How to cite: *Angew. Chem. Int. Ed.* **2025**, e18027
doi.org/10.1002/anie.202518027

Two-Electron Water Oxidation Reaction

Universal Catalyst Design Framework for Electrochemical Hydrogen Peroxide Synthesis Facilitated by Local Atomic Environment Descriptors

Zhijian Liu, Yan Liu, Yuqi Zhang, Yeyu Deng, Zhong Zheng, Ruth Knibbe, Tianxiang Gao, Mingzhe Li, Ziye Wang, Bingqian Zhang, Xue Jia, Di Zhang, Heng Liu, Xuqiang Shao, Zhengyang Gao, Li Wei, Hao Li,* and Weijie Yang**

Abstract: A universal design framework for high-performance catalysts remains challenging due to diverse structures and active sites. We developed a framework integrating weighted atom-centered symmetry function (wACSF) descriptors with machine learning, microkinetic modeling, and high-throughput screening. The wACSF descriptors unify geometric and chemical characteristics of active sites across different catalyst families. ML models trained on wACSF accurately predicted adsorption free energies of hydroxyl (ΔG_{OH*}, $R^2 = 0.84$) and oxygen (ΔG_{O*}, $R^2 = 0.91$) for intermetallic alloys, metal oxides, perovskites, and single-atom catalysts in the two-electron water oxidation reaction ($2e^-$ WOR). Density functional theory and microkinetic modeling yielded a universal $2e^-$ WOR volcano model that agreed well with experiments. High-throughput screening with ML-predicted ΔG_{OH*} identified $LiScO_2$, which achieved 90% H_2O_2 Faradaic efficiency at 2.2 V vs. reversible hydrogen electrode (RHE) with 168-hour stability (82%–86% retention). Experimental activity ($\log(j) = 1.56$) matched theoretical predictions ($\log(j) = 1.28$) within 5% deviation at 2.4 V_RHE. This universal framework provides a general paradigm for rational catalyst design and is implemented in the Digital Catalysis Platform (*DigCat*), enabling efficient discovery across diverse material classes and electrochemical reactions.

Introduction

The sustainable electrochemical approach to hydrogen peroxide (H_2O_2) synthesis is regarded as a cost-effective way for H_2O_2 production, presenting an environmentally friendly and energy-efficient method for harnessing renewable energy sources.[1,2] This method also provides a robust energy storage solution by enabling the conversion of electrical energy into chemical energy in the form of H_2O_2, thereby bridging energy generation with environmental sustainability.[3]

Electrochemical H_2O_2 synthesis can follow two two-electron pathways: the oxygen reduction reaction ($2e^-$ ORR) and the water oxidation reaction ($2e^-$ WOR). The two routes are complementary; $2e^-$ ORR is widely studied and technologically mature but can suffer from O_2 mass-transfer and $4e^-$ selectivity losses, whereas $2e^-$ WOR avoids external O_2 supply and can simplify separation, offering potential economic advantages under certain scenarios (process-dependent electrolyte, reactor, and separation choices)[4] So far, metal oxides are the predominant anode materials for $2e^-$ WOR in H_2O_2 production because some of them have shown excellent stability under oxidizing and alkaline conditions.[5] Various metal oxides, such as TiO_2,[6] SnO_2,[7] Sb_2O_3,[8] ZnO,[9] $BiVO_4$,[10] $CaSnO_3$,[11] $CuWO_4$,[12] and $LaAlO_3$,[13] have been reported for H_2O_2 generation via $2e^-$ WOR. However, these metal oxides still exhibit relatively low intrinsic activity and conductivity; some of them are associated with relatively large overpotentials (*i.e.*, >1.0 V) to overcome

[*] Z. Liu, Y. Liu, T. Gao, M. Li, Z. Wang, B. Zhang, Z. Gao, W. Yang
School of Energy and Power Engineering, North China Electric Power University, Baoding 071003, P.R. China
E-mail: yangwj@ncepu.edu.cn

Y. Zhang, X. Shao
Department of Computer Science, North China Electric Power University, Baoding 071003, P.R. China

Y. Deng, L. Wei
School of Chemical and Biomolecule Engineering, The University of Sydney, Sydney, New South Wales 2006, Australia
E-mail: l.wei@sydney.edu.au

Z. Zheng, R. Knibbe
School of Mechanical and Mining Engineering, The University of Queensland, Brisbane 4072, Australia

X. Jia, D. Zhang, H. Liu, H. Li
Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980–8577, Japan
E-mail: li.hao.b8@tohoku.ac.jp

Additional supporting information can be found online in the Supporting Information section

the energy barriers, which in turn results in relatively low H_2O_2 yields (typically <5 μmol min^{-1} cm^{-2}).[14–16] A large-scale search for promising 2e⁻ WOR catalysts is as pressing as ever.

Previous research has established that the binding energies of different reactive intermediates (O*, OH*, and OOH*) serve as suitable descriptors for the activity and selectivity of the WOR pathway. Catalysts with optimal OH* binding are characteristic of the active 2e⁻ WOR. To achieve selectivity, the catalysts should exhibit weak binding to O* in order to suppress the 4e⁻ WOR pathway and prevent the further oxidation of H_2O_2 into O_2.[2] Although some progress has been achieved in understanding 2e⁻ WOR, most of them predominantly focused on elucidating the relationship between the theoretical overpotential and adsorption free energy of hydroxyl (ΔG_{OH*}) from a thermodynamic perspective using a theoretical "limiting-potential model".[2,17] However, only analyzing thermodynamics for high-electrode-potential electrocatalysis may sometimes lead to a large discrepancy between theory and experimental observations. For example, back to 2007, study by Vassilev and Koper[18] on the ORR performance has already revealed that density functional theory (DFT) based thermodynamic analysis alone cannot fully explain experimental observations. Besides, overpotential is often difficult to be well-defined in experiments, making it hard to perform a direct benchmarking analysis between theory and experiments to validate the model accuracy. To overcome these challenges, microkinetic modeling, which reveals the intricate mechanisms of chemical reactions occurring on the catalyst surface, has emerged as a powerful method to help enhance the understanding of each step within the reaction process. By considering the essential information of both kinetics and thermodynamics, microkinetic modeling can more precisely elucidate the specific reaction pathways, identify key intermediates, and clarify the relationship between reactant bonding strengths and the catalyst's performance, which can in turn predict the key electrochemical indicators (*e.g.*, current density, turnover frequency, Tafel slope) that can be well benchmarked with experimental measurements.[19] Nørskov and colleagues[20] developed the microkinetic volcano model (*i.e.*, the volcano activity model predicting the current density as a function of electrode potential and binding energies of key reaction intermediates) by considering the essential kinetics of ORR, leading to excellent agreement with experimental observations on transition metals in terms of current density at the potentials of interest. This model leads to much higher accuracy than the classic "limiting-potential volcano" for ORR developed two decades ago.[21] Similarly, microkinetic volcano model for 4e⁻ WOR (*i.e.*, the oxygen evolution reaction, OER) was also developed, which has offered profound insights into the electrochemical OER and filled in many knowledge gaps that the classic OER overpotential model[22] cannot explain. Unfortunately, to the best of our knowledge, understanding 2e⁻ WOR was still mainly rely on the limiting-potential model[23,24] – a more precise microkinetic model for the 2e⁻ WOR process has yet to be established.

Furthermore, even a precise microkinetic model is developed, screening promising catalyst candidates still heavy rely on DFT to directly compute the reaction descriptors (*e.g.*, adsorption free energies of key reaction intermediates on a catalyst surface), which is time-consuming and computationally expensive.[25,26] In recent years, machine learning (ML) has emerged as a powerful tool to accelerate catalyst design and screening.[27,28] However, most ML models and descriptors currently available are primarily tailored to specific catalyst families, such as alloys[29] or metal oxides,[9] limiting their generalizability. One commonly used descriptor in machine learning potential is the Atomic Center Symmetry Function (ACSF), which captures geometric features of atomic local environments.[30] ACSF has proven effective for extracting and describing the atomic-scale features of various materials, making it a promising tool for developing universal catalyst descriptors that can apply to different catalyst systems. However, while ACSF provides valuable insights into the geometric structure, its lack of the incorporation of chemical information at the active site makes it unreliable for predicting adsorption energies of reactants across different catalyst systems, limiting its applicability in the field of catalyst design. This underscores the need for a universal descriptor capable of comprehensively describing both the geometric and chemical characteristics of active sites across diverse catalyst families, such as metal alloys, metal oxides, and single-atom catalysts (SACs).

Motivated by the current limitations in catalyst design, herein we develop a universal and transferable catalyst design framework that integrates a newly constructed set of weighted atom-centered symmetry function (wACSF) descriptors with machine learning and microkinetic modeling. Unlike conventional ACSF-based approaches that mainly describe the geometric environment of atoms, the proposed wACSF descriptors incorporate both geometric and chemical features, including the intrinsic activity characteristics of the active-center atoms. This dual representation enables a unified description of diverse catalytic systems, overcoming the poor transferability of traditional ML-assisted methods. By applying the same predefined symmetry function parameters across various materials, wACSF maintains consistent dimensionality, thereby demonstrating its universality across diverse catalyst families, as exemplified by intermetallic alloys, metal oxides, perovskites, and SACs. Based on these descriptors, XGBoost regression (XGBR) models were constructed, enabling precise prediction of ΔG_{OH*} and ΔG_{O*} with high accuracy. Coupled with DFT and microkinetic analysis, a refined volcano model for the 2e⁻ WOR was further established, enabling rapid identification of high-performance electrocatalysts using ML-predicted adsorption energies. This combined framework not only enhances computational efficiency but also provides a generalizable and chemically meaningful strategy for rational catalyst design. Experimental validation using $LiScO_2$, which was predicted as the top-performing catalyst, shows excellent agreement with theoretical results, confirming the accuracy and robustness of our framework.

