Journal of Materials Chemistry A

# Machine Learning for Design Principles for Single Atom Catalysts towards Electrochemical reactions

| | |
|---|---|
| Journal: | *Journal of Materials Chemistry A* |
| Manuscript ID | TA-REV-03-2022-002039.R1 |
| Article Type: | Review Article |
| Date Submitted by the Author: | 23-May-2022 |
| Complete List of Authors: | Tamtaji, Mohsen; Hong Kong University of Science and Technology<br>Gao, Hanyu; HKUST,<br>Hossain, Md Delowar; Stanford University, Chemical Engineering<br>Galligan, Patrick Ryan ; The Hong Kong University of Science and Technology<br>Wong, Hoilun ; The Hong Kong University of Science and Technology<br>Liu, Zhenjing; Hong Kong University of Science and Technology, Department of Chemical and Biomolecular Engineering,<br>Liu, Hongwei; Hong Kong University of Science and Technology<br>Cai, Yuting; The Hong Kong University of Science and Technology, The Chemical and Biological Engineering<br>Goddard III, William; California Institute of Technology,<br>Luo, Zhengtang; Hong Kong University of Science and Technology, Department of Chemical and Biomolecular Engineering, |
| | |

# Machine Learning for Design Principles for Single Atom Catalysts towards Electrochemical reactions

*Mohsen Tamtaji,[1] Hanyu Gao,[1] Md Delowar Hossain,[1] Patrick Ryan Galligan,[1] Hoilun Wong,[1] Zhenjing Liu,[1] Hongwei Liu,[1] Yuting Cai,[1] William A. Goddard III,[2] and Zhengtang Luo,[1*]*

[1]Department of Chemical and Biological Engineering, Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology, William Mong Institute of Nano Science and Technology, and Hong Kong Branch of Chinese National Engineering Research Center for Tissue Restoration and Reconstruction, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, 999077, P.R. China

[2]Materials and Process Simulation Center (MSC), MC 139-74, California Institute of Technology, Pasadena CA, 91125, USA

Email: keztluo@ust.hk

**Abstract**

Machine learning (ML) integrated density functional theory (DFT) calculations have recently been used to accelerate the design and discovery of heterogeneous catalysts such as single atom catalysts (SACs) through the establishment of deep structure-activity relationships. This review provides recent progress in the ML-aided rational design of heterogeneous catalysts with the focus on SACs in terms of structure-activity relationships, feature importance analysis, high-throughput screening, stability, and metal-support interactions for electrochemistry. Support vector machine (SVM), random forest regression (RFR), and deep neural networks (DNN) along with atomic properties are mainly used for the designing of SACs. The ML results shown that the number of electrons in d orbital, oxide formation enthalpy, ionization energy, Bader charge, d-band center, and enthalpy of vaporization are mainly the most important parameters for the defining of the structure-activity relationships for electrochemistry. However, the black-box nature of ML techniques occasionally makes a physical interpretation of descriptors, such as Bader charge, d-band center, and enthalpy of vaporization, non-trivial. At current stage, ML application is limited by the lack of a large and high-quality database. Future perspectives on the development of a large database and a generalized ML algorithm for SACs design are discussed to give insights for further studies in this field.

**Keywords:** structure-activity relationships, high-throughput screening, reduction reaction, stability, metal-support interactions, QM calculations, Density Functional Theory

## 1. Introduction

Heterogeneous catalysts play important roles in the synthetizing of high-value chemicals through thermal, electrochemical, and photochemical reactions. Designing improved catalysts requires deep understanding on how composition and processing affect the properties at the interface, but their progress is hindered due to the complexity in experimental and theoretical invesitigations.[1] Thus the successes have often involved time- and resource-consuming trial-and-error experimental and theoretical investigations. On the other hand, recent advances in Quantum Mechanics (QM) calculations provides accurate information about how molecules react at the interface to form various products, but QM calculations are limited in the size of the system and the time scale of the simulations. In order to discover new catalysts for specific applications, a combination of time-consuming experimental and QM studies is used to develop atomic level understanding of the fundamental mechanisms and to develop preparation-structure or structure-activity relationships. Accordingly, there is a huge demand for the accelerated discovery of novel catalysts with desired activities. Machine Learning (ML)[2–4] as a data-intensive tool can accelerate time-consuming experimental and QM studies to predict the catalytic activity in a vast dimensional space of heterogeneous catalysis.

**Figure 1** illustrates the general workflow for the integration of QM calculations and ML for the accelerated discovery of heterogeneous catalysts and single atom catalysts (SACs). The predicted data from QM calculations and features vector is used to design and train ML algorithms. Trained ML algorithms will then be used for not only the prediction of the optimal activity of heterogeneous catalysts, but also for performing feature importance analysis. Subsequently, optimized catalysts will be used for the desired reaction to produce valuable chemicals and fuels. Although the ML-assisted prediction of a single physical property such as formation energies[5] and band gaps[6] is widely applied for the purpose of materials discovery,[7–10] its application for heterogeneous catalyst design and discovery[11,12] is still in its early stage.[13] Here, ML as a supportive tool, aims to guide, not to replace experiments and QM calculations in the search for ideal catalyst.[14] However, the main hurdles for employing ML in heterogeneous catalyst design are the lack of a consistent database, the lack of a universal ML algorithm, and the existence of only a few descriptors as input features for ML.[15]

Herein, we review recent works reporting the incorporation of ML to QM calculations (typically density functional theory (DFT) calculations) and experiments to accelerate heterogeneous catalyst

design and discovery for various reactions. Recent review papers have summarized the recent studies on the application of ML for the catalytic reactions,[16–20] reaction prediction,[21] discovery of catalysts[13,22–27], inverse design of catalysts,[28] and catalysis informatics[29,30]. In this review paper, we focus mainly on the different aspects of ML in experimental and theoretical studies with an emphasis on the limitations and hurdles of ML in heterogeneous catalyst design. Inspired by the application of ML in heterogenous catalysts design, we continuous with a comprehensive review on the application of ML in SACs design and discovery with an emphasis on ML algorithms, different SACs, environmental effects, stability, support-metal interaction, structure-activity relationships, and high-throughput screening. Recent findings on the input features of ML and their importance for different electrochemical reactions will be reviewed, where the isolated electrons in d orbitals has been demonstrated to play key role in nitrogen reduction reaction (NRR).[31] Subsequently, the application of different ML algorithms in several examples including $O_2$ reduction reaction (ORR), $O_2$ evolution reaction (OER), $CO_2$ reduction reaction ($CO_2$RR), NRR, and $H_2$ evolution reaction (HER) will be provided to demonstrate the potential application of ML for the design and discovery of SACs for electroreduction reactions. Finally, a summary and future perspectives in the area of ML-guided SACs and DACs discovery are provided and discussed.

## 2. Machine learning (ML) algorithms

The most important ML algorithms applied for the establishment of deep structure-activity relationship are normally support vector machine (SVM), random forest regression (RFR), deep neural networks (DNN), sure independence screening and sparsifying (SISSO), and Gaussian process regression (GPR). As shown in **Figure 2a**, SVM as a binary classification and regression algorithm, classifies data points into two distinct categories by using hyperplanes.[32] The SVM assigns each point of training data to one of two classes and minimizes the error between the classes by dividing the categories using a hyperplane, that maximize the margin around the hyperplane. The hyperplane is completely defined by the data points that are closest to the plane and between the support vectors. SVM can be also used in mapping the non-separable data through the radial basis function (RBF) kernel by transforming a real space to a higher-dimensional space through several hyperplanes:[33,34]

$$\hat{f}(x) = \sum_{x}^{N} \omega_k G(x - x_k) \quad (3)$$

in which G is a radially symmetric function of its argument, $G(r)=\phi(|r|)$, x is the vector of joint angles or other parameters describing the current pose of the skeleton, $x_k$ is the pose of the $k^{th}$ example, and $\omega_k$ represents the different weights of each vertex coefficient. SVM is highly efficient in terms of memory usage; however, the boundary between categories may become obscured at high number of training data points. SVM can also create both the linear and non-linear model, which the latter one is based on a kernel-based regression technique.[35] When comparing SVMs and kernel ridge regression (KRR) algorithm, no big performance differences are to be expected. Usually, SVMs arrive at a sparser representation, which can be of advantage; however, their performance relies on a good setting of the C and $\gamma$ hyperparameters for SVM method and the $\alpha$ and $\gamma$ hyperparameters for KRR method. Normally, SVM method leads to faster predictions and consume less memory, whereas KRR method leads to less fitting time for large datasets. Nevertheless, because of the generally low computational cost of both algorithms, these differences are rarely significant for relatively small data points. Unfortunately, neither method is feasible for large datasets as the size of the kernel matrix scales quadratically increases with the number of data points.[36]

In comparison with other algorithms, random forest regression (RFR) needs fewer hyperparameters with higher robustness.[37] In fact, as shown in **Figure 2b**, RFR algorithm acts as a aggregated decision tree algorithm to lower the bias by reaching a collective decision.[38] The issue with the RFR is that it is not accurate for the out of sample predictions especially in case of small training data points.[39] Furthermore, feature importance analysis can be easily obtained after the training of the RFR, SVM, and KRR algorithms.[40] Similar to SVM and KRR methods, deep neural networks (DNN) algorithm has the potential to learn systems nonlinearity. As shown in **Figure 2c,** DNN is a mimic of the combination of neurons inside the human brain which is composed of several interconnected neurons in several layers. Similar to SVM and KRR methods, the number of neurons and layers as the hyperparameters for DNN should be optimized concerning the quality and accuracy of the output results for minimizing the loss functions such as root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE).[41–43]

Compared with other ML techniques, the SISSO algorithm possesses high convenience and accuracy, while the fitting formulae generated by the SISSO model possess high efficiency and

portability.[44] As shown in **Figure 2d**, GPR is a Bayesian approach to bringing waves to the ML area and works well with a small number of input data to provide uncertainty measurements on the predictions.[45]

ML techniques can be also applied as the text mining tools to gather the large numbers of already available QM calculations and experimental data in the literature, construct readily available databases applicable in the deep analysis, and study the preparation-structure-activity relationships. ML techniques for text mining can be categorized to supervised, unsupervised, and semi-supervised techniques.[46,47] Supervised and semi-supervised algorithms such as neural networks and transfer learning can be used for text classification, information extraction, and analyzing the data, while unsupervised algorithms such as expectation-maximization (EM) mostly are used for text clustering, summarization, and dimensionality reduction.[47]

## 3. Inspiration from heterogeneous catalysts design

Although ML techniques are widely used for the design of heterogenous catalysts, but its application to single atom catalysts (SACs) is in its infancy. Therefore, in accordance with the trends in ML-aided heterogeneous catalyst design which are discussed in this section, we will continue with the ML-aided design of SACs in the section 4. Integration of ML with experimental- and QM-predicted data is widely used along with atomic and structural properties as the input features to predict the properties of heterogeneous catalysts.[48–52] For example, a ML algorithm was trained based on experimental data and structural properties as the input features to optimize singlet oxygen ($^1O_2$) quantum yields of core-shell plasmonic photocatalysts applicable in organic synthesis and photodynamic therapy (PDT).[53] In addition, a ML model was trained based on DFT calculations data to predict and screen the surface reactivity of bimetallic alloys using atomic properties as the input features.[54] To shed light on the integration of ML with experiments and QM studies for the heterogeneous catalyst design and discovery, more details are provided in the following sections.

### 3.1. Integration of ML with experiments

Learning from experimental data is the earliest application of ML in heterogeneous catalyst design for electrocatalysis, photochemistry, and biocatalysis.[55–62] ML models can be trained based on experimental data to optimize the performance, decrease the number of experiments, therefore to

accelerate high-throughput experimentation.[63,64] The input features for ML models can be synthesis and reaction operation conditions, to predict the catalytic performance.[65] For example, a ML algorithm was used to calculate the yields of dioctyladipate synthesis by implementing the substrate molar ratio, enzyme amount, temperature, and reaction time as the input features.[66] Adaptive learning was applied to find high-activity $AA'B_2O_6$ cubic perovskite catalysts for OER by establishing a relationship between the electronic structure properties as the input features and the OER activity of the perovskite catalysts. It was revealed that the orbital electronic structure characteristics of the B-site ion is an important factor for OER.[51] Also a multi-output support vector regression (SVR) as the ML algorithm was applied to predict the selectivity and conversion of methane oxidation.[67] Likewise, ML allows the optimization of experimental data to increase the efficiency of heterogenous catalysts for the selective oxidation of methane.[68] In addition, ML was applied on experimental data to predict the activity and selectivity of bimetallic metal catalysts with TM-Pt-Pt(111) and Pt-TM-Pt(111) architectures for ethanol reforming.[69]

One of the disadvantages of ML models is that they are only applicable for the specific systems and are not transferable from one to another experiment due to the lack of consistent data and the presence of hidden variables for each specific experiment.[70,71]

To overcome this issue, ML can be applied to analyze available data in the literature through data mining processes[72,73] to extract and analyze previously published experimental data for future heterogeneous catalyst discovery.[74–76] For example, ML was used to extract the data for the synthesis of oxide materials from 12000 scientific articles.[77] In addition, several works have recently reported data mining from the literature for the ML-assisted design and discovery of new heterogeneous catalysts for oxidative coupling of methane.[78–83] **Figure 3** shows the workflow for the summary of data mining sequence from literature. It starts with a query search to find related papers from metadatabase, following by downloading and classifying the papers.[46,84] The classified papers can be used for text mining using several ML algorithm such as KRR, RFR, XGB, SVR, XGB, ETR, and ANN to extract the data. The extracted data can be used for regression, classification, and/or clustering purposes. For example, several ML algorithms such as Extreme Gradient Boosting (XGB), Random Forest Regressor (RFR), and extra trees regression (ETR) were used to analyze the literature data for the oxidative coupling of methane on metal supported catalysts to discover new heterogeneous catalysts.[85,86] Similarly, the statistical analysis of available data in the literature for CO oxidation, water-gas shift reaction, and oxidative coupling

of methane reactions was performed using several ML algorithms such as Kernel Ridge Regression (KRR), RFR, XGB, and SVR for the heterogeneous catalyst discovery. Through the feature importance analysis, reaction temperature was revealed as the key parameter for the three investigated reactions.[87] Very recently, suitable catalysts for environmental applications were discovered based on available data in the literature, from which binary and ternary element catalysts such as $Mn_xCo_y$ and $Zr_xMn_yCr_z$ were identified and optimized through ML for high $NO_x$ conversion. Artificial neural network (ANN) was used to predict $NO_x$ conversion efficiency as a function of temperature and the element molar ratio. The conversion reaches a maximum around 300 ℃ for the ternary element catalysts. Also, the loading amount of Zr was found to play an important role due to the fact that the $Cr^{5+}$ species can reduce as the Zr loading amount increases, which can subsequently lower the $NO_x$ conversion efficiency.[88] In addition, a ML algorithm along with 27 descriptors was applied to 2228 experimental data obtained from the literature[89] to predict activity of heterogeneous catalysts which reveals that temperature is the most important descriptor for the water-gas shift reaction.[90]

Moreover, learning from a large database in nanoscience can be used for rapid design and discovery of new heterogeneous catalysts using ML.[91] However, the obtained dataset from the literature is mostly incomplete and inconsistent, which limits the application of ML. In order to generate a consistent database for the training of ML algorithms, high-throughput experimentation can be performed. As a result, high-throughput experimentation for oxidative coupling of methane was performed for 20 catalysts and 216 reaction conditions to produce a consistent dataset for ML to accurately predict $C_2$ yields.[92] From the feature importance analysis, temperature, in the range of 700 to 900 ℃, is the most important parameter in comparison with other parameters such as flow rate of Argon, flow rate of $O_2$, flow rate of $CH_4$, contact time, and composition of catalyst.

