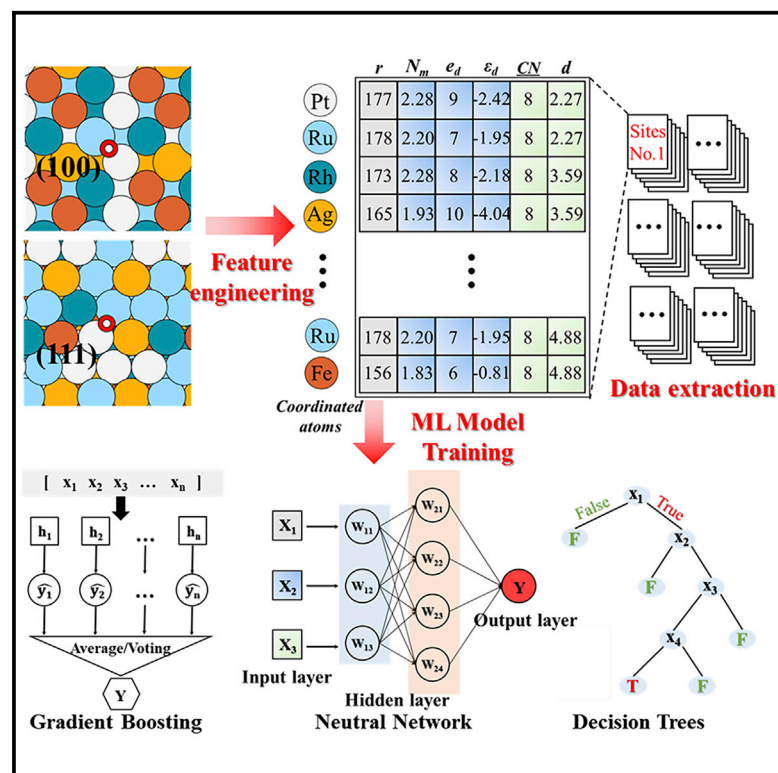


# Patterns

## Machine-learning-assisted discovery of highly efficient high-entropy alloy catalysts for the oxygen reduction reaction

### Graphical abstract



### Authors

Xuhao Wan, Zhaofu Zhang, Wei Yu, Huan Niu, Xiting Wang, Yuzheng Guo

### Correspondence

yguo@whu.edu.cn

### In brief

High-entropy alloys have great potential to become optimal heterogeneous catalysts due to vast chemical space. With the aid of machine learning, the catalytic performance of millions of reactive sites on HEA surfaces can be explored theoretically to overcome the disadvantages of traditional trial-and-error experiments. Our method offers rational guidance for the design of highly efficient HEA catalysts on component elements and composition ratio.

### Highlights

- The catalytic activity of six types of high-entropy alloys was studied theoretically
- Machine learning shows great potential to tackle the huge chemical space of HEAs
- The well-trained model can accurately predict catalytic performance of HEAs
- A strategy to improve catalytic activity by tuning HEA compositions was proposed



## Article

# Machine-learning-assisted discovery of highly efficient high-entropy alloy catalysts for the oxygen reduction reaction

Xuhao Wan,<sup>1,4</sup> Zhaofu Zhang,<sup>2,3,4</sup> Wei Yu,<sup>1</sup> Huan Niu,<sup>1</sup> Xiting Wang,<sup>1</sup> and Yuzheng Guo<sup>1,5,\*</sup>

<sup>1</sup>School of Electrical Engineering, Wuhan University, Wuhan, Hubei 430072, China

<sup>2</sup>The Institute of Technological Sciences, Wuhan University, Wuhan, Hubei 430072, China

<sup>3</sup>Department of Engineering, Cambridge University, Cambridge CB2 1PZ, UK

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: [yguo@whu.edu.cn](mailto:yguo@whu.edu.cn)

<https://doi.org/10.1016/j.patter.2022.100553>

**THE BIGGER PICTURE** Benefiting from huge chemical space, high-entropy alloys (HEAs) show great potential as heterogeneous catalysts for different reactions. However, vast chemical space makes it extremely difficult to comprehensively study HEAs by traditional trial-and-error experiments. Therefore, a machine-learning-assisted theoretical method is proposed to investigate the oxygen reduction reaction (ORR) catalytic activity of millions of reactive sites on HEA surfaces. The well-performed gradient boosting regression (GBR) model with high accuracy, generalizability, and simplicity is constructed by reasonable data extraction and feature engineering, which can accurately predict the catalytic activities of millions of reactive sites on HEA surfaces. Finally, one strategy to engineer the HEA surface structure by tuning the metal element component ratio is proposed, which doubles the amount of high-activity sites.



**Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

## SUMMARY

High-entropy alloys (HEAs) have recently been applied in the field of heterogeneous catalysis benefiting from vast chemical space. However, huge chemical space also brings extreme challenges for the comprehensive study of HEAs by traditional trial-and-error experiments. Therefore, the machine learning (ML) method is presented to investigate the oxygen reduction reaction (ORR) catalytic activity of millions of reactive sites on HEA surfaces. The well-performed ML model is constructed based on the gradient boosting regression (GBR) algorithm with high accuracy, generalizability, and simplicity. In-depth analysis of the results demonstrates that adsorption energy is a mixture of the individual contributions of coordinated metal atoms near the reactive site. An efficient strategy is proposed to further boost the ORR catalytic activity of promising HEA catalysts by optimizing the HEA surface structure, which recommends a highly efficient HEA catalyst of Ir<sub>48</sub>Pt<sub>74</sub>Ru<sub>30</sub>Rh<sub>30</sub>Ag<sub>74</sub>. Our work offers a guide to the rational design and nanostructure synthesis of HEA catalysts.

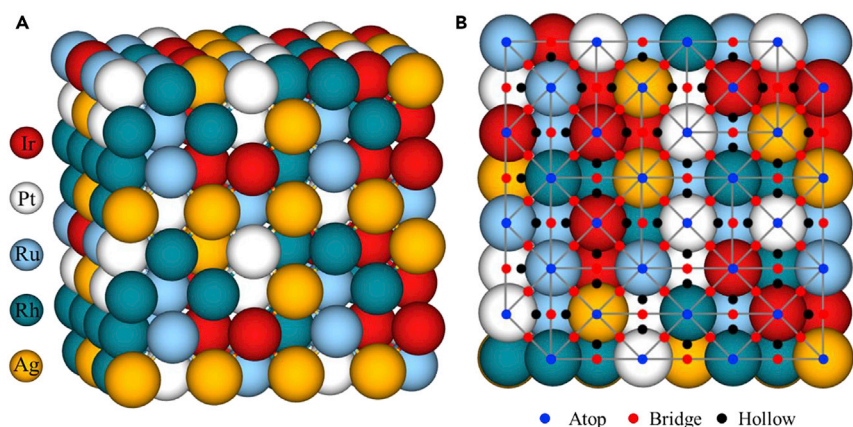
## INTRODUCTION

High-entropy alloys (HEAs) are multi-principal-component alloys that consist of five or more elements, with each element at near-equimolar proportion, which commonly are the complex solid solution.<sup>1–5</sup> HEAs are stable as the number of elements species and completely disordered atom positions make the system en-

trophy higher. The alloys have attracted the world's attention after first being discovered in 2004,<sup>1</sup> benefiting from their easily tunable mechanical property such as elasticity modulus, hardness, and strength of extension.<sup>6,7</sup>

Recently, HEAs have been applied in the field of catalysis as their huge chemical space from the enormous number of elements combinations and inherent surface complexity make it





**Figure 1. The typical structure and reactive sites of HEA IrPtRuRhAg**

(A) The geometric structure of the equimolar IrPtRuRhAg HEA with an fcc crystal configuration. (B) The schematic diagram of finding all possible sites on HEA surface, including atop (blue), bridge (red), and hollow (black) by the Delaunay triangulate algorithm.<sup>43</sup>

possible to achieve higher activity, selectivity, and stability as catalysts.<sup>8–10</sup> Thus far, HEAs have been studied as highly efficient catalysts in many different reactions, including hydrogen evolution reaction (HER),<sup>11</sup> oxygen reduction reaction (ORR),<sup>12–14</sup> carbon dioxide reduction reaction (CO<sub>2</sub>RR),<sup>8,15</sup> and methanol oxidation,<sup>16,17</sup> which are mostly experimental studies.

In the past several decades, theoretical simulations have been widely used to better understand the catalytic processes and directly design highly efficient catalysts before experiments.<sup>18–20</sup> The huge configuration space of HEAs can provide a surface with a very large number of unique reactive sites. However, millions of different active-site environments of HEAs and current limited computing implementation ability extremely complicate the theoretical research on the catalytic performance of all of the sites by only applying the traditional density functional theory (DFT) method. To break the bottleneck of computing power, one of the most reasonable strategies is to discover new algorithms or improve original algorithms, which can remarkably reduce total computational cost. As a result, machine learning (ML) attracts attention around the globe, as it is very helpful in these two directions.<sup>21–23</sup> The ML method can dramatically reduce the computational cost of the traditional DFT method and it can maintain high accuracy when predicting the catalytic activities. In addition, ML methods can reveal the intrinsic descriptor of catalytic reactions by elucidating the nonlinear relationship between the structure and properties of materials.<sup>24,25</sup> As a result, the state-of-the-art ML-assisted theoretical computations method has become a rising star in the catalysis field.<sup>26,27</sup> In 2015, Ma et al. realized high-throughput screening of highly efficient CO<sub>2</sub> electroreduction catalysts by combining ML and theoretical calculation.<sup>28</sup> Many researchers have developed and spread the ML-assisted theoretical method to complicated chemical systems and different reactions such as MXene ordered binary alloy for HER,<sup>29</sup> dual-metal-site catalysts (DMSCs) for ORR,<sup>30</sup> and perovskite for oxygen evolution reactions (OER).<sup>31</sup> ML has also been widely applied in many studies on HEAs such as searching for HEAs with large hardness<sup>32</sup> and phase prediction of HEAs.<sup>33</sup> This significant research inspires us to explore and predict the catalytic performance of HEAs by the advanced ML-assisted theoretical method.