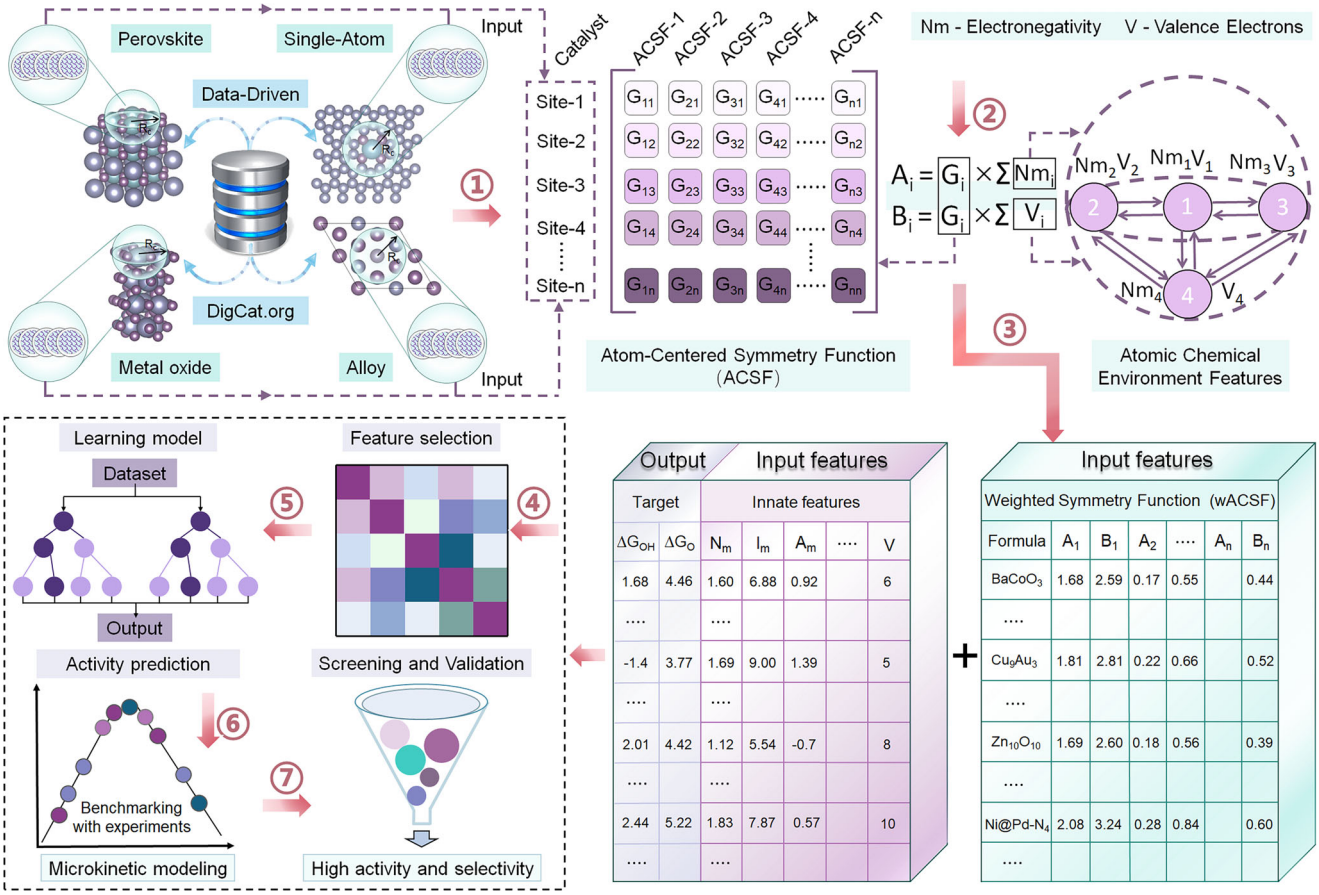

Figure 1. Overview of the catalyst design framework developed in this study. This framework comprises 1) feature extraction from the Digital Catalysis Platform (*DigCat*), 2) feature calculation, 3) data generation, 4) feature selection via Pearson correlation and recursive elimination, 5) ML training and testing, 6) microkinetic modeling and benchmarking analysis with available experimental data from the *DigCat* database, and 7) ML-accelerated catalyst screening.

Results

Overview of the Universal Framework

The proposed framework is outlined in Figure 1, adhering to the steps detailed therein. First, four typical types of materials are selected (Figure 1, **Step 1**), including metal alloys, metal oxides and perovskites, and SACs, from the recently developed *DigCat* database (www.DigCat.org). Afterward, the radial and angular symmetry function values (G_i) of active sites on various catalysts were calculated under different combinations of symmetry function parameters (Note 1 and Table S1).[31] Subsequently, the chemical environment features of the atoms were constructed by summing the electronegativity (Nm) and valence electron count (V) of the central and coordinating atoms of the catalysts.[32] The calculated G_i were then multiplied by the chemical environment features (ΣNm_i and ΣV_i) (Figure 1, **Step 2**). This process involves differentiating the contributions of G_i for each catalyst through weighting based on electronegativity and valence electron count, resulting in the "wACSF" descriptors (Supporting Information Note 2). The two types of weighted symmetry function values (A_i and B_i) were merged with the intrinsic feature values of the catalysts and used as the independent variables (Figure 1, **Step 3**). Specifically, we incorporated six intrinsic features (electronegativity, atomic number, atomic radius, first ionization energy, electron affinity and valence electrons) that depend solely on the primary metal and can be obtained from the periodic table (Supporting Information Dataset 1). Including these intrinsic features is helpful for developing models that involve adsorption sites across diverse metals. The adsorption free energy data of ΔG_{OH^*} and ΔG_{O^*} served as the target values (*i.e.*, the outputs of ML), establishing a connection between the adsorption characteristics at the catalytic sites and their numerical fingerprints. This process is fully automated through the implementation of our Python code (Supporting Information Note 3 and **Code Availability Section**). Subsequently, feature analysis and recursive elimination are conducted to discard redundant features (Figure 1, **Step 4**). Advanced ML algorithms are then employed to develop the effective model (Figure 1, **Step 5**). The optimal ML model will be used to predict the adsorption free energies of new catalysts. These adsorption free energies will be incorporated into a microkinetic volcano model, which will be benchmarked against experimental data to evaluate the predictive accuracy of the microkinetic model (Figure 1,

Step 6). Finally, potential high-performance catalysts will be rapidly identified by the microkinetic model through a screening process with the input of ML-predicted adsorption free energies (Figure 1, **Step 7**).

Database and Feature

We have assembled an adsorption free energy dataset with 962 materials from the *DigCat* database, comprising 426 intermetallic alloys, 160 metal oxides (excluding perovskites), 250 perovskites, and 126 SACs (Supporting Information Dataset 2). These datasets were originally extracted from published literature,[33–36] encompassing a variety of surfaces, adsorption sites, and values of adsorption free energies (*i.e.*, ΔG_{OH^*} and ΔG_{O^*}). The adsorption energy distributions of ΔG_{O^*} and ΔG_{OH^*} for four catalyst families were further examined to assess dataset balance (Figures S1 and S2). The violin plots show that all classes have broad and comparable energy ranges, confirming that the dataset is well balanced without dominance from any specific material type. Although the datasets were compiled from literature sources employing different computational parameters and exchange–correlation functionals, these functionals were selected according to the characteristics of each material class, which is known to yield adsorption energies in closer agreement with experimental values. A key advantage of this feature design strategy is that by applying the same set of predefined symmetry function parameters to all structures, these descriptors maintain consistent dimensionality across various materials. Therefore, each dataset was standardized to include 134 descriptors (Table S2), resulting in a 962 × 134 feature matrix with two endpoint vectors, ΔG_{OH^*} and ΔG_{O^*}.

To identify the key features that most strongly affect model performance, we applied recursive feature elimination (RFE) combined with Pearson correlation analysis (Supporting Information Notes 4 and 5).[37,38] This process yielded 21 important features for ΔG_{OH^*} and 19 for ΔG_{O^*} (Figure S3), including both atomic intrinsic properties and wACSF descriptors (Supporting Information Dataset 3). To minimize redundancy and mitigate overfitting, descriptors with correlation coefficients greater than 0.8 were further removed (Figure S4).[39] After retraining with the refined feature set, the model retained high predictive accuracy with only a slight decrease in R^2 and RMSE (Table S3), confirming that the selected descriptors provide sufficient information while ensuring model simplicity and robustness.

Performance and Validation of ML Models

For ML modeling, first, the entire dataset of adsorption energies was randomly partitioned into a training set (80%) and a test set (20%) (Supporting Information Note 6). The dataset was randomly partitioned into training and test sets using a fixed random seed to ensure reproducibility while maintaining statistical randomness. The training set was used for hyperparameter tuning through 10-fold cross-validation, while the unseen test set was reserved for independent validation. To ensure accurate prediction of adsorption free energies, multiple ML algorithms were evaluated, including random forest regression (RFR),[40] kernel ridge regression (KRR),[41] XGBR,[42] ridge regression (RR),[43] k-nearest neighbors (KNN),[44] gradient boosting regression (GBR),[45] support vector regression (SVR),[46] and extra trees regression (ETR)[47] (Supporting Information Note 7). For each model, multiple parameter combinations were tested to minimize bias in performance comparison. As summarized in Note 8 and Figure S5, models trained with RFR, XGBR, GBR, and KRR achieved higher predictive accuracy, while ETR, SVR, KNN, and RR showed inferior performance. Among them, XGBR exhibited the best overall accuracy, robustness, and scalability, effectively handling non-linear correlations and outliers while maintaining computational efficiency.[48,49] Therefore, XGBR was selected for developing the final ML models for ΔG_{OH^*} and ΔG_{O^*}, based on its consistently higher predictive accuracy and robustness across validation datasets. The optimized hyperparameter settings for the final models are provided in Table S4.

