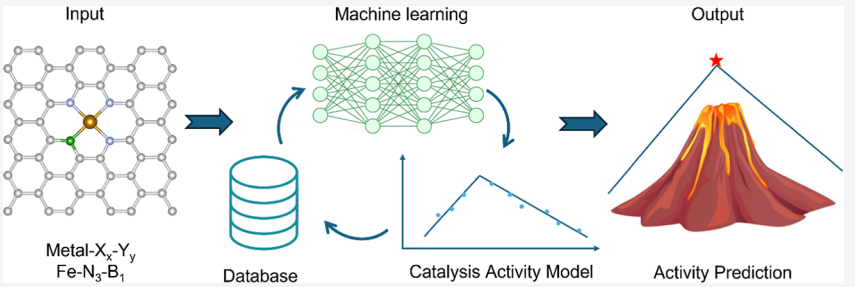
THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC
Article

Computational Single-Atom Catalyst Database Empowers the Machine Learning Assisted Design of High-Performance Catalysts

Mingye Huang, Ruiyang Shi, Heng Liu, Wenjun Ding, Jiahang Fan, Binghui Zhou, Bo Da, Zhengyang Gao,* Hao Li,* and Weijie Yang*

Cite This: J. Phys. Chem. C 2025, 129, 5043−5053
Read Online



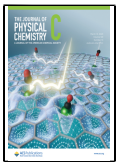
ABSTRACT: The data-driven strategy has emerged as an important approach for the rapid screening of high-performance single-atom catalysts (SACs). However, the lack of a comprehensive SACs database seriously hinders the widespread application of this strategy. Herein, we construct a public SACs database comprising 1197 samples via doping nonmetallic atoms (B, N. O, P, and S) in the coordination environment and regulating 3d metal centers (Ti, V, Cr, Mn, Fe, Co, Ni, Cu, and Zn). Based on density functional theory calculations, the electronic structural properties (i.e., Bader charge and d-band center) and binding energies are obtained. According to the binding energy calculations, 657 stable catalyst configurations are identified. Subsequently, the corresponding adsorption energies for O_2, O, and NO are calculated. Moreover, machine learning (ML) models, specifically extreme gradient boosting regression (XGBR), random forest regression, and support vector regression, are trained to predict the electronic structure and the adsorption energies of O. Among these models, XGBR demonstrates the highest predictive accuracy, with a mean squared error less than 0.35. We successfully integrate ML models based on this SACs database and catalytic volcano model. Through this framework, the catalytic activities of 1261 4d SACs in the oxidation of NO and Hg^0 are quickly predicted. Rh_1B_4 and $Rh_1C_2S_2$ are identified as potential catalysts for the oxidation of NO and Hg^0, with the respective energy barriers of 1.01 and 2.59 eV for Rh_1B_4, and 1.03 and 2.61 eV for $Rh_1C_2S_2$. These values are significantly lower than those of previously reported SACs. We anticipate that this public SACs database and ML-based activity prediction framework can provide new pathways for the rapid screening of highly active SACs for various catalytic reactions.

1. INTRODUCTION

Single-atom catalysts (SACs) have emerged as highly promising catalysts due to their uniform reactive active centers, which offers significant advantages such as high atomic utilization, high activity, and superior selectivity.[1−3] These features make SACs exhibit excellent performance in many crucial reactions across the energy and environment fields.[4−7] The catalytic activity of SACs can be effectively tuned by doping nonmetallic atoms and substituting active center atoms to regulate the coordination environment of SACs. Theoretical calculations have shown that varying the number and type of coordinated atoms in SACs can significantly tune their catalytic activity for different reactions.[8] Currently, regulating the microenvironment of SACs becomes a research hotspot for designing potential SACs with high activity for different catalytic reactions.[9−11]

For instance, Li et al.[12] optimized the catalytic activity of single-atom Mn catalyst for Hg^0 oxidation through introducing O- and N-coordination by experimental approach. Similarly, doping different amounts of B and N atoms into single-atom Co catalyst can effectively regulate the catalytic oxidation of CO. Additionally, the catalytic activity for oxygen evolution reactions (OER) was improved through introducing B, N, O, P, and S into single-atom Fe catalyst. Tang et al.[13] demonstrated that varying the number of B and N atoms in single-atom Co catalysts can effectively regulate the CO

Received: January 22, 2025
Revised: February 19, 2025
Accepted: February 19, 2025
Published: March 3, 2025



https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

oxidation activity. Yang et al.[14] introduced nonmetallic elements such as B, N, O, P, and S into single-atom Fe catalysts to investigate their OER activities; theoretical calculations showed that $Fe_1B_1C_1N_2$-pen coordinated with B and N exhibited the lowest OER overpotential (0.40 V) among 122 catalysts. Furthermore, Yang et al.[15] found that doping nonmetallic atoms into single-atom Co catalysts can effectively regulate the adsorption energy for O_2 and O, and they identified several SACs with high catalytic activity for NO and Hg^0. Among them, the energy barriers of $Co_1P_2C_2$-pen in the rate-determining steps of NO and Hg^0 oxidation are 0.98 and 2.14 eV, respectively.

Previous research has demonstrated that regulating the coordination environment of SACs (by doping nonmetallic atoms and changing the type of metal atoms in the active center) is an effective way to regulate their catalytic performance. This presents a new approach for the construction of a publicly accessible SACs database. However, the challenge lies in the fact that there are a great many combinations between the central site and the coordination elements. Therefore, the traditional development mode of preparation-testing is no longer efficient, necessitating the adoption of more efficient and rapid research methods.

Machine learning (ML), as a powerful tool, can greatly enhance the efficiency of catalyst design. For example, Nishimura et al.[16] employed high-throughput screening and literature data sets combined with ML methods to design a multicomponent La_2O_3-based methane oxidative coupling catalyst. Similarly, Khatamirad et al.[17] created a database of oxygen vacancy formation energies using high-throughput calculations, and explored a large number of potential materials with improved performance for CO_2 hydrogenation to methanol. Jia et al.[19,20] used the Materials Project database and density functional theory (DFT) calculations to investigate oxygen reduction reaction (ORR) activity of stable metal oxide materials under acidic conditions and performed subsequent experimental validations. Luo et al.[21] developed an XGBoost ML model for chromium diphosphate (Cr-PNP) catalysts through high-throughput feature selection, predicting the selectivity for ethylene tris/tetramerization, enabling data-driven design of Cr-PNP catalysts. Zhou et al.[22] introduced a data-driven high-throughput method combining theoretical calculations with topology-based multiscale convolution and ML algorithms, predicting two highly active catalysts, Pt_1C_4 and $Sc_1C_1N_3$, for hydrogen evolution reaction (HER) after training on 25 data sets and predicting on 168 SACs. We may employ the same methodology to predict the SACs capable of effectively catalyzing the reactions involving NO and Hg^0.

However, the lack of a comprehensive SACs database poses a significant challenge to the rapid screening of SACs under data-driven strategies. Therefore, building an available SACs database will provide important theoretical support for the rapid screening of highly active SACs under data-driven strategies. To overcome this limitation, herein, we establish a SACs database through high-throughput calculations and developed a data-mechanism-driven framework for predicting the catalytic activity of SACs. This framework integrates the database, ML, and catalytic volcano models. First, 1197 SACs structures are constructed by regulating their coordination environments. Then, electronic structure properties, binding energy, adsorption energy, and other properties of the catalyst are obtained through high-throughput DFT calculations. Subsequently, an open-source SACs database is constructed using the Streamlit application tool. Finally, extreme gradient boosting regression (XGBR), random forest regression (RFR), and support vector regression (SVR) ML models are trained based on the database. Among these, XGBR is proven the most accurate in predicting the electronic structure properties and O adsorption energies. Therefore, we develop a framework based on ML combining the XGBR model for predicting the adsorption energy of O atoms with a catalytic activity volcano model. This framework enables rapid prediction of the catalytic oxidation activity of SACs for NO and Hg^0. Among 1261 SACs with 4d metals, Rh_1B_4 and $Rh_1C_2S_2$ are precisely screened as potential catalysts, demonstrating the effectiveness of this DFT-based database and ML framework for SAC design.

