2 evolution from MoS_2 nanocatalysts," Science **317**, 100–102 (2007).
[99]M. Li, K. Duanmu, C. Wan, T. Cheng, L. Zhang, S. Dai, W. Chen, Z. Zhao, P. Li, and H. Fei, "Single-atom tailoring of platinum nanocatalysts for high-performance multifunctional electrocatalysis," Nat. Catal. **2**, 495–503 (2019).
[100]B. Ruqia and S. I. Choi, "Pt and Pt–Ni$(OH)_2$ electrodes for the hydrogen evolution reaction in alkaline electrolytes and their nanoscaled electrocatalysts," ChemSusChem **11**, 2643–2653 (2018).
[101]S. Back, M. S. Yeom, and Y. Jung, "Active sites of Au and Ag nanoparticle catalysts for CO_2 electroreduction to CO," ACS Catal. **5**, 5089–5096 (2015).
[102]W. Zhu, R. Michalsky, O. n Metin, H. Lv, S. Guo, C. J. Wright, X. Sun, A. A. Peterson, and S. Sun, "Monodisperse Au nanoparticles for selective electrocatalytic reduction of CO_2 to CO," J. Am. Chem. Soc. **135**, 16833–16836 (2013).
[103]A. Salehi-Khojin, H.-R. M. Jhong, B. A. Rosen, W. Zhu, S. Ma, P. J. Kenis, and R. I. Masel, "Nanoparticle silver catalysts that show enhanced activity for carbon dioxide electrolysis," J. Phys. Chem. C **117**, 1627–1632 (2013).
[104]Z. Yang, F. E. Oropeza, and K. H. Zhang, "P-block metal-based (Sn, In, Bi, Pb) electrocatalysts for selective reduction of CO_2 to formate," APL Mater. **8**, 060901 (2020).
[105]Y. Hori, I. Takahashi, O. Koga, and N. Hoshi, "Selective formation of C_2 compounds from electrochemical reduction of CO_2 at a series of copper single crystal electrodes," J. Phys. Chem. B **106**, 15–17 (2002).
[106]K. Schouten, Y. Kwon, C. Van Der Ham, Z. Qin, and M. Koper, "A new mechanism for the selectivity to C1 and C2 species in the electrochemical reduction of carbon dioxide on copper electrodes," Chem. Sci. **2**, 1902–1909 (2011).
[107]D. H. Mok, H. Li, G. Zhang, C. Lee, K. Jiang, and S. Back, "Data-driven discovery of electrocatalysts for CO_2 reduction using active motifs-based machine learning," Nat. Commun. **14**, 7303 (2023).
[108]M. Ren, X. Guo, S. Zhang, and S. Huang, "Design of graphdiyne and Holey graphyne-based single atom catalysts for CO_2 reduction with interpretable machine learning," Adv. Funct. Mater. **33**, 2213543 (2023).
[109]A. Chen, X. Zhang, L. Chen, S. Yao, and Z. Zhou, "A machine learning model on simple features for CO_2 reduction electrocatalysts," J. Phys. Chem. C **124**, 22471–22478 (2020).
[110]M. Zafari, D. Kumar, M. Umer, and K. S. Kim, "Machine learning-based high throughput screening for nitrogen fixation on boron-doped single atom catalysts," J. Mater. Chem. A **8**, 5209–5216 (2020).
[111]Z. W. Chen, Z. Lu, L. X. Chen, M. Jiang, D. Chen, and C. V. Singh, "Machine-learning-accelerated discovery of single-atom catalysts based on bidirectional activation mechanism," Chem Catal. **1**, 183–195 (2021).
[112]B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, and W.-J. Yin, "Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts," Nat. Commun. **11**, 3513 (2020).
[113]R. Miyazaki, K. S. Belthle, H. Tüysüz, L. Foppa, and M. Scheffler, "Materials genes of CO_2 hydrogenation on supported cobalt catalysts: An artificial intelligence approach integrating theoretical and experimental data," J. Am. Chem. Soc. **146**, 5433–5444 (2024).