The cutoff radius (*Rc*) in the wACSF descriptors,[50] which defines the local atomic environment, was optimized to improve model performance (Figure S6). Different *Rc* values were tested, and the optimal parameters were identified as approximately 7 Å for ΔG_{OH^*} ($R^2 = 0.84$, RMSE = 0.52 eV) and 5 Å for ΔG_{O^*} ($R^2 = 0.91$, RMSE = 0.65 eV). ΔG_{OH^*} requires a larger *Rc* than ΔG_{O^*}, which is plausibly because the bonding strength of an adsorbate with a high electron affinity (*e.g.*, radical adsorbates like OH) is generally influenced by longer-range interactions.[51] These values represent the convergence of performance metrics with increasing *Rc*, indicating that adsorption free energies are mainly governed by the local environment while avoiding unnecessary computational cost.

Figure 2a–d shows the comparison between the ML-predicted adsorption free energies from the optimal XGBR model and those calculated by DFT. For ΔG_{OH^*}, the R^2 between DFT and ML predictions are 0.93 and 0.84, respectively in the training and test sets. The RMSE, used as the loss function to evaluate optimization errors, is 0.32 eV for the training set and 0.52 eV for the test set (Figure 2a,b). For ΔG_{O^*}, the R^2 values are 0.97 for the training set and 0.91 for the test set. The corresponding RMSE values are 0.36 eV for the training set and 0.65 eV for the test set (Figure 2c,d). The above results clearly demonstrate that these ML models can provide high accuracies in predicting adsorption free energies.

To further show the high accuracy of our method, Table S5 compares the performance of our model with previously reported methods for predicting adsorption energies.[49,52–58] Although our model yields slightly higher numerical errors than some task-specific ML models reported in the literature, it achieves comparable accuracy while providing much broader applicability. However, those approaches rely on system-specific descriptors that are highly tailored to particular catalyst families, which limits their transferability and general applicability across different material systems. In contrast, our framework employs wACSF descriptors that couple geometric and chemical characteristics of active sites while maintaining consistent feature dimensionality across diverse materials. This framework allows the model

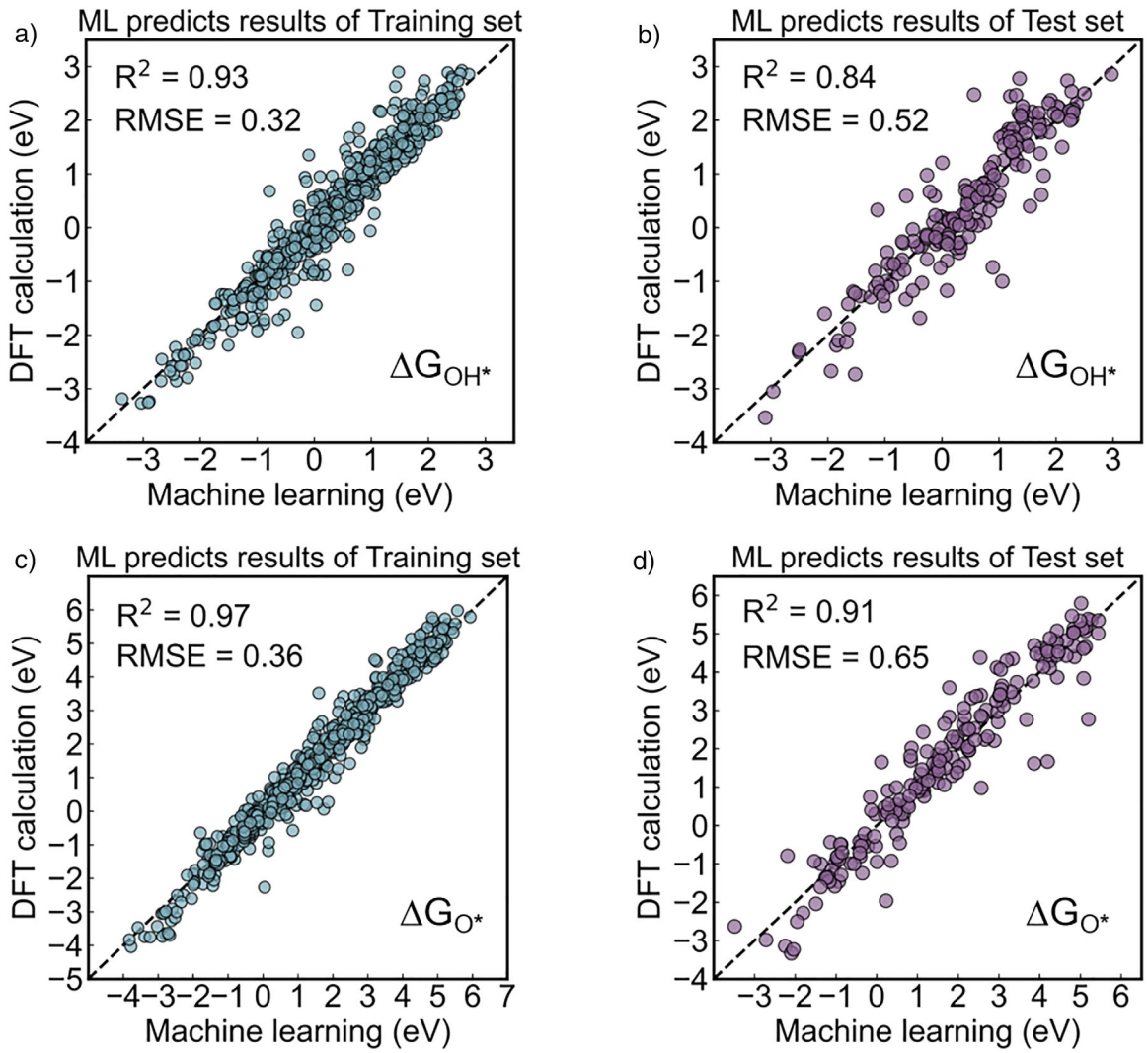

Figure 2. Evaluation of ML regression performance for adsorption free energy predictions. Panels a), b) display the comparison between the ΔG_{OH*} calculated by DFT and predicted by the ML model, respectively in the training and test sets. Panels c), d) display the comparison between the ΔG_{O*} calculated by DFT and ΔG_{O*} predicted by the ML model, respectively in the training and test sets.

to generalize beyond specific catalyst families and reliably predict ΔG_{OH*} and ΔG_{O*} involving more complex surface intermediates. This combination of high accuracy and strong generalization demonstrates that our approach provides a universal and chemically meaningful foundation for catalyst screening across diverse materials.

Importance of Descriptors and their Physicochemical Interpretations

To interpret how different features influence the predicted adsorption free energies, SHAP (SHapley Additive exPlanations) analysis was performed (Figure 3 and Supporting Information Note 9).[56] SHAP values quantify the contribution of each feature to ΔG_{OH*} and ΔG_{O*}, where broader distributions indicate higher feature sensitivity. For ΔG_{OH*}, the three most important descriptors are Nm (central atom electronegativity; importance 0.50), B6 (radial symmetry function weighted by coordination valence electrons; 0.19), and B21 (angular symmetry function weighted by coordination valence electrons; 0.18). For ΔG^{O*}, Nm (1.02), B19 (0.50), and B4 (0.32) show the highest importance. Therefore, accurately describing materials' performance in ML modeling requires considering the inherent properties of the central metal atom.[57]

The strong contribution of Nm demonstrates that adsorption energies are closely linked to the electron-donating and -accepting ability of the central atom. Lower Nm values correspond to weaker electron-withdrawing ability, leading to stronger adsorption, whereas higher Nm values favor weaker adsorption. This trend aligns with previous studies showing that differences in electronegativity modulate charge redistribution and adsorption affinity.[58] Similarly, the weighted radial (B4, B6) and angular (B19, B21) symmetry functions exhibit high SHAP value dispersion, indicating their crucial role in describing coordination geometry and electronic interactions between the central atom