2. COMPUTATION AND MODELING METHODS

In this work, spin-polarized density functional theory (DFT) calculations were performed using the Vienna ab initio simulation package (VASP 5.4.4).[23−25] In detail, the projected augmented wave (PAW) method was employed to describe the interactions between the nucleus and valence electrons, while the Perdew−Burke−Ernzerhof (PBE) functional was used for electronic exchange and correlation.[26−28] To more accurately describe the interaction between reactants and catalyst surfaces, dipole correction and DFT-D3 were applied.[29] A $5 \times 3\sqrt{3} \times 1$ graphene substrate model ($12.33 \times 12.88 \times 20.00$ Å) was constructed, with a vacuum layer set at 20 Å to prevent interactions between mirror images.[30] The kinetic energy cutoff was set to 450 eV for all calculations, and a $2 \times 2 \times 1$ k-point grid was used for geometry optimizations. The force convergence criterion during geometric optimization was set to −0.02 eV/Å. For electronic self-consistent calculations, a k-point grid of $4 \times 4 \times 1$ was used, with a convergence criterion of 10^{-5} eV.

Bader charge analysis was performed to calculate the amount of charge of metal atoms in SACs.[31] The formulas for calculating the catalyst's d-band center (ε_d), system electronegativity (X), binding energy (E_{bind}), and adsorption energy (E_{ads}) are provided in the Supporting Information eqs S1−S4.

A publicly accessible SACs database was created using the Streamlit application,[32] while feature generation based on SACs coordination atom composition was accomplished using the Matminer package.[33] The XGBR, RFR, and SVR algorithms used in this study were implemented using the Scikit Learn library,[34−37] with hyperparameters optimized through a grid search approach and 5-fold cross-validation.[38] The performance of ML models was evaluated using the coefficient of determination (R-squared, R^2), mean squared error (MSE), and mean absolute error (MAE).[39] The corresponding calculation formulas are shown in the Supporting Information eqs S5−S7.

3. RESULTS AND DISCUSSION

3.1. Catalyst Design and High-Throughput Computations. *3.1.1. Catalyst Design.* The design of the coordination environment for SACs mainly involves two aspects: the substitution of metal center and the doping of nonmetallic elements that coordinate with metal centers. In this study, 3d transition metals (Ti, V, Cr, Mn, Fe, Co, Ni, Cu, and Zn) are used as metal center for SACs, whereas nonmetallic elements (B, N, O, P, and S) are doped in a controlled manner to regulate the coordination environment of SACs. The coordination of nonmetallic elements and metal

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

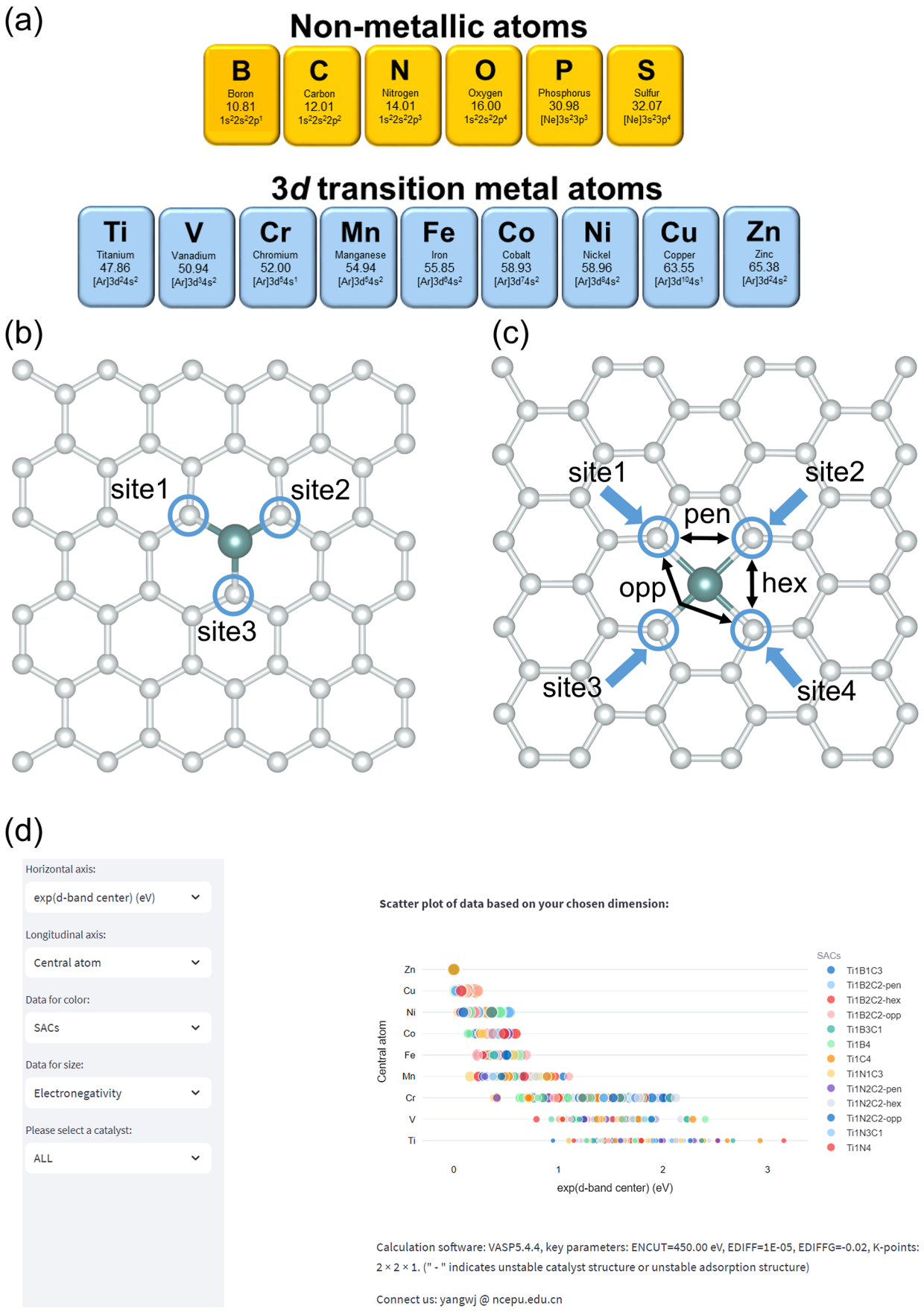


Figure 1. (a) Nonmetallic elements and metal centers in the catalyst coordination environment; (b,c) schematic representation of the coordination environments in single and double vacancy catalysts, respectively; (d) main interface of the publicly available online database application, showing the information on 1197 SACs.