In this work, the stability of six types of quinary HEAs was investigated at first. Then, the ORR volcano curve between the ORR catalytic activity of reactive sites on HEA surfaces and

OH\* absorption energies was established. The excellent ML model was constructed by reasonable data extraction, feature engineering, and model validation processes to predict the OH\* absorption energies of millions of reactive sites on different crystal facets of HEAs with high accuracy. Finally, the predicted results and the ML models were further analyzed to find highly efficient HEA catalysts and reveal their ORR activity origin.

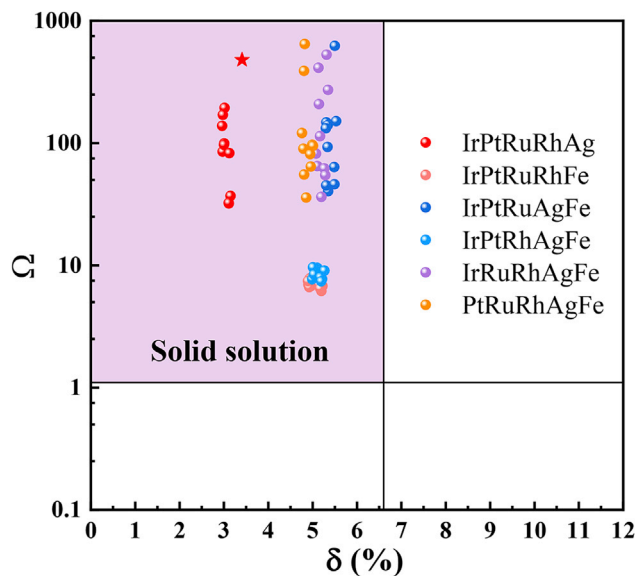
## RESULTS AND DISCUSSION

### The components and stabilities of HEAs

Six transition metal (TM) elements are under consideration as the constituent elements of quinary HEAs—Ir, Pt, Ru, Rh, Ag, and Fe. They can form six types of HEAs. These elements are chosen because of a similar atomic radius and close lattice constant with the same crystal structure of face-centered cubic (fcc), as listed in Table S1. In addition, their corresponding elementary metals have been reported as being highly efficient catalysts for ORR.<sup>34–39</sup> The HEAs of fcc are constructed to study in this work because six selected elements mostly have fcc structures. The constructed HEA IrPtRuRhAg are shown in Figure 1A as an example, with 256 atoms in the supercell. There are several HEAs with a similar structure as above that had been synthesized and show promising catalytic activities.<sup>40–42</sup>

The high disorder resulting from the completely random configuration space of HEAs can increase the mixed configuration entropy of HEAs, which is beneficial for the formation of a stable single-phase solid solution structure rather than fragile intermetallic compounds. Whether HEAs can form stable solid solutions is closely related to some thermodynamic properties of HEAs such as atomic radius, energy differences, and configurational mixing entropy. According to the Hume-Rothery rules, the difference between atom radii and the ratio of formation enthalpy to entropy can evaluate the stability in a straight way.<sup>44–46</sup> Hereafter, the atom radii difference factor ( $\delta$ ) and the ratio of formation enthalpy to entropy ( $\Omega$ ) are used to describe the phase of HEAs. The detailed computation methods of the  $\delta$  and  $\Omega$  parameters are listed in Note S1.

Instead of simulating a large number of metal atoms in a whole crystal lattice, we calculated smaller 32-atom supercells with periodic boundary conditions, considering the huge computation cost of *ab initio* calculations of so many extremely disordered metal atoms. The approximation of a supercell had a negligible influence on the DFT-calculated values, as shown in Note S2 and Figure S1. In addition, for each type of HEA, 10 different HEAs with different structure and composition ratios of elements that are adjusted in the range of  $\pm 15\%$  are studied to evaluate the



**Figure 2. The distribution of  $\delta$  and  $\Omega$  parameters of HEAs**

The HEAs within the purple area are more likely to form a solid solution. Six types of quinary HEAs are studied, and 10 different component ratios are considered for each type of HEA. The highly efficient ORR catalyst  $\text{Ir}_{48}\text{Pt}_{74}\text{Ru}_{30}\text{Rh}_{30}\text{Ag}_{74}$  is marked by a red star.

stability of HEAs more reliably and thoroughly. That is, 60 HEAs with different structures are studied. The detailed component and ratios of 60 HEAs are listed in Table S2. As shown in Figure 2, all 60 HEAs studied fall within the solid solution phase area, the criteria of which are defined as that the parameter  $\delta$  is smaller than 6.6% and  $\Omega$  is larger than 1.1 simultaneously.<sup>46</sup> The HEAs matching the requirements are quite likely to form a single solid solution owing to the subtle lattice structure and mixing energy.<sup>45</sup>

### The ORR activity volcano of HEA surfaces

After confirming the stabilities of HEAs, the possible active sites for ORR were studied. As shown in Figure 1B, the Delaunay triangulate algorithm<sup>43</sup> was applied to find all of the possible sites on the HEA surface, including atop, bridge, and hollow sites for this high-throughput research. Stability analysis of reactive intermediates on the HEA surface demonstrated that the intermediate adsorbates are mostly unstable on atop and hollow sites, but only stable on bridge sites for different HEA surfaces, which is consistent with previous experimental studies.<sup>47,48</sup> Therefore, only bridge sites were considered as possible reactive sites on two fcc-structured HEA Miller index surfaces, (100) and (111), as they are the most common surfaces.<sup>14</sup>

Then, we focused on the catalytic activities for ORR of different reactive sites on HEA surfaces. Based on the Sabatier rule,<sup>49</sup> the absorption energies of reaction intermediates are generally good descriptors for catalytic activity. Accordingly, the volcano curve between adsorption energies of intermediates and catalytic activities for ORR can be established, which makes it simple to explore the activities of millions of possible active sites on HEA surfaces. To obtain a comprehensive understanding of the ORR catalytic activity on HEA, the intermediates ( $\text{OH}^*$ ,

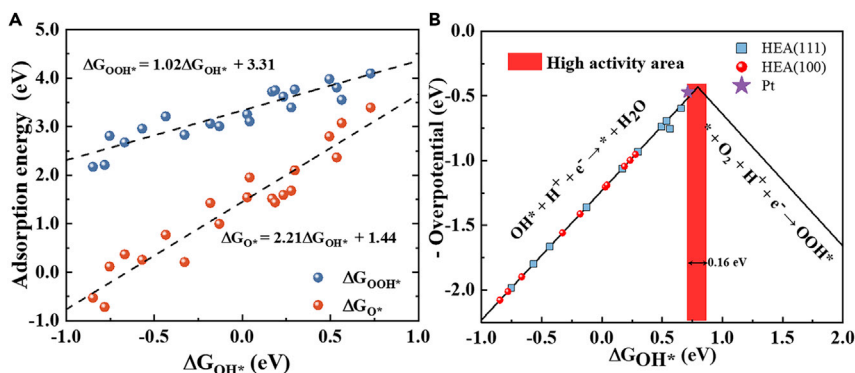
$\text{O}^*$ , and  $\text{OOH}^*$ ) involved in the reaction process of 20 different sites on different HEA surfaces experienced systematic and adequate DFT optimizations. The 20 sites were equally and randomly selected from different bridge sites and Miller index surfaces to guarantee the reliability of the results.

The previous research demonstrates similar TM–O interactions between TM and ORR intermediates ( $\text{OH}$ ,  $\text{O}$ , and  $\text{OOH}$ ) will lead to similar adsorption behaviors on metal surfaces,<sup>50</sup> meaning the adsorption energies of  $\text{OH}$ ,  $\text{O}$ , and  $\text{OOH}$  on metal surfaces are likely to have a similar trend. As shown in Figure 3A, the results show that the adsorption energies of  $\text{OH}$ ,  $\text{O}$ , and  $\text{OOH}$  on HEA surfaces possess apparent scaling relations, as discussed above. Specifically, the relationship between  $\Delta G_{\text{O}^*}$  and  $\Delta G_{\text{OH}^*}$  can be expressed as  $\Delta G_{\text{O}^*} = 2.21 \cdot \Delta G_{\text{OH}^*} + 1.44$ , with a high coefficient of determination ( $R^2$ ) of 0.91, while  $\Delta G_{\text{OOH}^*}$  and  $\Delta G_{\text{OH}^*}$  are  $\Delta G_{\text{OOH}^*} = 1.02 \cdot \Delta G_{\text{OH}^*} + 3.31$ , with a  $R^2$  of 0.85. Based on the aforementioned scaling relationship, the volcano curve was established by choosing  $\Delta G_{\text{OH}^*}$  as the descriptor for activity, as illustrated in Figure 3B. According to the Sabatier rule, both the too strong and too weak adsorption of reaction intermediates on surfaces harm catalytic performance; because too strong adsorption prevents the desorption process and makes the catalysts poisoned, whereas too weak adsorption impedes the activation of intermediates. Therefore, the sites with the best catalytic activity are the peak of the volcano, where the adsorption energies of intermediates are moderate. As Figure 3B shows, the chosen reactive sites all lie to the left of the volcano, and the reference point of Pt (111) is very close to the peak. Our motivation then changed to find those sites on HEA surfaces with higher activity than Pt (111), namely to cross over Pt (111) and achieve the peak of the ORR volcano ( $\sim 0.16$  eV weaker adsorption energy than Pt (111)).