ML also has the great potential to alter the current form of conventional experiments and increase the efficient heterogeneous catalysts discovery through automation.[93–95] In fact, ML-assisted robots can help to accelerate high-throughput experimentation without human interaction.[96–99] As a result, a ML-guided robot was used to carry out 688 experiments within an experimental space of ten variables, 1000 times faster than manual approaches. The ML-assisted high-throughput experimentation revealed a new photocatalyst mixture with six times more activity.[100]

### 3.2. Integration of ML with Quantum Mechanics (QM)

Learning from Quantum Mechanics (QM) is highly desired due to existence of enormous amounts of quantitative QM-predicted data as training dataset for ML. The trained ML can be used for accelerated and accurate prediction of catalytic properties and adsorption energies of reaction intermediates.[101] Using the adsorption energies as the key parameter, the reaction barrier can be predicted, the reaction mechanism can be investigated, and the desired catalyst can be discovered. For example, the local similarity kernel and Bayesian linear regression as ML algorithms were used for predicting adsorption energies of NO, O, and N on the $Rh_1$–$xAu_x$ alloy, based on the nanoparticle composition and size.[102,103] The findings were used to predict the rate of NO decomposition on RhAu nanoparticles which indicates a maximum for catalytic activity at the particle diameter of 2.0 nm. In addition, structure-activity relationships was established for predicting CO and H adsorption energies based on structural properties using active learning across reaction intermediates.[104,105] In fact, an automated screening approach through integration and optimization of ML was presented to guide DFT calculations for predicting catalytic activity.[105] The feasibility of this approach was demonstrated by screening various alloys combining 31 elements, which resulted in 131 candidate surfaces across 54 alloys being identified for $CO_2$RR and identification of 258 surfaces across 102 alloys for HER.[104,105] Likewise, active learning was then used to accelerate the screening of CO adsorption energy on Cu based components.[106]

ML-predicted adsorption energies of reaction intermediates were also used for investigation and optimization of the reaction network of the syngas reaction (CO + $H_2$) over Rh(111) catalysts at 573 K and 1 atm. The Gaussian process regression (GPR) as a ML algorithm was trained based on a few DFT calculations to predict the adsorption energies for all intermediates in the reaction network. A probable reaction network from syngas to acetaldehyde was revealed by using a simple classifier to select the potential rate-limiting steps, where only predicted potential rate-limiting steps were analyzed via further DFT calculations.[107]

ML was also trained on the DFT-calculated data to accelerate the prediction of adsorption energies of H and $CH_x$ intermediates on Cu-based alloys using 12 properties as the input features. Amongst several ML algorithms, the ETR algorithm resulted in the highest accuracy. Based on the feature importance analysis, the surface energy, element group, and melting point were identified to be the most important parameters for predicting adsorption energies.[108] In addition, ML was applied for predicting adsorption energies of different intermediates on the metal alloys.[109] ML was also used to predict adsorption energies of H on the $Ni_2P(0001)$ surfaces. From the feature engineering

perspective, the Ni-Ni bond length is the key parameter for HER activity, where a higher Ni-Ni bond length leads to lower HER activity.[110] Similarly, ML was used to predict the adsorption energies of CO on bimetallic alloys, where the feature engineering analysis resulted in the d-band shape and sp-band filling as key parameters.[111,112] Furthermore, to accurately predict the d-band as one of the most important parameters in CO adsorption, a GBR model was applied to several individual 3d, 4d, and 5d transition metal structures and their binary alloys for both the cases of metal impurities and overlayer-covered metal surfaces.[113,114] Recently, ML was integrated with DFT calculations to predict adsorption energies of various molecules on metal oxide surfaces. Feature importance analysis indicates that the highest occupied molecular orbital (HOMO) of the adsorbates and the metal oxide surface energy are the most important parameters for molecular adsorption.[115] ML in combination with DFT calculations was used for the prediction of adsorption energies of 12 elements on 38 metal surfaces by using SVR, RFR, and multi-layer perceptron regression (MLPR).[116]

Integration of ML and QM can also be performed to accelerate the discovery and high-throughput screening of heterogeneous catalysts. For example, ML integrated DFT calculations were used to accelerate discovery and high-throughput screening of 2D MXenes for HER.[117,118] SVR, GPR, RFR, and AdaBoost were used as ML algorithms to accelerate prediction of $\Delta G_{H*}$, based on the distance between the nearest neighbor O atoms as well as surface oxygen-metal bond length are the most important parameters.[117] Similarly, several ML models, such as DNN, KRR, SVM, and RFR, were used to accelerate high-throughput screening of $\Delta G_{H*}$ by using several elemental properties as the input features. RFR led to the highest accuracy, with the lowest RMSE of 0.27 eV for the test data. The feature importance analysis shows that HER performance is highly dependent on charge and structural properties. S- and $Os_2B$-terminated $Sc_{n+1}N_n$ (n =1, 2, 3) were revealed as appropriate catalysts for HER with $\Delta G_{H*}$ near to zero and satisfactory hydrogen coverages. It was also shown that S functional groups are of great importance in regulating the HER performance. This is because filling antibonding states with electrons weakens the adsorption of H*, which is a key step for HER.[118]

For spinel structures, the ML model was used to accurately calculate the energy difference between the centers of the oxygen p and metal d bands to identify the better spinel oxide catalysts for OER. It was shown that a $[Mn]T[Al_{0.5}Mn_{1.5}]O\text{-}O_4$ spinel catalyst has the optimal energy difference for high activity, as confirmed by experimental observations.[119] ML was also applied to optimize

$TiO_2$-supported Re and zeolite catalysts for methylation of aromatic hydrocarbons.[120] Similarly, ML was applied on the DFT-calculated data to predict how strain in platinum core-shell nanocatalysts can improve the ORR activity. It was revealed that the optimal strain depends on the nanoparticle size rather than bimetallic material composition and shell thickness.[121]

As with experimental data, there is a large amount of QM-predicted data in the literature that can be mined for the purposes of ML analysis to commence a new direction using a large database in the rational design of heterogeneous catalysis and SACs.[122] For example, ML was applied on literature data for $CO_2$ hydrogenation.[123] In addition, a dataset of 37,000 structures from the Catalysis-Hub database,[124] containing 11 adsorbates on 2000 metal alloy surfaces was used for training a Graph neural network (GNN) to predict adsorption energy based on the relaxed structures.[125]

ML can be also used for investigating reaction mechanisms and finding active sites for reactions. For instance, the LASSO ML algorithm was trained on DFT-calculated data for predicting the methane activation mechanism on Rutile metal oxides.[126] It was revealed that the energy of methane activation decreased if the reacted atoms including O, C, H, and metal atoms could be placed in the same plane. In addition, ML was combined with multi-scale simulations and QM to identify the performance of surface sites on Au nanoparticles as well as dealloyed Au surfaces for $CO_2$RR.[127] Based on ML results, surface defect is responsible for the high performance of Au surfaces. Similarly, ML was applied to DFT-calculated data to discover active bimetallic facets for $CO_2$RR.[128] It was revealed that most facets of nickel gallium bimetallics lead to similar activity on Ni surfaces.

ML integrated DFT calculations is able to predict the surface segregation energies of bimetallic catalysts through the establishment of structure-activity relationships.[129] ML was also used for the prediction of reaction barriers on a variety of surface[130] and for the discovery of phase diagram applicable in electrochemical reactions.[131] In addition, symbolic regression as a ML technique in combination with QM calculations was used to accelerate the discovery of new perovskite catalysts with excellent OER activity. The ratio of octahedral factor to tolerance factor ($\mu/t$) was revealed as a simple and important descriptor for the discovery of perovskite catalysts.[132]

## 4. Single atom catalysts (SACs)

Along with the studies mentioned above on the heterogeneous catalysis, single atom catalysts (SACs) have recently been applied to several photochemical and electroreduction reactions to produce a wide range of chemicals.[133–136] The unique properties and high atom-utilization efficiency of SACs make them interesting and promising.[137–139] With these increased applications, the rational design of SACs has come into forefront to generate improvements in efficiency and feasibility of optimizing the desired products.[140] DFT calculations are widely used for the rational design of SACs with efficient activity, selectivity, and stability. DFT calculations, however, are time-consuming and computationally expensive[141,142] because the complexity of structure-activity relationships requires performing a large numbers of non-trivial DFT calculations in a large parameter space, including SAC type, environmental coordination, and reactants.[143] On the other hand, ML is considered as a fast, accurate, inexpensive,[144] and supportive tool[145] to predict the properties of SACs towards their rational design.[146–148] As shown in **Figure 3**, using ML, one can apply the available datasets from QM and DFT calculations to construct readily available databases applicable in the deep analysis and establishment of preparation-structure-activity relationships. The stablished relationships can be used to predict adsorption energy ($E_{ads}$) or Gibbs free energy ($\Delta G$) of various reaction intermediates adsorbed on SACs to discover more active and selective SACs. Once enough high quality databases are provided, a reliable ML model can be trained and constructed to address the electroreduction challenge.[149,150] ML in combination with DFT calculations commences a new direction for rapid and low cost rational design of SACs predicted to optimal electroreduction catalytic activity.[151,152] For example, several works have used ML to design single atom alloy catalysts (SAACs) with excellent stability and activity by predicting the $E_{ads}$, $\Delta G$, or aggregation energies.[153–156] ML can be also used for the interpretation of characterization of SACs.[157,158] For example, as shown in **Figure 4**, ML techniques have been used to interpret the EXAFS spectra based on which edge sites (zigzag or armchair) are responsible for the HER activity of cobalt SAC embedded in graphene.[146] In the following section, incorporation of ML in the acceleration of structure-activity relationship, feature engineering, high-throughput screening, and stability of SACs is broadly discussed. As the application of SACs in thermal and electrochemical reactions is presented in the recent review paper,[159] we only focus on the progress of ML for the designing of SACs and DACs especially for electrochemical reactions.

### 4.1. Structure-activity relationship and feature engineering

ML is a strong tool[160] to provide a fundamental understanding of structure sensitivity[136,161,162] through establishing deep relationships between catalytic activity and structural as well as atomic properties based on mechanisms and similarities in SACs.[13,32,163] ML is considered as a new direction for the rational design of SACs by exploring feature importance analysis for electroreduction reactions to introduce more perceptions into the origin of activity and stability of SACs.[164–166] For example, ML integrated DFT was applied to establish a relationship between various descriptors and hydrogen adsorption free energy ($\Delta G_{H*}$) for HER by altering the size and dimensionality of the nitrogen-doped 2D-carbon substrate for the 3d, 4d, and 5d transition metals (TM) as SACs.[167] The sure independent screening and sparsifying operator (SISSO) as the supervised ML algorithm was applied with 10 input features including d-state center ($\varepsilon_d$), covalent radius ($r_{cov}$), Bader charge (q), number of occupied d states ($d_{occ}$), Zunger radius ($r_d$), number of valance electrons ($N_e$), ionization energy (IE), electronegativity (EN), and formation energy of single atom sites ($E_f$). Our evaluation on this work using support vector machine (SVM) algorithm is shown through **Figure 6a**, demonstrating that number of occupied d states ($d_{occ}$) and Bader charge (q) are the most important parameters for HER. Using SISSO algorithm, the following general descriptor for HER activity containing four properties was obtained, in which EN is electronegativity of the SACs:

$$\Delta G_H = -1.032\left(\frac{\varepsilon_d}{q}\right) + 13.424\left(\frac{1}{r_{cov}}\right) + 1.726(\varepsilon_d \times EN) - 0.045 d_{occ}^2 - 9.241 \quad (1)$$

Similarly, several atomic properties were implemented as input features to establish structure-activity relationships and predict OER overpotential of SACs on carbon substrates. The full connection neural network (FCNN) ML algorithm trained using DFT-calculated data leads to an accurate prediction of overpotentials with relative error of 6.49% and a 130,000 times reduction in computational time. It was revealed that d-electron count ($d_e$), the atomic radius of metal ($At_R$), and electron affinity (EA) are the most important parameters for OER overpotential. Moreover, an intrinsic descriptor ($\phi$) that defines the overpotential of SACs based on their intrinsic atomic properties was proposed using the ML and DFT:[168]

$$\phi = IE_1 d_e At_M\left(\frac{EN_M}{At_{RM}} + \frac{N_C EN_C}{At_{RC}}\right) \quad (2)$$

where $EN_C$, $At_{RC}$, and $N_C$ are the electronegativity of carbon, atomic radius of carbon, and the nearest neighbor carbon atoms, respectively. $EN_M$, $IE_1$, and $At_M$ are the electronegativity of metal, first ionization energy, and atomic mass, respectively.

In another work, atomic properties such as electronegativity, electron affinity, and radii of the metal atoms were considered as input features to reveal ORR activity for heterobimetallic SACs. Using RFR, the origin of ORR activity of SACs was investigated experimentally or by establishing structure-activity relationships based on DFT-calculated data.[169] Similarly, atomic properties were used to predict the catalytic activity of SACs and bi-atom catalysts for $CO_2$RR. Based on results from the GBR algorithm, Ag-MoPc was revealed as an excellent electrocatalyst with a limiting potential of -0.33 V.[170] Subsequently, the data from the abovementioned work was used as an example to evaluate the efficiency of a DFT-ML hybrid program for catalysis programming.[171]

In order to observe the effect of substrates on the activity and stability of SACs, the combination of atomic and structural properties should be considered as input features for the training of ML algorithms. Therefore, several atomic as well as structural properties were used to establish structure-activity relationships for discovery and design of bifunctional rhodium SACs on defective g-$C_3N_4$ for OER and ORR using the GBR ML algorithm.[172] The atomic and structural properties include TM bond length and coordination atoms (dTM-N1, dTM-C1, and dTM-C2), the d-band center ($\varepsilon_d$), the charge transfer of TM atoms ($Q_e$), the electronegativity (EN), the electron affinity (EA), the first ionization energy ($IE_1$), the radius of TM atom ($At_R$), and the number of TM-d electrons ($d_e$). As shown in **Figure 5a**, the GBR model predicts $\Delta G_{*OH}$ with an $R^2 = 0.99$ with a low RMSE = 0.03 eV. However, this work included only 16 points of input data, which is insufficient. Feature importance analysis revealed that first ionization energy ($IE_1$) and the charge transfer of transition metal atoms ($Q_e$) are the key features (**Figure 6b**). The most important descriptor $IE_1$, the energy needed to remove one or more electrons from a neutral atom to form a positively charged ion (which increases from left to right in each period) affects the OER and ORR activities.

Similarly, atomic and structural properties including the number of electrons in d orbitals, the oxide formation enthalpy, the Pauling electronegativity of the metal atom, the sum of Pauling electronegativity of surrounding atoms, and the average pKa values of the surrounding atoms were used to establish structure-activity relationships. To do this, the RFR algorithm was applied based on DFT-calculated data for 104 SACs embedded in graphene including M@$C_3$, M@$C_4$,

M@pyridine-$N_4$, and M@pyrrole-$N_4$. RFR algorithm revealed that the number of electrons in d orbitals is the most important parameter for ORR, OER, and HER. The trained RFR algorithm was employed to predict the activity of 260 graphene-based SACs (M@$N_xC_y$), through which, it was revealed that Fe@pyrrole-$N_1C_3$ and Fe@pyrrole-$N_2C_2$ were more active than Fe@pyridine-$N_1C_3$ and Fe@pyridine-$N_2C_2$.[173]

Comparably, 8 atomic and structural properties including oxide formation enthalpy ($H_{f,\ ox}$), the number of electrons in d/p orbitals ($dp_e$), electron affinity (EA), electronegativity (EN), number of coordinated N atoms ($N_N$), first ionization energy ($IE_1$) of the central atoms, the sum of the electronegativity of neighboring C and N atoms ($S_{EN}$), and the distance ratio ($D_R$) were used to establish the structure-activity relationship for two electron ORR using RFR. **Figure 5b** shows the comparison of ML- and DFT-predicted $\Delta G_{O*}$ for this system. Through the feature importance analysis of 8 intrinsic features, it was revealed that the oxide formation enthalpy ($H_{f,\ ox}$) and the number of electrons in d/p orbitals ($dp_e$) are the most important parameters for determining $\Delta G_{O*}$ of SACs (**Figure 6c**).[174] The feature importance analysis implies that metals like Ag, Au, and Pd with a weaker affinity for oxygen, can remarkably decrease band hybridization between the oxygen and metal, leading to enhanced $H_2O_2$ selectivity.

As the complexity of SAC structures increases, new and general descriptors will be needed for establishing the correct structure-activity relationships. For example, the number of isolated electrons in d-orbitals, obtained from a bidirectional activation mechanism, was suggested as a new input feature for ML algorithm, which introduces new insights for the rational design of SACs. It was shown that this new descriptor is the most important parameter for NRR, while the electron affinity of metal atom was shown to be the most important parameter for HER. ML using this new input features was therefore used to accelerate the computational screening, design, and discovery of SACs by establishing the structure-activity relationship on 126 SACs for NRR, validated by experimental studies and DFT calculations.[31]

Unlike SACs, the geometry of dual atom catalysts (DACs) is more complex and the synergetic effect between two metal atoms plays important role in the performance. In other words, the linear relationships for the DACs are significantly weakened, demonstrating that the DACs' activity requires new descriptors to consider the effects of both metals in the structure. Therefore, in order to consider the synergetic effect of two metals, ML integrated DFT was used to identify the structure-activity relationship of DACs embedded on nitrogen-doped graphene for ORR. **Figure**

**4c** shows the ML- and DFT-predicted limiting potentials and also the feature importance on the limiting potentials using the random forest regression (RFR) model.[175] The feature importance analysis indicates that the average distance between metal and N atoms ($M_{12}$-N), the distance between metal atoms ($M_1$-$M_2$), and the outer electron number of metal atoms ($N_{e,O}$) are the most important factors on the ORR limiting potentials (**Figure 6d**).