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

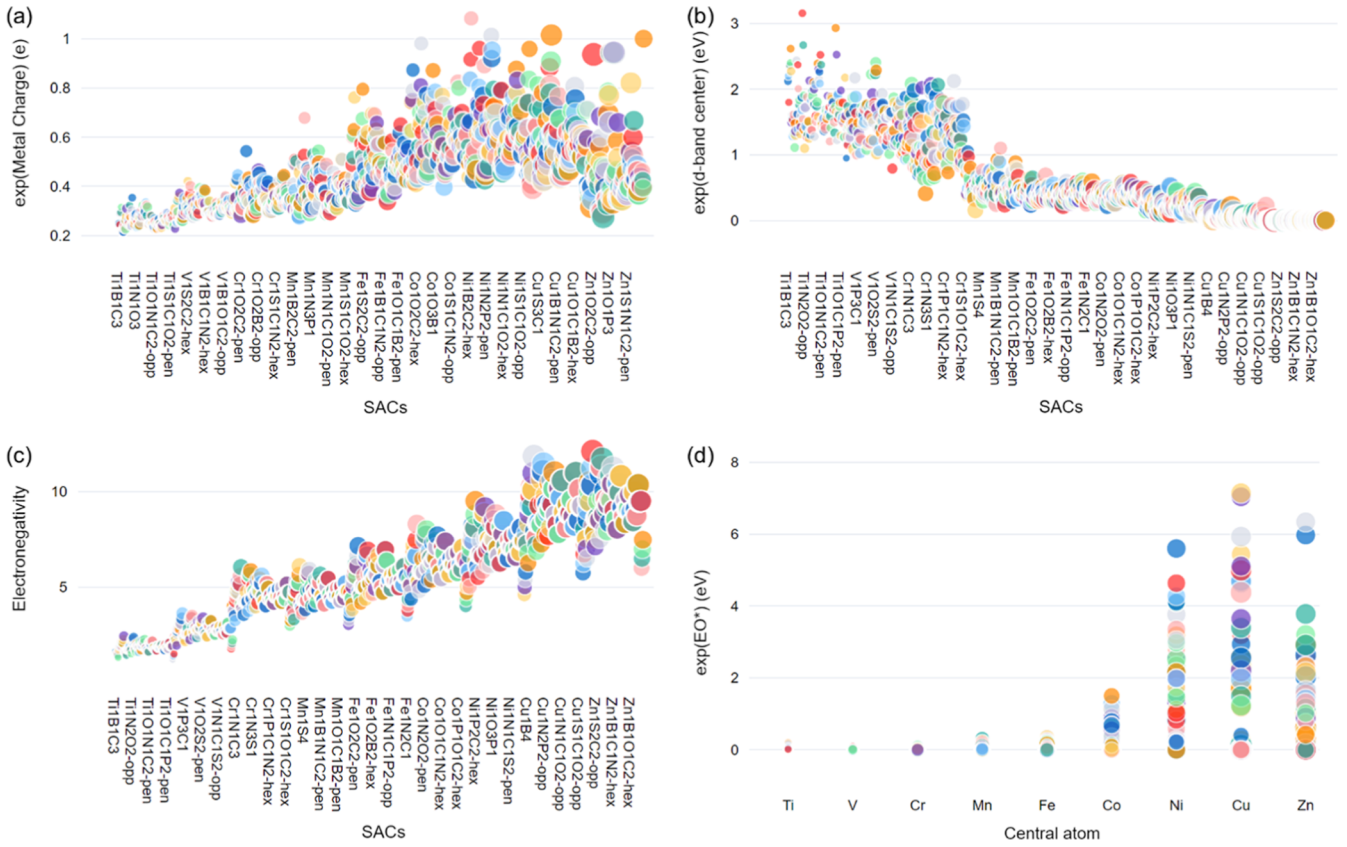


Figure 2. (a−c) Represents the trends of Bader charge (using the natural index values), d-band center (using the natural index values), and system electronegativity with different types of SACs, respectively; (d) the distribution of O adsorption energy (using the natural index values) as the change of central metal atoms in SACs. The size and color of the data in the figure represent the electronegativity of the system and the type of single-atom catalyst, respectively. These visualizations can be customized based on the specific needs of users.

centers in the catalyst's coordination environment is illustrated in Figure 1a, while the coordination environments of single and double vacancy catalysts are shown in Figures 1b,c.

It should be noted that in the case of double-vacancy catalysts, when two identical nonmetallic atoms are doped, three different configurations can arise due to the structural characteristics of graphene. Based on the relative positions of two identical nonmetallic atoms, these configurations are referred to as -pen, -hex, and -opp[40] (Figure 1c). This phenomenon does not occur in single vacancy catalysts. A total of 1197 SACs structures are constructed by adjusting the metal center and the coordinated nonmetal elements.

3.1.2. High-Throughput Computations. Through high-throughput DFT computations, we obtain the geometrically optimized structures of 1197 catalysts. Subsequently, electronic structural properties including metal atom's Bader charge, d-band center, and system electronegativity, are determined through electronic self-consistent calculations. Binding energy calculations are then performed, and by comparing these with the cohesive energy of the corresponding metal block, 657 stable structures are identified among these 1197 catalysts.[41] The adsorption energies of stable catalysts for O_2, O, and NO are also calculated.

It should be noted that during the O_2 adsorption process on the catalyst surface, two different adsorption configurations—side adsorption and end adsorption—can occur, with their corresponding adsorption energies labeled as $E_{ads}(O_{2side})$ and $E_{ads}(O_{2end})$, respectively. An example of the adsorption configuration of O_2 on $Fe_1N_3O_1$ is provided in Figure S1, illustrating both the side and end adsorption configurations. Finally, the data obtained from these DFT calculations are organized and summarized for further analysis.

3.2. Construction of Catalyst Online Database. The public SACs database (with the URL available in https://catalysis-ncepu-hvkydg736ykqeq26d5gxrn.streamlit.app/) is constructed using the Streamlit application, an open-source tool for data science applications based on the Python language framework.[32] Streamlit provides a user-friendly platform for creating online interactive applications, allowing data scientists and ML engineers to quickly and conveniently share their data results with others. The main interface of the online SACs database is displayed in Figure 1d. Through the dropdown menus in the gray option bar on the left, users can select four dimensions (horizontal axis, vertical axis, scatter point color, and scatter point size) to visualize catalyst information. Users can also hover the mouse cursor over any scatter point position to view detailed information about the corresponding SAC. To handle the issue of displaying both positive and negative numerical values in the interactive interface, data (except for Bader charge values) is transformed using the natural exponential function (exp) for consistent representation.

3.3. Properties of the Catalysts. *3.3.1. Bader Charge Analysis.* The Bader charge of the metal atoms in SACs reflects the amount of charge transfer between the metal atoms and the graphene substrate. DFT calculation results reveal that during the formation process of graphene-based SACs, the metal atoms on the catalyst surface lose electrons to varying degrees,