[114]Y. Guo, X. He, Y. Su, Y. Dai, M. Xie, S. Yang, J. Chen, K. Wang, D. Zhou, and C. Wang, "Machine-learning-guided discovery and optimization of additives in preparing Cu catalysts for CO_2 reduction," J. Am. Chem. Soc. **143**, 5755–5762 (2021).
[115]X. Jiang, Y. Wang, B. Jia, X. Qu, and M. Qin, "Using machine learning to predict oxygen evolution activity for transition metal hydroxide electrocatalysts," ACS Appl. Mater. Interfaces **14**, 41141–41148 (2022).
[116]K. Wan, H. Wang, and X. Shi, "Machine learning-accelerated high-throughput computational screening: Unveiling bimetallic nanoparticles with peroxidase-like activity," ACS Nano **18**, 12367–12376 (2024).
[117]J. R. Lunger, J. Karaguesian, H. Chun, J. Peng, Y. Tseo, C. H. Shan, B. Han, Y. Shao-Horn, and R. Gómez-Bombarelli, "Towards atom-level understanding of metal oxide catalysts for the oxygen evolution reaction with machine learning," npj Comput. Mater. **10**, 80 (2024).
[118]C. Liu, Q. Li, J. Zhang, Y. Jin, D. R. MacFarlane, and C. Sun, "Conversion of dinitrogen to ammonia on Ru atoms supported on boron sheets: A DFT study," J. Mater. Chem. A **7**, 4771–4776 (2019).
[119]Z. W. Ulissi, A. J. Medford, T. Bligaard, and J. K. Nørskov, "To address surface reaction network complexity using scaling relations machine learning and DFT calculations," Nat. Commun. **8**, 14621 (2017).
[120]H. Chun, E. Lee, K. Nam, J.-H. Jang, W. Kyoung, S. H. Noh, and B. Han, "First-principle-data-integrated machine-learning approach for high-throughput searching of ternary electrocatalyst toward oxygen reduction reaction," Chem Catal. **1**, 855–869 (2021).
[121]S. Kapse, N. Barman, and R. Thapa, "Identification of ORR activity of random graphene-based systems using the general descriptor and predictive model equation," Carbon **201**, 703–711 (2023).
[122]T. A. Batchelor, J. K. Pedersen, S. H. Winther, I. E. Castelli, K. W. Jacobsen, and J. Rossmeisl, "High-entropy alloys as a discovery platform for electrocatalysis," Joule **3**, 834–845 (2019).
[123]Z. W. Chen, Z. Gariepy, L. Chen, X. Yao, A. Anand, S.-J. Liu, C. G. Tetsassi Feugmo, I. Tamblyn, and C. V. Singh, "Machine-learning-driven high-entropy alloy catalyst discovery to circumvent the scaling relation for CO_2 reduction reaction," ACS Catal. **12**, 14864–14871 (2022).
[124]X. Guo, S. Lin, J. Gu, S. Zhang, Z. Chen, and S. Huang, "Simultaneously achieving high activity and selectivity toward two-electron O_2 electroreduction: The power of single-atom catalysts," ACS Catal. **9**, 11042–11054 (2019).
[125]M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, and P. Brodersen, "Accelerated discovery of CO_2 electrocatalysts using active machine learning," Nature **581**, 178–183 (2020).

[126]Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, and T. F. Jaramillo, "Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO_2 reduction," ACS Catal. **7**, 6600–6608 (2017).
[127]M. Sun and B. Huang, "Direct Machine Learning Predictions of C3 Pathways," Adv. Energy Mater. **14**, 2400152 (2024).
[128]N. Artrith, Z. Lin, and J. G. Chen, "Predicting the activity and selectivity of bimetallic metal catalysts for ethanol reforming using machine learning," ACS Catal. **10**, 9438–9444 (2020).
[129]K. Lodaya, N. D. Ricke, K. Chen, and T. Van Voorhis, "Machine learning identification of active sites in graphite-conjugated catalysts," J. Phys. Chem. C **127**, 2303–2313 (2023).