In order to shed more light on the structure-activity relationships, the effect of different intermediates should also be considered on the activity of SACs. Therefore, in addition to atomic and structural properties, the properties of intermediates were also considered as input features for training the RFR algorithm to calculate the binding energies of H*, OH*, O*, and OOH* on SACs embedded in nitrogen-doped graphene using 1700 DFT-calculated data points. Based on the feature importance analysis, the type of intermediate was found to be one of the most important features.[176]

The input features with high feature importance can be used for descriptor-based SACs design to predict adsorption energies. For example, descriptor-based design was used to predict adsorption energies of intermediates on SACs embedded in graphitic carbon nitride (g-$C_3N_4$), g-CN, and g-$C_2N$. It was shown that Ni@g-CN, Cu@g-CN, and Co@$C_2N$ are excellent SACs for $CO_2$RR.[177] It was also shown that catalytic activities are highly related to $\Delta G_{OH*}$, $\Delta G_{OCH*}$, the number of electrons in d orbitals, and the TM enthalpy of vaporization.

The descriptors can be also used for establishing volcano-shaped relationships[178] from which the candidate SACs for various electrocatalytic reactions can be found.[179] Therefore, a new intrinsic descriptor based on the bonding, topology, and electronic structure of SACs embedded in carbon supports, shown through **Figure 7a**, was defined as follows:[180]

$$\emptyset = \frac{N_e EN}{I_R} \quad (3)$$

In which $N_e$, EN, and $I_R$ are the valence electron number, electronegativity, and ionic radius of central metals, respectively. This descriptor was used for volcano plots of overpotential, onset potential, and Faraday efficiency, shown through **Figure 7b-d**, indicating two definitive volcanoes in the plot for overpotential with Ti and Co located at the summits. Another descriptor to consider the effect of supports was also introduced as follows:[140]

$$\varphi = d_e \frac{EN_M + a(N_N EN_N + N_C EN_C)}{EN_{O/H}} \quad (4)$$

In which $EN_N$ and $EN_C$, $N_N$, $N_C$, and $d_e$ represent the electronegativity of N atoms, the electronegativity of C atoms, the number of nearest-neighbor N atoms, the number of nearest-neighbor C atoms, and valence electrons in d orbitals, where α is the correction coefficient. These descriptors were used to predict adsorption energies of different intermediates for $CO_2RR$. Moreover, these descriptors were used for volcano plots of onset potential and overpotential with Ni and Pt located at the summits of volcano plots.

However universal and appropriate descriptors are still insufficient, as input features to establish structure-activity relationships for all types of SACs, supports, and electroreduction reactions.[181] Therefore, a large number of DFT calculations and ML analyses are still needed to screen different descriptors for each reaction system.[182]

### 4.2. High throughput computational screening for SACs

DFT calculations have been applied for high-throughput screenings of SACs,[96,183–187] where, for example, S was found to be the best dopant in graphene-based Co SACs for HER.[188] ML, however, can accelerate the screening of SACs and decrease the computational cost and time by screening for similarities in SACs and establishing deep structure-activity relationships.[147,189–191] Therefore, integration of ML algorithms and DFT calculations has been performed for the rapid and high-throughput screening of SACs.[192] For example, ML combined DFT calculations were employed to screen and design MBenes-based SACs for HER. $\Delta G_{H^*}$ values were calculated accurately via SVM by using atomic and structural features. The Bader charge transfer of the surface metal was revealed as the most important parameter for HER activity. Stable Co2B2 and Mn/Co2B2 were also identified as efficient HER catalysts because $|\Delta G_{H^*}|<0.15$ eV.[193] In addition, the screening of SACs embedded on MXenes was performed using ML and DFT calculations to show the ability of ML to screen new candidates with excellent performance.[194] It shows that the HER catalytic activity is dependent on the synergistic effect between single metal atoms and substrates. In addition, the bag-tree algorithm supervised ML technique was applied for the separation of DFT-calculated data and converse prediction of HER performance.[195] ML integrated DFT calculations were applied to accelerate the discovery and screening of TM and lanthanide (Ln) metals for SACs embedded in graphdiyne, based on adsorption energies, adsorption trend, electronic structures, reaction pathway, and active sites.

In addition to HER, ML algorithms were employed based on DFT-calculated data for the fast-screening of efficient NRR and $CO_2RR$ electrocatalysts.[105] For instance, graph-based convolutional neural network (GCNN) was applied for the accelerated screening of SACs for NRR. The results shows superior NRR selectivity over HER with overpotentials of 0.44 V, 0.40 V, 0.24 V, 0.60 V, 0.17 V, 0.17 V, 0.64 V, 0.37 V and 0.58 V, respectively for SACs embedded in MBenes, defect-engineered 2D-materials, and 2D p-conjugated polymer, TaB, NbTe2, NbB, HfTe2, MoB, MnB, HfSe2, TaSe2 and Nb.[196] A deep neural network (DNN) was used for rapid and high throughput screening of efficient SACs embedded on boron-doped graphene for NRR. The adsorption and free energies were calculated using the light gradient boosting machine (LGBM) model based on the bonding characteristics and structural properties as input features. The feature importance analysis was also provided for nitrogen fixation, revealing that TM coordination number and the number of hydrogen atoms are the key parameters.[197] Extreme gradient boosting regression (XGBR) was implemented as a supervised ML algorithm to screen $\Delta G_{CO*}$ and $\Delta G_{H*}$ for 1060 SACs embedded in metal-nonmetal co-doped graphene using simple features for $CO_2RR$.[198] Based on feature importance analysis, the Pauling electronegativity (E_M), covalent radius (M_cov), and first ionization energy of metal atoms (1E_M) are the most important parameters on $\Delta G_{CO*}$.

### 4.3. Stability of SACs

SAC stability is the prerequisite for the constructing high-activity SACs, which should be considered by studying metal-support interactions, aggregation energies, and adsorbate-induced structural changes.[199–202] In other words, constructing a strong coordination environment for achieving SACs with strong metal-support interactions is highly desirable and can be achieved by increasing either the anchoring capability of supports or the number of anchor sites.[203] The former can be performed by optimizing the coordination environment and the coordination atoms. The latter can be achieved by introducing intrinsic defects and structural engineering through controlling its size and morphology.

In this regard, ML can be applied as the new guideline to efficiently synthesize highly-loaded-yet-stable SACs with strong metal-support interactions.[36,204] For example, ML integrated DFT calculations were employed to correlate the stability of SACs embedded on oxide supports to the binding energy ($E_{bind}$) and cohesive energy of the bulk metal ($E_c$). Assisted by ML methods, it was

found that the diffusion activation barrier ($E_a$) correlates with $E^2_{bind}/E_c$ in the physical descriptor space,[205] while $E_{bind}$ was previously explored to be correlated to $E_c$.[206]

Designed SACs should be thermodynamically stable with lowest energy state. Therefore, thermodynamic stability and optimal combination of dual atomic catalysts embedded in graphdiyne were also investigated by using the d-band center modifications and formation stability. Using gaussian process regression (GPR) as the ML algorithm with seven input features, the potential f-d orbital coupling was found as the most important factor in tuning the d-band center with high stability.[33] Based on these results, the combination of lanthanide metals and transition metals leads to appropriate stability and activity. The thermodynamic stability of SAACs was also investigated in terms of aggregation energies and adsorbate (O*)-caused changes in structure by using ML algorithms trained with DFT-calculated data for 38 different SAACs on Cu support. A GPR model was applied on the aggregation energy and O* adsorption energies with a MAE of 0.092 and 0.091 eV, respectively. Moreover, the GPR model is extendable to other substrates, adsorbates, and larger cluster sizes to address the huge number of degrees of freedom and decrease the calculation time.[207]

The zero-valence stability and electron transfer ability of SACs should be also investigated on the stability by considering the redox process between transition metals and graphdiyne support using ML and DFT. It was indicated that amongst transition metals, Co, Pd, and Pt show high stability of zero-valence SACs based on the difference of energy barrier between gaining and losing electrons.[208] The Fuzzy C-Means (FCM) as a unsupervised ML algorithm was used for the separation of DFT-calculated data. The developed ML algorithm has also been applied to create a database capable of screening out SACs embedded in graphdiyne.[208] The different number and directions of electron transfer between the transition metals and graphdiyne were also analyzed, finding that the initial one-electron transfer is the most difficult one.

Very recently, the stability of the SAACs configuration based on a ML based approach was examined to investigate the tendency of the promoter atom to diffuse into the bulk material, form surface clusters, or avoid alloying with the host.[209] Decision trees, neural networks (NN), and support vector machines (SVM) with atomic properties as the input feature, were used to analyze DFT-calculated data. Then, a physical bond counting model was combined with a kernel ridge regression (KRR) algorithm to expand the domain where the model is useful.

The stability and activity of SACs embedded in $N_xC_y$ ($TM@N_xC_y$) was screened and explored in terms of structure, coordination, formation energy, structural and electrochemical stability, electronic properties, electrical conductivity, and reaction mechanism for HER, OER, and ORR using DFT and ML-based descriptors.[210] Among various $TM@N_xC_y$ SACs, the $TM@N_2C_2$ shows higher electrochemical catalytic performance, tends to be more easily formed, and possesses longer durability without aggregation or dissolution. In the $TM@N_2C_2$ templates, Ni/Ru/Rh/Pt show low HER overpotentials. The ML-based descriptors indicate superior HER, OER, and ORR performances of $TM@N_2C_2$ compared to those of bench-mark noble metal catalysts. It was shown for the first time that both TM and carbon atoms participates in H adsorption.

**Table 1** shows the summary of applied ML algorithms and their applications in SACs designing through the input features engineering and feature importance analysis. The list of abbreviation for **Table 1** are presented in **Table 2**. As shown in **Table 1**, SVM, KRR, RFR, and DNN are mostly used as the supervised ML algorithm for the design of SACs to describe the relationship between the input features and SACs activity. All the mentioned algorithms are normally applied in Scikit-learn[211]. Atomic properties are mainly used as the input features for the designing of SACs from which the number of electrons in d orbital and enthalpy of vaporization are usually the most important input features for ML algorithms. However, the application of ML is limited by the lack of not only a large and high-quality database but also a generalized ML algorithm for further studies in this field.

Moreover, based on **Table 1**, the d-band center, enthalpy of vaporization, Bader charge, ionization energy, electron affinity, covalent radius, the electron numbers in d orbital, formation energy, oxide formation enthalpy, etc. mainly are used as the key descriptors to describe the catalytic activity of SACs. Still, one of the main hurdles for employing ML in heterogeneous catalyst design is the lack of appropriate descriptors as input features for ML. An appropriate descriptor needs to simultaneously possess: (1) physical interpretation, (2) high simplicity, and (3) relatively high feature importance. To some extent, the black-box nature of ML techniques occasionally makes a physical interpretation of descriptors, such as d-band center and enthalpy of vaporization, non-trivial. In particular, the d-band center is widely adopted as an efficient descriptor,[212] typically with high feature importance to describe the reactivity of SACs. However, the d levels of atomically dispersed metal atoms on a graphene substrate may not form a band that makes evaluating the position of the d-band center impossible. Therefore, frontier molecular orbitals and density of

states (DOS) seem more appropriate descriptors than the d-band center.[213] However, obtaining frontier molecular orbital and DOS requires time-consuming DFT calculations, making this descriptor not worthwhile. In fact, the simplicity of descriptors requires using metal atom and substrate properties, being readily obtained without needing time-consuming DFT calculations. In contrast to Bader charge and DOS, descriptors such as atomic number, number of electrons in d orbital, ionization energy, and coordination number of metal atoms possess simplicity requirements.

## 5. Summary and future prospective

Recently ML has gained much interest for rational deign of heterogenous catalysts due to its potential for robust and fast prediction of catalysts properties by establishing structure-activity relationships. High throughput screening and feature importance analysis can be achieved through deep structure-activity establishment. However, ML is still at an early stage for designing of heterogeneous catalysts. In this review, high throughput screening and feature importance analysis using ML are provided as the guidelines for heterogeneous catalysts screening and discovery. Although much research has been carried out on the application of ML to improve the activity and stability of heterogeneous catalysts and SACs, there are still challenges to be resolved, requiring additional studies as follows:

(1) There remains room for ML to investigate the catalytic performances and strability,[214–217] and improve calculated parameters for stable SACs.[218,219] In addition, ML technique can help to investigate the hybridization of SACs,[220] atomic interface effect,[221] and the aggregation energy [207]. Moreover, SACs face such challenges as low metal loading, low selectivity and activity, and the lack of catalytic mechanisms.[137] Therefore ML can help the community to understand the reaction pathways and the catalytic mechanisms[222–226] to improve selectivity and activity of high-loaded SACs on graphene supports.[227–230] In addition, there is a clear need for ML to consider environmental effects, interfacial engineering, SAC coverage, and the potential for agglomeration. ML can be used for the synthesis of high-loaded SACs, multi-metal SACs, and multi atom catalysts.[231,232] In other words, since the structure-activity relationships for nanoclusters and DACs is much more complicated than SACs,[233] it will be useful to apply ML for predicting adsorption energies for them using new descriptors to consider the synergetic effect of several metals.[234]

(2) ML techniques continue to improve for studying adsorption energies, overpotentials, and metal-support interactions for various SACs, but the field of predictive SAC synthesis to guide experiments is much needed. Because SACs face tedious preparation processes,[8,235,236] ML can accelerate high-throughput experimentation for synthesis and characterizations of SACs.[191,237–242] ML can be also applied to predict Faraday Efficiency and onset potentials to help understand the volcano plots.

(3) A major hurdle for developing ML-aided heterogeneous catalyst design is the lack of sufficient and consistent datasets, data scarcity, bias, and noise from both experiments and QM calculations, which is a high priority to avoid overfitting.[48] In order to solve this issue, active learning and transfer learning can be applied which are efficient in compensating for the lack in data. In other words, having a large database composed of DFT-calculated and experimental data is required to train the generalized ML algorithm for systematic and comprehensive discovery of SACs. We expect that in the near future, with a huge database and a universal ML algorithm, the applicability of theoretical calculations for electroreduction reactions using SACs will be improved greatly.[243] In addition, the vast parameter space for dynamic catalysts requires applying ML to screen candidate catalysts by predicting the regions with high selectivity and operability.[244] The effect of coordination number, coordination atoms, designed bond length, and bond angle on the current density, overpotentials, and reaction mechanism should be considered through ML.[245–247]

(4) ML has the potential to predict the properties of SACs very quickly and accurately, but its application has been limited to specific systems using various ML algorithms. Therefore, a fair comparison to assess the strengths and best use of different ML algorithms is needed. Also, similar to ML-aided retrosynthesis and reaction planning,[72,248,249] a strong need is for development of a universal (generalized) ML algorithm that changes ML from a supportive tool to a surrogate tool for SACs design. This universal ML algorithm should be extended to widespread SACs and supports for all electroreduction reactions toward efficient and cost-effective potential SACs to balance between the activity and stability.[250]

(5) As shown in **Table 1**, two-dimensional (2D) materials leading to reduced computational cost due to their simplicity in structure. However, three-dimensional (3D) materials, such as Oxides and Nitrides,[251,252] play major role in catalysis and should be heavily investigated by existing or new ML algorithms.

## Conflicts of interest

These authors respectfully declare that, there are no conflicts of interest to acknowledge for this research.

## Acknowledgements

Z.L. acknowledge supports by the RGC (16304421), the Innovation and Technology Commission (ITC-CNERC14SC01), Guangdong Science and Technology Department (Project#:2020A0505090003), Research Fund of Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology (No. 2020B1212030010), IER foundation (HT-JD-CXY-201907), and Shenzhen Special Fund for Central Guiding the Local Science and Technology Development (2021Szvup136). Technical assistance from the Materials Characterization and Preparation Facilities of HKUST is greatly appreciated. W.A.G. acknowledges support by the DOE Liquid Sunlight Alliance (LiSA) (DE-SC0021266) and the US National Science Foundation (NSF CBET-2005250).