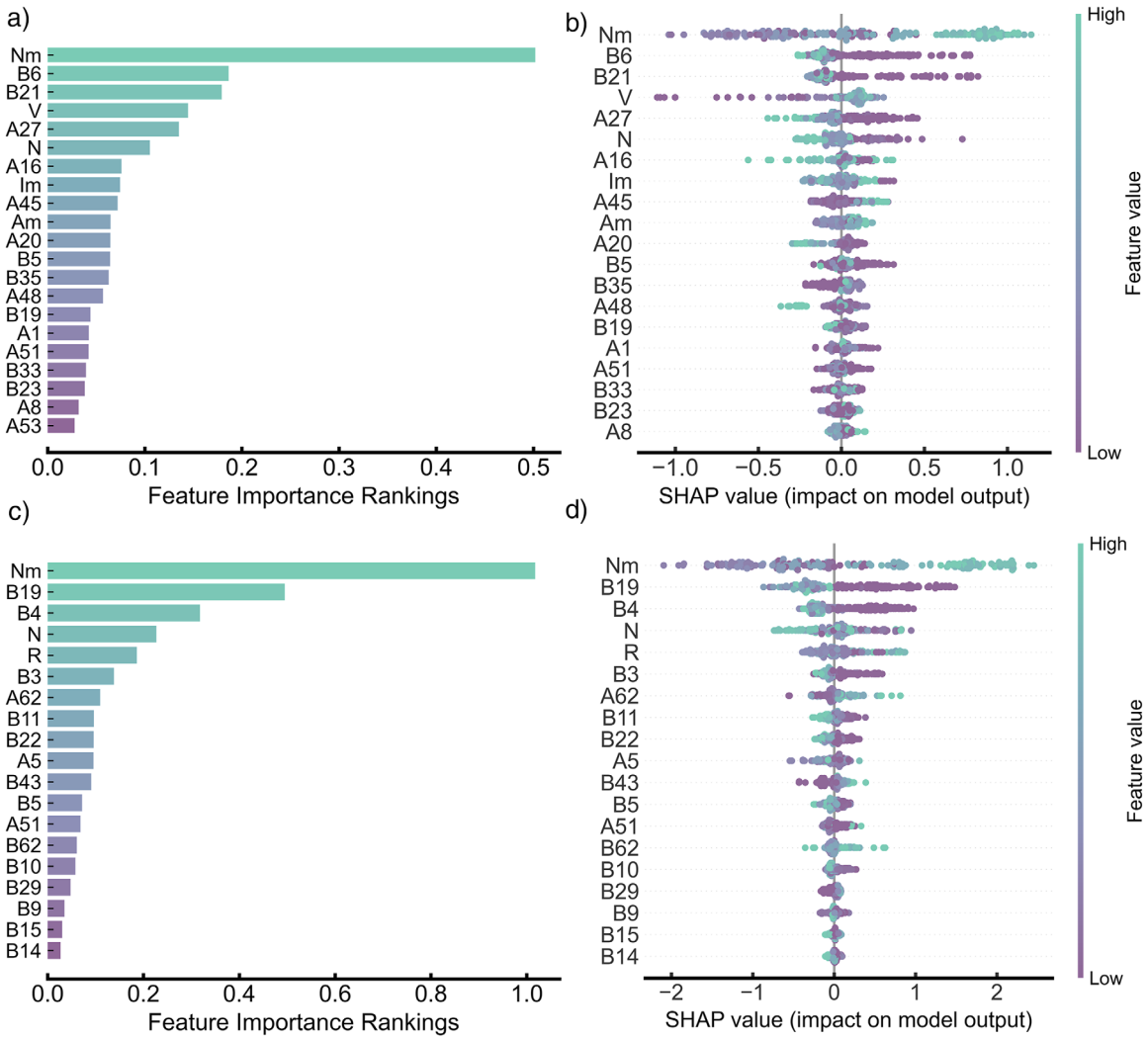

Figure 3. SHAP analysis of feature importance for various materials based on the XGBR model. a), c) Feature importance based on average SHAP values for the a) ΔG_{OH*} model and c) the ΔG_{O*} model. b), d) SHAP sensitivity analysis of individual features for the b) ΔG_{OH*} model and d) the ΔG_{O*} model, illustrating the impact of each feature on the model's output.

and neighboring atoms. These descriptors capture both radial and angular contributions of coordination atoms and reflect how local atomic environments affect adsorption strength.[38,50]

In summary, the top three important descriptors (*i.e.*, electronegativity, radial distribution, and angular distribution) comprehensively reflect the local chemical environment and electronic distribution between the central atom and coordination atoms from different physicochemical perspectives. These key features are new descriptors proposed by this study, highlighting their role as effective universal descriptors and their critical importance in predicting adsorption energies.

Microkinetic Modeling and Catalyst Screening

To accurately predict the activities of 2e⁻ WOR catalysts, a microkinetic model was developed to describe the relationship between the exchange current density (j_0) and the OH* binding free energy (ΔG_{OH*}) (Supporting Information Note 10). Figure 4a shows the simulated microkinetic volcano, benchmarked against experimental data from the *DigCat* database[59] and literatures (Table S6). The simulated microkinetic volcano model reproduces the experimental activity trends reported in the literature for 2e⁻ WOR catalysts, confirming the reliability of the kinetic analysis. Figure 4b compares the microkinetic volcano with the thermodynamic volcano derived from the limiting potential (Supporting Information Note 11). Both exhibit consistent trends, with peaks appearing at $\Delta G_{OH*} \approx 1.76$–$1.77$ eV. While the limiting-potential model captures general activity tendencies, it oversimplifies the reaction mechanism. In contrast, the derived microkinetic activity volcano plot provides a more accurate description of the rate-limiting steps and thus serves as a practical guideline for the subsequent rational design of high-performance 2e⁻ WOR catalysts.

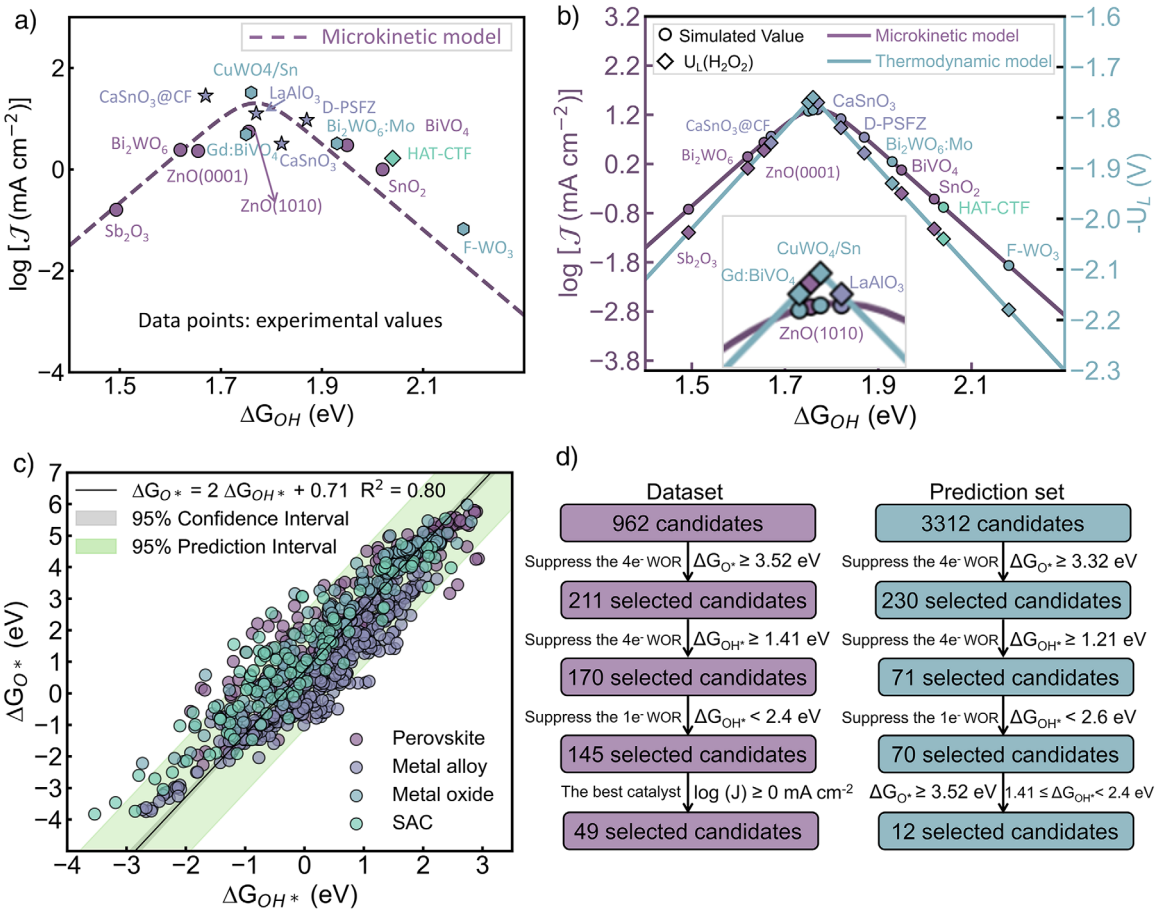

Figure 4. Microkinetic and Thermodynamic Volcano Models for 2e⁻ WOR and the Catalyst Screening Process. a) Simulated microkinetic volcano model for 2e⁻ WOR at 2.4 V versus RHE. The data points were extracted from experiments (Table S6), categorized by catalyst type: purple circles denote metal oxides, blue pentagrams denote perovskites, cyan hexagonal diamonds denote metal-doped metal oxides, and blue-green diamonds denote metal-free organic network materials. Each symbol corresponds to a specific catalyst type as indicated by the consistent color scheme. b) Microkinetic and thermodynamic volcano models, illustrating the theoretical limiting potentials (represented by diamonds) and simulated exchange current densities (represented by circles) for various catalysts. The matching colors of circles and diamonds indicate the same catalyst type. c) Scaling relation between ΔG_{OH*} and ΔG_{O*} for 962 catalyst surfaces. d) Workflow of the data screening process, illustrating the number of candidates selected after each screening step.

At an operating potential of 2.4 V versus reversible hydrogen electrode (RHE), catalysts such as $CaSnO_3$@CF,[60] $CuWO_4$/Sn,[61] $LaAlO_3$/FTO,[14] ZnO/FTO (1010),[9] and Gd-doped $BiVO_4$ (Gd:$BiVO_4$)[10] exhibit the highest simulated current densities near the apex of the microkinetic volcano model (Figure 4a). These catalysts are therefore identified as the best-performing 2e⁻ WOR electrocatalysts under the modeled conditions, and this trend is consistent with the experimental data reported in the *DigCat* database. In addition to conventional metal oxides, perovskites such as $Pr_{1.0}Sr_{1.0}Fe_{0.75}Zn_{0.25}O_{4-\delta}$ (D-PSFZ),[62] metal-free organic networks like covalent triazine frameworks hexaazatriphenylenes (HAT-CTF),[63] and doped or modified metal oxides also show promising performance. These further validate the microkinetic model's accuracy and applicability, demonstrating exceptionally good consistency with experimental results across various types of materials.