### ML process

Although the volcano curve properly bridges the catalytic activity and the adsorption energies of  $\text{OH}^*$ , it is almost impossible to calculate the adsorption energies of millions of reactive sites on HEA. Therefore, ML was introduced to link the local atomic environment around the sites with the adsorbate strength, which makes it feasible to discover the activity of millions of sites on HEAs. At first, 360 reactive sites on HEAs with different Miller index surfaces and component elements were randomly selected, and the adsorption energies of  $\text{OH}^*$  intermediates on them were obtained by DFT calculations to construct the original dataset for building ML models. The 360 sites can be divided into 12 groups based on 2 different Miller index surfaces and 6 different quinary HEAs; there are 30 sites for each group. The specific types of 12 groups and the specific positions of 360 sites can be found in Note S3 and Figure S2.

Then, some easily obtainable physical and chemical properties were selected to reasonably describe the local atomic environment of reactive sites by feature engineering. It is the most significant step as it determines the highest accuracy of the ML model.<sup>22,27</sup> We abstracted the science problem of describing the local environment of sites into generating the dataset, which was characteristics of coordinated atoms of two metal atoms that make up bridge sites, and the characteristics were reasonable physical and chemical features of each coordinated atom. There are several common strategies to determine



**Figure 3. The volcano curve between  $\Delta G_{OH^*}$  and overpotential**

(A) Scaling relations between the adsorption energies of reaction intermediates ( $\Delta G_{O^*}$  versus  $\Delta G_{OH^*}$  in red, while  $\Delta G_{OOH^*}$  versus  $\Delta G_{OH^*}$  in blue) on HEA (100) and (111) surfaces.

(B) The ORR volcano curve of reactive sites on HEA surfaces between overpotential and  $\Delta G_{OH^*}$ . The Pt (111) is marked as a purple star as the reference point. The high activity area is marked in red.

features: They should comprehensively describe the atomic, electronic structures, and the local environment of reactive sites, and they should be physically intuitive to guarantee the model's robustness.<sup>22,27</sup> Based on these strategies, prior knowledge, and previous studies,<sup>21,51</sup> 11 features were selected: atomic radius ( $r$ ), relative mass ( $M$ ), atomic number ( $AN$ ), Pauli electronegativity ( $N_m$ ), electron number of d orbital ( $e_d$ ), d-band center of the corresponding pure metal surface ( $\varepsilon_d$ ), electron affinity ( $A_m$ ), first ionization energy ( $I_m$ ), oxide formation enthalpy ( $H_f$ ), generalized coordination number ( $CN$ ) of the coordinated atoms, and the distance between the coordinated atoms and  $OH^*$  ( $d$ ). Among these features,  $r$ ,  $M$ , and  $AN$  are atomic structure features;  $N_m$ ,  $e_d$ ,  $\varepsilon_d$ ,  $A_m$ ,  $I_m$ , and  $H_f$  are electronic structure features;  $CN$  and  $d$  are local environment features. Considering that wrapper strategy is time-consuming and embedded strategy is only applicable for specific algorithms, filter strategy was chosen to determine final features. The standard of the filter strategy was that the final features should be relatively independent and have larger variance. In addition, as the features that describe the local environment are simple enough and they are not relative, only atomic features and electronic structure features are filtered.

First, the Pearson matrix between the features was calculated as shown in Figure S3. The results indicate that all of the atomic features have a strong correlation. There are also two groups of relevant electronic structure features. The first group are  $N_m$ ,  $A_m$ , and  $I_m$ , while the second are  $\varepsilon_d$  and  $H_f$ . Then, the variance of these relevant features was obtained and painted in three colors at the bottom of Figure S3. Among the three groups,  $r$ ,  $N_m$ , and  $\varepsilon_d$  show the highest variance, so other features are excluded as the features with higher variance are easier to learn by ML models. It should be noticed that the variance is from the features after normalization; therefore, their absolute values are relatively small. In general, six features are selected as the final input features, including  $r$ ,  $N_m$ ,  $e_d$ ,  $\varepsilon_d$ ,  $CN$ , and  $d$ , as listed in Table S3. It means the whole input dataset is a  $[360 * n * 6]$  matrix, where  $n$  is 15 for sites on HEA (111) and 14 for HEA (100) surface, as shown in Figure S4 and Table 1. The values of the  $r$ ,  $M$ ,  $N_m$ ,  $A_m$ ,  $I_m$ , and  $e_d$  come from the PTable database.<sup>52</sup>

After completing the process of feature engineering and data extraction, the pre-processed dataset was used to train seven different ML models with seven different regression algorithms, including gradient boosted regression (GBR),<sup>53</sup> feedforward neural network (FNN),<sup>54</sup> and Random forest regression (RFR),<sup>55</sup> sup-

port vector regression (SVR),<sup>56</sup> k-neighbor regression (KNN),<sup>57</sup> kernel ridge regression (KRR),<sup>58</sup> and least absolute shrinkage and selection operator regression (LASSO),<sup>59</sup> as illustrated in Figure 4. Then, the first round of model selection was performed by simply tuning the hyperparameters of 7 ML models and comparing model metrics on the test set after 4-fold cross-validation. The results were demonstrated in Figure S5, and it indicated that the model performance of the GBR and RFR models were evidently better than the other 5 models as they have the lowest root-mean-square error (RMSE) (both  $<0.15$ ) and the highest  $R^2$  score (both  $>0.95$ ). The model performance of FNN was relatively bad, but it was more likely to achieve better performance after a comprehensive parameters-tuning process. As a result, the GBR, RFR, and FNN algorithms were selected to establish better ML models. Hereafter, 500-time repeated 4-fold cross-validation was applied to evaluate model performance so as to reduce the random effect of dataset split and the risk of overfitting.<sup>27</sup> Table S4 indicates that 500 times is enough to eliminate the sampling error and guarantee the generalization ability of ML models. By manually adjusting the hyperparameters over and over again, the GBR, FNN, and RFR models achieved the best of themselves, and the hyperparameters were tuned as listed in Table S5. A more detailed flow chart is illustrated in Figure S6.

The results of model training and testing are demonstrated in Figure 5. As shown in Figures 5A and 5B, the GBR and RFR models both performed well on the train and test set with lower RMSE ( $\sim 0.1$  eV) and higher  $R^2$  score ( $\sim 0.95$ ). The average metrics of 500-time repeated 4-fold cross-validation on train and test set were close, indicating no risk of overfitting. In addition, the error bar of the two models was relatively short, suggesting good model robustness. As for the well-known FNN model, the model performance was bad and the error bar was long, showing the unstable character of the FNN model. The poor accuracy of the FNN model can be attributed to the limited size of the original dataset from DFT calculations and the difficult hyperparameter tuning process. For the well-performed GBR and RFR models, the metrics on the test set of the GBR model were better with lower RMSE (0.112 eV) and higher  $R^2$  score (0.961) than RFR (RMSE of 0.129 eV and  $R^2$  of 0.948). Therefore, the GBR model was the optimal ML model that can exactly describe the underlying pattern of the local atomic environment and the ORR catalytic activities of reactive sites on HEA surfaces.

The parity plot of the GBR model applied in the whole dataset is clearly shown in Figure 5C, and the GBR-predicted absorption energies are consistent with those energies from DFT calculations. Nearly all of the dots fell into the  $\pm 0.2$  eV deviation area,

**Table 1. An example table of the input data to establish ML models of 5 reactive sites on HEAs surfaces**

Sites	A <sub>1</sub>	A <sub>2</sub>	A <sub>n</sub>	A <sub>15</sub>	ΔG <sub>OH*</sub> (eV)
1	[180, 2.20, 7, -2.25, 8, 2.28]	[177, 2.28, 9, -2.42, 8, 2.28]	...	[0, 0, 0, 0, 0, 0]	0.0373
2	[156, 7.83, 6, -0.81, 8, 2.40]	[180, 2.20, 7, -2.25, 8, 2.40]		[0, 0, 0, 0, 0, 0]	-0.6672
3	[173, 2.28, 8, -2.18, 9, 2.22]	[177, 2.28, 9, -2.42, 9, 2.22]		[177, 2.28, 9, -2.42, 9, 4.91]	-0.5263
4	[178, 2.20, 7, -1.95, 9, 2.22]	[180, 2.20, 7, -2.25, 9, 2.22]		[173, 2.28, 8, -2.18, 9, 4.88]	0.1562
5	[156, 7.83, 6, -0.81, 9, 2.23]	[173, 2.28, 8, -2.18, 9, 2.23]		[180, 2.20, 7, -2.25, 9, 4.88]	-0.6126

A<sub>n</sub> denotes the coordinated metal atoms in the order of distance. The values in the square brackets correspond to  $r$ ,  $N_m$ ,  $e_d$ ,  $\epsilon_d$ ,  $CN$ , and  $d$ . The first 2 sites are on HEA (100), while the latter 3 are on HEA (111). The data of A<sub>15</sub> atom are empty for sites 1 and 2 because the number of coordinated atoms for sites on HEA (100) is only 14.

indicating the effective-trained GBR model achieves high accuracy in prediction via learning significant information about the underlying pattern of the local environment and adsorbate strength on HEA surfaces. The learning curve of the GBR model was plotted in Figure 5D, suggesting that there was no risk of overfitting as the RMSE on the train and test set gradually converged with the increase in training size and the convergence values were close to one another. The high efficiency of the ML-assisted method is shown in Note S4.