**References**:


1. Liu, Y., Esan, O. C., Pan, Z. & An, L. Machine learning for advanced energy materials. *Energy AI* **3**, 100049 (2021).
2. Lemm, D., von Rudorff, G. F. & von Lilienfeld, O. A. Machine learning based energy-free structure predictions of molecules, transition states, and solids. *Nat. Commun.* **12**, 1–10 (2021).
3. Guo, Y. *et al.* Machine-Learning-Guided Discovery and Optimization of Additives in Preparing Cu Catalysts for CO2Reduction. *J. Am. Chem. Soc.* **143**, 5755–5762 (2021).
4. Baum, Z. J. *et al.* Artificial Intelligence in Chemistry: Current Trends and Future Directions. *J. Chem. Inf. Model.* **61**, 3197–3212 (2021).
5. Faber, F., Lindmaa, A., Anatole Von Lilienfeld, O. & Armiento, R. Crystal Structure Representations for Machine Learning Models of Formation Energies. doi:10.1002/qua.24917.
6. Pilania, G. *et al.* Machine learning bandgaps of double perovskites OPEN. *Nat. Publ. Gr.* (2015) doi:10.1038/srep19375.
7. Huang, S.-D., Shang, C., Zhang, X.-J. & Liu, Z.-P. Material discovery by combining stochastic surface walking global optimization with a neural network †. (2017) doi:10.1039/c7sc01459g.
8. Han, Y. *et al.* Machine-learning-driven synthesis of carbon dots with enhanced quantum yields. *ACS Nano* **14**, 14761–14768 (2020).
9. Yin, H. *et al.* The data-intensive scientific revolution occurring where two-dimensional materials meet machine learning. *Cell Reports Phys. Sci.* **2**, 100482 (2021).
10. Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *J. Mater.* **3**, 159–177 (2017).
11. Kito, S. & Hattori, T. Neural network as a tool for catalyst development. *Catal. Today* **23**, 347–355 (1995).
12. Kito, S., Hattori, T. & Murakami, Y. Estimation of the acid strength of mixed oxides by a neural network. *Ind. Eng. Chem. Res.* **31**, 979–981 (1992).
13. Toyao, T. *et al.* Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **10**, 2260–2297 (2020).
14. Chen, C. *et al.* A Critical Review of Machine Learning of Energy Materials. *Adv. Energy*

*Mater.* **10**, 1–36 (2020).
15. Chanussot, L. *et al.* Open Catalyst 2020 (OC20) Dataset and Community Challenges. (2021) doi:10.1021/acscatal.0c04525.
16. Funes-Ardoiz, I. & Schoenebeck, F. Established and Emerging Computational Tools to Study Homogeneous Catalysis—From Quantum Mechanics to Machine Learning. *Chem* **6**, 1904–1913 (2020).
17. Li, Z., Wang, S. & Xin, H. Toward artificial intelligence in catalysis. *Nat. Catal.* **1**, 641–642 (2018).
18. Schlexer Lamoureux, P. *et al.* Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem* **11**, 3581–3601 (2019).
19. Freeze, J. G., Kelly, H. R. & Batista, V. S. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. (2019) doi:10.1021/acs.chemrev.8b00759.
20. Keith, J. A. *et al.* Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. (2021) doi:10.1021/acs.chemrev.1c00107.
21. Xu, J., Cao, X.-M. & Hu, P. Perspective on computational reaction prediction using machine learning methods in heterogeneous catalysis. *Phys. Chem. Chem. Phys* **23**, 11155 (2021).
22. Yang, W., Fidelis, T. T. & Sun, W.-H. Machine Learning in Catalysis, From Proposal to Practicing. (2019) doi:10.1021/acsomega.9b03673.
23. Erdem Günay, M. & Yıldırım, R. Recent advances in knowledge discovery for heterogeneous catalysis using machine learning. *Catal. Rev.* **63**, 120–164 (2021).
24. Liu, W. *et al.* catalysts Molecular Dynamics and Machine Learning in Catalysts. **11**, 1129 (2021).
25. Yu, Z. & Huang, W. Accelerating Optimizing the Design of Carbon-based Electrocatalyst Via Machine Learning. (2021) doi:10.1002/elan.202100224.
26. Guan, Y. *et al.* Machine Learning in Solid Heterogeneous Catalysis: Recent Developments, Challenges and Perspectives. *Chem. Eng. Sci.* **248**, 117224 (2021).
27. Orupattur, N. V., Mushrif, S. H. & Prasad, V. Catalytic materials and chemistry development using a synergistic combination of machine learning and ab initio methods. *Comput. Mater. Sci.* **174**, 109474 (2020).

28. Freeze, J. G., Kelly, H. R. & Batista, V. S. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. *Chem. Rev.* **119**, 6595–6612 (2019).
29. Medford, A. J., Ross Kunz, M., Ewing, S. M., Borders, T. & Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. (2018) doi:10.1021/acscatal.8b01708.
30. Takahashi, K. *et al.* Title: The Rise of Catalyst Informatics: Towards Catalyst Genomics The Rise of Catalyst Informatics: Towards Catalyst Genomics. doi:10.1002/cctc.201801956.
31. Chen, Z. W. *et al.* Machine-learning-accelerated discovery of single-atom catalysts based on bidirectional activation mechanism. *Chem Catal.* 1–13 (2021) doi:10.1016/j.checat.2021.03.003.
32. Soyemi, A. & Szilvási, T. Trends in computational molecular catalyst design. *Dalt. Trans.* (2021) doi:10.1039/d1dt01754c.
33. Sun, M. *et al.* Self-Validated Machine Learning Study of Graphdiyne-Based Dual Atomic Catalyst. *Advanced Energy Materials* vol. 11 (2021).
34. Haywood, A. *et al.* Kernel Methods for Predicting Yields of Chemical Reactions. (2021).
35. Noble, W. S. What is a support vector machine? *Nat. Biotechnol.* **24**, 1565–1567 (2006).
36. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, (2019).
37. Svetnik, V. *et al.* Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **43**, 1947–1958 (2003).
38. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
39. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science (80-. ).* **360**, 186–190 (2018).
40. Strobl, C., Boulesteix, A.-L., Zeileis, A. & Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* **8**, 1–21 (2007).
41. Malek, A. *et al.* A Data-Driven Framework for the Accelerated Discovery of CO2

Reduction Electrocatalysts. *Front. Energy Res.* **9**, 1–15 (2021).
42. Palkovits, S. A Primer about Machine Learning in Catalysis-A Tutorial with Code ChemCatChem. *ChemCatChem* **12**, 3995–4008 (2020).
43. Huang, B. & Anatole Von Lilienfeld, O. Ab Initio Machine Learning in Chemical Compound Space. (2021) doi:10.1021/acs.chemrev.0c01303.
44. Wei, A., Ye, H., Guo, Z. & Xiong, J. SISSO-assisted prediction and design of mechanical properties of porous graphene with a uniform nanopore array. *Nanoscale Adv.* **4**, 1455–1463 (2022).
45. Rasmussen, C. E. Gaussian processes in machine learning. in *Summer school on machine learning* 63–71 (Springer, 2003).
46. Gupta, T., Zaki, M., Krishnan, N. M. A. & Mausam. MatSciBERT: A Materials Domain Language Model for Text Mining and Information Extraction. 1–11 (2021) doi:10.1038/s41524-022-00784-w.
47. Žižka, J., Dařena, F. & Svoboda, A. *Text mining with machine learning: principles and techniques*. (CRC Press, 2019).
48. Gambo, Y. *et al.* Catalyst design and tuning for oxidative dehydrogenation of propane – A review. *Appl. Catal. A Gen.* **609**, 117914 (2021).
49. Kitchin, J. R. Machine learning in catalysis. *Nat. Catal.* **1**, 230–232 (2018).
50. Goldsmith, B. R., Esterhuizen, J., Liu, J. X., Bartel, C. J. & Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **64**, 2311–2323 (2018).
51. Li, Z., Achenie, L. E. K. & Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electrocatalysts. **27**, 2021 (2020).
52. Li, X.-T., Chen, L., Wei, G.-F., Shang, C. & Liu, Z.-P. Sharp Increase in Catalytic Selectivity in Acetylene Semihydrogenation on Pd Achieved by a Machine Learning Simulation-Guided Experiment. (2020) doi:10.1021/acscatal.0c02158.
53. Tamtaji, M. *et al.* Singlet Oxygen Photosensitization Using Graphene-Based Structures and Immobilized Dyes: A Review. *ACS Appl. Nano Mater.* **4**, 7563–7586 (2021).
54. Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-Learning-Augmented Chemisorption Model for CO2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).
55. Mohammed, M. L. *et al.* Optimisation of alkene epoxidation catalysed by polymer

supported Mo(VI) complexes and application of artificial neural network for the prediction of catalytic performances. *Appl. Catal. A Gen.* **466**, 142–152 (2013).

56. Sasaki, M., Hamada, H., Kintaichi, Y. & Ito, T. *Application of a neural network to the analysis of catalytic reactions Analysis of NO decomposition over Cu/ZSM-5 zeolite*. *Applied Catalysis A: General* vol. 132 (1995).
57. Selvaratnam, B. & Koodali, R. T. Machine learning in experimental materials chemistry. *Catal. Today* (2020) doi:10.1016/j.cattod.2020.07.074.
58. Tao, Q., Xu, P., Li, M. & Lu, W. Machine learning for perovskite materials design and discovery. *npj Comput. Mater.* **7**, 1–18 (2021).
59. Shi, L., Chang, D., Ji, X. & Lu, W. Using Data Mining to Search for Perovskite Materials with Higher Specific Surface Area. *J. Chem. Inf. Model.* **58**, 2420–2427 (2018).
60. Mowbray, M. *et al.* Machine learning for biochemical engineering: A review. *Biochem. Eng. J.* **172**, 108054 (2021).
61. Coşgun, A., Günay, M. E. & Yıldırım, R. Exploring the critical factors of algal biomass and lipid production for renewable fuel production by machine learning. *Renew. Energy* **163**, 1299–1317 (2021).
62. Alper Tapan, N., Yıldırım, R. & Erdem Günay, M. Analysis of past experimental data in literature to determine conditions for high performance in biodiesel production. *Biofuels, Bioprod. Biorefining* **10**, 422–434 (2016).
63. McCullough, K., Williams, T., Mingle, K., Jamshidi, P. & Lauterbach, J. High-throughput experimentation meets artificial intelligence: A new pathway to catalyst discovery. *Phys. Chem. Chem. Phys.* **22**, 11174–11196 (2020).
64. Desgranges, C. & Delhommelle, J. Towards a machine learned thermodynamics: Exploration of free energy landscapes in molecular fluids, biological systems and for gas storage and separation in metal-organic frameworks. *Mol. Syst. Des. Eng.* **6**, 52–65 (2021).
65. Torrisi, S. B. *et al.* ARTICLE Random forest machine learning models for interpretable X-ray absorption near-edge structure spectrum-property relationships. doi:10.1038/s41524-020-00376-6.
66. Basyaruddin Abdul Rahman, M. *et al.* Application of Artificial Neural Network for Yield Prediction of Lipase-Catalyzed Synthesis of Dioctyl Adipate. doi:10.1007/s12010-008-

8465-z.

67. Miyazato, I., Nguyen, T. N., Takahashi, L., Taniike, T. & Takahashi, K. Representing Catalytic and Processing Space in Methane Oxidation Reaction via Multioutput Machine Learning. *J. Phys. Chem. Lett. 2021* **12**, 814 (2021).

68. Kunkel, B., Kabelitz, A., Buzanich, A. G. & Wohlrab, S. Increasing the Efficiency of Optimized V-SBA-15 Catalysts in the Selective Oxidation of Methane to Formaldehyde by Artificial Neural Network Modelling. *Catalysts* **10**, 1411 (2020).

69. Artrith, N., Lin, Z. & Chen, J. G. Predicting the Activity and Selectivity of Bimetallic Metal Catalysts for Ethanol Reforming using Machine Learning. *ACS Catal.* **22**, 41 (2021).

70. Ma, S. & Liu, Z.-P. Machine Learning for Atomic Simulation and Activity Prediction in Heterogeneous Catalysis: Current Status and Future. (2020) doi:10.1021/acscatal.0c03472.

71. Yang, T. *et al.* High-Throughput Identification of Exfoliable Two-Dimensional Materials with Active Basal Planes for Hydrogen Evolution. *Cite This ACS Energy Lett* **2020**, (2313).

72. Schwaller, P. *et al.* Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* **5**, 1572–1583 (2019).

73. Coley, C. W., Green, W. H. & Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).

74. Zavyalova, U., Holena, M., Schlçgl, R. & Baerns, M. Statistical Analysis of Past Catalytic Data on Oxidative Methane Coupling for New Insights into the Composition of High-Performance Catalysts. doi:10.1002/cctc.201100186.

75. Hyun Woo Kim, S. *et al.* As featured in: Reaction condition optimization for non-oxidative conversion of methane using artificial intelligence †. *React. Chem. Eng* (2021) doi:10.1039/d0re00378f.

76. Wulf, C. *et al.* A Unified Research Data Infrastructure for Catalysis Research – Challenges and Concepts. 3223–3236 (2021) doi:10.1002/cctc.202001974.

77. Kim, E. *et al.* Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater* **29**, 31 (2017).

78. Ohyama, J. *et al.* Catalysis Science & Technology Direct design of active catalysts for low

temperature oxidative coupling of methane via machine learning and data mining †. *Cite this Catal. Sci. Technol* **11**, 524 (2021).

79. Takahashi, L. *et al.* Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts †. (2021) doi:10.1039/d1sc04390k.
80. Takahashi, K., Takahashi, L., Nguyen, T. N., Thakur, A. & Taniike, T. Multidimensional Classification of Catalysts in Oxidative Coupling of Methane through Machine Learning and High-Throughput Data. *J. Phys. Chem. Lett.* **11**, 6819–6826 (2020).
81. Nguyen, T. N. *et al.* Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane. *ACS Catal.* **11**, 1797–1809 (2021).
82. Nakanowatari, S. *et al.* Extraction of Catalyst Design Heuristics from Random Catalyst Dataset and their Utilization in Catalyst Development for Oxidative Coupling of Methane. *ChemCatChem* **13**, 3262–3269 (2021).
83. Ohyama, J., Nishimura, S. & Takahashi, K. Data Driven Determination of Reaction Conditions in Oxidative Coupling of Methane via Machine Learning. *ChemCatChem* **11**, 4307–4313 (2019).
84. Wang, W. *et al.* Automated pipeline for superalloy data by text mining. *npj Comput. Mater.* **8**, 1–12 (2022).
85. Nishimura, S. *et al.* Revisiting Machine Learning Predictions for Oxidative Coupling of Methane (OCM) based on Literature Data ChemCatChem. *ChemCatChem* **12**, 5888–5892 (2020).
86. Mine, S. *et al.* Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts. (2021) doi:10.1002/cctc.202100495.
87. Suzuki, K. *et al.* Statistical Analysis and Discovery of Heterogeneous Catalysts Based on Machine Learning from Diverse Published Data. *ChemCatChem* **11**, 4537–4547 (2019).
88. Chen, Y., Li, R., Suo, H. & Liu, C. Evaluation of a Data-Driven, Machine Learning Approach for Identifying Potential Candidates for Environmental Catalysts: From Database Development to Prediction. (2021) doi:10.1021/acsestengg.1c00125.
89. Odabaşi, Ç., Günay, M. E. & Yildirim, R. Knowledge extraction for water gas shift reaction over noble metal catalysts from publications in the literature between 2002 and 2012. *Int. J. Hydrogen Energy* **39**, 5733–5746 (2014).