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

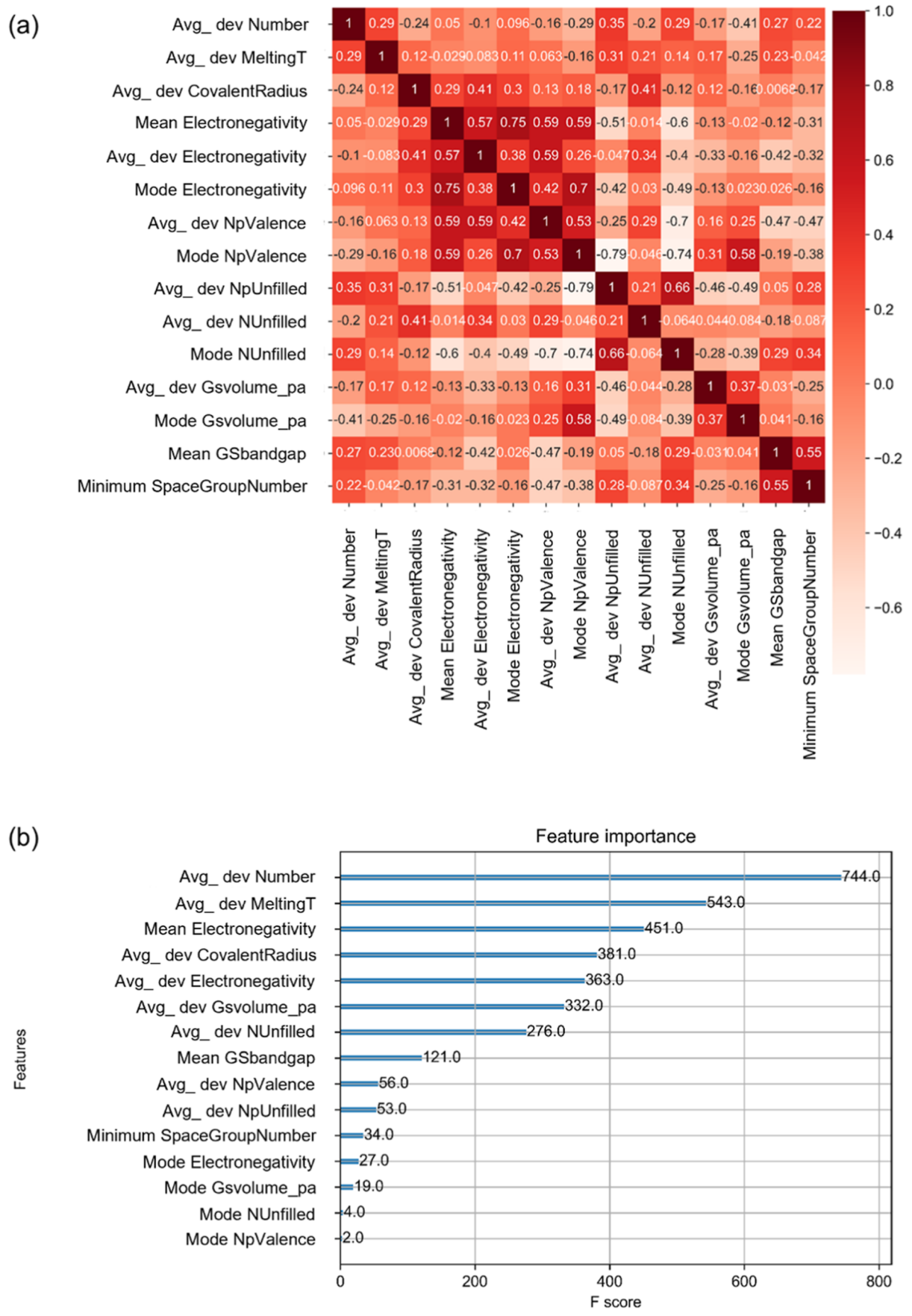


Figure 3. (a) Thermal correlation heatmap of the 15 weakly correlated Magpie features; (b) the feature importance scores for predicting Bader charge using the XGBR model.

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

indicating that the electrons from the metal atom are transferred to the graphene substrate. Figure 2a shows the trend of the Bader charge of metal atoms across different SAC types in the database. As shown in Figure 2a, with increasing atomic number of the metal atoms, the amount of charge transfer follows a volcano trend, initially decreasing and then increasing. SACs with Co, Ni, and Cu as the central metal atoms exhibit relatively smaller charge transfers, indicating that electron transfer from these three metal atoms to the graphene substrate is more challenging during catalyst formation. In comparison, varying the metal atoms has a greater impact on the Bader charge transfer of SACs than modifying the coordinated nonmetallic atoms. Similar findings regarding charge transfer trends in metal atoms within different SACs have been reported in previous studies.[42]

3.3.2. d-Band Center. In heterogeneous catalysis, the d-band theory proposed by Hammer and Nørskov in 1995 has proven to be a practical framework for predicting the catalytic activity of transition metals.[43−46] The d-band theory suggests that the catalytic activity of transition metal surfaces is mainly related to the d-band characteristics of the metal, particularly the d-band center (ε_d), which is an effective descriptor for catalytic performance. Therefore, determining the d-band center of the metal atom at the active site of a catalyst is crucial for guiding the design and synthesis of highly efficient catalysts. It should be specifically noted that we took spin polarization into account and calculated the d band center using the spin up state.[46]

As shown in Figure 2b, we plotted different types of SACs on the horizontal axis and their corresponding ε_d values on the vertical axis in the database. It can be observed that as the atomic number of the central metal atoms in the SACs increase, the value of ε_d tends to approach zero. Specifically, as the atomic number increases from Ti to Zn, ε_d gradually shifts from positive to negative, with the magnitude of the negative value increasing. This suggests that the d-band center is mainly determined by the type of metal atom. Previous studies[42] on SACs coordinated with N atoms have a similar conclusion.

3.3.3. System Electronegativity. For graphene-based SACs, previous studies introduced the concept of system electronegativity (X).[47] As shown in Figure 2c, the types of SACs are represented on the horizontal axis in the database, and the electronegativity of the catalyst system is plotted on the vertical axis. It can be observed that as the atomic number of the central metal atoms increases from Ti to Zn, the system electronegativity (X) also increases, indicating that the electronegativity of SACs is mainly influenced by the type of central metal atoms, with the coordinated nonmetallic atoms exerting a relatively minor effect on X. For instance, as the central metal atom changes from Ti to Zn, the system electronegativity across all SACs ranges from 1.22 to 12.11. Moreover, for 133 SAC systems with different graphene substrates coordinated with Ti atoms, the electronegativity varies between 1.22 and 2.44. This further demonstrates that the type of metal atom plays the primary role in determining the system's electronegativity, with limited influence from the coordinating atoms.

3.3.4. O Adsorption Energy. Previous studies[48−51] have shown that the adsorption energy of O atoms ($E_{ads}(O)$) plays an important role in describing catalyst activity and guiding the selection of highly active catalysts. As shown in Figure 2d, the types of metal atoms at the catalyst center are plotted on the x-axis, while $E_{ads}(O)$ is shown as the natural index on the y-axis. It can be observed that as the atomic number of the central metal atoms increases, $E_{ads}(O)$ decreases gradually. The $E_{ads}(O)$ of SACs centered on Ni, Cu, and Zn are significantly affected by the doping of coordinating nonmetallic atoms. For example, single-atom Ti catalysts with various nonmetallic atom coordination exhibit $E_{ads}(O)$ ranging from −1.63 to −4.63 eV. The $E_{ads}(O)$ of single-atom Cu catalysts varies from 1.97 to −3.67 eV, indicating a stronger influence from different nonmetallic coordinating substrates. This trend suggests that SACs with Ni, Cu, and Zn centers exhibit a weakening adsorption of O atoms by the central metal atoms, highlighting the influence of nonmetallic atom coordination on the adsorption energy.

3.4. ML Prediction of Catalyst Properties. Previous reports[18,52−56] have demonstrated that the electronic structure properties of catalysts are key factors in describing their activity. In addition, as a simple adsorbate, the adsorption energy of O plays a crucial role in predicting the activity of specific catalytic reactions.[49,50,57,58] To enhance the prediction of these properties, we construct ML models for XGBR, RFR, and SVR to predict the electronic structure properties and O adsorption energy of catalysts.

3.4.1. Feature Engineering. Based on the coordination environments, atomic composition information on 1197 catalysts is contained in the graphene-based SACs database. Using the "ElementProperty" function from the "matminer.featurers.composition" module in the Matminer package, structural features are generated based on the atomic composition of the catalyst's coordination environment.[33] A total of 132 features were generated for SACs using the Magpie method,[59] with detailed descriptions of each feature provided in Table S1. For the 132 generated Magpie features, this work removed constants, zeros, and invalid features with obvious linear relationships, ultimately resulting in the selection of 24 preliminary Magpie features for constructing ML data set. In the subsequent updates of the database, we will take more electrochemical stability assessments into consideration, such as the dissolution potential.[60]

Pearson correlation analysis was performed on these 24 Magpie features, and 15 features with absolute correlation values below 0.8 were identified. A heatmap of the Pearson correlation coefficients between these 15 features, shown in Figure 3a, indicates that there are no strong linear relationships between the remaining features.