[130]X.-G. Zhao, Q. Yang, Y. Xu, Q.-Y. Liu, Z.-Y. Li, X.-X. Liu, Y.-X. Zhao, and S.-G. He, "Machine learning for experimental reactivity of a set of metal clusters toward C–H activation," J. Am. Chem. Soc. **146**, 12485–12495 (2024).
[131]F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the Kohn-Sham equations with machine learning," Nat. Commun. **8**, 872 (2017).
[132]A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, "Solving the electronic structure problem with machine learning," npj Comput. Mater. **5**, 22 (2019).
[133]R. Jinnouchi and R. Asahi, "Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm," J. Phys. Chem. Lett. **8**, 4279–4283 (2017).
[134]P. S. Rice, Z.-P. Liu, and P. Hu, "Hydrogen coupling on platinum using artificial neural network potentials and DFT," J. Phys. Chem. Lett. **12**, 10637–10645 (2021).
[135]Y. Sun, H. Liao, J. Wang, B. Chen, S. Sun, S. J. H. Ong, S. Xi, C. Diao, Y. Du, and J.-O. Wang, "Covalency competition dominates the water oxidation structure–activity relationship on spinel oxides," Nat. Catal. **3**, 554–563 (2020).
[136]E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, "Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning," Phys. Rev. B **99**, 064114 (2019).
[137]R. Jinnouchi, F. Karsai, and G. Kresse, "On-the-fly machine learning force field generation: Application to melting points," Phys. Rev. B **100**, 014105 (2019).
[138]D. Chen, P.-L. Kang, and Z.-P. Liu, "Active site of catalytic ethene epoxidation: Machine-learning global pathway sampling rules out the metal sites," ACS Catal. **11**, 8317–8326 (2021).
[139]D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, "Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron," Phys. Rev. Mater. **2**, 013808 (2018).
[140]B. W. Chen, B. Wang, M. B. Sullivan, A. Borgna, and J. Zhang, "Unraveling the synergistic effect of Re and Cs promoters on ethylene epoxidation over silver catalysts with machine learning-accelerated first-principles simulations," ACS Catal. **12**, 2540–2551 (2022).
[141]B. Grabowski, Y. Ikeda, P. Srinivasan, F. Körmann, C. Freysoldt, A. I. Duff, A. Shapeev, and J. Neugebauer, "Ab initio vibrational free energies including anharmonicity for multicomponent alloys," npj Comput. Mater. **5**, 80 (2019).
[142]T. Kostiuchenko, F. Körmann, J. Neugebauer, and A. Shapeev, "Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials," npj Comput. Mater. **5**, 55 (2019).
[143]E. Meshkov, I. Novoselov, A. Shapeev, and A. Yanilkin, "Sublattice formation in CoCrFeNi high-entropy alloy," Intermetallics **112**, 106542 (2019).
[144]R. Jinnouchi, F. Karsai, and G. Kresse, "Machine learning-aided first-principles calculations of redox potentials," npj Comput. Mater. **10**, 107 (2024).
[145]Z. Li, J. R. Kermode, and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," Phys. Rev. Lett. **114**, 096405 (2015).
[146]A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, "Machine learning unifies the modeling of materials and molecules," Sci. Adv. **3**, e1701816 (2017).
[147]C. Liu, Q. Li, J. Zhang, Y. Jin, D. R. MacFarlane, and C. Sun, "Theoretical evaluation of possible 2D boron monolayer in N_2 electrochemical conversion into ammonia," J. Phys. Chem. C **122**, 25268–25273 (2018).
[148]J. Liu, H. Zhang, J. Wang, Y. Xie, Y. Gao, C. Sun, L. Wang, and X. Zong, "Boosting the internal electron transfer of two-dimensional hybrid perovskites via fluorination of organic cation towards enhanced photocatalytic hydrogen production," Appl. Catal. B **352**, 124018 (2024).
[149]N. Zhang, B. Yang, K. Liu, H. Li, G. Chen, X. Qiu, W. Li, J. Hu, J. Fu, and Y. Jiang, "Machine learning in screening high performance electrocatalysts for CO_2 reduction," Small Methods **5**, 2100987 (2021).