### The ORR activity of HEA surfaces

The well-performed GBR model empowered us to reliably obtain accurate absorption energies of reactive sites, thereby exploring catalytic activities of HEAs from the huge chemical space. Hereafter, 12,000 different bridge sites on HEAs with different Miller index surfaces and component elements were randomly and equally generated, which means there were 2,000 sites for each type of HEA, divided into 1,000 sites for (100) surfaces and 1,000 for (111). Then, these sites were fed into the well-performed GBR model, and the output absorption energies were collected and sorted instantly.

The frequency distribution of OH\* absorption energies of sites on HEA (100) and (111) surfaces are illustrated in Figure 6. As shown, the distribution of energies is continuous, suggesting the possibility that the HEA catalytic activity can be fine-tuned by adjusting the coordination environments of reactive sites. Apparent and individual peaks emerge on the total frequency distribution, inspiring us to classify all energies points according to the types of bridge sites—namely the types of the two metal atoms of sites—to decompose the frequency distribution and discover the intrinsic mechanism. The result demonstrates that the decomposed (colorful) sites type distribution peaks have similar shapes and widths, indicating the effect of the coordinated atoms on the adsorbate strength of reaction intermediates possess similar physical origins. These decomposed peaks own an independent distribution center, which is also the average absorption energy of one type of bridge site. As for the identical metal reactive sites with the XX pattern, the average absorption energies follow the same order for both the (100) and (111) surfaces. The order is FeFe < RuRu < RhRh < IrIr < PtPt < AgAg, which is consistent with the oxygen intermediates absorption energies trend of the corresponding monometallic surfaces. For the dual metal reactive sites with the XY pattern, their absorption energies distribute between the absorption energies distribution center of the XX and YY sites. For example, the absorption energies of PtFe sites all lie in the middle of PtPt and

FeFe distribution center. In addition, the frequency of XX sites is relatively lower than XY sites as a result of the XY sites consisting of XY and YX combinations, while the XX sites have only one combination possibility.

As shown in Figure 6, we noted that the absorption energies of the same sites on HEA (100) are more negative than HEA (111) and the differences were ~0.4 eV. To directly compare the absorption energies distribution of different Miller index surfaces and study the ORR activity of HEAs, the total frequency distribution of two types of surfaces and the volcano curve were painted together in Figure 7A. The maximum, minimum, and means values of the absorption energies of HEA (100) are all lower than HEA (111), and it can be concluded that lower coordinated environments result in stronger OH\* binding on the HEA surface and it can be attributed to high coordination environments mean stronger electronic coupling between metal atoms and it weakens the binding strength of adventitious adsorbate. In addition, only part of the reactive sites on HEA (111) lies in the high activity area and it consists of AgAg, PtAg, and part of the IrAg sites.

To reveal the intrinsic mechanism of the OH\* absorption by model analysis, the feature importance<sup>27</sup> in the coordinated atom dimension and physical or chemical property dimension were studied and plotted in Figures 7B and 7C, respectively. The specific values of feature importance are listed in Tables S6 and S7. As for the coordinated atom dimension, the atoms are sorted and numbered in order from close to far—A<sub>1</sub> and A<sub>2</sub> are two bridge site atoms. Figure 7B demonstrates that the feature importance of two bridge metal atoms is relatively high and close to each other (39.29% and 39.47%), indicating two bridge metal atoms are equivalent and the most significant coordinated atoms to influence OH\* absorption, which is consistent with physical intuition. For the same reason, the third and fourth atoms have close feature importance. It is noteworthy that the feature importance presents the downward trend with the increase in the distance between the coordinated atom and oxygen intermediate, indicating that closer coordination atoms have a stronger influence on the adsorbate strength of the OH intermediate. Interestingly, the 15<sup>th</sup> coordinated atom features an abnormal increase as it can determine the surface is (100) or (111), confirming that the abstract ML model can reflect scientific fact. The specific position of coordinated atoms is displayed in Figure S4.

For the physical and chemical properties dimension,  $\epsilon_d$  is the most important descriptor for OH\* absorption, with a high feature importance of 71.18%, which is in agreement with the theory that the d-band center has a direct and strong relation to adsorbate

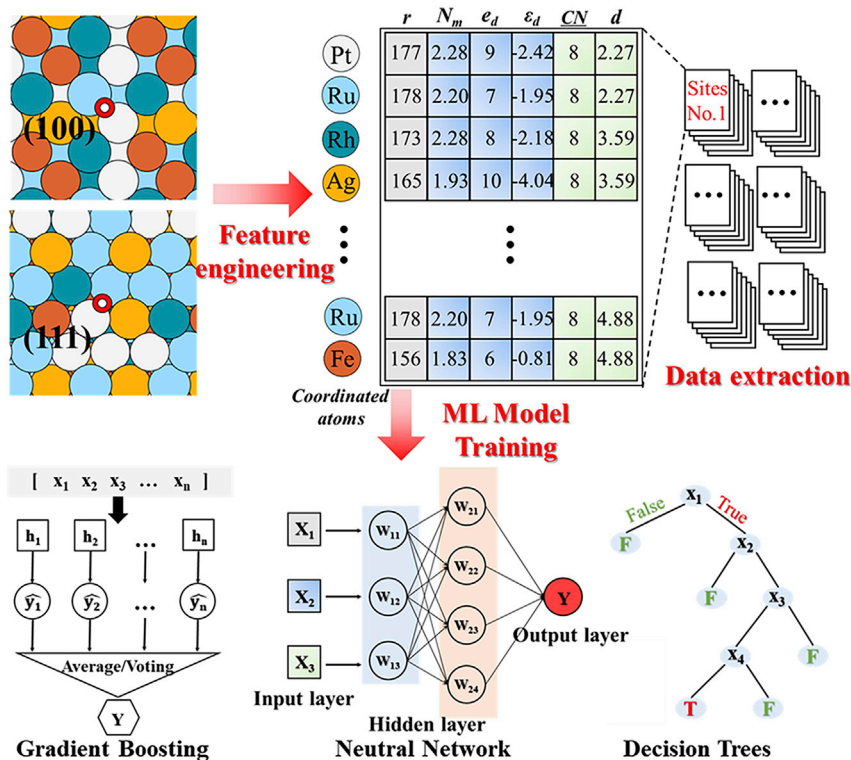


Figure 4. The scheme of the machine learning process

corresponding metal XX sites are correlative, and more negative  $\varepsilon_d$  leads to weaker binding strength, which is consistent with the d-band center theory.

Based on these advanced analyses, it can be concluded that adsorption energy on the HEA surface is a mixture of the individual contributions from all ambient coordinated metal atoms; the two bridge metal atoms which directly bonded to OH\* are the dominant factor to determine the adsorption energy of adsorbent ORR intermediate; the closer the coordinated atom is to OH\*, the stronger its influence. As a result, the decomposed peaks according to the type of two atoms of bridge sites own individual distribution center and the absorption energies of XY sites lie between XX and YY sites. In addition, the coordination environments of two XY bridge atoms are similar in the statistical view, leading to similar shape and width of the decomposed peaks. The

strength.<sup>60,61</sup> The feature importance of  $CN$  and  $d$  are, respectively, 7.04% and 11.24%, suggesting that they are relatively more important than the other three descriptors (<5%). Considering the cruciality of  $\varepsilon_d$ , we compared it with the average absorption energies of XX pattern HEA sites. As shown in Figure S7, the changing trend of  $\varepsilon_d$  and average absorption energies of the

model analysis of feature importance also obeys these rules. The importance of two bridge metal atoms occupies ~80%, which is far larger than other coordinated atoms.  $\varepsilon_d$  possesses ~70% importance, as it represents the absorption ability of a single coordinated atom, while  $CN$  and  $d$  possess ~20%, as they reflect the strength of the single contribution.