The importance of the 15 features was further evaluated, as shown in Figure 3b. The ranking order of the importance scores of the 15 structural features is consistent across the predicted variables. Among them, "Avg_dev Number" has the highest feature importance score, while "Mode NpValence" corresponds to the lowest. It indicates that "Avg_dev Number" has the greatest impact on the performance of the model, while "Mode NpValence" has the smallest. Therefore, "Mode NpValence" was removed, leaving 14 structural features for machine learning model training, as shown in Table 1.

3.4.2. Data Set Construction. To construct a prediction model for the electronic structure properties of catalysts, we utilized data including 14 Magpie features corresponding to 1197 graphene-based SACs in the database, Bader charges of metal atoms, d-band centers of metal atoms, and system electronegativity. For predicting adsorption energy, we built a data set that included 14 Magpie features of 657 stable catalysts and the adsorption energy of O. The data set was then split into training and test sets in a 70:30 partition ratio.

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

Table 1. 14 Magpie Features Generated by the Matminer Package and Their Descriptions

no.	magpie features	descriptions
1	Avg_dev number	average deviation of atomic number
2	Avg_dev meltingT	average deviation of melting temperature
3	mean electronegativity	average electronegativity
4	Avg_dev CovalentRadius	average deviation of covalent radius
5	Avg_dev GSvolume_pa	average deviation of DFT-computed volume of elemental solid
6	Avg_dev electronegativity	average deviation of electronegativity
7	Avg_dev NUnfilled	average deviation of number of unfilled valence orbitals
8	mean GSbandgap	mean DFT band gap of elemental solid
9	Avg_dev NpUnfilled	average deviation of unfilled number of p orbitals
10	Avg_dev NpValence	average deviation of filled number of p orbitals
11	minimum space group number	minimum space group number
12	mode electronegativity	mode electronegativity
13	mode GSvolume_pa	mode DFT-computed volume of elemental solid
14	mode NUnfilled	mode number of unfilled valence orbitals

3.4.3. ML Models. We applied grid optimization to fine-tune the hyperparameters of the prediction models for the electronic structure properties and O atom adsorption energy of catalysts, using XGBR, RFR, and SVR algorithms. The hyperparameters for each model are shown in Tables S2−S13. Figure 4 represents the performance of the XGBR model in predicting the electronic structural properties of catalysts and the adsorption energy of O. The performance of RFR and SVR models is shown in Figures S2−S3.

Table 2 shows the performance of three machine learning models. Among them, the XGBR prediction model demonstrates the best prediction performance, showing the highest R^2, lowest MSE, and lowest MAE on the test set. In contrast, SVR has the worst prediction performance with the lowest R^2, and the highest MSE and MAE. The XGBR model constructed in this work has better predictive performance on catalyst properties compared to RFR and SVR models.

To further demonstrate the predictive advantages of the XGBR model developed in this work, a comparison was made between the previous used prediction models and the models proposed in this paper, as shown in Table 3. According to prior reports on predicting material properties, the XGBR model established in this paper for predicting the electronic structure properties and E_{ads}(O) of SACs demonstrates a relatively higher R^2 and lower MSE and MAE values. This indicates that the XGBR model provides superior prediction accuracy compared to the earlier works.

3.4.4. Predicting Catalytic Activity Based on ML. Previous studies[49,50] have demonstrated that the adsorption energy of O can serve as an indicator of the catalytic activity of SACs in various catalytic reactions. Therefore, in this work we combine the XGBR model, which predicts the adsorption energy of O atoms, with the activity volcano plot model for the catalytic oxidation of NO and Hg^0.[49,50] This has led to the development of a machine learning based framework for catalytic activity prediction (AML), enabling rapid prediction of SACs' catalytic activity for NO and Hg^0 oxidation, as shown in Figures 5b and S4. In this AML framework, the catalytic activity of SACs can be conveniently predicted by simply inputting the coordination

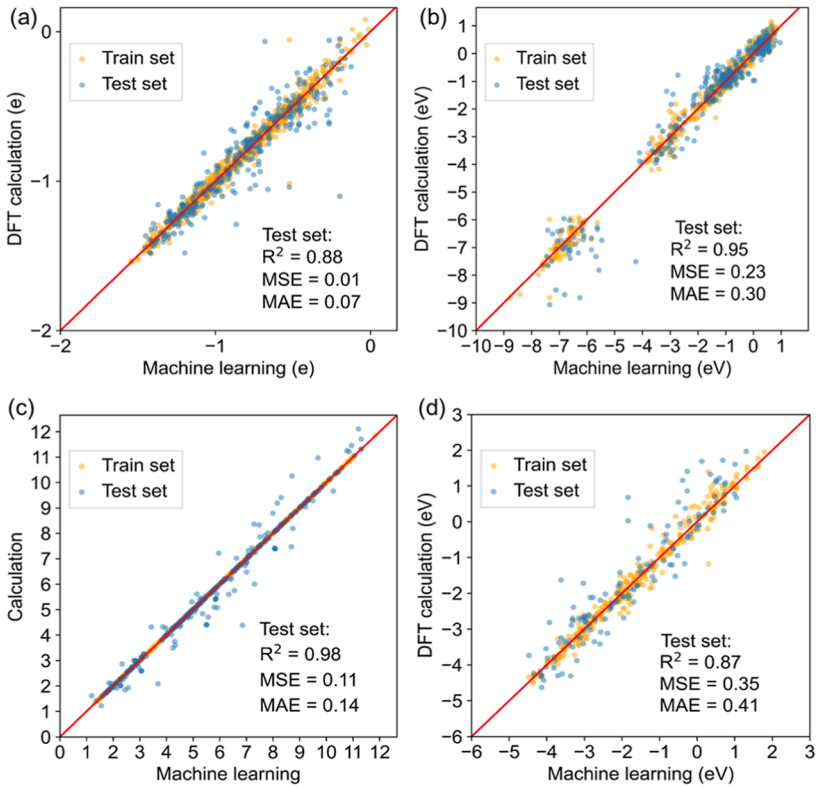


Figure 4. (a−d) Represent the XGBR prediction models for the metal Bader charge, metal d-band center, system electronegativity, and O adsorption energy of SACs, respectively.

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

Table 2. Predictive Performance of the Three Models on the Electronic Structure Properties and O Adsorption Energy of Catalysts (Note: The Table Shows the R^2/MSE/MAE of the Model on the Test Set)

model	Bader charge	d-band center	X	$E_{ads}(O)$
XGBR	0.88/0.01/0.07	0.95/0.23/0.31	0.98/0.11/0.14	0.87/0.35/0.41
RFR	0.87/0.01/0.08	0.95/0.28/0.33	0.96/0.27/0.14	0.86/0.33/0.43
SVR	0.67/0.03/0.13	0.69/1.57/0.64	0.67/2.12/0.72	0.66/0.79/0.60

Table 3. Comparison of the Predictive Performance Between Models from Previous Literature and the XGBR Model Proposed in This Work (Note: The Table Shows the R^2/MSE/MAE for Each Model on the Test Set; an Asterisk "*" Indicates That the Evaluation Value was Not Reported in the Literature)

algorithm	system	Bader charge	d-band center	$E_{ads}(O)$
ABR[61]	SACs	0.73/0.02/*		
RFR (V)[62]	metal oxides	0.83/*/0.08		
RFR (Cr)[62]	metal oxides	0.85/*/0.06		
RFR (Mn)[62]	metal oxides	0.80/*/0.06		
RFR (Fe)[62]	metal oxides	0.80/*/0.09		
RFR (Co)[62]	metal oxides	0.80/*/0.07		
GBR[63]	metals/bimetals		*/0.25/*	
GPR[64]	perovskites			*/0.90/0.67
RFR[65]	DACs/BNCs			0.80/*/0.50
SVR[65]	DACs/BNCs			0.79/*/0.32
K-NN[65]	DACs/BNCs			0.60/*/0.90
XGBR[this work]	SACs	0.88/0.01/0.07	0.95/0.23/0.31	0.87/0.35/0.41