90. Smith, A., Keane, A., Dumesic, J. A., Huber, G. W. & Zavala, V. M. A machine learning framework for the analysis and prediction of catalytic activity from experimental data. *Appl. Catal. B Environ.* **263**, 118257 (2020).
91. Brown, K. A., Brittman, S., Jariwala, D. & Celano, U. Machine Learning in Nanoscience: Big Data at Small Scales. *Nano Lett.* (2021) doi:10.1021/acs.nanolett.9b04090.
92. Nhat Nguyen, T. *et al.* High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. (2021) doi:10.1021/acscatal.9b04293.
93. Tang, H., Hosein, A. & Mattioli-Belmonte, M. Traditional Chinese Medicine and orthopedic biomaterials: Host of opportunities from herbal extracts. *Mater. Sci. Eng. C* **120**, 111760 (2021).
94. Shi, Y., Prieto, P. L., Zepel, T., Grunert, S. & Hein, J. E. Automated Experimentation Powers Data Science in Chemistry. *Acc. Chem. Res* **54**, 31 (2021).
95. Szymanski, N. J. *et al.* Toward autonomous design and synthesis of novel inorganic materials. **8**, 2169 (2021).
96. Xiao, Y., Shen, C. & Hadaeghi, N. Quantum Mechanical Screening of 2D MBenes for the Electroreduction of CO 2 to C1 Hydrocarbon Fuels . *J. Phys. Chem. Lett.* **12**, 6370–6382 (2021).
97. Stach, E. *et al.* Autonomous experimentation systems for materials development: A community perspective. *Matter* **4**, 2702–2726 (2021).
98. Stroyuk, O. *et al.* 2100169 (1 of 11) High-Throughput Robotic Synthesis and Photoluminescence Characterization of Aqueous Multinary Copper-Silver Indium Chalcogenide Quantum Dots. (2021) doi:10.1002/ppsc.202100169.
99. Moses, O. A. *et al.* Integration of data-intensive, machine learning and robotic experimental approaches for accelerated discovery of catalysts in renewable energy-related reactions. *Mater. Reports Energy* **1**, 100049 (2021).
100. Burger, B. *et al.* A mobile robotic chemist. *Nature* **583**, (2020).
101. Ge, L. *et al.* Predicted Optimal Bifunctional Electrocatalysts for the Hydrogen Evolution Reaction and the Oxygen Evolution Reaction Using Chalcogenide Heterostructures Based on Machine Learning Analysis of in Silico Quantum Mechanics Based High Throughput Screening. *J. Phys. Chem. Lett.* **11**, 869–876 (2020).
102. Jinnouchi, R., Hirata, H. & Asahi, R. Extrapolating Energetics on Clusters and Single-

Crystal Surfaces to Nanoparticles by Machine-Learning Scheme. (2017) doi:10.1021/acs.jpcc.7b08686.

103. Jinnouchi, R. & Asahi, R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. *J. Phys. Chem. Lett.* **8**, 4279–4283 (2017).

104. Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution. *Nat. Catal.* **1**, 696–703 (2018).

105. Huang, Y. *et al.* Mechanistic Understanding and Design of Non-noble Metal-based Single-atom Catalysts Supported on Two-dimensional Materials for CO2 Electroreduction. *J. Mater. Chem. A* (2021).

106. Zhong, M. *et al.* Accelerated discovery of CO2 electrocatalysts using active machine learning. *Nature* **581**, 178–183 (2020).

107. Ulissi, Z. W., Medford, A. J., Bligaard, T. & Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **8**, (2017).

108. Toyao, T. *et al.* Toward Effective Utilization of Methane: Machine Learning Prediction of Adsorption Energies on Metal Alloys. *J. Phys. Chem. C* **122**, 8315–8326 (2018).

109. Esterhuizen, J. A., Goldsmith, B. R. & Linic, S. Theory-Guided Machine Learning Finds Geometric Structure-Property Relationships for Chemisorption on Subsurface Alloys. *Chem* **6**, 3100–3117 (2020).

110. Wexler, R. B., Mark, J., Martirez, P. & Rappe, A. M. Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni 2 P from Nonmetal Surface Doping Interpreted via Machine Learning. *J. Am. Chem. Soc* **140**, 4678–4683 (2018).

111. Li, Z., Ma, X. & Xin, H. Feature engineering of machine-learning chemisorption models for catalyst design. *Catal. Today* **280**, 232–238 (2017).

112. Li, Z., Wang, S., Chin, W. S., Achenie, L. E. & Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. *J. Mater. Chem. A* **5**, 24131–24138 (2017).

113. Takigawa, I., Shimizu, K.-I., Tsuda, K. & Takakusagi, S. Machine-learning prediction of the d-band center for metals and bimetals. (2016) doi:10.1039/c6ra04345c.

114. Tanaka, I. *Nanoinformatics*. (Springer Nature, 2018).

115. Liu, C. *et al.* Frontier Molecular Orbital Based Analysis of Solid-Adsorbate Interactions

over Group 13 Metal Oxide Surfaces. *J. Phys. Chem. C* **124**, 15355–15365 (2020).

116. Liu, Z. H., Shi, T. T. & Chen, Z. X. Machine learning prediction of monatomic adsorption energies with non-first-principles calculated quantities. *Chem. Phys. Lett.* **755**, 137772 (2020).
117. Wang, X. *et al.* Accelerating 2D MXene catalyst discovery for the hydrogen evolution reaction by computer-driven workflow and an ensemble learning strategy. *J. Mater. Chem. A* **8**, 23488–23497 (2020).
118. Zheng, J. *et al.* High-Throughput Screening of Hydrogen Evolution Reaction Catalysts in MXene Materials. *J. Phys. Chem. C* **124**, 13695–13705 (2020).
119. Sun, Y. *et al.* Covalency competition dominates the water oxidation structure–activity relationship on spinel oxides. *Nat. Catal.* **3**, 554–563 (2020).
120. Ting, K. W. *et al.* Catalytic Methylation of m-Xylene, Toluene, and Benzene Using CO2 and H2 over TiO2-Supported Re and Zeolite Catalysts: Machine-Learning-Assisted Catalyst Optimization. *ACS Catal.* **11**, 5829–5838 (2021).
121. Rück, M., Garlyyev, B., Mayr, F., Bandarenka, A. S. & Gagliardi, A. Oxygen Reduction Activities of Strained Platinum Core–Shell Electrocatalysts Predicted by Machine Learning. *J. Phys. Chem. Lett.* **11**, 1773–1780 (2020).
122. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, (2016).
123. Yang, Q. *et al.* Revealing property-performance relationships for efficient CO2 hydrogenation to higher hydrocarbons over Fe-based catalysts: Statistical analysis of literature data and its experimental validation. *Appl. Catal. B Environ.* **282**, 119554 (2021).
124. Mamun, O., Winther, K. T., Boes, J. R. & Bligaard, T. High-throughput calculations of catalytic properties of bimetallic alloy surfaces. doi:10.1038/s41597-019-0080-z.
125. Fung, V., Zhang, J., Juarez, E. & Sumpter, B. G. ARTICLE Benchmarking graph neural networks for materials chemistry. doi:10.1038/s41524-021-00554-0.
126. Xu, J., Cao, X.-M. & Hu, P. Improved Prediction for the Methane Activation Mechanism on Rutile Metal Oxides by a Machine Learning Model with Geometrical Descriptors. (2019) doi:10.1021/acs.jpcc.9b08939.

127. Chen, Y., Huang, Y., Cheng, T. & Goddard, W. A. Identifying Active Sites for CO2 Reduction on Dealloyed Gold Surfaces by Combining Machine Learning with Multiscale Simulations. *J. Am. Chem. Soc.* **141**, 11651–11657 (2019).
128. Ulissi, Z. W. *et al.* Machine-learning methods enable exhaustive searches for active Bimetallic facets and reveal active site motifs for CO2 reduction. *ACS Catal.* **7**, 6600–6608 (2017).
129. Ologunagba, D. & Kattel, S. Machine learning prediction of surface segregation energies on low index bimetallic surfaces. *Energies* **13**, 20–25 (2020).
130. Singh, A. R., Rohr, B. A., Gauthier, J. A. & Nørskov, J. K. Predicting Chemical Reaction Barriers with a Machine Learning Model. *Catal. Letters* **149**, 2347–2354 (2019).
131. Ulissi, Z. W., Singh, A. R., Tsai, C. & Nørskov, J. K. Automated Discovery and Construction of Surface Phase Diagrams Using Machine Learning. *J. Phys. Chem. Lett.* **7**, 3931–3935 (2016).
132. Weng, B. *et al.* Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts. *Nat. Commun.* **11**, 1–8 (2020).
133. Ding, S., Hülsey, M. J., Pérez-Ramírez, J. & Yan, N. Transforming Energy with Single-Atom Catalysts. *Joule* **3**, 2897–2929 (2019).
134. Hossain, M. D., Huang, Y., Yu, T. H., Goddard, W. A. & Luo, Z. Reaction mechanism and kinetics for CO2 reduction on nickel single atom catalysts from quantum mechanics. *Nat. Commun.* **11**, 1–14 (2020).
135. Gu, Y. *et al.* Atomic-Scale Tailoring and Molecular-Level Tracking of Oxygen-Containing Tungsten Single-Atom Catalysts with Enhanced Singlet Oxygen Generation. *ACS Appl. Mater. Interfaces* 1–10 (2021) doi:10.1021/acsami.1c09016.
136. Tamtaji, M. *et al.* Singlet Oxygen Photosensitization Using Graphene-Based Structures and Immobilized Dyes : A Review. *ACS Appl. Nano Mater.* **4**, 7563–7583 (2021).
137. Chen, C., Zhang, Z., Li, G., Li, L. & Lin, Z. Recent Advances on Nanomaterials for Electrocatalytic CO2Conversion. *Energy and Fuels* **35**, 7485–7510 (2021).
138. Sun, K. *et al.* Electrochemical Oxygen Reduction to Hydrogen Peroxide via a Two-Electron Transfer Pathway on Carbon-Based Single-Atom Catalysts. *Adv. Mater. Interfaces* **8**, 1–16 (2021).
139. Wang, A., Li, J. & Zhang, T. Heterogeneous single-atom catalysis. *Nat. Rev. Chem.* **2**, 65–

81 (2018).
140. Xu, H., Cheng, D., Cao, D. & Zeng, X. C. A universal principle for a rational design of single-atom electrocatalysts. *Nat. Catal.* **1**, 339–348 (2018).
141. Hossain, M. D. *et al.* Rational Design of Graphene-Supported Single Atom Catalysts for Hydrogen Evolution Reaction. *Adv. Energy Mater.* **9**, 1–10 (2019).
142. Li, L., Chang, X., Lin, X., Zhao, Z. J. & Gong, J. Theoretical insights into single-atom catalysts. *Chem. Soc. Rev.* **49**, 8156–8178 (2020).
143. Liu, D., He, Q., Ding, S. & Song, L. Structural Regulation and Support Coupling Effect of Single-Atom Catalysts for Heterogeneous Catalysis. *Advanced Energy Materials* vol. 10 (2020).
144. Dobbelaere, M. R., Plehiers, P. P., Van de Vijver, R., Stevens, C. V. & Van Geem, K. M. Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats. *Engineering* (2021) doi:10.1016/j.eng.2021.03.019.
145. Wang, K. *et al.* Metal-free nitrogen -doped carbon nanosheets: A catalyst for the direct synthesis of imines under mild conditions. *Green Chem.* **21**, 2448–2461 (2019).
146. Liu, X. *et al.* Identifying the Activity Origin of a Cobalt Single-Atom Catalyst for Hydrogen Evolution Using Supervised Learning. *Adv. Funct. Mater.* **2100547**, 1–9 (2021).
147. Wu, L., Guo, T. & Li, T. Rational design of transition metal single-Atom electrocatalysts: A simulation-based, machine learning-Accelerated study. *J. Mater. Chem. A* **8**, 19290–19299 (2020).
148. Wang, M. & Zhu, H. Machine Learning for Transition-Metal-Based Hydrogen Generation Electrocatalysts. *ACS Catal.* 3930–3937 (2021) doi:10.1021/acscatal.1c00178.
149. Kibria, M. G. *et al.* Electrochemical CO2 Reduction into Chemical Feedstocks: From Mechanistic Electrocatalysis Models to System Design. *Adv. Mater.* **31**, 1–24 (2019).
150. Kim, J., Kang, D., Kim, S. & Jang, H. W. Catalyze Materials Science with Machine Learning. *ACS Mater. Lett.* 1151–1171 (2021) doi:10.1021/acsmaterialslett.1c00204.
151. Chen, L., Xu, X., Yang, W. & Jia, J. Recent advances in carbon-based electrocatalysts for oxygen reduction reaction. *Chinese Chem. Lett.* **31**, 626–634 (2020).
152. Johnson, D., Qiao, Z. & Djire, A. Progress and Challenges of Carbon Dioxide Reduction Reaction on Transition Metal Based Electrocatalysts. (2021)

doi:10.1021/acsaem.1c01624.

153. Hannagan, R. T., Giannakakis, G., Flytzani-Stephanopoulos, M. & Sykes, E. C. H. Single-Atom Alloy Catalysis. *Chem. Rev.* **120**, 12044–12088 (2020).

154. Saxena, S., Khan, T. S., Jalid, F., Ramteke, M. & Haider, M. A. In silico high throughput screening of bimetallic and single atom alloys using machine learning and ab initio microkinetic modelling. *J. Mater. Chem. A* **8**, 107–123 (2020).

155. Dasgupta, A., Gao, Y., Broderick, S. R., Pitman, E. B. & Rajan, K. Machine Learning-Aided Identification of Single Atom Alloy Catalysts. *J. Phys. Chem. C* **124**, 14158–14166 (2020).

156. Hoyt, R. A. *et al.* Machine Learning Prediction of H Adsorption Energies on Ag Alloys. *J. Chem. Inf. Model.* **59**, 1357–1365 (2019).

157. Mitchell, S. *et al.* Automated Image Analysis for Single-Atom Detection in Catalytic Materials by Transmission Electron Microscopy. *J. Am. Chem. Soc.* (2022) doi:10.1021/jacs.1c12466.

158. Xiang, S. *et al.* Solving the structure of 'single-atom' catalysts using machine learning-assisted XANES analysis. *Phys. Chem. Chem. Phys.* **24**, 5116–5124 (2022).

159. Zhang, J. *et al.* Single-atom catalysts for thermal- and electro-catalytic hydrogenation reactions. *J. Mater. Chem. A* **10**, 5743–5757 (2022).

160. Jorner, K., Tomberg, A., Bauer, C., Sköld, C. & Norrby, P. O. Organic reactivity from mechanism to machine learning. *Nat. Rev. Chem.* **5**, (2021).

161. Gao, D., Liu, T., Wang, G. & Bao, X. Structure Sensitivity in Single-Atom Catalysis toward CO2Electroreduction. *ACS Energy Lett.* **6**, 713–727 (2021).

162. Zahrt, A. F. *et al.* Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science (80-. ).* **363**, (2019).

163. Zhang, N. *et al.* Single-atom site catalysts for environmental catalysis. *Nano Res.* **13**, 3165–3182 (2020).

164. Hart, G. L. W., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nat. Rev. Mater.* **6**, (2021).

165. Ying, Y., Fan, K., Luo, X., Qiao, J. & Huang, H. Unravelling the origin of bifunctional OER/ORR activity for single-atom catalysts supported on C 2 N by DFT and machine learning . *J. Mater. Chem. A* 16860–16867 (2021) doi:10.1039/d1ta04256d.