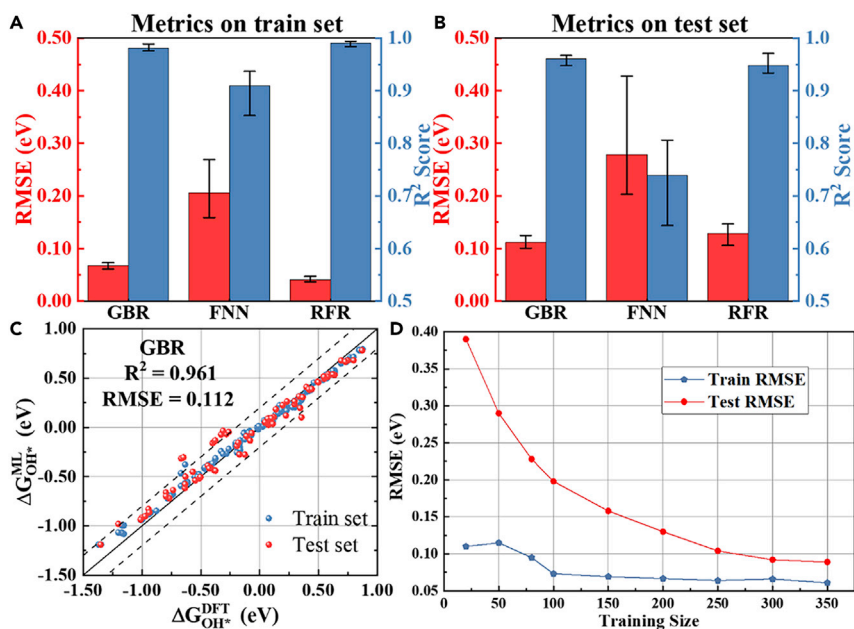
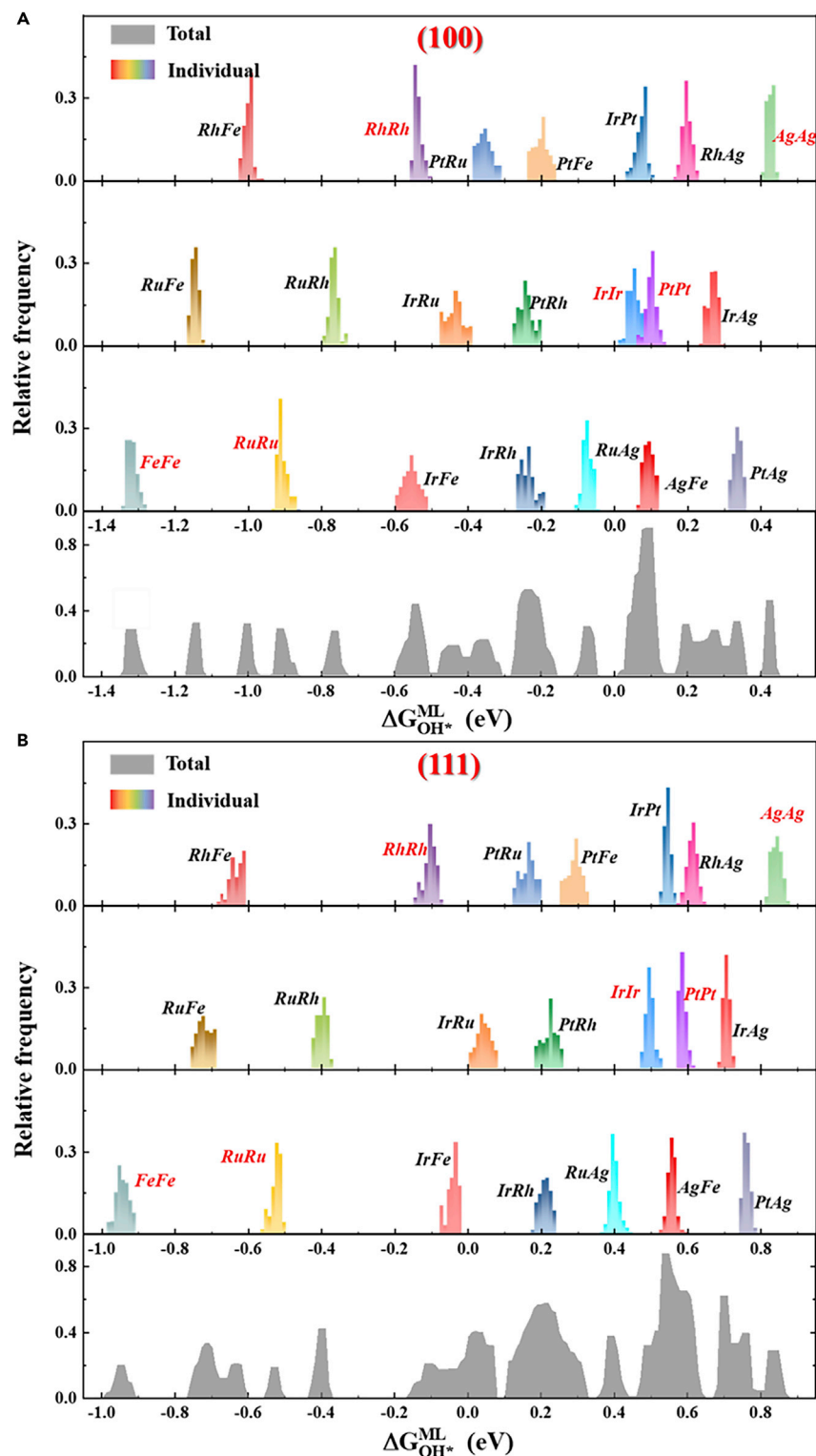


Figure 5. Model performance of 3 different ML models

(A and B) Comparison of the RMSE and the R<sup>2</sup> score of 3 ML models on (A) the train set and (B) the test set. The error bar denotes the range of RMSE and R<sup>2</sup> in the process of 500-time repeated 4-fold cross-validation.

(C) Parity plot of DFT-calculated  $\Delta G_{OH^*}$  with those predicted by the GBR model and model performance metrics of the optimal GBR model. Dotted lines indicate  $\pm 0.2$  eV deviation.

(D) The learning curve of the GBR model.



**Figure 6. The frequency distribution of OH\* adsorption energies of 12,000 reactive sites** (A) The sites on HEA (100) surface. (B) HEA (111). The bottom row shows the total energy distribution with a gray color, while the top 3 rows show the decomposed varicolored peaks according to the identity of the 2 bridge site atoms.

HEA IrPtRuRhAg, IrPtRuAgFe, and IrPtRhAgFe, including the key elements Ag, Pt, and Ir, are promising highly efficient ORR catalysts in all six types of HEAs studied in this work. Hereafter, 10,000 reactive sites on the (111) surface were randomly generated for each HEA belonging to the above 3; then, they were fed into the ML model to obtain their corresponding absorption energies. The active sites coverage (the proportion of the active sites between all of the reactive sites) of IrPtRuRhAg, IrPtRuAgFe, and IrPtRhAgFe are 16.59%, 15.02%, and 15.84%, respectively. Among them, IrPtRuRhAg is the optimal HEA catalyst, owing to more active sites as the coordination environments including Ru and Rh are more likely than RuFe and RhFe environments to weaken the OH absorption energies of IrAg sites, as listed in Table S8. However, even for the best HEA IrPtRuRhAg catalyst, the ORR activity is limited because of relatively low active sites coverage. As a result, one strategy to enhance the ORR activity of HEA catalysts was proposed by optimizing the surface so that the likelihood of finding highly efficient active sites with desired OH\* absorption energies can be enlarged.

The strategy is to adjust the metal element component ratio of HEAs to enlarge the number of active sites. Taking the optimal HEA IrPtRuRhAg as the research target, we adjusted its component ratio and investigated the variation of the active sites. Considering the AgAg, PtAg, and part of IrAg sites are active sites, the component ratio of Ir was fixed to maintain the amount of the possible IrAg active sites and not to occupy the configurational space of key elements Pt and Ag. The ratio of Pt and Ag was enlarged, while Ru and Rh were minimized in the same

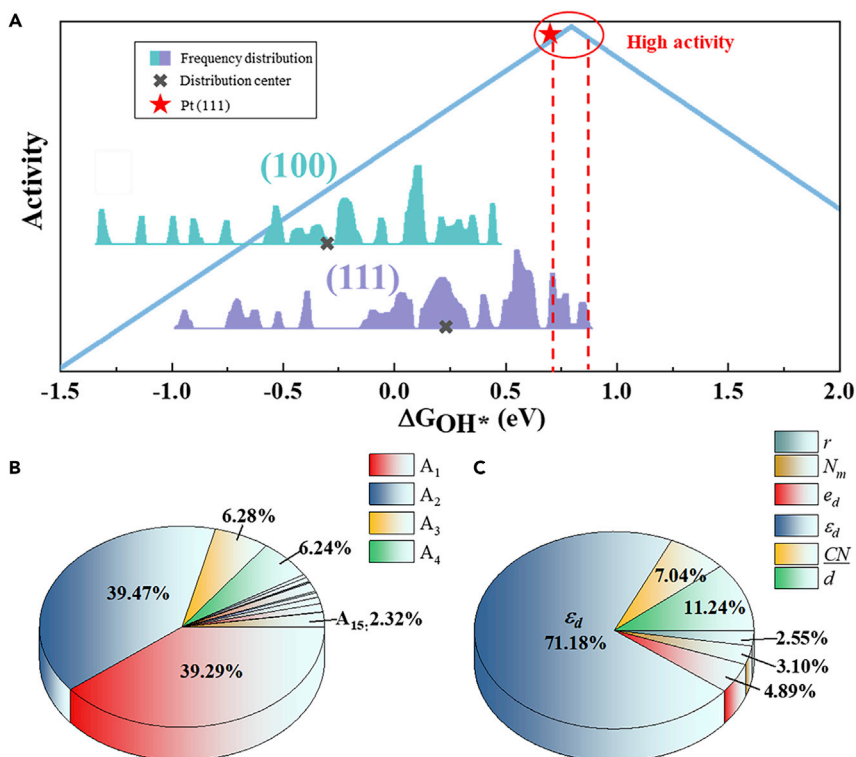
### Optimizing HEA for activity enhancement

As discussed in the last section, AgAg, PtAg, and part of IrAg sites on the HEA (111) surface show favorable OH\* absorption energy, which is commonly called active sites and located around the peak of the volcano curve. Therefore, equimolar

proportion to increase the amount of highly efficient AgAg and PtAg active sites.