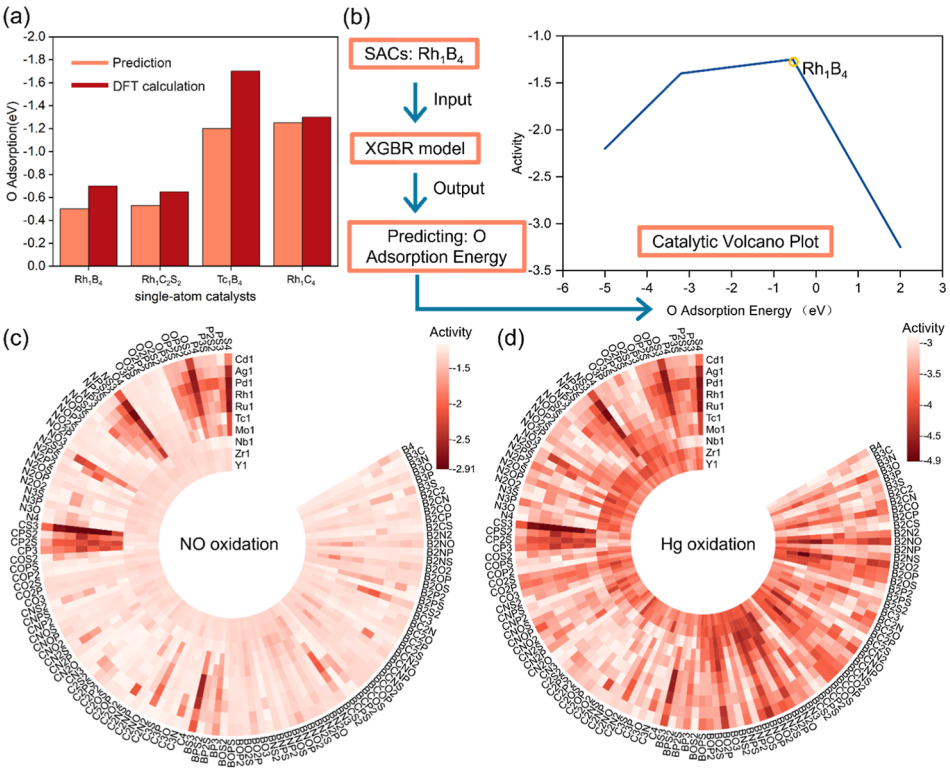


Figure 5. (a) Comparison of the predicted adsorption values for SACs with corresponding DFT calculated values. (b) Taking Rh_1B_4 as an example, the AML framework is demonstrated to quickly predict the catalytic activity for NO oxidation. (c) Prediction of the catalytic activity for NO oxidation across 1261 SACs with 4d metals. (d) Prediction of the catalytic activity for Hg^0 oxidation across 1261 SACs with 4d metals.

environment, such as Rh_1B_4. Based on this input, this framework automatically calculates the adsorption energy of O using the XGBR model. The predicted O adsorption energy is then integrated with the catalytic activity volcano plot to determine the catalytic activity of the SACs. Detailed information on AML can be found in the code availability section.

Based on this powerful framework, the catalytic activity for NO and Hg^0 oxidation of 1261 SACs containing 4d metals are rapidly calculated, as shown in Figure 5c,d. Since the accuracy of the predicted catalytic activity depends on the adsorption energy of O, the predicted adsorption energy of seven representative SACs is compared with DFT calculation results, as shown in Figure 5a. Although there are some discrepancies

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

between the predicted values and DFT results, this can be attributed to differences in the electronic structures between the training database (based on 3d metals) and the prediction data (based on 4d metals). Despite these differences, the AML framework remains fast and efficient for catalytic activity prediction. Among the 1261 SACs with 4d metals, Rh_1B_4 and $Rh_1C_2S_2$ are accurately identified as potential catalysts within minutes, indicating that this framework, based on DFT calculation database, is powerful for the design of SACs.

4. CONCLUSION

By regulating the coordination environment of SACs, a public database containing 1197 types of SACs were constructed through high-throughput DFT calculations. This public SAC database includes 657 stable SACs and their corresponding electronic structure properties (e.g., atom charge, electronegativity, d-band center) and adsorption energy of key reaction intermediate species. A ML model was developed using XGBR to rapidly predict the electronic structure properties and the O adsorption energy. The model demonstrates high predictive accuracy, with an $R^2 > 0.87$, a mean squared error (MSE) < 0.35, and a mean absolute error (MAE) < 0.41.

Through integrating our public SAC database and ML model, we developed a powerful framework for predicting the catalytic activity of SACs in the oxidation of NO and Hg^0 based on a previously reported theoretical catalytic volcano plot. This framework, referred to as AML, enabled the rapid screening of unreported SACs, identifying Rh_1B_4 and $Rh_1C_2S_2$ as potential catalysts from 1,261 candidates within a few minutes. The energy barriers for Rh_1B_4 in catalyzing NO and Hg^0 oxidation are 1.01 and 2.59 eV, respectively, while those for $Rh_1C_2S_2$ are 1.03 and 2.61 eV, respectively. More importantly, the publicly available SAC database, ML model, and the catalytic activity prediction framework reported in this work can provide new avenues for the rapid screening of highly active SACs for key reactions in the field of energy and environmental catalysis.

■ ASSOCIATED CONTENT

Data Availability Statement

The codes of AML project are available in https://github.com/Weijie-Yang/AML. The data of SACs database is available in https://catalysis-ncepu-hvkydg736ykqeq26d5gxrn.streamlit.app/.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.5c00491.

Equations, explanation of machine learning variables, computational model, research flowchart (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Zhengyang Gao − *Department of Power Engineering, North China Electric Power University, Baoding, Hebei 071003, China; Hebei Key Laboratory of Low Carbon and High-Efficiency Power Generation Technology, North China Electric Power University, Baoding, Hebei 071003, China*; Email: gaozhyan@163.com

Hao Li − *Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan; Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan*; orcid.org/0000-0002-7577-1366; Email: li.hao.b8@tohoku.ac.jp

Weijie Yang − *Department of Power Engineering, North China Electric Power University, Baoding, Hebei 071003, China; Hebei Key Laboratory of Low Carbon and High-Efficiency Power Generation Technology, North China Electric Power University, Baoding, Hebei 071003, China*; orcid.org/0000-0002-0232-1129; Email: yangwj@ncepu.edu.cn

Authors

Mingye Huang − *Department of Power Engineering, North China Electric Power University, Baoding, Hebei 071003, China; Hebei Key Laboratory of Low Carbon and High-Efficiency Power Generation Technology, North China Electric Power University, Baoding, Hebei 071003, China*

Ruiyang Shi − *Department of Power Engineering, North China Electric Power University, Baoding, Hebei 071003, China; Hebei Key Laboratory of Low Carbon and High-Efficiency Power Generation Technology, North China Electric Power University, Baoding, Hebei 071003, China*

Heng Liu − *Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan*

Wenjun Ding − *Department of Computer, North China Electric Power University, Baoding, Hebei 071003, China*

Jiahang Fan − *Department of Computer, North China Electric Power University, Baoding, Hebei 071003, China*

Binghui Zhou − *Department of Power Engineering, North China Electric Power University, Baoding, Hebei 071003, China; Hebei Key Laboratory of Low Carbon and High-Efficiency Power Generation Technology, North China Electric Power University, Baoding, Hebei 071003, China*

Bo Da − *Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan*; orcid.org/0000-0002-0785-8662

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.5c00491

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China (nos. 52176104 and 52006073), JSPS KAKENHI (no. JP24K23069) and Ensemble Grants for Early Career Researchers 2024. Part of the calculations in this study were performed on the Numerical Materials Simulator at NIMS.

■ REFERENCES

(1) Wang, A.; Li, J.; Zhang, T. Heterogeneous single-atom catalysis. *Nat. Rev. Chem* **2018**, *2* (6), 65−81.