Selectivity is another key criterion for 2e⁻ WOR catalyst screening. Based on thermodynamic considerations (Note 11 and Figure S7), catalysts should exhibit $\Delta G_{OH*} < 2.4$ eV to suppress OH· radical formation and $\Delta G_{O*} \geq 3.52$ eV to favor H_2O_2 formation over OER. Figure 4c presents the scaling relationship between ΔG_{O*} and ΔG_{OH*} ($\Delta G_{O*} = 2\Delta G_{OH*} + 0.71$). The slope of the proportional line for the catalyst was found to be ~2, aligning with the previously reports for metal oxide catalysts.[63,64] This relationship defines the lower bound of $\Delta G_{OH*} \approx 1.41$ eV. Accordingly, systems meeting $1.41 \text{ eV} \leq \Delta G_{OH*} < 2.4$ eV and $\Delta G_{O*} \geq 3.52$ eV were selected as the most promising candidates for 2e⁻ WOR.

Theoretical Selection and Validation of Highly Active and Selective 2e⁻ WOR Catalysts

The workflow for the screening process is exemplified in Figure 4d. Initially, a preliminary screening was conducted on 962 catalysts, resulting in the selection of 145 candidates that met the criteria (Supporting Information Dataset 4). These

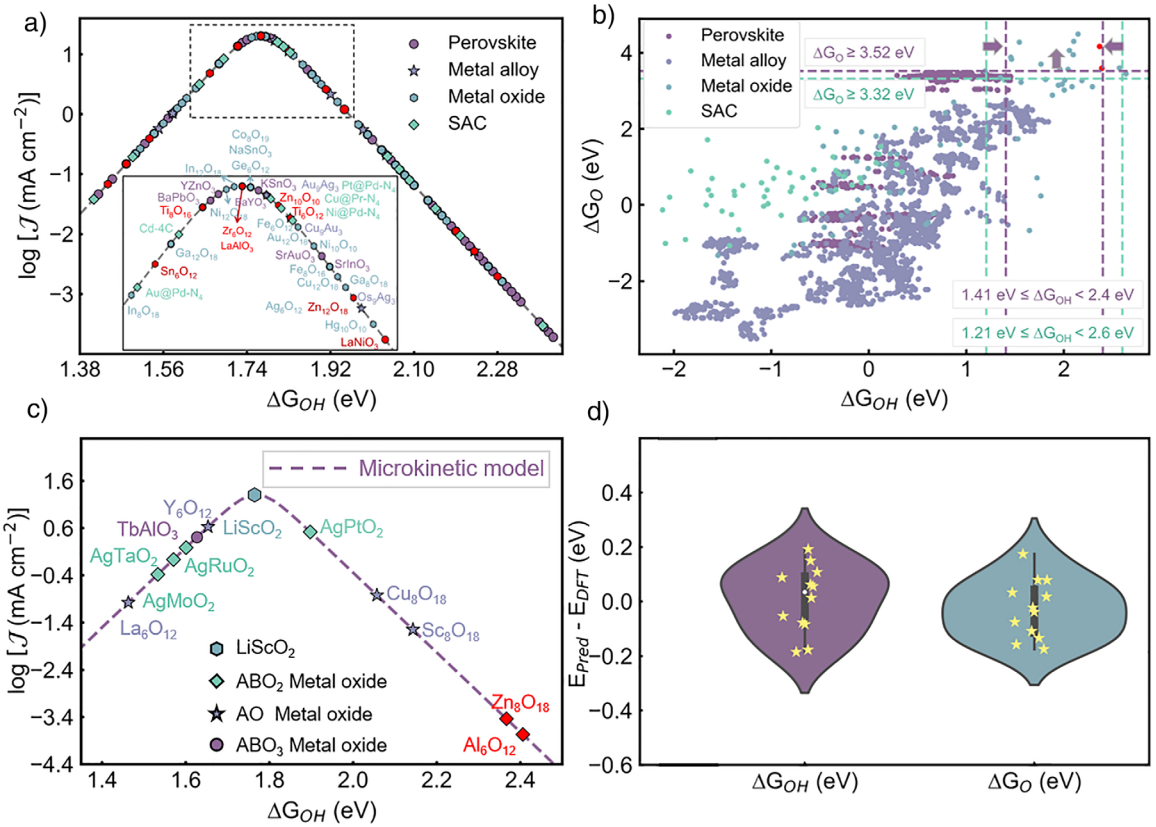

Figure 5. Performance Analysis and Validation of the Selected Catalysts. a) Microkinetic volcano model predicts the 2e⁻ WOR performance of 145 materials from the preliminary screening process in the *DigCat* database. b) Selectivity plot for the ML-predicted adsorption free energies of 3312 materials. c) Microkinetic volcano model plotted with the 12 materials selected from the prediction from (b). d) Violin plot showing the distribution of differences between the predicted and DFT-calculated adsorption free energies for the 12 materials selected from (c).

catalysts demonstrated OH* adsorption energies within an optimal range of 1.41 to 2.4 eV, with ΔG_{OH^*} values consistent with the conditions required for high selectivity in H_2O_2 production. After applying a more stringent criterion, specifically log(J) ≥ 0 mA cm^{-2}, the candidates were shortlisted to 49 (Supporting Information Dataset 4).

These 145 catalysts were then subjected to further microkinetic modeling analysis (Figure 5a). After additional literature search, catalysts marked with stars in the figure, such as $LaAlO_3$, $Zn_{10}O_{10}$, and Ti_6O_{12}, have also been synthesized in experiments and demonstrated excellent 2e⁻ WOR activities.[5,9,13] This suggests a high likelihood of successful experimental perspective for these ML-predicted catalysts, demonstrating that our predictive model and the screening process are reliable and effective in identifying efficient catalysts with promising practical applications.

To perform a larger-scale catalyst screening, a total of 3312 new structures were created across four categories of catalysts: 2381 metal alloys, 311 metal oxides (excluding perovskites), 570 perovskites, and 50 SACs, to predict ΔG_{OH^*} and ΔG_{O^*} (Supporting Information Dataset 4). Boron-doped diamond (BDD) is a representative catalyst for electrochemical H_2O_2 synthesis, but it was not included in this study because the active sites and electrochemical mechanisms of pure sp^2/sp^3 carbon materials remain under debate, which makes adsorption-energy-based modeling less reliable.[65] In principle, however, the proposed wACSF framework is equally applicable to such systems and can be adapted through reparameterization in future studies. Similarly, a preliminary screening was performed on these 3312 materials in the prediction set to validate the selectivity and activity screening criteria. Figure 5b depicts the adsorption free energies predicted by the ML model developed above. To enhance screening accuracy, a prediction error margin of 0.20 eV was considered. This adjustment aims to minimize uncertainty in model predictions and identify candidates with greater practical potential. The screening criteria were set as 1.21 eV ≤ ΔG_{OH^*} < 2.6 eV and $\Delta G_{O^*} \geq 3.32$ eV, leading to the finalist with 70 suitable catalysts (Supporting Information Dataset 4), including metal oxides and perovskites, and metal alloys. These candidates have potential for H_2O_2 production in the 2e⁻ WOR process. Notably, Zn_8O_{18}[36] and Al_2O_3[66] have been experimentally confirmed to exhibit 2e⁻ WOR catalytic activity. This demonstrates that our screening criteria effectively identify materials with practical catalytic potential and provides valuable references for future catalyst development and optimization.

Based on the established adsorption-energy relationships, stringent screening criteria (1.41 eV ≤ ΔG_{OH^*} < 2.4 eV and $\Delta G_{O^*} \geq 3.52$ eV) were applied, yielding 12 promising catalysts (Supporting Information Dataset 4). These catalysts were further analyzed using the microkinetic model (Figure 5c).

All the selected catalysts exhibited 2e⁻ WOR activity and selectivity, with those marked in red having been experimentally validated for catalytic activity. $LiScO_2$ showed the highest activity, occupying the top regions of both models, demonstrating their exceptional catalytic performance and selectivity. DFT validation results indicate that the differences between the DFT-calculated and ML-predicted adsorption free energies of ΔG_{OH^*} and ΔG_{O^*} (Supporting Information Note 12) are less than 0.2 eV (Figure 5d), confirming the high accuracy of the model. Moreover, we further examined the uncertainty of these predictions, as shown in Figure S8. The DFT-calculated and ML-predicted adsorption free energies (ΔG_{OH^*} and Δ_{GO^*}) show strong consistency, with all data points falling within the 95% confidence interval of the fitted correlation. In addition, t-SNE analyses of the feature-space distribution indicate that these 12 predicted catalysts are located within or close to the dense region of the training dataset (Figures S9 and S10), confirming that the screened catalysts lie within the well-learned region of the model and that the predictions are statistically reliable.

These 12 predicted catalysts were compared with literature-reported values (Table S7), revealing that most of them, including $LiScO_2$, $TbAlO_3$ and Y_6O_{12}, all of which exhibit η_{WOR} overpotentials close to 0 V, with catalytic activity and selectivity comparable to the best-performing catalysts reported in the literature, such as ZnO(1010), WO_3, and $CaSnO_3$.[9,11,67] This result also highlights the potential of single-atom catalysts (SACs) as effective 2e⁻ WOR catalysts, alongside metal oxides and perovskites. Overall, the screening framework effectively identifies non-precious, high-performance 2e⁻ WOR catalysts and provides a reliable theoretical basis for guiding subsequent studies on H_2O_2 electrosynthesis.