166. Zheng, G. *et al.* High-Throughput Screening of a Single-Atom Alloy for Electroreduction of Dinitrogen to Ammonia. *ACS Appl. Mater. Interfaces* **13**, 16336–16344 (2021).
167. Fung, V., Hu, G., Wu, Z. & Jiang, D. E. Descriptors for Hydrogen Evolution on Single Atom Catalysts in Nitrogen-Doped Graphene. *J. Phys. Chem. C* **124**, 19571–19578 (2020).
168. Wu, L., Guo, T. & Li, T. Machine learning-accelerated prediction of overpotential of oxygen evolution reaction of single-atom catalysts. *iScience* **24**, 102398 (2021).
169. Zhu, X. *et al.* Activity Origin and Design Principles for Oxygen Reduction on Dual-Metal-Site Catalysts: A Combined Density Functional Theory and Machine Learning Study. *J. Phys. Chem. Lett.* **10**, 7760–7766 (2019).
170. Wan, X. *et al.* Machine-Learning-Accelerated Catalytic Activity Predictions of Transition Metal Phthalocyanine Dual-Metal-Site Catalysts for CO2Reduction. *J. Phys. Chem. Lett.* **12**, 6111–6118 (2021).
171. Wan, X., Zhang, Z., Yu, W. & Guo, Y. A density-functional-theory-based and machine-learning-accelerated hybrid method for intricate system catalysis. *Mater. Reports Energy* **1**, 100046 (2021).
172. Niu, H. *et al.* Single-Atom Rhodium on Defective g-C3N4: A Promising Bifunctional Oxygen Electrocatalyst. *ACS Sustain. Chem. Eng.* **9**, 3590–3599 (2021).
173. Lin, S., Xu, H., Wang, Y., Zeng, X. C. & Chen, Z. 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).
174. Guo, X. *et al.* Simultaneously Achieving High Activity and Selectivity toward Two-Electron O2 Electroreduction: The Power of Single-Atom Catalysts. *ACS Catal.* **9**, 11042–11054 (2019).
175. Deng, C. *et al.* Understanding activity origin for the oxygen reduction reaction on bi-atom catalysts by DFT studies and machine-learning. *J. Mater. Chem. A* **8**, 24563–24571 (2020).
176. Melisande Fischer, J. *et al.* Accurate prediction of binding energies for two-dimensional catalytic materials using machine learning. *ChemCatChem* **12**, 5109–5120 (2020).
177. Yuan, H., Li, Z., Zeng, X. C. & Yang, J. Descriptor-Based Design Principle for Two-Dimensional Single-Atom Catalysts: Carbon Dioxide Electroreduction. *J. Phys. Chem.*

*Lett.* **11**, 3481–3487 (2020).
178. Meyer, B., Sawatlon, B., Heinen, S., Anatole Von Lilienfeld, O. & Emence Corminboeuf, C. Machine learning meets volcano plots: computational discovery of cross-coupling catalysts †. (2018) doi:10.1039/c8sc01949e.
179. Pablo-García, S., García-Muelas, R., Sabadell-Rendón, A. & López, N. Dimensionality reduction of complex reaction networks in heterogeneous catalysis: From linear-scaling relationships to statistical learning techniques. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1–16 (2021) doi:10.1002/wcms.1540.
180. Gong, L. *et al.* Catalytic Mechanisms and Design Principles for Single-Atom Catalysts in Highly Efficient CO2 Conversion. *Advanced Energy Materials* vol. 9 (2019).
181. Yu, Z., Xu, H. & Cheng, D. Design of Single Atom Catalysts. *Adv. Phys. X* **6**, (2021).
182. Guan, X., Gao, W. & Jiang, Q. Design of bimetallic atomic catalysts for CO2reduction based on an effective descriptor. *J. Mater. Chem. A* **9**, 4770–4780 (2021).
183. Ling, C. *et al.* A General Two-Step Strategy–Based High-Throughput Screening of Single Atom Catalysts for Nitrogen Fixation. *Small Methods* vol. 3 (2019).
184. Song, W., Fu, L., He, C., Xie, K. & Guo, Y. Computational screening of 3d transition metal atoms anchored on the defective graphene for efficient electrocatalytic N2 fixation. *ChemPhysChem* (2021) doi:10.1002/cphc.202100257.
185. Tang, Y. *et al.* Nitrogen and boron coordinated single-atom catalysts for low-temperature CO/NO oxidations. *J. Mater. Chem. A* **9**, 15329–15345 (2021).
186. Gao, F., Wei, Y., Du, J. & Jiang, G. Theoretical screening of 2D materials supported transition-metal single atoms as efficient electrocatalysts for hydrogen evolution reaction. *Materialia* **18**, 101168 (2021).
187. Wang, Y. *et al.* High-throughput screening of carbon-supported single metal atom catalysts for oxygen reduction reaction. *Nano Res.* **12**, (2021).
188. Ren, C. *et al.* Relative Efficacy of Co− X4 Embedded Graphene (X= N, S, B, and P) Electrocatalysts towards Hydrogen Evolution Reaction: Is Nitrogen Really the Best Choice? *ChemCatChem* **12**, 536–543 (2020).
189. Prezhdo, O. V. Advancing Physical Chemistry with Machine Learning. *J. Phys. Chem. Lett.* **11**, 9656–9658 (2020).
190. Han, Z. K. *et al.* Single-atom alloy catalysts designed by first-principles calculations and

artificial intelligence. *Nat. Commun.* **12**, 1–9 (2021).

191. Liu, J. *et al.* Recent Progress in Non-Precious Metal Single Atomic Catalysts for Solar and Non-Solar Driven Hydrogen Evolution Reaction. *Advanced Sustainable Systems* vol. 4 (2020).
192. Yang, Z., Gao, W. & Jiang, Q. A machine learning scheme for the catalytic activity of alloys with intrinsic descriptors. *J. Mater. Chem. A* **8**, 17507–17515 (2020).
193. Sun, X. *et al.* Machine-learning-accelerated screening of hydrogen evolution catalysts in MBenes materials. *Appl. Surf. Sci.* **526**, 146522 (2020).
194. Liang, H., Xu, M. & Asselin, E. A Study of Two-Dimensional Single Atom-Supported MXenes as Hydrogen Evolution Reaction Catalysts Using DFT and Machine Learning. *ChemRxiv* (2021) doi:10.26434/chemrxiv.14566656.v1.
195. Sun, M., Dougherty, A. W., Huang, B., Li, Y. & Yan, C. H. Accelerating Atomic Catalyst Discovery by Theoretical Calculations-Machine Learning Strategy. *Advanced Energy Materials* vol. 10 (2020).
196. Zafari, M., Nissimagoudar, A. S., Umer, M., Lee, G. & Kim, K. S. First principles and machine learning based superior catalytic activities and selectivities for N2reduction in MBenes, defective 2D materials and 2D π-conjugated polymer-supported single atom catalysts. *Journal of Materials Chemistry A* vol. 9 9203–9213 (2021).
197. Zafari, M., Kumar, D., Umer, M. & Kim, K. S. Machine learning-based high throughput screening for nitrogen fixation on boron-doped single atom catalysts. *J. Mater. Chem. A* **8**, 5209–5216 (2020).
198. Chen, A., Zhang, X., Chen, L., Yao, S. & Zhou, Z. A Machine Learning Model on Simple Features for CO2Reduction Electrocatalysts. *J. Phys. Chem. C* **124**, 22471–22478 (2020).
199. Yang, T. *et al.* Protecting Single Atom Catalysts with Graphene/Carbon-Nitride 'chainmail'. *J. Phys. Chem. Lett.* **10**, 3129–3133 (2019).
200. Rivera-Cárcamo, C. *et al.* Stabilization of Metal Single Atoms on Carbon and TiO2 Supports for CO2 Hydrogenation: The Importance of Regulating Charge Transfer. *Adv. Mater. Interfaces* **8**, 1–17 (2021).
201. Hensen, E. J. M., Vlachos, D. G., Wang, Y. & Su, Y. Q. Finite-temperature structures of supported subnanometer catalysts inferred via statistical learning and genetic algorithm-based optimization. *ACS Nano* **14**, 13995–14007 (2020).

202. Serp, P. Cooperativity in supported metal single atom catalysis. *Nanoscale* **13**, 5985–6004 (2021).
203. Zhang, H., Lu, X. F., Wu, Z. P. & Lou, X. W. D. Emerging Multifunctional Single-Atom Catalysts/Nanozymes. *ACS Cent. Sci.* **6**, 1288–1301 (2020).
204. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
205. Su, Y. Q. *et al.* Stability of heterogeneous single-atom catalysts: a scaling law mapping thermodynamics to kinetics. *npj Comput. Mater.* **6**, (2020).
206. O'Connor, N. J., Jonayat, A. S. M., Janik, M. J. & Senftle, T. P. Interaction trends between single metal atoms and oxide supports identified with density functional theory and statistical learning. *Nat. Catal.* **1**, 531–539 (2018).
207. Lu, Z., Yadav, S. & Singh, C. V. Predicting aggregation energy for single atom bimetallic catalysts on clean and O∗ adsorbed surfaces through machine learning models. *Catal. Sci. Technol.* **10**, 86–98 (2020).
208. Sun, M. *et al.* Mapping of atomic catalyst on graphdiyne. *Nano Energy* **62**, 754–763 (2019).
209. Rao, K. K., Do, Q. K., Pham, K., Maiti, D. & Grabow, L. C. Extendable Machine Learning Model for the Stability of Single Atom Alloys. *Top. Catal.* **63**, 728–741 (2020).
210. Ha, M. *et al.* Tuning metal single atoms embedded in NxCy moieties toward high-performance electrocatalysis. *Energy Environ. Sci.* **14**, 3455–3468 (2021).
211. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
212. Di Liberto, G., Cipriano, L. A. & Pacchioni, G. Role of Dihydride and Dihydrogen Complexes in Hydrogen Evolution Reaction on Single-Atom Catalysts. *J. Am. Chem. Soc.* **143**, 20431–20441 (2021).
213. Liberto, G. Di, Cipriano, L. A. & Pacchioni, G. Universal Principles for the Rational Design of Single Atom Electrocatalysts ? Handle with Care. (2022) doi:10.1021/acscatal.2c01011.
214. Zhang, J., Yang, H. & Liu, B. Coordination Engineering of Single-Atom Catalysts for the Oxygen Reduction Reaction: A Review. *Adv. Energy Mater.* **11**, 1–20 (2021).
215. Li, X. *et al.* Microenvironment modulation of single-atom catalysts and their roles in

electrochemical energy conversion. *Sci. Adv.* **6**, 1–20 (2020).
216. Wang, Q. *et al.* Recent Advances in Strategies for Improving the Performance of CO 2 Reduction Reaction on Single Atom Catalysts . *Small Sci.* **1**, 2000028 (2021).
217. Li, L. *et al.* Recent Developments of Microenvironment Engineering of Single-Atom Catalysts for Oxygen Reduction toward Desired Activity and Selectivity. **2103857**, 41–47 (2021).
218. Li, J. *et al.* Highly Active and Stable Metal Single-Atom Catalysts Achieved by Strong Electronic Metal-Support Interactions. *J. Am. Chem. Soc.* **141**, 14515–14519 (2019).
219. Li, Z. *et al.* Metal-support interactions in designing noble metal-based catalysts for electrochemical CO2 reduction: Recent advances and future perspectives. *Nano Res.* **12**, (2021).
220. Mitchell, S. & Pérez-Ramírez, J. Single atom catalysis: a decade of stunning progress and the promise for a bright future. *Nat. Commun.* **11**, 10–12 (2020).
221. Jiang, Z. *et al.* Atomic interface effect of a single atom copper catalyst for enhanced oxygen reduction reactions. *Energy Environ. Sci.* **12**, 3508–3514 (2019).
222. Ju, W. *et al.* Unraveling Mechanistic Reaction Pathways of the Electrochemical CO2 Reduction on Fe-N-C Single-Site Catalysts. *ACS Energy Lett.* **4**, 1663–1671 (2019).
223. Fan, Q. *et al.* Electrochemical CO2 reduction to C2+ species: Heterogeneous electrocatalysts, reaction pathways, and optimization strategies. *Mater. Today Energy* **10**, 280–301 (2018).
224. Askins, E. J. *et al.* Toward a mechanistic understanding of electrocatalytic nanocarbon. *Nat. Commun.* **12**, 1–15 (2021).
225. Wang, F., Xie, W., Yang, L., Xie, D. & Lin, S. Revealing the importance of kinetics in N-coordinated dual-metal sites catalyzed oxygen reduction reaction. *J. Catal.* **396**, 215–223 (2021).
226. Cheng, T., Xiao, H. & Goddard, W. A. Reaction Mechanisms for the Electrochemical Reduction of CO2 to CO and Formate on the Cu(100) Surface at 298 K from Quantum Mechanics Free Energy Calculations with Explicit Water. *J. Am. Chem. Soc.* **138**, 13802–13805 (2016).
227. Back, S., Lim, J., Kim, N. Y., Kim, Y. H. & Jung, Y. Single-atom catalysts for CO2 electroreduction with significant activity and selectivity improvements. *Chem. Sci.* **8**,

1090–1096 (2017).
228. Zheng, T. *et al.* Large-Scale and Highly Selective CO 2 Electrocatalytic Reduction on Nickel Single-Atom Catalyst. *Joule* **3**, 265–278 (2019).
229. Song, W., Fu, L., He, C. & Xie, K. Carbon-Coordinated Single Cr Site for Efficient Electrocatalytic N2 Fixation. *Adv. Theory Simulations* **4**, 2100044 (2021).
230. Li, M. *et al.* Heterogeneous Single-Atom Catalysts for Electrochemical CO2 Reduction Reaction. *Adv. Mater.* **32**, 1–24 (2020).
231. Lu, J. *et al.* Scalable two-step annealing method for preparing ultra-high-density single-atom catalyst libraries. *Nat. Nanotechnol.* (2021).
232. Jiao, L. *et al.* Non-Bonding Interaction of Neighboring Fe and Ni Single-Atom Pairs on MOF-Derived N ‐ Doped Carbon for Enhanced CO 2 Electroreduction. (2021) doi:10.1021/jacs.1c08050.
233. Hunter, M. A., Fischer, J. M. T. A., Yuan, Q., Hankel, M. & Searles, D. J. Evaluating the Catalytic Efficiency of Paired, Single-Atom Catalysts for the Oxygen Reduction Reaction. *ACS Catal.* **9**, 7660–7667 (2019).
234. Guo, X. *et al.* Tackling the Activity and Selectivity Challenges of Electrocatalysts toward the Nitrogen Reduction Reaction via Atomically Dispersed Biatom Catalysts. *J. Am. Chem. Soc.* **142**, 5709–5721 (2020).
235. Doherty, F., Wang, H., Yang, M. & Goldsmith, B. R. Nanocluster and single-atom catalysts for thermocatalytic conversion of CO and CO2. *Catal. Sci. Technol.* **10**, 5772–5791 (2020).
236. Williams, T., McCullough, K. & Lauterbach, J. A. Enabling Catalyst Discovery through Machine Learning and High-Throughput Experimentation. *Chem. Mater.* **32**, 157–165 (2020).
237. Thakkar, A. *et al.* Artificial intelligence and automation in computer aided synthesis planning. *React. Chem. Eng.* **6**, 27–51 (2021).
238. Eyke, N. S., Koscher, B. A. & Jensen, K. F. Toward Machine Learning-Enhanced High-Throughput Experimentation. *Trends Chem.* **3**, 120–132 (2021).
239. Li, X. *et al.* Combining machine learning and high-throughput experimentation to discover photocatalytically active organic molecules. *Chem. Sci.* (2021) doi:10.1039/d1sc02150h.

240. Lo Dico, G., Nuñez, Á. P., Carcelén, V. & Haranczyk, M. Machine-learning-accelerated multimodal characterization and multiobjective design optimization of natural porous materials. *Chem. Sci.* **12**, 9309–9317 (2021).
241. Abbasi, K. *et al.* Dimensional Stacking for Machine Learning in ToF-SIMS Analysis of Heterostructures. (2020) doi:10.1002/admi.202001648.
242. Higgins, K. *et al.* Exploration of Electrochemical Reactions at Organic-Inorganic Halide Perovskite Interfaces via Machine Learning in In Situ Time-of-Flight Secondary Ion Mass Spectrometry. (2020) doi:10.1002/adfm.202001995.
243. Liu, L. & Corma, A. Metal Catalysts for Heterogeneous Catalysis: From Single Atoms to Nanoclusters and Nanoparticles. *Chem. Rev.* **118**, 4981–5079 (2018).
244. Shetty, M. *et al.* The Catalytic Mechanics of Dynamic Surfaces: Stimulating Methods for Promoting Catalytic Resonance. (2020) doi:10.1021/acscatal.0c03336.
245. Jing, H. *et al.* Electronics and coordination engineering of atomic cobalt trapped by oxygen-driven defects for efficient cathode in solar cells. *Nano Energy* **89**, 106365 (2021).
246. Wang, Y. *et al.* Regulating the coordination structure of metal single atoms for efficient electrocatalytic CO2reduction. *Energy Environ. Sci.* **13**, 4609–4624 (2020).
247. Kou, Z., Zang, W., Wang, P., Li, X. & Wang, J. Single atom catalysts: A surface heterocompound perspective. *Nanoscale Horizons* **5**, 757–764 (2020).
248. Shen, Y. *et al.* Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Prim.* **1**, (2021).
249. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
250. Huang, X. *et al.* Applying machine learning to balance performance and stability of high energy density materials. *iScience* **24**, 102240 (2021).
251. Lang, R. *et al.* Single-Atom Catalysts Based on the Metal-Oxide Interaction. *Chem. Rev.* **120**, 11986–12043 (2020).
252. Li, R. *et al.* Single atoms supported on metal oxides for energy catalysis. *J. Mater. Chem. A* **10**, 5717–5742 (2022).
253. Wang, D. *et al.* Accelerated prediction of Cu-based single-atom alloy catalysts for CO2 reduction by machine learning. *Green Energy Environ.* (2021) doi:10.1016/j.gee.2021.10.003.