(2) Cheng, N.; Zhang, L.; Doyle-Davis, K.; Sun, X. Single-Atom Catalysts: From Design to Application. *Electrochem. Energy Rev.* **2019**, *2* (4), 539−573.

(3) Yang, X.-F.; Wang, A.; Qiao, B.; Li, J.; Liu, J.; Zhang, T. Single-Atom Catalysts: A New Frontier in Heterogeneous Catalysis. *Acc. Chem. Res.* **2013**, *46* (8), 1740−1748.

(4) Zhang, D.; Hirai, Y.; Nakamura, K.; Ito, K.; Matsuo, Y.; Ishibashi, K.; Hashimoto, Y.; Yabu, H.; Li, H. Benchmarking pH-field coupled microkinetic modeling against oxygen reduction in large-scale Fe−azaphthalocyanine catalysts. *Chem. Sci.* **2024**, *15* (14), 5123−5132.

(5) Wang, Y.; Zhang, D.; Sun, B.; Jia, X.; Zhang, L.; Cheng, H.; Fan, J.; Li, H. Divergent Activity Shifts of Tin-Based Catalysts for Electrochemical CO2 Reduction: pH-Dependent Behavior of Single-Atom versus Polyatomic Structures. *Angew. Chem., Int. Ed.* **2025**, *64* (8), No. e202418228.
(6) Yu, Z.; Zhang, D.; Wang, Y.; Liu, F.; She, F.; Chen, J.; Zhang, Y.; Wang, R.; Zeng, Z.; Song, L.; et al. Spin Manipulation of Heterogeneous Molecular Electrocatalysts by an Integrated Magnetic Field for Efficient Oxygen Redox Reactions. *Adv. Mater.* **2024**, *36* (45), 2408461.
(7) Yu, B.; Li, H.; White, J.; Donne, S.; Yi, J.; Xi, S.; Fu, Y.; Henkelman, G.; Yu, H.; Chen, Z.; et al. Tuning the catalytic preference of ruthenium catalysts for nitrogen reduction by atomic dispersion. *Adv. Funct. Mater.* **2020**, *30* (6), 1905665.
(8) Zhang, D.; Wang, Z.; Liu, F.; Yi, P.; Peng, L.; Chen, Y.; Wei, L.; Li, H. Unraveling the pH-dependent oxygen reduction performance on single-atom catalysts: from single-to dual-sabatier optima. *J. Am. Chem. Soc.* **2024**, *146* (5), 3210−3219.
(9) Jiao, L.; Wang, J.; Jiang, H.-L. Microenvironment Modulation in Metal−Organic Framework-Based Catalysis. *Acc. Mater. Res.* **2021**, *2* (5), 327−339.
(10) Guo, Z.; Liu, C.; Sun, C.; Xu, J.; Li, H.; Wang, T. Tuning the Coordination Environment of Single-Atom Iron Catalysts Towards Effective Nitrogen Reduction. *ChemCatChem* **2023**, *15* (14), No. e202300669.
(11) Khedr, G. E.; Fawzy, S. M.; Sharafeldin, I. M.; Allam, N. K. Designing N, P-doped graphene surface-supported Mo single-atom catalysts for efficient conversion of nitrogen into ammonia: a computational guideline. *Nanoscale Advances* **2024**, *6* (16), 4160−4166.
(12) Li, J.; Xu, H.; Liao, Y.; Qiu, Y.; Yan, N.; Qu, Z. Atomically Dispersed Manganese on a Carbon-Based Material for the Capture of Gaseous Mercury: Mechanisms and Environmental Applications. *Environ. Sci. Technol.* **2020**, *54* (8), 5249−5257.
(13) Tang, Y.; Chen, W.; Shi, J.; Wang, Z.; Cui, Y.; Teng, D.; Li, Y.; Feng, Z.; Dai, X. Nitrogen and boron coordinated single-atom catalysts for low-temperature CO/NO oxidations. *J. Mater. Chem. A* **2021**, *9* (27), 15329−15345.
(14) Yang, W.; Zhou, B.; Jia, Z.; Wu, C.; Wei, L.; Gao, Z.; Li, H. Coordination Engineering of Single-Atom Iron Catalysts for Oxygen Evolution Reaction. *ChemCatChem* **2022**, *14* (22), No. e202201016.
(15) Yang, W.; Zhou, B.; Chen, L.; Shi, R.; Li, H.; Liu, X.; Gao, Z. Coordination engineering for single-atom catalysts in bifunctional oxidation NO and mercury. *Fuel* **2023**, *349*, 128751.
(16) Nishimura, S.; Le, S. D.; Miyazato, I.; Fujima, J.; Taniike, T.; Ohyama, J.; Takahashi, K. High-throughput screening and literature data-driven machine learning-assisted investigation of multi-component La2O3-based catalysts for the oxidative coupling of methane. *Catal. Sci. Technol.* **2022**, *12* (9), 2766−2774.
(17) Khatamirad, M.; Fako, E.; De, S.; Muller, M.; Boscagli, C.; Baumgarten, R.; Ingale, P.; d'Alnoncourt, R. N.; Rosowski, F.; Schunk, S. A. Data-driven Design of Enhanced In-based Catalyst for CO_2 to Methanol Reaction. . *ChemCatChem.* **2023**, *15*, 1−7.
(18) Zheng, X.; Peng, L.; Li, L.; Yang, N.; Yang, Y.; Li, J.; Wang, J.; Wei, Z. Role of non-metallic atoms in enhancing the catalytic activity of nickel-based compounds for hydrogen evolution reaction. *Chem. Sci.* **2018**, *9* (7), 1822−1830.
(19) Jia, X.; Li, H. Data mining of stable, low-cost metal oxides as potential electrocatalysts. *Artif. Intell. Chem.* **2024**, *2* (1), 100065.
(20) Jia, X.; Yu, Z.; Liu, F.; Liu, H.; Zhang, D.; Campos dos Santos, E.; Zheng, H.; Hashimoto, Y.; Chen, Y.; Wei, L.; et al. Identifying Stable Electrocatalysts Initialized by Data Mining: Sb2WO6 for Oxygen Reduction. *Advanced Science* **2023**, *11* (5), 2305630.
(21) Luo, Z.; Peng, J.; Mu, Y.; Sun, L.; Zhu, Z.; Liu, Z. Design of Cr-PNP catalysts for ethylene tri-/tetramerization assisted by a data-driven approach. *J. Catal.* **2023**, *428*, 115127.
(22) Zhou, L.; Tian, P.; Zhang, B.; Xuan, F.-Z. Data-driven rational design of single-atom materials for hydrogen evolution and sensing. *Nano Res.* **2024**, *17* (4), 3352−3358.
(23) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(24) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B:Condens. Matter Mater. Phys.* **1999**, *59* (3), 1758−1775.
(25) Argaman, N.; Makov, G. Density functional theory: An introduction. *Am. J. Phys.* **2000**, *68* (1), 69−79.
(26) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B:Condens. Matter Mater. Phys.* **1996**, *54* (16), 11169−11186.
(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(28) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew−Burke−Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999**, *110* (11), 5029−5036.
(29) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
(30) Xiao, G.; Lu, R.; Liu, J.; Liao, X.; Wang, Z.; Zhao, Y. Coordination environments tune the activity of oxygen catalysis on single atom catalysts: A computational study. *Nano Res.* **2022**, *15* (4), 3073−3081.
(31) Yu, M.; Trinkle, D. R. Accurate and efficient algorithm for Bader charge integration. *J. Chem. Phys.* **2011**, *134* (6), 064111.
(32) Khorasani, M.; Abdou, M.; Fernández, J. H. *Web Application Development with Streamlit*; Apress, 2022; Vol. *136*, p 154102.
(33) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.
(34) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM, 2016; pp 785−794.
(35) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(36) Zhang, F.; O'Donnell, L. J. Support vector regression. *Mach. Learn.* **2020**, 123−140.
(37) Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn: Machine Learning Without Learning the Machinery. *GetMobile Mob. Comput. Commun.* **2015**, *19* (1), 29−33.
(38) Cerqueira, T. F. T.; Sanna, A.; Marques, M. A. L. Sampling the Materials Space for Conventional Superconducting Compounds. *Adv. Mater.* **2023**, *36* (1), 2307085.
(39) Chicco, D.; Warrens, M. J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, No. e623.
(40) Liu, M.; Liu, C.; Gouse Peera, S.; Liang, T. Catalytic oxidation mechanism of CO on FeN2-doped graphene. *Chem. Phys.* **2022**, *559*, 111536.
(41) Janthon, P.; Kozlov, S. M.; Vines, F.; Limtrakul, J.; Illas, F. Establishing the Accuracy of Broadly Used Density Functionals in Describing Bulk Properties of Transition Metals. *J. Chem. Theory Comput.* **2013**, *9* (3), 1631−1640.
(42) Yang, W.; Xu, S.; Ma, K.; Wu, C.; Gates, I. D.; Ding, X.; Meng, W.; Gao, Z. Geometric structures, electronic characteristics, stabilities, catalytic activities, and descriptors of graphene-based single-atom catalysts. *Nano Mater. Sci.* **2020**, *2* (2), 120−131.
(43) Hammer, B.; Norskov, J. K. Electronic factors determining the reactivity of metal surfaces. *Surf. Sci.* **1995**, *343* (3), 211−220.
(44) Hammer, B.; Norskov, J. K. Why gold is the noblest of all the metals. *Nature* **1995**, *376* (20), 238−240.
(45) Hammer, B.; Nørskov, J. K. Theoretical Surface Science and Catalysis—Calculations and Concepts. *Adv. Catal.* **2000**, *45*, 71−129.