[150]S. Roy, Y. Liu, M. Topsakal, E. Dias, R. Gakhar, W. C. Phillips, J. F. Wishart, D. Leshchev, P. Halstenberg, and S. Dai, "A holistic approach for elucidating local structure, dynamics, and speciation in molten salts with high structural disorder," J. Am. Chem. Soc. **143**, 15298–15308 (2021).
[151]A. Leitherer, A. Ziletti, and L. M. Ghiringhelli, "Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning," Nat. Commun. **12**, 6234 (2021).
[152]M. Shen, G. Li, D. Wu, Y. Yaguchi, J. C. Haley, K. G. Field, and D. Morgan, "A deep learning based automatic defect analysis framework for *in-situ* TEM ion irradiations," Comput. Mater. Sci. **197**, 110560 (2021).
[153]Z. Cheng, C. Wang, X. Wu, and J. Chu, "Review in situ transmission electron microscope with machine learning," J. Semicond. **43**, 081001 (2022).
[154]J. P. Horwath, D. N. Zakharov, R. Mégret, and E. A. Stach, "Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images," npj Comput. Mater. **6**, 108 (2020).
[155]J. Timoshenko and B. Roldan Cuenya, "*In situ*/operando electrocatalyst characterization by x-ray absorption spectroscopy," Chem. Rev. **121**, 882–961 (2021).
[156]V. Venkatasubramanian, "The promise of artificial intelligence in chemical engineering: Is it here, finally?," AIChE J. **1**, 65 (2019).
[157]Z. Yu and W. Huang, "Accelerating optimizing the design of carbon-based electrocatalyst via machine learning," Electroanalysis **34**, 599–607 (2022).
[158]J. Timoshenko, C. J. Wrasman, M. Luneau, T. Shirman, M. Cargnello, S. R. Bare, J. Aizenberg, C. M. Friend, and A. I. Frenkel, "Probing atomic distributions in mono- and bimetallic nanoparticles by supervised machine learning," Nano Lett. **19**, 520–529 (2019).
[159]A. A. Gudaa, S. A. Gudaa, K. A. Lomachenkob, M. A. Soldatova, I. A. Pankina, A. V. Soldatova, L. Bragliad, A. L. Bugaeva, A. Martinia, M. Signorilec, E. Groppoc, A. Piovanoc, E. Borfecchiae, and C. Lambertia, "Quantitative structural determination of active sites from *in situ* and operando XANES spectra: From standard *ab initio* simulations to chemometric and machine learning approaches," Catalysis Today **336**, 3–21 (2019).
[160]X. Liu, L. Zheng, C. Han, H. Zong, G. Yang, S. Lin, A. Kumar, A. R. Jadhav, N. Q. Tran, and Y. Hwang, "Identifying the activity origin of a cobalt single-atom catalyst for hydrogen evolution using supervised learning," Adv. Funct. Mater. **31**, 2100547 (2021).
[161]Z. X. Li, X. P. Fu, C. Ma, W. W. Wang, J. C. Liu, and C. J. Jia, "Identifying the key structural features of Ni-based catalysts for the CO_2 methanation reaction," J. Catal. **436**, 115585 (2024).
[162]A. Martini, D. Hursán, J. Timoshenko, M. Rüscher, F. Haase, C. Rettenmaier, E. Ortega, A. Etxebarria, and B. Roldan Cuenya, "Tracking the evolution of single-atom catalysts for the CO_2 electrocatalytic reduction using operando x-ray absorption spectroscopy and machine learning," J. Am. Chem. Soc. **145**, 17351–17366 (2023).
[163]A. Mistry, A. A. Franco, S. J. Cooper, S. A. Roberts, and V. Viswanathan, "How machine learning will revolutionize electrochemical sciences," ACS Energy Lett. **6**, 1422–1431 (2021).