We consider that the GBR ML model constructed by the data from nearly equimolar HEAs may be inaccurate when predicting activities of HEAs with different compositions of the elements.





**Figure 7. Further analysis of differences between HEA(100) and HEA(111) and the feature importance of the GBR model**

Frequency distribution of  $OH^*$  adsorption energies of reactive sites on 2 different Miller index surfaces and the volcano curve (A). The mean values of absorption energies, high ORR activity area, and the position of Pt (111) on the volcano curve are marked. The feature importance obtained from the well-performed GBR model in (B) coordinated atom dimension and (C) physical or chemical property.

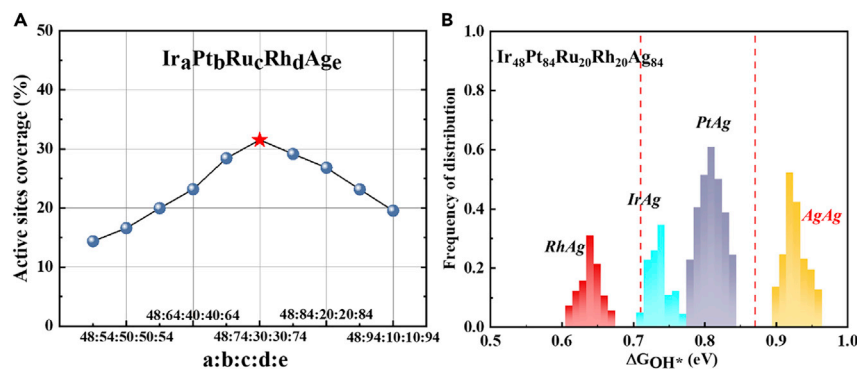
particular single input data is not absolutely related to the composition.

As for each HEA IrPtRuRhAg with different metal element component ratios, 10,000 reactive sites on (111) surfaces were generated and fed into the GBR model for corresponding  $OH^*$  absorption energies. The variation of the active site coverage of these HEAs is illustrated in Figure 8A, demonstrating that the activity sites' coverage increases first and decreases later with the variation of component ratio. The peak of active sites coverage is Ir<sub>48</sub>Pt<sub>74</sub>Ru<sub>30</sub>Rh<sub>30</sub>Ag<sub>74</sub>, with a coverage of 31.54%, approximately twice

Herein, an active learning strategy was used to make the GBR model more reasonable for predicting the activities of HEAs with different compositions. First, 10 HEA IrPtRuRhAg were generated with random compositions, and 1 site was randomly selected for each HEA. Second, the  $OH^*$  adsorption energies of these 10 sites were calculated and compared with the predicted values from the GBR model based on equimolar HEAs. Then, the 10 new sites were regarded as the new data points to train the new GBR model by adding new data points to the original dataset. Finally, 10 new sites were generated, evaluated the performance of the new GBR model, and retrained the GBR model until its performance achieved the standard.

The metric to evaluate model performance is RMSE and the standard to stop the iteration is 0.1 as it is close to the stable metric on the test set for the GBR model for equimolar HEAs, as shown in Figure 5D. The iteration results were painted in Figure S8. It shows that the RMSE of the new GBR model has a gradual downward trend and it achieves the standard after 5 complete active learning loops with a low RMSE of 0.098 eV. The specific information of 60 new sites on HEAs with different compositions can be found in Table S9. The result demonstrates that the strategy can effectively improve the model performance when predicting HEAs with different compositions by tuning the original GBR model. In addition, even the RMSE in iteration 1 shows a small difference with RMSE of the GBR model for equimolar HEAs, indicating that the error of applying the same GBR model to predict the activities of HEAs with different composition is small. As a result, the GBR model based on equimolar HEAs also possesses the knowledge of HEAs with different compositions and it can be attributed to the fact that the whole input dataset is statistically related to the elements composition, but the

the equimolar HEA IrPtRuRhAg, indicating the efficiency of the strategy. The results are interesting because the increase in activity sites coverage is expected, but the decrease is not predicted. For the decrease, we surmise that the reason is that the new coordination environments with far more Pt and Ag atoms than equimolar HEA make the adsorbate strength between OH and promising active AgAg, PtAg, and IrAg sites weakened immoderately, so their statistical distribution moves away from the high activity area. To prove the hypothesis, the  $OH^*$  adsorption energies distribution around the high activity area of the reactive sites on typical HEA Ir<sub>48</sub>Pt<sub>84</sub>Ru<sub>20</sub>Rh<sub>20</sub>Ag<sub>84</sub> (which shows decreasing ORR activity) are shown in Figure 8B. As shown, the  $OH^*$  adsorption energies of the promising active AgAg, PtAg, and IrAg sites change to more positive and the distribution of them moves to the right as a whole because of the weakening effect on absorption strength from new coordination environments. The AgAg sites even move out of the high activity area, which is the main reason for the anomalous decrease in catalytic activity. In addition, it is noteworthy that the increase in  $OH^*$  adsorption energies of active sites makes more sites closer to the vertex of the volcano, which means that more single active sites feature lower energy barriers and better activity. By the effective strategy, the ORR catalytic activity of equimolar HEA IrPtRuRhAg successfully enhances more than double. The stability of Ir<sub>48</sub>Pt<sub>74</sub>Ru<sub>30</sub>Rh<sub>30</sub>Ag<sub>74</sub> is then checked and it is more likely to form a stable solid solution, as shown in Figure 2 (star). DFT calculations were performed to compare the GBR-predicted and accurate values of 10 sites on HEA Ir<sub>48</sub>Pt<sub>74</sub>Ru<sub>30</sub>Rh<sub>30</sub>Ag<sub>74</sub>, and the error is small enough to determine the high activity of recommend catalysts, as listed in Table S10. In general, the discovery of the highly



**Figure 8. The activity results of HEAs with different compositions**

(A) The variation trend of the active sites coverage of HEA IrPtRuRhAg with the change of metal element component ratio.

(B) Frequency distribution of OH\* adsorption energies around the high activity area of the reactive sites on the (111) surface of the typical HEA Ir<sub>48</sub>Pt<sub>84</sub>Ru<sub>20</sub>Rh<sub>20</sub>Ag<sub>84</sub>, which shows decreasing ORR activity.

efficient Ir<sub>48</sub>Pt<sub>74</sub>Ru<sub>30</sub>Rh<sub>30</sub>Ag<sub>74</sub> HEA catalyst offers rational guidance for the experimental nanostructure synthesis of excellent ORR HEA catalysts such as the determination of component elements, composition ratio, and crystal phase.

## Conclusions

In this study, considering that the HEA surface can provide a near-continuum of oxygen intermediate adsorption energies distribution benefit from their huge configurational and chemical space, the potential of highly active ORR electrocatalysts of six types of quinary HEAs were investigated with the aid of ML. The stability of these HEAs was confirmed because they quite likely to form a stable single solid solution according to the Hume-Rothery rule. The ORR volcano curve was established based on the Sabatier rule, providing the way to study the catalytic activity by the key ORR descriptor absorption energies of the OH\* intermediates. Then, the well-performed GBR model with high accuracy, generalizability, and simplicity was constructed by reasonable data extraction, feature engineering, and model validation process. Using this excellent GBR model, millions of reactive sites on HEA surfaces with different coordination combinations, which are impossible to study by traditional DFT calculations or experiments, can be predicted with high fidelity. The ML-predicted results and further model analysis demonstrate that the adsorption energy on HEA surfaces is approximately a mixture of the individual contributions of the metal atoms near the reactive site. Finally, a strategy to engineer the HEA surface structure by tuning the metal element component ratio so that the adsorption energies distribution can be closer to the peak of the volcano to enlarge the ORR catalytic activity was proposed, which doubles the ORR activity. Our proposed DFT-ML scheme demonstrates its ability and potential to become the pioneer in the field of HEA catalysis as it can conquer the extremely broad configurational and chemical space, as well as offer a rational and direct guide on the practical nanostructure synthesis of highly efficient HEA catalysts.

## EXPERIMENTAL PROCEDURES

### Resource availability

#### Lead contact

Requests for further information should be directed to and will be fulfilled by the lead contact, Yuzheng Guo, PhD ([yguo@whu.edu.cn](mailto:yguo@whu.edu.cn)).

### Materials availability

This study did not generate chemical reagents.

### Data and code availability

All original data and code are available in the online repository at GitHub (<https://github.com/XuhaoWan/HEA-ORR>) and at Zenodo (<http://doi.org/10.5281/zenodo.6666342>).