Experimental Verification of the Predicted High-Performance Catalysts

To verify our theoretical prediction, $LiScO_2$ was selected as the target compound because its predicted properties place it at the top of the volcano plot for catalytic activity, indicating superior performance among the candidates. It was prepared by a solid-state reaction method following a reported method (Supporting Information Note 13).[68] Inductively coupled plasma atomic emission spectroscopy measurement confirmed the product is of stoichiometry. X-ray diffraction (XRD) pattern displayed in Figure 6a suggests that the as-prepared sample is tetragonal with space group of I41/amd, showing good agreement to the standard card (PDF#76–1701). At low magnification, transmission electron microscope (TEM) characterization suggests that the as-prepared $LiCoO_2$ catalysts are infused from smaller particles (Figure S11). Under high-resolution observation, clear lattice fringes can be identified, which extended into the bulk crystal, as shown in Figure 6b. Further magnification of the red-rectangle section in panel (b) is displayed in Figure 6c, where the lattice fringes can be assigned to the $LiScO_2$ (101) plain with a d-spacing of *c.a.* 0.382 nm. Moreover, a less crystallized surface layer with a thickness of ~1 nm can be further identified, as marked by the red dashed lines. This layer can be attributed to surface passivation, which is further confirmed by surface sensitive X-ray photoelectron spectroscopy (XPS) measurement, as discussed later. Energy-dispersive X-ray elemental mapping results (Figure 6d) show the uniform distribution of Sc and O elements in the sample. Li is not shown since it could not be accurately detected by EDX.

X-ray photoelectron spectroscopy (XPS, Figure 6e) survey scan further confirmed the existence of Li, Sc and O in the sample. High-resolution spectra collected at the Li 1 *s*, Sc 2*p* and O 1 *s* regions are further analyzed and displayed in Figure 6f. Li 1 *s* exhibited a single peak at ~53.4 eV, corresponding to Li^+. Deconvolution of Sc 2*p* spectrum resulted in a pair of peaks with a peak area ratio of 1.67 in both Sc $2p_{3/2}$ and $2p_{1/2}$ regions. The peaks at 401.0 and 405.2 eV can be assigned to the Sc in oxide form and the peaks with higher binding energies (402.4 and 406.8 eV) can be assigned to the formation of hydrated Sc-OOH phase at the surface of the sample.[69] This observation also agrees with the O 1 *s* spectrum, which can be further deconvoluted into three peaks, i.e., the Sc-O in oxides (~529.8 eV), hydroxide and hydrated species (531.8 eV), and small amount of surface adsorbed water (533.8 eV). Sush a hydrated surface corresponds to the amorphous layer found under TEM observation.

The WOR performance of the catalysts are assessed in an argon (Ar)-saturated 0.5 M $KHCO_3$ electrolyte by a rotary ring-disk electrode (RRDE) method (Note 13 and Figure S12). A heterogenous molecular catalyst was also synthesized as a reference for comparison by loading nickel (II) phthalocyanine molecules on purified carbon nanotube (NiPc/CNT) following our reported method,[70] which, according to our model, preferentially follows 4e⁻ WOR pathway. The top panel of Figure 6g shows the linear sweep voltammetry (LSV) curves of the $LiScO_2$ and NiPc/CNT catalysts. In the weak alkaline electrolyte (pH = 9.0), both catalysts exhibited similar onset potential on the disk electrode while the current density of $LiScO_2$ could quickly surpass that of the NiPc/CNT. Meanwhile, a huge difference in ring current density can be observed, where the generated H_2O_2 via 2e⁻ WOR can be readily reduced by the cathodically biased Pt ring electrode. $LiScO_2$ reached a maximum FE_{H2O2} of about 90% at 2.2 V_{RHE} and maintained over 80% up to 2.4 V_{RHE}. In contrast, NiPc/CNT favored the 4e⁻ WOR pathway, showing a FE_{H2O2} below 4%. These two catalysts, representing high ($LiScO_2$) and low (NiPc/CNT) performance systems, were experimentally examined to validate the predictive reliability of the proposed framework. As shown in Figure 6h, for $LiScO_2$ with a ΔG_{OH^*} of 1.75 eV, the experimental and predicted 2e⁻ WOR partial current densities at 2.4 V_{RHE} are nearly identical. A similar consistency was observed for the low-activity NiPc/CNT reference catalyst, confirming that the model can accurately distinguish between efficient and inefficient materials. The ML model was trained using static adsorption energies, which capture the intrinsic catalytic properties of materials. While environmental effects such as surface restructuring and electrolyte interactions are not explicitly included, the predicted results remain in excellent agreement with experiments (Figures 4a and 6h), confirming

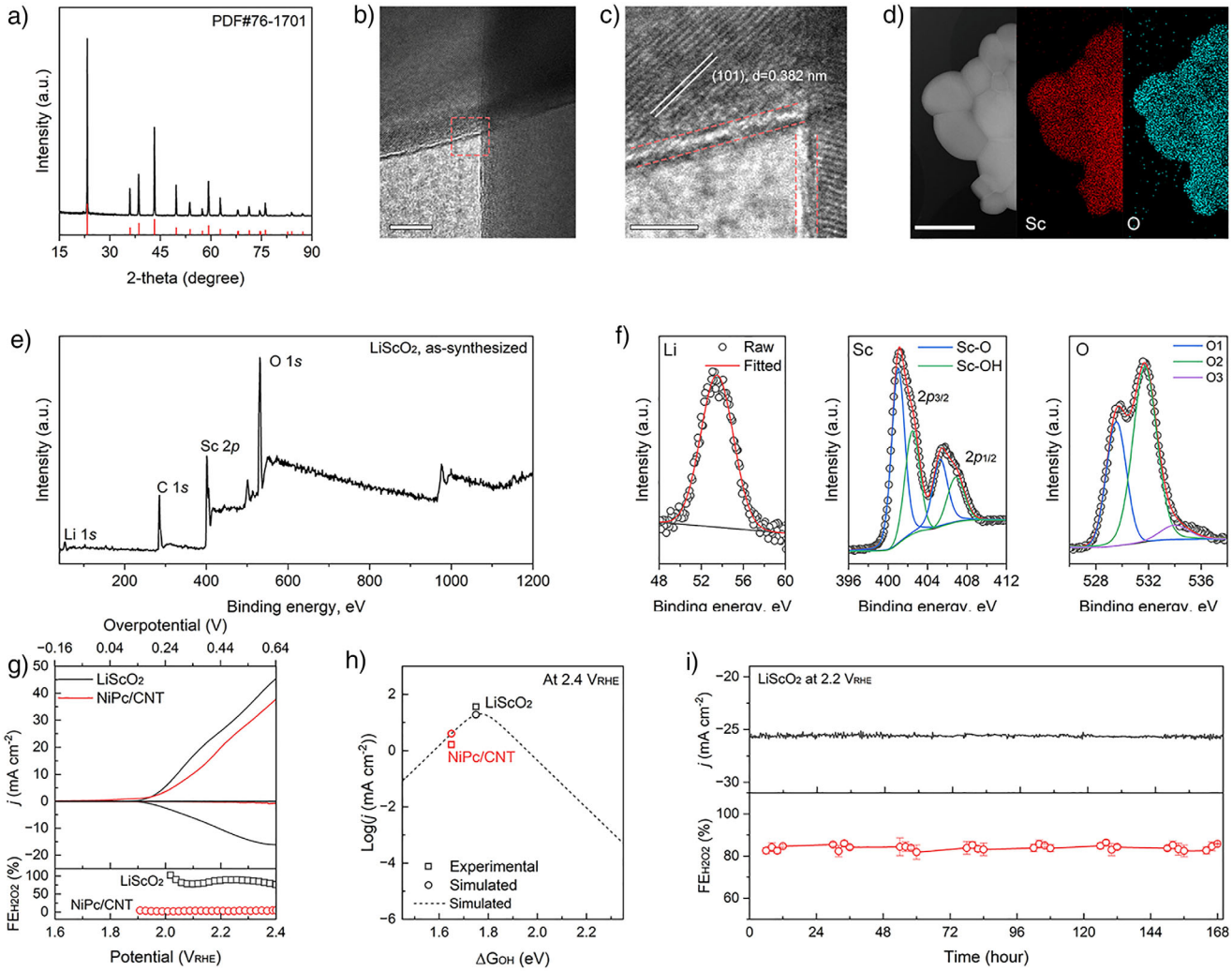

Figure 6. Experimental validation of theoretical prediction. a) XRD pattern. b) A high-resolution TEM image of the catalyst collected near its surface. Scale bar = 10 nm. c) The magnified rectangle region in panel (b). Scale bar = 4 nm. d) A STEM image and its corresponding EDX elemental mapping results. Scale bar = 1 μm. e) XPS survey scan and f) the high-resolution spectra of Li 1*s*, Sc 2*p*, and O 1*s* features. g) WOR LSV curve (top) and the corresponding H_2O_2 Faradaic efficiency (FE_{H2O2}) of $LiScO_2$ and a reference NiPc/CNT catalyst. Tests were performed in an Ar-saturated 0.5 M $KHCO_3$ electrolyte with electrode rotating at 1600 rpm. Ring electrode was kept at 0.4 VRHE. h) Performance comparison between experimental results and theoretical predictions. i) Long-term stability test performed in a flow cell electrolyzer using 0.5 M $KHCO_3$ electrolyte and the periodically assessed FE_{H2O2}.

that the framework effectively describes the key activity trends relevant to catalyst design.