**Table 1.** Summary of ML algorithms and their applications in SACs' design. List of abbreviations are presented in **Table 2**.

| # | Support/substrate | ML algorithms | Reaction | Purpose | Input features | Most important features | Year | Ref. |
|---|---|---|---|---|---|---|---|---|
| **1** | $CeO_2$, $TiO_2$, MgO, ZnO, $SeTiO_3$, $MoS_2$, and graphene | LASSO, elastic net, ridge | --- | stability | Ec, $Ec^{-1}$, $Ec^{0.5}$, $Ec^{-0.5}$, $Ec^2$, $Ec^{-2}$, ln(Ec), Eb, $Eb^{-1}$, $Eb^{0.5}$, $Eb^{-0.5}$, $Eb^2$, $Eb^{-2}$, ln(Eb), $Eb^2$/Ec | $(E_b)^2/E_c$ | 2020 | 205 |
| **2** | graphdiyne (bi-atom catalysts) | GPR | --- | optimal combination of metals for high stability | --- | potential f-d orbital coupling | 2021 | 33 |
| **3** | graphdiyne | FCM | --- | clustering the data | EA, EN, $Q_e$, $\varepsilon_d$, etc. | --- | 2019 | 208 |
| **4** | Cu, Ru, Rh, Pd, Ag, Re, Os, Ir, Pt, and Au | GKR, SVM, GPR | --- | aggregation energy and $\Delta G_{O^*}$ | $At_N$, $At_{wt}$, $At_{PN}$, $At_{GN}$, $At_R$, EN, IE, EA, $B_{01,O^*}$, etc. | $At_R$, EN, and $At_{GN}$ | 2020 | 207 |
| **5** | transition metals | DT, SVM, NN, hybrid KRR | --- | stability | $At_N$, $At_{wt}$, $At_{GN}$, $At_R$, $r_{cov}$, $P_{EN}$, $IE_1$, $E_f$, $d_e$, etc. | --- | 2020 | 209 |
| **6** | $N_xC_y$ | GNB, LR, KNN, radius neighbor classifier, support vector classifier, NN, DT, RFR, ETR, and GBR | HER, OER, and ORR | stability and activity | $At_N$, $IE_1$, etc. | --- | 2021 | 210 |
| **7** | graphene | KRR, RFR, NN, SISSO | HER | $\Delta G_{H^*}$ | $\varepsilon_d$, $r_{cov}$, q, $d_{unocc}$, $d_{occ}$, N, $r_d$, $E_f$, IE, EN | $d_{occ}$ and q | 2020 | 167 |
| **8** | graphene (dual atom catalysts) | RFR | ORR | $U_L$ | $M_1$-$M_2$, $M_{12}$-N, $At_R$, $N_{e,O}$, $P_{EN}$, $IE_1$, EA of two metals | $M_{12}$-N, $M_1$-$M_2$, and $N_{e,O}$ | 2020 | 175 |
| **9** | carbon | FCNN | OER | $\eta$ | $At_R$, $d_e$, EN, EA, and $IE_1$ | $d_e$, $At_R$, and EA | 2021 | 168 |
| **10** | g-$C_3N_4$ | GBR | OER and ORR | $\Delta G_{OH^*}$ | $\varepsilon_d$, $Q_e$, EN, EA, $IE_1$, $At_R$, and $d_e$, etc. | $IE_1$ and $Q_e$ | 2021 | 172 |
| **11** | graphene | RFR | HER, ORR, OER | $U_L$ | $d_e$, $H_{f,ox}$, $P_{EN}$, the sum of $P_{EN}$, etc. | $d_e$ | 2020 | 173 |
| **12** | 2D materials | LSBoost | HER and $N_2RR$ | $\Delta G$ | EN, EA, IE, and $d_{iso,e}$, etc. | $d_{iso,e}$,for NRR and EA for HER | 2021 | 31 |
| **13** | graphene | RFR and SVM | --- | $\Delta G_{H^*}$, $\Delta G_{OH^*}$, $\Delta G_{O^*}$, and $\Delta G_{OOH^*}$ | --- | adsorbate type | 2020 | 176 |
| **14** | 2D materials | RFR | ORR | $\Delta G_{O^*}$ | $H_{f,ox}$, $dp_e$, EN, EA, $IE_1$, $N_N$, $S_{EN}$, etc. | $H_{f,ox}$ and $dp_e$ | 2019 | 174 |
| **15** | graphene | NN | HER | EXAFS spectra | experimental EXAFS spectrum | --- | 2021 | 146 |
| **16** | g-$C_3N_4$, CN, and C2N | ETR method | $CO_2RR$ | $\Delta G_{OH^*}$ and $\Delta G_{OCH^*}$ | $At_N$, $d_e$, $At_R$, EN, $H_{vap}$, IE, and EA | $d_e$ and $H_{vap}$ | 2020 | 177 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **17** | transition metals | SVM, KRR, GBR, GPR, DTR, ETR, RFR, ABR, MLPR, KNR | $CO_2RR$ | $\Delta G_{CO*}$, $\Delta G_{CHO*}$, $\Delta G_{COOH*}$, $\Delta G_{HCOO*}$, and $\Delta G_{COH*}$ | EN, $N_e$, and ratio of EN and $N_e$ | ratio of EN and $N_e$ | 2020 | [192] |
| **18** | Au (111) | RFR | $N_2RR$ | $\Delta G_{N2*}$ | $At_R$, EN, EA, $At_{GN}$, $d_e$ | $At_{GN}$ | 2021 | [166] |
| **19** | MBenes | SVM | HER | $\Delta G_{H*}$ | q, $At_R$ of C, N, and B elements, molar ratio, $At_R$, and EA of metal | q | 2020 | [193] |
| **20** | MXenes | SVM, RFR, ANN, LASSO, KNN, Bayesian | HER | $\Delta G_{H*}$ and E | --- | Molar volume of surface element | 2021 | [194] |
| **21** | MBenes and 2D-materials | LGBM | $N_2RR$ | $\Delta G_{N2*}$ | --- | N-N bond length | 2021 | [196] |
| **22** | graphene | extreme GBR | $CO_2RR$ and HER | $\Delta G_{CO*}$ | --- | --- | 2020 | [198] |
| **23** | graphdiyne | Bag-tree algorithm | HER | $\Delta G_{H*}$ | --- | --- | 2020 | [195] |
| **24** | graphdiyne | DNN and LGBM | $N_2RR$ and HER | $\Delta G$ | EN, $At_N$, $At_R$, $N_N$, CN, etc. | CN | 2020 | [197] |
| **25** | $C_2N$, $C_1N_1$, and $C_1S_1$ | RFR | ORR and OER | $\Delta G_{O*}$ | $At_N$, $At_R$, $N_{e,O}$, EN, $IE_1$, EA, $S_{EN}$, $H_{f,ox}$ | $H_{f,ox}$ and $N_{e,O}$ | 2021 | [165] |
| **26** | Cu | GBR, SVM, RFR | $CO_2RR$ | $\Delta G_{CO*}$ | --- | $N_e$ | 2021 | [253] |

1

**Table 2. List of abbreviations for Table 1.**

| Abbreviation | Explanation | Abbreviation | Explanation |
|---|---|---|---|
| GPR | gaussian process regression | Ec, Eb | cohesive energy of bulk metals, binding energy |
| GKR | gaussian kernel regression | $At_N$, $At_{wt}$, $At_R$ | atomic number, atomic weight, atomic radius |
| GNB | gaussian naive bayes | $At_{PN}$, $At_{GN}$ | period number, group number |
| SVM | support vector machine | EN | electronegativity |
| LASSO | least absolute shrinkage and selection operator | $P_{EN}$ | Pauling electronegativity |
| SISSO | sure independence screening and sparsifying operator | $S_{EN}$ | sum of the electronegativity of coordinated atoms such as N and C |
| FCM | fuzzy C-Means | IE, $IE_1$ | ionization energy, first ionization energy |
| GBR | gradient boosting regression | EA | electron affinity |
| LGBM | light gradient boosting machine | $\varepsilon_d$ | d-states center |
| LR | logistic regression | $r_{cov}$ | covalent radius |
| KRR | kernel ridge regression | $r_d$ | Zunger radius |
| RFR | random forest regression | $N_{e,O}$ | outer electron number |
| ERT | extremely randomized trees | $d_{occ,\ e}$ | number of occupied d states |
| NN | neural network | $d_e$ | the electron numbers of d orbital |
| FCNN | full connection neural network | $d_{iso,\ e}$ | isolated electrons in d orbitals |
| DNN | deep neural network | $dp_e$ | adjusted electron numbers of d/p orbital |
| ANN | artificial neural network | $N_e$ | number of valance electrons |
| KNN | k-nearest neighbors | $E_f$ | formation energy of single atom site |
| LSBoost | least-squares boosting | $H_{f,\ ox}$ | oxide formation enthalpy |
| DT | decision tree | $H_{Vap}$ | enthalpy of vaporization |
| DTR | decision tree regression | q, $Q_e$ | Bader charge, charge transfer of metal atoms |
| ETR | extra tree regression | CN | coordination number |
| ABR | adaptive boost regression | $N_N$ | number of coordinated N atoms |
| TPOT | tree-based pipeline optimization tool | $M_1$-$M_2$ | the distance of two metal atoms |
| MLPR | multilayer perceptron regression | $M_{12}$-N | the average distance between two metal atoms and the coordinated N atoms |
| KNR | k-neighbor regression | | |
| SAC | Single atom catalyst | $\eta$ | overpotential |
| SAAC | Single atom alloy catalyst | $\Delta G$ | Gibbs free energies |
| | | E | Adsorption energies |
| | | $U_L$ | limiting potential |
| | | $V_{onset}$ | onset potential |

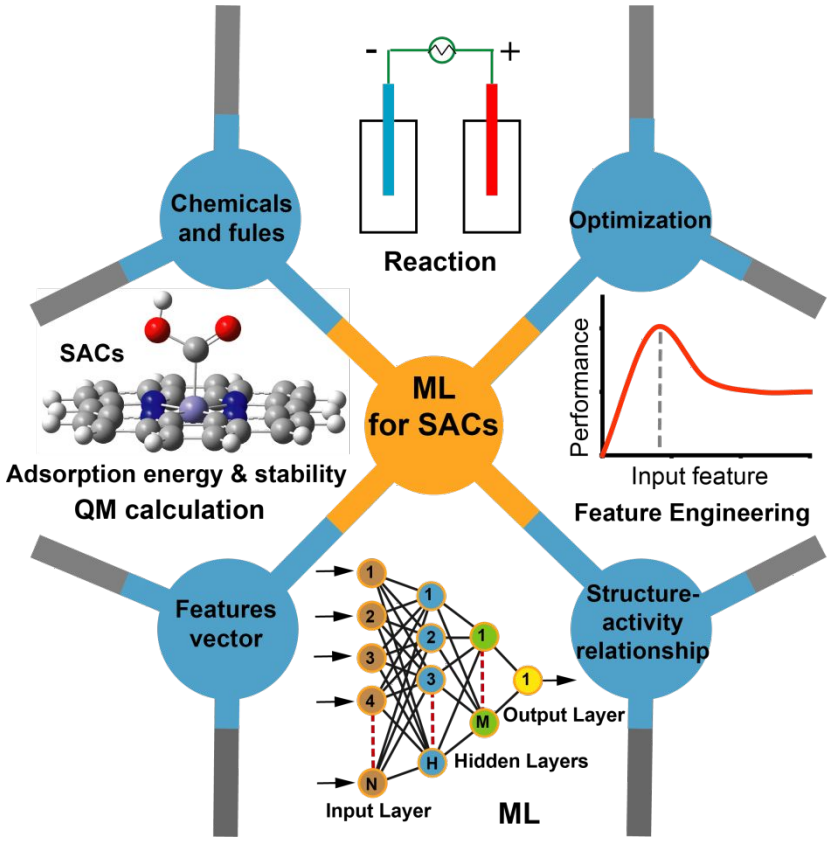


**Figure 1. The general workflow for the integration of QM calculations and ML for the rational design of heterogeneous catalysts.** The process contains several steps: data generation using QM calculations, training of ML, optimization, and feature importance analysis, and using designed catalysts to produce chemicals and fuels.

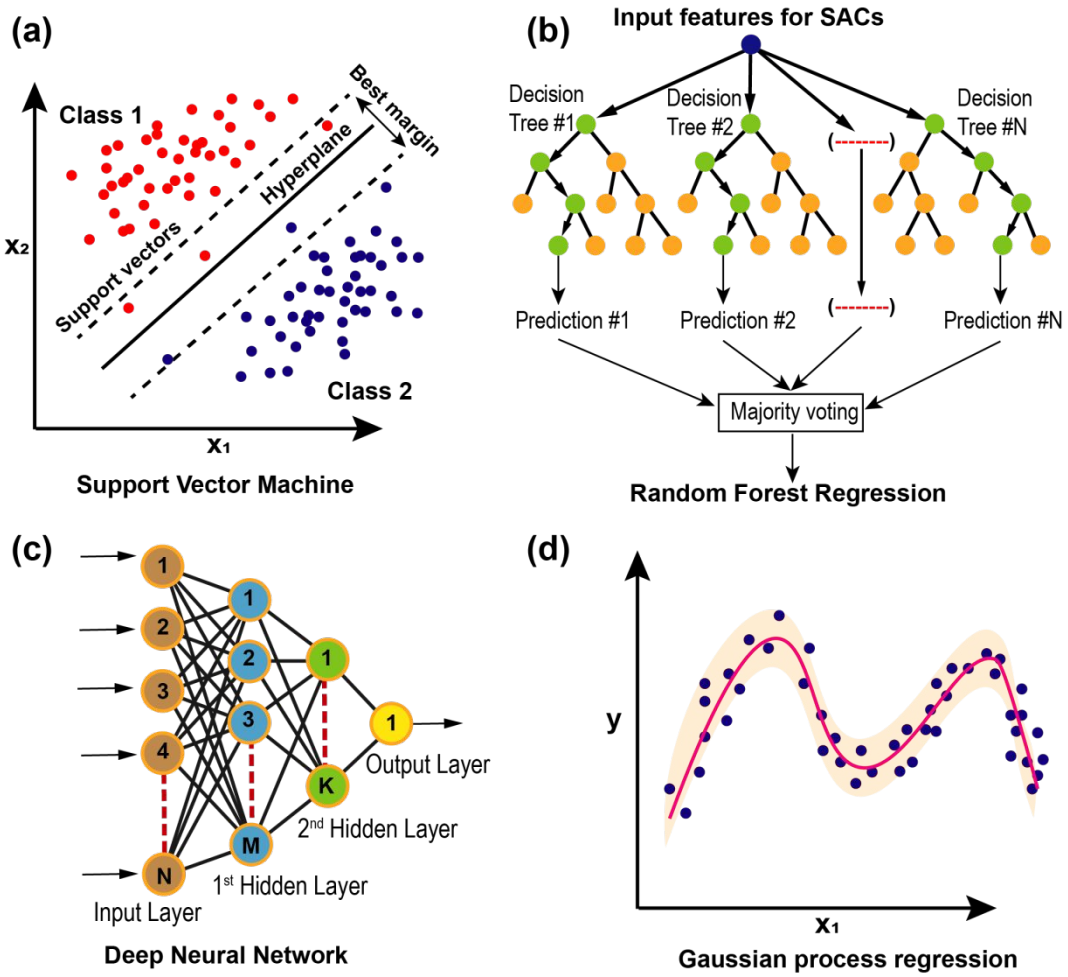


**Figure 2. Machine Learning Algorithms.** (a) Schematic of SVM algorithm. The hyperplane divides SACs into two distinct classes based on the largest distance of the data points placed between the support vectors. Class 1 and class 2 (red and blue circles) show the SACs with similar properties based on features $x_1$ and $x_2$. (b) Schematic of RFR. orange and green circles represent decision nodes containing 'if/then' statements. The result that is predicted by the highest number of decision trees (majority voting) is given as the output of the RFR. (c) Schematic of DNN. Circles are representing neurons in the input, hidden, and output layers of the DNN. Neurons are interconnected using the black lines. (d) Schematic of GPR algorithm. Predicted mean (red line) and confidence interval (light orange interval) for GPR algorithm trained based on input dataset (blue dots).