[164]K. Higgins, M. Lorenz, M. Ziatdinov, R. K. Vasudevan, A. V. Ievlev, E. D. Lukosi, O. S. Ovchinnikova, S. V. Kalinin, and M. Ahmadi, "Exploration of electrochemical reactions at organic–inorganic halide perovskite interfaces via machine learning in in situ time-of-flight secondary ion mass spectrometry," Adv. Funct. Mater. **30**, 2001995 (2020).
[165]J. Liu and F. Ciucci, "The Gaussian process distribution of relaxation times: A machine learning tool for the analysis and prediction of electrochemical impedance spectroscopy data," Electrochim. Acta **331**, 135316 (2020).
[166]J. Song and M. Z. Bazant, "Electrochemical impedance imaging via the distribution of diffusion times," Phys. Rev. Lett. **120**, 116001 (2018).
[167]J. E. Saal, A. O. Oliynyk, and B. Meredig, "Machine learning in materials discovery: Confirmed predictions and their underlying approaches," Annu. Rev. Mater. Res. **50**, 49–69 (2020).

[168]E. Kim, Z. Jensen, A. van Grootel, K. Huang, M. Staib, S. Mysore, H.-S. Chang, E. Strubell, A. McCallum, and S. Jegelka, "Inorganic materials synthesis planning with literature-trained neural networks," J. Chem. Inf. Model. **60**, 1194–1201 (2020).
[169]J. Cannarella and C. B. Arnold, "State of health and charge measurements in lithium-ion batteries using mechanical stress," J. Power Sources **269**, 7–14 (2014).
[170]A. Widodo, M.-C. Shim, W. Caesarendra, and B.-S. Yang, "Intelligent prognostics for battery health monitoring based on sample entropy," Expert Syst. Appl. **38**, 11763–11769 (2011).
[171]R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli, "SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates," Phys. Rev. Mater. **2**, 083802 (2018).
[172]R. Ouyang, E. Ahmetcik, C. Carbogno, M. Scheffler, and L. M. Ghiringhelli, "Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO," J. Phys. Mater. **2**, 024002 (2019).
[173]P. Xu, X. Ji, M. Li, and W. Lu, "Small data machine learning in materials science," npj Comput. Mater. **9**, 42 (2023).
[174]J. Li, N. Wu, J. Zhang, H.-H. Wu, K. Pan, Y. Wang, G. Liu, X. Liu, Z. Yao, and Q. Zhang, "Machine learning-assisted low-dimensional electrocatalysts design for hydrogen evolution reaction," Nano-Micro Lett. **15**, 227 (2023).
[175]M. F. Uddin, J. Lee, S. Rizvi, and S. Hamada, "Proposing enhanced feature engineering and a selection model for machine learning processes," Appl. Sci. **8**, 646 (2018).
[176]L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, and S. Zoupanos, "Materials Cloud, a platform for open computational science," Sci. Data **7**, 299 (2020).
[177]F. Dinic, K. Singh, T. Dong, M. Rezazadeh, Z. Wang, A. Khosrozadeh, T. Yuan, and O. Voznyy, "Applied machine learning for developing next-generation functional materials," Adv. Funct. Mater. **31**, 2104195 (2021).
[178]J. A. Keith, J. R. McKone, J. D. Snyder, and M. H. Tang, "Deeper learning in electrocatalysis: Realizing opportunities and addressing challenges," Curr. Opin. Chem. Eng. **36**, 100824 (2022).
[179]J. Yuan, Z. Qin, H. Huang, X. Gan, Z. Wang, Y. Yang, S. Liu, A. Wen, C. Bi, and B. Li, "Progress in the prognosis of battery degradation and estimation of battery states," Sci. China Mater. **67**, 1014–1028 (2024).
[180]W. J. Kort-Kamp, M. Ferrandon, X. Wang, J. H. Park, R. K. Malla, T. Ahmed, E. F. Holby, D. J. Myers, and P. Zelenay, "Adaptive learning-driven high-throughput synthesis of oxygen reduction reaction Fe–N–C electrocatalysts," J. Power Sources **559**, 232583 (2023).