The stability of $LiScO_2$ was further evaluated in a flow cell using a 0.5 M $KHCO_3$ electrolyte at 2.2 V_{RHE} (Figure S13). The catalyst loaded on carbon paper exhibited a stable current over 168 hours (Figure 6i, top), while the Faradaic efficiency for H_2O_2 remained steady at 82~86% (Figure 6i, bottom). The $LiScO_2$ catalyst could afford an average production rate of 0.40 ± 0.006 mmol cm^{-2} h^{-1} (Figure S14). Post-test XPS (Figure S15) and TEM (Figure S16) analyses revealed no detectable structural degradation, confirming the excellent durability of $LiScO_2$. To further validate the catalytic kinetics, a Tafel analysis gave a slope of 69 mV dec^{-1} (Figure S17), and the turnover frequency (TOF), assuming all Sc atoms as active sites, was estimated to be 0.078 s^{-1} at 2.2 V_{RHE}. These results collectively confirm the high intrinsic activity and stability of $LiScO_2$, consistent with the predictions of our framework.

Conclusion

In this study, we have established a universal and extensible framework for rational catalyst design, integrating database automation, machine learning (ML), and microkinetic modeling through the development of weighted atom-centered symmetry function (wACSF) descriptors. Unlike conventional geometry-only descriptors, the wACSF incorporates both geometric and chemical features, including the intrinsic activity characteristics of active-site atoms, enabling a unified and transferable representation of local environments across

diverse catalyst families. This framework employs wACSF descriptors together with advanced ML models to accurately capture the correlations between local atomic environments and adsorption free energies (ΔG_{OH^*} and ΔG_{O^*}) for a wide range of catalysts, such as metal alloys, metal oxides, perovskites, and single-atom catalysts (SACs). Combined with high-accuracy microkinetic volcano modeling, it provides a powerful predictive tool for identifying high-performance $2e^-$ water oxidation reaction ($2e^-$ WOR) catalysts across multiple material systems.

Experimental validation of the top-predicted $LiScO_2$ catalyst demonstrated exceptional $2e^-$ WOR performance with 90% H_2O_2 Faradaic efficiency at 2.2 V_{RHE}, maintaining > 80% selectivity up to 2.4 V_{RHE} with appreciable current densities. The catalyst exhibited 168-hour operational stability in flow cell tests (82%–86% FE_{H2O2} retention), with post-test characterization confirming structural integrity through preserved lattice spacing (0.382 nm (101) plane) and surface hydration layer (~1 nm). Notably, the experimental activity metric ($\log(j_0) = 1.56$ mA cm^{-2}) aligned within 5% deviation from theoretical predictions ($\Delta G_{OH^*} = 1.75$ eV, $\log(j_0) = 1.28$ mA cm^{-2}), establishing the tightest theory-experiment correlation reported for $2e^-$ WOR catalysts to date.

In summary, this work presents a generalizable and data-driven framework that unifies structural, chemical, and microkinetic information for catalyst design. By integrating wACSF, ML, microkinetic modeling, and an automation framework built on the Digital Catalysis Platform (*DigCat*), the proposed framework enables cross-material catalyst screening and mechanistic insight within a unified design paradigm. This approach not only accelerates the discovery of high-performance catalysts but also provides a transferable foundation for extending data-driven design principles to other catalytic reactions, such as ORR and OER, where ΔG_{OH^*} and ΔG_{O^*} remain universal descriptors of reactivity.

Author Contributions

Z.L. and Z.G. mentored, edited and financed the work. All authors contributed to manuscript proofreading. Y.L., B.Z., T.G., and X.S. participated in the collection and establishment of the machine learning dataset. Y.L.,Y.Z., and M.L. contributed to the coding work. X.J. performed detailed revisions of the machine learning work. H.L., D.Z., and Z.W. contributed to the revision and editing of the manuscript. W.Y. conceptualized the study. W.Y., H.L., and L.W. provided guidance. Y.L. was the major contributor to the writing of the manuscript. Y.D., Z.Z., and R.K. contributed to the section of experiment validation. All authors have read and approved the final manuscript.

Acknowledgements

This work was funded by the Outstanding Youth Team Project Fund (No. 2752023YQ001), the Distinguished Youth Science Foundation of Hebei Province (No. E2023502015), the National Natural Science Foundation of China (Nos. 52176104 and 52006073), JSPS KAKENHI (Nos. JP25H01508, JP25K01737, JP25K17991, JP24K23068, and JP24K23069), Australian Research Council Future Fellowship (FT210100218), and Linkage Project (LP230200886). The authors acknowledge the Center for Computational Materials Science, Institute for Materials Research, Tohoku University for the use of MASAMUNE-IMR (Nos. 202412-SCKXX-0211, 202412-SCKXX-0209, and 202412-SCKXX -0203), the Institute for Solid State Physics (ISSP) at the University of Tokyo, and National Computational Infrastructure (NCI, NCMAS-2025–116) for the computational resources.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

All relevant data in this study are available in the Supporting information.

Code Availability

The custom code developed in this study is available at https://github.com/Weijie-Yang/FG-LAED. The microkinetic models of $2e^-$ WOR is available in our CatMath platform: https://catmath.cloud/. All the experimental data and computational structures can also be found in our DigCat database: https://www.digcat.org/.

Keywords: AI for catalyst design · Atomic descriptors · Hydrogen peroxide electrosynthesis · Machine learning · Water oxidation

[1] X. Shi, S. Siahrostami, G.-L. Li, Y. Zhang, P. Chakthranont, F. Studt, T. F. Jaramillo, X. Zheng, J. K. Nørskov, *Nat. Commun.* **2017**, *8*, 701, https://doi.org/10.1038/s41467-017-00585-6.
[2] X. Shi, S. Back, T. M. Gill, S. Siahrostami, X. Zheng, *Chem* **2021**, *7*, 38–63.
[3] S. Siahrostami, A. Verdaguer-Casadevall, M. Karamad, D. Deiana, P. Malacrida, B. Wickman, M. Escudero-Escribano, E. A. Paoli, R. Frydendal, T. W. Hansen, I. Chorkendorff, I. E. L. Stephens, J. Rossmeisl, *Nat. Mater.* **2013**, *12*, 1137–1143, https://doi.org/10.1038/nmat3795.
[4] S. C. Perry, D. Pangotra, L. Vieira, L.-I. Csepei, V. Sieber, L. Wang, C. Ponce de León, F. C. Walsh, *Nat. Rev. Chem.* **2019**, *3*, 442–458, https://doi.org/10.1038/s41570-019-0110-6.
[5] K. Fuku, Y. Miyase, Y. Miseki, T. Gunji, K. Sayama, *Chemistry Select* **2016**, *1*, 5721–5726.
[6] J. Li, D. Solanki, Q. Zhu, X. Shen, G. Callander, J. Kim, Y. Li, H. Wang, S. Hu, *J. Mater. Chem. A* **2021**, *9*, 18498–18505, https://doi.org/10.1039/D1TA05451A.

[7] W. Guo, Y. Xie, S. Tang, B. Yu, X. Lian, G. Henkelman, X. Liu, *Appl. Surf. Sci.* **2022**, *596*, 153634, https://doi.org/10.1016/j.apsusc.2022.153634.

[8] W. Guo, Y. Xie, Y. Liu, S. Shang, X. Lian, X.i Liu, *Appl. Surf. Sci.* **2022**, *606*, 155006, https://doi.org/10.1016/j.apsusc.2022.155006.

[9] S. R. Kelly, X. Shi, S. Back, L. Vallez, S. Y. Park, S. Siahrostami, X. Zheng, J. K. Nørskov, *ACS Catal.* **2019**, *9*, 4593–4599, https://doi.org/10.1021/acscatal.8b04873.

[10] J. H. Baek, T. M. Gill, H. Abroshan, S. Park, X. Shi, J. Nørskov, H. S. Jung, S. Siahrostami, X. Zheng, *ACS Energy Lett.* **2019**, *4*, 720–728, https://doi.org/10.1021/acsenergylett.9b00277.

[11] S. Y. Park, H. Abroshan, X. Shi, H. S. Jung, S. Siahrostami, X. Zheng, *ACS Energy Lett.* **2019**, *4*, 352–357, https://doi.org/10.1021/acsenergylett.8b02303.

[12] L. Li, K. Xiao, P. K. Wong, Z. Hu, J. C. Yu, *ACS Appl. Mater. Interfaces* **2022**, *14*, 7878–7887, https://doi.org/10.1021/acsami.1c20834.

[13] J. Baek, Q. Jin, N. S. Johnson, Y. Jiang, R. Ning, A. Mehta, S. Siahrostami, X. Zheng, *Nat. Commun.* **2022**, *13*, 7256, https://doi.org/10.1038/s41467-022-34884-4.

[14] S. Mavrikis, S. C. Perry, P. K. Leung, L. Wang, C. Ponce de León, *ACS Sustainable Chem. Eng.* **2021**, *9*, 76–91, https://doi.org/10.1021/acssuschemeng.0c07263.

[15] K. Dong, Y. Lei, H. Zhao, J. Liang, P. Ding, Q. Liu, Z. Xu, S. Lu, Q. Li, X. Sun, *J. Mater. Chem. A* **2020**, *8*, 23123–23141, https://doi.org/10.1039/D0TA08894C.

[16] Y. Wang, X. Lian, Y. Zhou, W. Guo, H. He, *New J. Chem.* **2021**, *45*, 8958–8964, https://doi.org/10.1039/D1NJ00637A.

[17] P. Chaudhary, I. Evazzade, R. Belosludov, V. Alexandrov, *ChemCatChem* **2023**, *15*, e202300055, https://doi.org/10.1002/cctc.202300055.