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053

(46) Bhattacharjee, S.; Waghmare, U. V.; Lee, S.-C. An improved d-band model of the catalytic activity of magnetic transition metal surfaces. *Sci. Rep.* **2016**, *6* (1), 35916.
(47) Xu, H.; Cheng, D.; Cao, D.; Zeng, X. C. A universal principle for a rational design of single-atom electrocatalysts. *Nat. Catal.* **2018**, *1* (5), 339−348.
(48) Grabow, L. C.; Hvolbæk, B.; Nørskov, J. K. Understanding Trends in Catalytic Activity: The Effect of Adsorbate−Adsorbate Interactions for CO Oxidation Over Transition Metals. *Top. Catal.* **2010**, *53* (5−6), 298−310.
(49) Yang, W.; Chen, X.; Feng, Y.; Wang, F.; Gao, Z.; Liu, Y.; Ding, X.; Li, H. Understanding trends in the mercury oxidation activity of single-atom catalysts. *Environ. Sci.: Nano* **2022**, *9* (6), 2041−2050.
(50) Yang, W.; Feng, Y.; Chen, X.; Wu, C.; Wang, F.; Gao, Z.; Liu, Y.; Ding, X.; Li, H. Understanding Trends in the NO Oxidation Activity of Single-Atom Catalysts. *J. Environ. Chem. Eng.* **2022**, *10* (6), 108744.
(51) Cheng, J.; Hu, P. Utilization of the Three-Dimensional Volcano Surface To Understand the Chemistry of Multiphase Systems in Heterogeneous Catalysis. *J. Am. Chem. Soc.* **2008**, *130*, 10868−10869.
(52) Deng, T.; Cen, C.; Shen, H.; Wang, S.; Guo, J.; Cai, S.; Deng, M. Atom-Pair Catalysts Supported by N-Doped Graphene for the Nitrogen Reduction Reaction: d-Band Center-Based Descriptor. *J. Phys. Chem. Lett.* **2020**, *11* (15), 6320−6329.
(53) Zhou, Y.; Zhou, Z.; Shen, R.; Ma, R.; Liu, Q.; Cao, G.; Wang, J. Correlating electrocatalytic oxygen reduction activity with d-band centers of metallic nanoparticles. *Energy Storage Mater.* **2018**, *13*, 189−198.
(54) Toyoda, E.; Jinnouchi, R.; Hatanaka, T.; Morimoto, Y.; Mitsuhara, K.; Visikovskiy, A.; Kido, Y. The d-Band Structure of Pt Nanoclusters Correlated with the Catalytic Activity for an Oxygen Reduction Reaction. *J. Phys. Chem. C* **2011**, *115* (43), 21236−21240.
(55) Li, H.; Li, X.; Wang, P.; Zhang, Z.; Davey, K.; Shi, J. Q.; Qiao, S.-Z. Machine Learning Big Data Set Analysis Reveals C−C Electro-Coupling Mechanism. *J. Am. Chem. Soc.* **2024**, *146* (32), 22850−22858.
(56) Li, H.; Hao, J.; Qiao, S.-Z. AI-Driven Electrolyte Additive Selection to Boost Aqueous Zn-Ion Batteries Stability. *Adv. Mater.* **2024**, *36* (49), 2411991.
(57) Li, H.; Cao, A.; Nørskov, J. K. Understanding trends in ethylene epoxidation on group IB metals. *ACS Catal.* **2021**, *11* (19), 12052−12057.
(58) Li, H.; Abraham, C. S.; Anand, M.; Cao, A.; Nørskov, J. K. Opportunities and challenges in electrolytic propylene epoxidation. *J. Phys. Chem. Lett.* **2022**, *13* (9), 2057−2063.
(59) Zhou, Y.; Gao, J.; Gui, Y.; Wen, J.; Wang, Y.; Huang, X.; Cheng, J.; Liu, Q.; Wang, Q.; Wei, C. Prediction of formation energies of UCr4C4-type compounds from Magpie feature descriptor-based machine learning approaches. *Opt. Mater.:X* **2022**, *16*, 100196.
(60) Kapse, S.; Narasimhan, S.; Thapa, R. Descriptors and graphical construction for in silico design of efficient and selective single atom catalysts for the eNRR. *Chem. Sci.* **2022**, *13* (34), 10003−10010.
(61) Wang, Z.; Jin, J.; Liu, M. Atomic Charge Schemes Comparison for Fe Single Atom in Graphitic Carbon: Insights from Quantum Simulations and Machine Learning. *J. Phys. Chem. C* **2023**, *127* (35), 17345−17354.
(62) Torrisi, S. B.; Carbone, M. R.; Rohr, B. A.; Montoya, J. H.; Ha, Y.; Yano, J.; Suram, S. K.; Hung, L. Random forest machine learning models for interpretable X-ray absorption near-edge structure spectrum-property relationships. *npj Comput. Mater.* **2020**, *6* (1), 109.
(63) Takigawa, I.; Shimizu, K.-i.; Tsuda, K.; Takakusagi, S. Machine-learning prediction of the d-band center for metals and bimetals. *RSC Adv.* **2016**, *6* (58), 52587−52595.
(64) Shambhawi, S.; Csányi, G.; Lapkin, A. A. Active Learning Training Strategy for Predicting O Adsorption Free Energy on Perovskite Catalysts using Inexpensive Catalyst Features. *Chem.:-Methods* **2021**, *1* (10), 444−450.
(65) Luo, Y.; Du, X.; Wu, L.; Wang, Y.; Li, J.; Ricardez-Sandoval, L. Machine-Learning-Accelerated Screening of Double-Atom/Cluster Electrocatalysts for the Oxygen Reduction Reaction. *J. Phys. Chem. C* **2023**, *127* (41), 20372−20384.

https://doi.org/10.1021/acs.jpcc.5c00491
J. Phys. Chem. C 2025, 129, 5043−5053