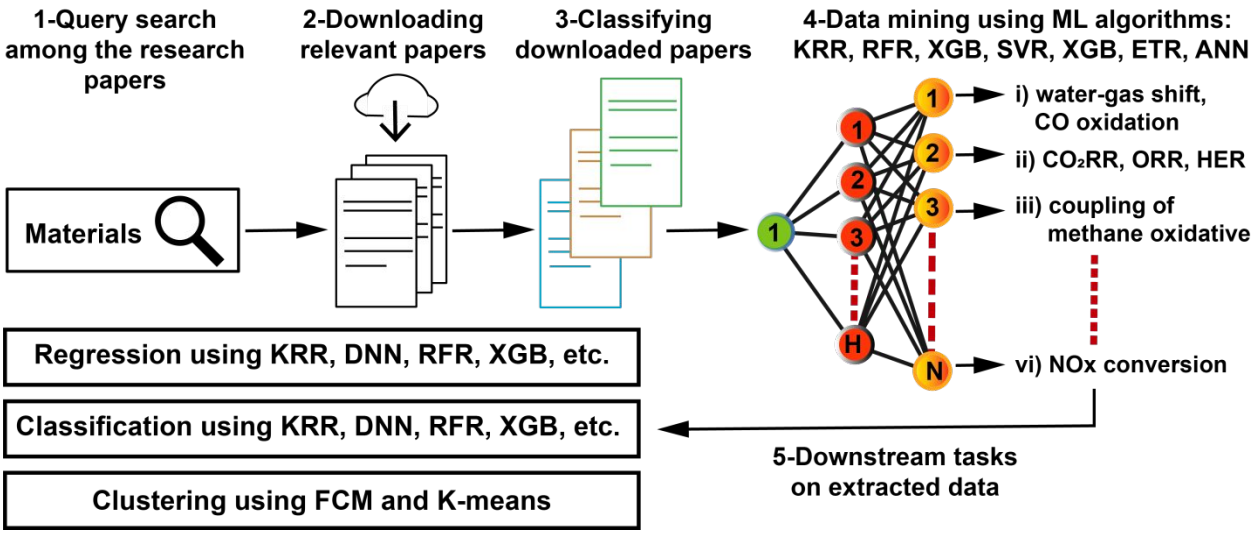


**Figure 3. The workflow for the data mining from literature.** Summary of data miming sequence from literature using several ML algorithms such as KRR, RFR, XGB, SVR, XGB, ETR, and ANN.

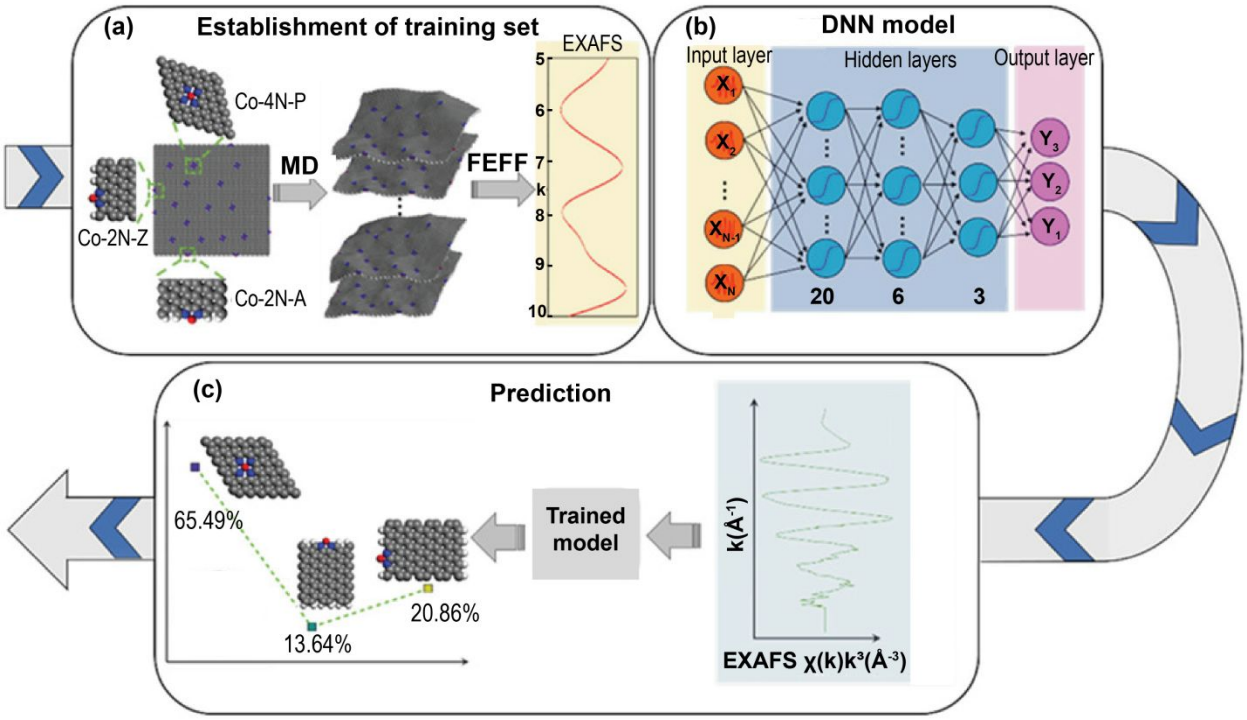


**Figure 4. ML for the interpretation of the EXAFS of Co-N doped grapheme.** (a) establishment of training data using MD-EXAFS calculations for Co-4N-P, Co-2N-A, and Co-2N-Z. (b) the architecture of the DNN composed of one input layer of the EXAFS spectrum, two hidden layers, and one output layer of the proportion vector. (c) The estimation of local structural proportion from the experimental EXAFS measurement. Reproduced with permission from ref. [146], copyright 2021, Wiley-VCH. Results show that ML is an appropriate and powerful tool for the interpretation of EXAFS.

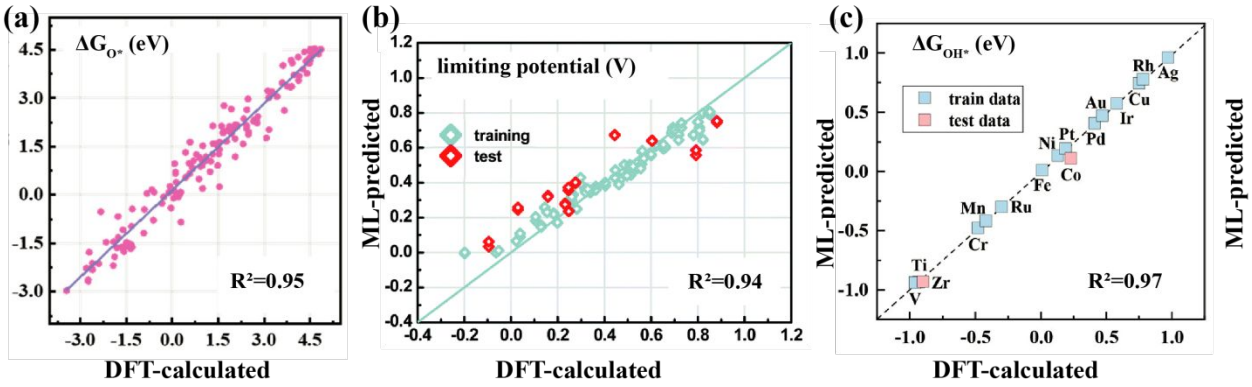


**Figure 5. Density functional theory (DFT)-based Machine Learning.** Comparison of ML- and DFT-predicted (a) $\Delta G_{O*}$ using RFR algorithm, Reproduced with permission from ref. [174], copyright 2019, American Chemical Society. (b) limiting potentials using RFR algorithm, Reproduced with permission from ref. [175], copyright 2020, Royal Society of Chemistry. (c) $\Delta G_{OH*}$ using GBR algorithm. Reproduced with permission from ref. [172], copyright 2021, American Chemical Society. Results indicate that ML can be used for the out-of-sample (test set) predictions of activity for SACs using the deep structure-activity relationships. However, the quantity of training dataset is not enough for having a generalized ML algorithm.

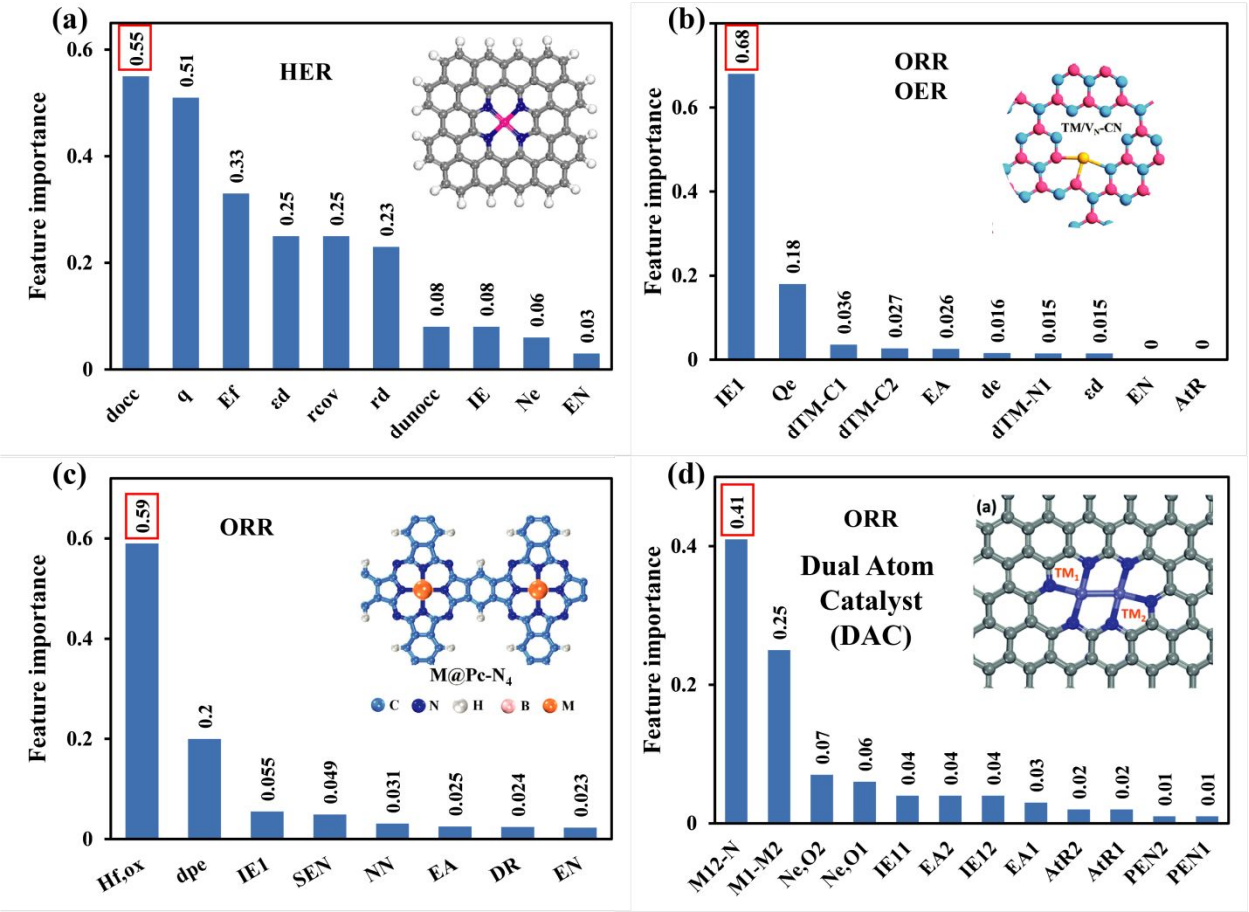


**Figure 6. Feature importance analysis.** (a) The feature importance for SACs embedded in nitrogen-doped graphene indicating that number of occupied d states ($d_{occ}$) and Bader charge (q) are the most important parameters for HER. Please note that this is our evaluation on ref. [167]. Reproduced with permission from ref. [167], copyright 2020, American Chemical Society. (b) The feature importance based on the GBR algorithm for rhodium SACs. Reproduced with permission from ref. [172], copyright 2021, American Chemical Society. First ionization energy ($IE_1$) and the charge transfer of TM atoms ($Q_e$) are the most important factors on the $\Delta G_{OH*}$. Inset shows the structure of rhodium SACs on defective g-$C_3N_4$ for OER and ORR. (c) The feature importance based on the RFR algorithm for SACs embedded on nitrogen-doped carbon supports. Reproduced with permission from ref. [174], copyright 2019, American Chemical Society. The oxide formation enthalpy ($H_{f,\ ox}$) and the adjusted electron numbers of d/p orbital ($dp_e$) are the most important factors on the $\Delta G_{O*}$. Inset shows the structure of SACs embedded on nitrogen-doped carbon supports for two-electron ORR. (d) The feature importance for dual atom catalysts (DACs) based on a RFR algorithm indicating that the average distance between metal atoms and the coordinated

N atoms ($M_{12}$–N), the distance between two metal atoms ($M_1$–$M_2$), and the outer electron number of metal atoms ($d_e$) are the most important factors on the ORR limiting potentials. Reproduced with permission from ref. [175], copyright 2020, Royal Society of Chemistry. Inset shows the structure of DACs embedded in nitrogen-doped graphene for ORR. Results indicate that feature engineering of SACs and DACs depends on the application and the type of substrate. Please see **Table 2** for the abbreviations.

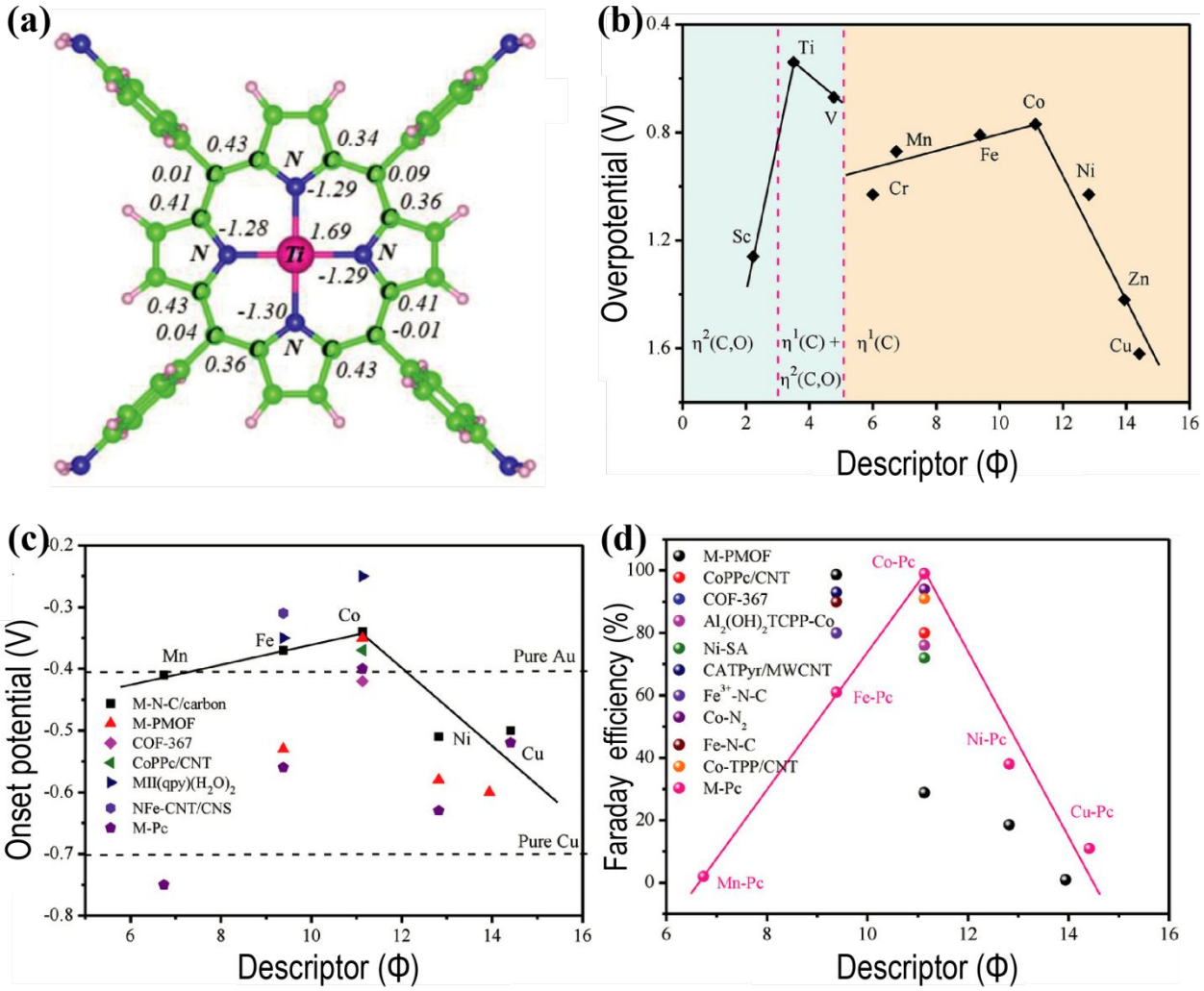


**Figure 7. Volcano plots.** (a) Structure of SACs embedded in nitrogen-doped graphene supports for the descriptor-based SACs design. Volcano plots for (b) overpotential (η), (c) onset potential ($V_{onset}$), and (d) Faraday efficiency (FE) based on the descriptor for SACs embedded in nitrogen-doped graphene supports. This indicates two definitive volcanoes in the plot for overpotential with Ti and Co located at the summits. Also, for the onset potential and Faraday efficiency, Co is in the summit of volcanoes with better CO2RR performance. Reproduced with permission from ref.[180], copyright 2019, Wiley-VCH.