### DFT calculations

In this work, spin-polarized DFT calculations were conducted with the Synopsys Quantum ATK simulation package.<sup>62</sup> Detailed computation methods and parameters were listed in [Notes S5, S6, and S7](#).<sup>63</sup>

### ML

The ML process was conducted by DMCP,<sup>27</sup> together with the Scikit-learn,<sup>64</sup> and PyTorch<sup>65</sup> package. The computational methods of RMSE and R<sup>2</sup> score are listed in [Note S8](#). Normalization is conducted for the original dataset.<sup>66</sup> The FNN model uses a three-layer neural network (input, hidden, and predict layer) and ReLU as the activation function. The specific hyperparameters of the ML models can be found in [Table S11](#).

### Limitations of the study

The study does not include experimental validation considering the workloads and difficulties of synthesizing the corresponding HEAs, exactly controlling their component ratio, and evaluating their ORR performance. The significance of the work is to guide later experimental studies, which can be ensured by the accuracy of DFT calculation and reasonable cross-validation process.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2022.100553>.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant no. 6217030786) and the Fundamental Research Funds for the Central Universities. The numerical calculations in this paper have been made on the supercomputing system in the Supercomputing Center of Wuhan University.

## AUTHOR CONTRIBUTIONS

X.W. and Y.G. created the concept and method; X.W. and Z.Z. wrote the original manuscript; X.W. and H.N. performed the DFT calculations and prepared the figures; X.W. and W.Y. wrote and debugged the code; X.W. and W.Y. the prepared supplemental information; Z.Z. and Y.G. reviewed and edited the manuscript; and Y. G. acquired the funding and supervised the project.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 11, 2022

Revised: May 16, 2022

Accepted: June 23, 2022

Published: August 2, 2022

## REFERENCES

- Yeh, J.W., Chen, S.K., Lin, S.J., Gan, J.Y., Chin, T.S., Shun, T.T., Tsau, C.H., and Chang, S.Y. (2004). Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303. <https://doi.org/10.1002/adem.200300567>.
- George, E.P., Raabe, D., and Ritchie, R.O. (2019). High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534. <https://doi.org/10.1038/s41578-019-0121-4>.
- Toda-Caraballo, I., and Rivera-Díaz-Castillo, P.E. (2015). Modelling solid solution hardening in high entropy alloys. *Acta Mater.* **85**, 14–23. <https://doi.org/10.1016/j.actamat.2014.11.014>.
- Miracle, D.B., and Senkov, O.N. (2017). A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511. <https://doi.org/10.1016/j.actamat.2016.08.081>.
- Zhang, W., Liaw, P.K., and Zhang, Y. (2018). Science and technology in high-entropy alloys. *Sci. China Mater.* **61**, 2–22. <https://doi.org/10.1007/s40843-017-9195-8>.
- Cantor, B., Chang, I., Knight, P., and Vincent, A. (2004). Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng., A* **375**, 213–218. <https://doi.org/10.1016/j.msea.2003.10.257>.
- Zhang, Y., Zuo, T.T., Tang, Z., Gao, M.C., Dahmen, K.A., Liaw, P.K., and Lu, Z.P. (2014). Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93. <https://doi.org/10.1016/j.pmatsci.2013.10.001>.
- Pedersen, J.K., Batchelor, T.A.A., Bagger, A., and Rossmeisl, J. (2020). High-entropy alloys as catalysts for the CO<sub>2</sub> and CO reduction reactions. *ACS Catal.* **10**, 2169–2176. <https://doi.org/10.1021/acscatal.9b04343>.
- Lu, Z., Chen, Z.W., and Singh, C.V. (2020). Neural network-assisted development of high-entropy alloy catalysts: decoupling ligand and coordination effects. *Matter* **3**, 1318–1333. <https://doi.org/10.1016/j.matt.2020.07.029>.
- Ludwig, A. (2019). Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *NPJ Comput. Mater.* **5**, 70. <https://doi.org/10.1038/s41524-019-0205-0>.
- Zhang, G., Ming, K., Kang, J., Huang, Q., Zhang, Z., Zheng, X., and Bi, X. (2018). High entropy alloy as a highly active and stable electrocatalyst for hydrogen evolution reaction. *Electrochim. Acta* **279**, 19–23. <https://doi.org/10.1016/j.electacta.2018.05.035>.
- Qiu, H.-J., Fang, G., Wen, Y., Liu, P., Xie, G., Liu, X., and Sun, S. (2019). Nanoporous high-entropy alloys for highly stable and efficient catalysts. *J. Mater. Chem.* **7**, 6499–6506. <https://doi.org/10.1039/c9ta00505f>.
- Löffler, T., Meyer, H., Savan, A., Wilde, P., Garzón Manjón, A., Chen, Y.T., Ventosa, E., Scheu, C., Ludwig, A., and Schuhmann, W. (2018). Discovery of a multinary noble metal-free oxygen reduction catalyst. *Adv. Energy Mater.* **8**, 1802269. <https://doi.org/10.1002/aenm.201802269>.
- Batchelor, T.A., Pedersen, J.K., Winther, S.H., Castelli, I.E., Jacobsen, K.W., and Rossmeisl, J. (2019). High-entropy alloys as a discovery platform for electrocatalysis. *Joule* **3**, 834–845. <https://doi.org/10.1016/j.joule.2018.12.015>.
- Nellaiappan, S., Katiyar, N.K., Kumar, R., Parui, A., Malviya, K.D., Pradeep, K.G., Singh, A.K., Sharma, S., Tiwary, C.S., and Biswas, K. (2020). High-entropy alloys as catalysts for the CO<sub>2</sub> and CO reduction reactions: experimental realization. *ACS Catal.* **10**, 3658–3663. <https://doi.org/10.1021/acscatal.9b04302>.
- Wang, A.-L., Wan, H.-C., Xu, H., Tong, Y.-X., and Li, G.-R. (2014). Quinary PdNiCoCuFe alloy nanotube arrays as efficient electrocatalysts for methanol oxidation. *Electrochim. Acta* **127**, 448–453. <https://doi.org/10.1016/j.electacta.2014.02.076>.
- Tsai, C.-F., Yeh, K.-Y., Wu, P.-W., Hsieh, Y.-F., and Lin, P. (2009). Effect of platinum present in multi-element nanoparticles on methanol oxidation. *J. Alloys Compd.* **478**, 868–871. <https://doi.org/10.1016/j.jallcom.2008.12.055>.
- Jacobsen, C.J.H., Dahl, S., Clausen, B.S., Bahn, S., Logadottir, A., and Nørskov, J.K. (2001). Catalyst design by interpolation in the periodic table: bimetallic ammonia synthesis catalysts. *J. Am. Chem. Soc.* **123**, 8404–8405. <https://doi.org/10.1021/ja010963d>.
- Strasser, P., Fan, Q., Devenney, M., Weinberg, W.H., Liu, P., and Nørskov, J.K. (2003). High throughput experimental and theoretical predictive screening of materials— a comparative study of search strategies for new fuel cell anode catalysts. *J. Phys. Chem. B* **107**, 11013–11021. <https://doi.org/10.1021/jp030508z>.
- Byskov, L.S., Nørskov, J.K., Clausen, B.S., and Topsøe, H. (1999). DFT calculations of unpromoted and promoted MoS<sub>2</sub>-based hydrodesulfurization catalysts. *J. Catal.* **187**, 109–122. <https://doi.org/10.1006/jcat.1999.2598>.
- Wan, X., Zhang, Z., Niu, H., Yin, Y., Kuai, C., Wang, J., Shao, C., and Guo, Y. (2021). Machine-learning-accelerated catalytic activity predictions of transition metal phthalocyanine dual-metal-site catalysts for CO<sub>2</sub> reduction. *J. Phys. Chem. Lett.* **12**, 6111–6118. <https://doi.org/10.1021/acs.jpcclett.1c01526>.
- Toyao, T., Maeno, Z., Takakusagi, S., Kamachi, T., Takigawa, I., and Shimizu, K.-i. (2019). Machine learning for catalysis informatics: recent applications and prospects. *ACS Catal.* **10**, 2260–2297. <https://doi.org/10.1021/acscatal.9b04186>.
- Dral, P.O. (2020). Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **11**, 2336–2347. <https://doi.org/10.1021/acs.jpcclett.9b03664>.
- O'Connor, N.J., Jonayat, A.S.M., Janik, M.J., and Senftle, T.P. (2018). Interaction trends between single metal atoms and oxide supports identified with density functional theory and statistical learning. *Nat. Catal.* **1**, 531–539. <https://doi.org/10.1038/s41929-018-0094-5>.
- Tran, K., and Ulissi, Z.W. (2018). Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution. *Nat. Catal.* **1**, 696–703. <https://doi.org/10.1038/s41929-018-0142-1>.
- Kitchin, J.R. (2018). Machine learning in catalysis. *Nat. Catal.* **1**, 230–232. <https://doi.org/10.1038/s41929-018-0056-y>.
- Wan, X., Zhang, Z., Yu, W., and Guo, Y. (2021). A density-functional-theory-based and machine-learning-accelerated hybrid method for intricate system catalysis. *Mater. Rep. Energy* **1**, 100046. <https://doi.org/10.1016/j.matre.2021.100046>.
- Ma, X., Li, Z., Achenie, L.E.K., and Xin, H. (2015). Machine-learning-augmented chemisorption model for CO<sub>2</sub> electroreduction catalyst screening. *J. Phys. Chem. Lett.* **6**, 3528–3533. <https://doi.org/10.1021/acs.jpcclett.5b01660>.
- Wang, X., Wang, C., Ci, S., Ma, Y., Liu, T., Gao, L., Qian, P., Ji, C., and Su, Y. (2020). Accelerating 2D MXene catalyst discovery for the hydrogen evolution reaction by computer-driven workflow and an ensemble learning strategy. *J. Mater. Chem.* **8**, 23488–23497. <https://doi.org/10.1039/d0ta06583h>.
- Zhu, X., Yan, J., Gu, M., Liu, T., Dai, Y., Gu, Y., and Li, Y. (2019). 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. <https://doi.org/10.1021/acs.jpcclett.9b03392>.
- Li, Z., Achenie, L.E.K., and Xin, H. (2020). An adaptive machine learning strategy for accelerating discovery of perovskite electrocatalysts. *ACS Catal.* **10**, 4377–4384. <https://doi.org/10.1021/acscatal.9b05248>.
- Wen, C., Zhang, Y., Wang, C., Xue, D., Bai, Y., Antonov, S., Dai, L., Lookman, T., and Su, Y. (2019). Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117. <https://doi.org/10.1016/j.actamat.2019.03.010>.