[181]S. N. Steinmann, Q. Wang, and Z. W. Seh, "How machine learning can accelerate electrocatalysis discovery and optimization," Mater. Horiz. **10**, 393–406 (2023).
[182]B. Ryu, L. Wang, H. Pu, M. K. Chan, and J. Chen, "Understanding, discovery, and synthesis of 2D materials enabled by machine learning," Chem. Soc. Rev. **51**, 1899–1925 (2022).
[183]Z. Sun, H. Yin, K. Liu, S. Cheng, G. K. Li, S. Kawi, H. Zhao, G. Jia, and Z. Yin, "Machine learning accelerated calculation and design of electrocatalysts for CO_2 reduction," SmartMat **3**, 68–83 (2022).
[184]E. Hu, C. Liu, W. Zhang, and Q. Yan, "Machine learning assisted understanding and discovery of CO_2 reduction reaction electrocatalyst," J. Phys. Chem. C **127**, 882–893 (2023).
[185]C. Sutton, M. Boley, L. M. Ghiringhelli, M. Rupp, J. Vreeken, and M. Scheffler, "Identifying domains of applicability of machine learning models for materials science," Nat. Commun. **11**, 4428 (2020).
[186]J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. Marques, "Predicting the thermodynamic stability of solids combining density functional theory and machine learning," Chem. Mater. **29**, 5090–5103 (2017).
[187]G. B. Goh, C. Siegel, A. Vishnu, and N. Hodas, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Association for Computing Machinery, 2018), pp. 302–310.
[188]Y. Zhang and C. Ling, "A strategy to apply machine learning to small datasets in materials science," npj Comput. Mater. **4**, 25 (2018).
[189]M. Umehara, H. S. Stein, D. Guevarra, P. F. Newhouse, D. A. Boyd, and J. M. Gregoire, "Analyzing machine learning models to accelerate generation of fundamental materials insights," npj Comput. Mater. **5**, 34 (2019).
[190]L. Fiedler, K. Shah, M. Bussmann, and A. Cangi, "Deep dive into machine learning density functional theory for materials science and chemistry," Phys. Rev. Mater. **6**, 040301 (2022).
[191]J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Network **61**, 85–117 (2015).
[192]Y. Duan, M. Andrychowicz, B. Stadie, O. J. Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba, "One-shot imitation learning," in *Advance Neural Information Processing Systems* (MIT Press, 2017), Vol. 30.
[193]A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," arXiv:1410.5401 (2014).
[194]N. Jankowski, W. Duch, and K. Grąbczewski, *Meta-Learning in Computational Intelligence* (Springer, 2011).
[195]Y. Hong, B. Hou, H. Jiang, and J. Zhang, "Machine learning and artificial neural network accelerated computational discoveries in materials science," WIREs Comput. Mol. Sci. **10**, e1450 (2020).
[196]Y. Hu, J. Chen, Z. Wei, Q. He, and Y. Zhao, "Recent advances and applications of machine learning in electrocatalysis," J. Mater. Inf. **3**, 18 (2023).
[197]J. A. Esterhuizen, B. R. Goldsmith, and S. Linic, "Interpretable machine learning for knowledge generation in heterogeneous catalysis," Nat. Catal. **5**, 175–184 (2022).
[198]J. A. Esterhuizen, B. R. Goldsmith, and S. Linic, "Theory-guided machine learning finds geometric structure-property relationships for chemisorption on subsurface alloys," Chem **6**, 3100–3117 (2020).
[199]J. Sun, A. Chen, J. Guan, Y. Han, Y. Liu, X. Niu, M. He, L. Shi, J. Wang, and X. Zhang, "Interpretable machine learning-assisted high-throughput screening for understanding NRR electrocatalyst performance modulation between active center and C-N coordination," Energy Environ. Mater. **7**, e12693 (2024).
[200]X. Lin, X. Du, S. Wu, S. Zhen, W. Liu, C. Pei, P. Zhang, Z.-J. Zhao, and J. Gong, "Machine learning-assisted dual-atom sites design with interpretable descriptors unifying electrocatalytic reactions," Nat. Commun. **15**, 8169 (2024).