33. Li, Y., and Guo, W. (2019). Machine-learning model for predicting phase formations of high-entropy alloys. *Phys. Rev. Mater.* 3, 095005. <https://doi.org/10.1103/physrevmaterials.3.095005>.
34. Antolini, E. (2014). Iridium as catalyst and cocatalyst for oxygen evolution/reduction in acidic polymer electrolyte membrane electrolyzers and fuel cells. *ACS Catal.* 4, 1426–1440. <https://doi.org/10.1021/cs4011875>.
35. Wu, J., and Yang, H. (2013). Platinum-based oxygen reduction electrocatalysts. *Acc. Chem. Res.* 46, 1848–1857. <https://doi.org/10.1021/ar300359w>.
36. Schmidt, T.J., Paulus, U.A., Gasteiger, H.A., Alonso-Vante, N., and Behm, R.J. (2000). Oxygen reduction on Ru<sub>1</sub>, Ru<sub>0.92</sub>Mo<sub>0.08</sub>SeO<sub>4</sub>, Ru/carbon, and Pt/carbon in pure and methanol-containing electrolytes. *J. Electrochem. Soc.* 147, 2620. <https://doi.org/10.1149/1.1393579>.
37. Cao, D., Wieckowski, A., Inukai, J., and Alonso-Vante, N. (2006). Oxygen reduction reaction on ruthenium and rhodium nanoparticles modified with selenium and sulfur. *J. Electrochem. Soc.* 153, A869. <https://doi.org/10.1149/1.2180709>.
38. Zhou, Y., Lu, Q., Zhuang, Z., Hutchings, G.S., Kattel, S., Yan, Y., Chen, J.G., Xiao, J.Q., and Jiao, F. (2015). Oxygen reduction at very low overpotential on nanoporous Ag catalysts. *Adv. Energy Mater.* 5, 1500149. <https://doi.org/10.1002/aenm.201500149>.
39. Li, W., Wu, J., Higgins, D.C., Choi, J.-Y., and Chen, Z. (2012). Determination of iron active sites in pyrolyzed iron-based catalysts for the oxygen reduction reaction. *ACS Catal.* 2, 2761–2768. <https://doi.org/10.1021/cs300579b>.
40. Chen, X., Si, C., Gao, Y., Frenzel, J., Sun, J., Eggeler, G., and Zhang, Z. (2015). Multi-component nanoporous platinum–ruthenium–copper–osmium–iridium alloy with enhanced electrocatalytic activity towards methanol oxidation and oxygen reduction. *J. Power Sources* 273, 324–332. <https://doi.org/10.1016/j.jpowsour.2014.09.076>.
41. Zhan, C., Xu, Y., Bu, L., Zhu, H., Feng, Y., Yang, T., Zhang, Y., Yang, Z., Huang, B., Shao, Q., and Huang, X. (2021). Subnanometer high-entropy alloy nanowires enable remarkable hydrogen oxidation catalysis. *Nat. Commun.* 12, 6261. <https://doi.org/10.1038/s41467-021-26425-2>.
42. Qin, Y.-C., Wang, F.-Q., Wang, X.-M., Wang, M.-W., Zhang, W.-L., An, W.-K., Wang, X.-P., Ren, Y.-L., Zheng, X., Lv, D.-C., and Ahmad, A. (2021). Noble metal-based high-entropy alloys as advanced electrocatalysts for energy conversion. *Rare Met.* 40, 2354–2368. <https://doi.org/10.1007/s12598-021-01727-y>.
43. Lee, D.-T., and Schachter, B.J. (1980). Two algorithms for constructing a Delaunay triangulation. *Int. J. Comput. Inf. Sci.* 9, 219–242. <https://doi.org/10.1007/bf00977785>.
44. Zhang, Y., Zhou, Y.J., Lin, J.P., Chen, G.L., and Liaw, P.K. (2008). Solid-solution phase formation rules for multi-component alloys. *Adv. Eng. Mater.* 10, 534–538. <https://doi.org/10.1002/adem.200700240>.
45. Yang, X., and Zhang, Y. (2012). Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Mater. Chem. Phys.* 132, 233–238. <https://doi.org/10.1016/j.matchemphys.2011.11.021>.
46. Mizutani, U. (2010). The Hume-Rothery rules for structurally complex alloy phases. In *Surface Properties and Engineering of Complex Intermetallics (World Scientific)*, pp. 323–399.
47. Chaves, A.S., Piotrowski, M.J., Guedes-Sobrinho, D., and Da Silva, J.L.F. (2015). Theoretical investigation of the adsorption properties of CO, NO, and OH on monometallic and bimetallic 13-atom clusters: the example of Cu<sub>13</sub>, Pt<sub>7</sub>Cu<sub>6</sub>, and Pt<sub>13</sub>. *J. Phys. Chem. A* 119, 11565–11573. <https://doi.org/10.1021/acs.jpca.5b08330>.
48. Forgie, R., Bugosh, G., Neyerlin, K.C., Liu, Z., and Strasser, P. (2010). Bimetallic Ru electrocatalysts for the OER and electrolytic water splitting in acidic media. *Electrochem. Solid State Lett.* 13, B36. <https://doi.org/10.1149/1.3290735>.
49. Che, M. (2013). Nobel prize in chemistry 1912 to sabatier: organic chemistry or catalysis? *Catal. Today* 218, 162–171. <https://doi.org/10.1016/j.cattod.2013.07.006>.
50. Rossmeisl, J., Logadottir, A., and Nørskov, J. (2005). Electrolysis of water on (oxidized) metal surfaces. *Chem. Phys.* 319, 178–184. <https://doi.org/10.1016/j.chemphys.2005.05.038>.
51. Li, Z., Wang, S., Chin, W.S., Achenie, L.E., and Xin, H. (2017). High-throughput screening of bimetallic catalysts enabled by machine learning. *J. Mater. Chem.* 5, 24131–24138. <https://doi.org/10.1039/c7ta01812f>.
52. Dayah, M. (2022). Ptable. <https://ptable.com/>.
53. Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 1189–1232.
54. Sanger, T.D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Network.* 2, 459–473. [https://doi.org/10.1016/0893-6080\(89\)90044-0](https://doi.org/10.1016/0893-6080(89)90044-0).
55. Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R. News* 2, 18–22.
56. Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. <https://doi.org/10.1007/bf00994018>.
57. Devroye, L., Györfi, L., Krzyżak, A., and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Ann. Stat.* 22, 1371–1385. <https://doi.org/10.1214/aos/1176325633>.
58. Vovk, V. (2013). Kernel ridge regression. In *Empirical inference (Springer)*, pp. 105–116.
59. Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* 101, 1418–1429. <https://doi.org/10.1198/016214506000000735>.
60. Norsko, J.K. (1990). Chemisorption on metal surfaces. *Rep. Prog. Phys.* 53, 1253–1295. <https://doi.org/10.1088/0034-4885/53/10/001>.
61. Nørskov, J. (1991). Electronic factors in catalysis. *Prog. Surf. Sci.* 38, 103–144. [https://doi.org/10.1016/0079-6816\(91\)90007-q](https://doi.org/10.1016/0079-6816(91)90007-q).
62. Smidstrup, S., Markussen, T., Vancraeyveld, P., Wellendorff, J., Schneider, J., Gunst, T., Verstichel, B., Stradi, D., Khomyakov, P.A., Vej-Hansen, U.G., et al. (2019). QuantumATK: an integrated platform of electronic and atomic-scale modelling tools. *J. Phys. Condens. Matter* 32, 015901. <https://doi.org/10.1088/1361-648x/ab4007>.
63. Niu, H., Wan, X., Wang, X., Shao, C., Robertson, J., Zhang, Z., and Guo, Y. (2021). Single-atom rhodium on defective g-C<sub>3</sub>N<sub>4</sub>: a promising bifunctional oxygen electrocatalyst. *ACS Sustainable Chem. Eng.* 9, 3590–3599. <https://doi.org/10.1021/acssuschemeng.0c09192>.
64. Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., and Mueller, A. (2015). Scikit-learn: machine learning without learning the machinery. *GetMobile: Mobile Comput. Commun.* 19, 29–33. <https://doi.org/10.1145/2786984.2786995>.
65. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., and Antiga, L. (2019). Pytorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32, 8026–8037.
66. Sola, J., and Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* 44, 1464–1468. <https://doi.org/10.1109/23.589532>.