[18] P. Vassilev, M. T. M. Koper, *J. Phys. Chem. C* **2007**, *111*, 2607–2613, https://doi.org/10.1021/jp064515+.

[19] D. Zhang, Y. Hirai, K. Nakamura, K. Ito, Y. Matsuo, K. Ishibashi, Y. Hashimoto, H. Yabu, H. Li, *Chem. Sci.* **2024**, *15*, 5123–5132, https://doi.org/10.1039/D4SC00473F.

[20] H. A. Hansen, V. Viswanathan, J. K. Nørskov, *J. Phys. Chem. C* **2014**, *118*, 6706–6718, https://doi.org/10.1021/jp4100608.

[21] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, H. Jónsson, *J. Phys. Chem. B* **2004**, *108*, 17886–17892.

[22] I. C. Man, H.-Y. Su, F. Calle-Vallejo, H. A. Hansen, J. I. Martínez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov, J. Rossmeisl, *ChemCatChem* **2011**, *3*, 1159–1165, https://doi.org/10.1002/cctc.201000397.

[23] J.-M. Hu, J.-Q. Zhang, C.-N. Cao, *Int. J. Hydrogen Energy* **2004**, *29*, 791–797, https://doi.org/10.1016/j.ijhydene.2003.09.007.

[24] J.-J. Zhang, J.-M. Hu, J.-Q. Zhang, C.-N. Cao, *Int. J. Hydrogen Energy* **2011**, *36*, 5218–5226, https://doi.org/10.1016/j.ijhydene.2011.01.131.

[25] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, Z. Ulissi, *ACS Catal.* **2021**, *11*, 6059–6072, https://doi.org/10.1021/acscatal.0c04525.

[26] P. S. Lamoureux, K. T. Winther, J. A. G. Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen, T. Bligaard, *ChemCatChem* **2019**, *11*, 3581–3601, https://doi.org/10.1002/cctc.201900595.

[27] C. Pan, M. Mahmoudabadbozchelou, X. Duan, J. C. Benneyan, S. Jamali, R. M. Erb, *J. Colloid Interface Sci.* **2022**, *611*, 29–38, https://doi.org/10.1016/j.jcis.2021.11.195.

[28] D. Seddon, E. A. Müller, J. T. Cabral, *J. Colloid Interface Sci.* **2022**, *625*, 328–339, https://doi.org/10.1016/j.jcis.2022.06.034.

[29] X. Li, B.o Li, Z.e Yang, Z. Chen, W. Gao, Q. Jiang, *J. Mater. Chem. A* **2022**, *10*, 872–880, https://doi.org/10.1039/D1TA09184K.

[30] J. Behler, P. Michele, *Phys. Rev. Lett.* **2007**, *98*, 14, https://doi.org/10.1103/PhysRevLett.98.146401.

[31] L. Li, H. Li, I. D. Seymour, L. Koziol, G. Henkelman, *J. Chem. Phys.* **2020**, *152*, 224102, https://doi.org/10.1063/5.0007391.

[32] X. Zhang, J. Liu, R. Li, X. Jian, X. Gao, Z. Lu, X. Yue, *J. Colloid Interface Sci.* **2023**, *645*, 956–963, https://doi.org/10.1016/j.jcis.2023.05.011.

[33] X. Wang, B. Xiao, Y. Li, Y. Tang, F. Liu, J. Chen, Y. Liu, *Appl. Surf. Sci.* **2020**, *531*, 147323, https://doi.org/10.1016/j.apsusc.2020.147323.

[34] H. Xu, D. Cheng, D. Cao, X. C. Zeng, *Nat. Catal.* **2024**, *7*, 207–218, https://doi.org/10.1038/s41929-023-01106-z.

[35] O. Mamun, K. T. Winther, J. R. Boes, T. Bligaard, *Scientific Data* **2019**, *6*, 76, https://doi.org/10.1038/s41597-019-0080-z.

[36] B. M. Comer, N. Bothra, J. R. Lunger, F. Abild-Pedersen, M. Bajdich, K. T. Winther, *ACS Catal.* **2024**, *14*, 5286–5296, https://doi.org/10.1021/acscatal.4c00111.

[37] H. Zhou, Z. Deng, Y. Xia, M. Fu, *Neurocomputing* **2016**, *216*, 208–215, https://doi.org/10.1016/j.neucom.2016.07.036.

[38] R. Su, X. Liu, L. Wei, *Brief. Bioinform.* **2020**, *21*, 687–698, https://doi.org/10.1093/bib/bbz021.

[39] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, S. P. Ong, *Adv. Energy Mater.* **2020**, *10*, 1903242, https://doi.org/10.1002/aenm.201903242.

[40] L. Breiman, *Mach. Learn.* **2001**, *45*, 5–32, https://doi.org/10.1023/A:1010933404324.

[41] C. Robert, *Chance* **2014**, *27*, 62–63, https://doi.org/10.1080/09332480.2014.914768.

[42] H. F. Jerome, *Ann. Stat.* **2001**, *29*, 1189–1232.

[43] A. K. M. E. Saleh, M. Arashi, B. M. G. Kibria, *Wiley Probability and Statistics Series*, Wiley, Hoboken, NJ, USA, **2019**.

[44] Z. Zhang, *Ann. Transl. Med.* **2016**, *4*, 218, https://doi.org/10.21037/atm.2016.03.37.

[45] A. Keprate, R. M. C. Ratnayake, Using gradient boosting regressor to predict stress intensity factor of a crack propagating in small bore piping. in *IEEE Int. Conf. Ind. Eng. Eng. Manag. (IEEM)* **2017**, 1331–1336, https://doi.org/10.1109/IEEM.2017.8290109.

[46] A. J. Smola, B. Schölkopf, *Stat. Comput.* **2004**, *14*, 199–222, https://doi.org/10.1023/B:STCO.0000035301.49549.88.

[47] P. Geurts, D. Ernst, L. Wehenkel, *Mach. Learn.* **2006**, *63*, 3–42, https://doi.org/10.1007/s10994-006-6226-1.

[48] X. J. Gao, J. Yan, J. J. Zheng, S. Zhong, X. Gao, *Adv. Healthcare Mater.* **2023**, *12*, e2202925, https://doi.org/10.1002/adhm.202202925.

[49] X. Zong, D. G. Vlachos, *J. Chem. Inf. Model.* **2022**, *62*, 4361–4368, https://doi.org/10.1021/acs.jcim.2c00872.

[50] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, P. Marquetand, *J. Chem. Phys.* **2018**, *148*, 241709, https://doi.org/10.1063/1.5019667.

[51] H. Li, J. K. Nørskov, *Phys. Chem. Chem. Phys.* **2020**, *22*, 26216–26222, https://doi.org/10.1039/D0CP04536E.

[52] T.-R. Wang, J.-C. Li, W. Shu, S.-L. Hu, R.-H. Ouyang, W.-X. Li. *Chin. J. Chem. Phys.* **2020**, *33*, 703–711, https://doi.org/10.1063/1674-0068/cjcp2004049.

[53] B. Deng, P. Chen, P. Xie, Z. Wei, S. Zhao, *Chem. Eng. Sci.* **2023**, *267*, 118368, https://doi.org/10.1016/j.ces.2022.118368.

[54] H. Xu, D. Cheng, D. Cao, X. C. Zeng, *Nat. Catal.* **2024**, *7*, 207–218, https://doi.org/10.1038/s41929-023-01106-z.

[55] W. Xu, M. Andersen, K. Reuter, *ACS Catal.* **2021**, *11*, 734–742, https://doi.org/10.1021/acscatal.0c04170.

[56] M. Baucells, R. Lejano, C.-Z. Qin, L. S. Shapley, (Born 1923). in *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK, London, **2017**, 1-9.

[57] Y. An, M. Ouyang, S. Kong, G. Wang, X. Chen, *Phys. Chem. Chem. Phys.* **2023**, *25*, 13289–13296, https://doi.org/10.1039/D2CP06067A.
[58] J. Hao, Z. Zhuang, K. Cao, G. Gao, C. Wang, F. Lai, S. Lu, P. Ma, W. Dong, T. Liu, M. Du, H. Zhu, *Nat. Commun.* **2022**, *13*, 2662, https://doi.org/10.1038/s41467-022-30379-4.
[59] Data Platform, **2024**. Available at: http://81.71.102.117:8080/. Accessed on: 22 September 2024.
[60] C. Zhang, R. Lu, C. Liu, L. Yuan, J. Wang, Y. Zhao, C. Yu, *Adv. Funct. Mater.* **2021**, *31*, 2100099, https://doi.org/10.1002/adfm.202100099.
[61] L. Li, L. Xu, A. W. M. Chan, Z. Hu, Y. Wang, J. C. Yu, *Chem. Mater.* **2022**, *34*, 63–71, https://doi.org/10.1021/acs.chemmater.1c02787.
[62] J. Behler, *J. Chem. Phys.* **2011**, *134*, 074106, https://doi.org/10.1063/1.3553717.
[63] H. Liu, D. Zhang, S. M. Holmes, C. D'Agostino, H. Li, *Chem. Sci.* **2023**, *14*, 9000–9009, https://doi.org/10.1039/D3SC01827J.
[64] H. Li, S. Kelly, D. Guevarra, Z. Wang, Y. Wang, J. A. Haber, M. Anand, G. T. K. K. Gunasooriya, C. S. Abraham, S. Vijay, J. M. Gregoire, J. K. Nørskov, *Nat. Catal.* **2021**, *4*, 463–468, https://doi.org/10.1038/s41929-021-00618-w.
[65] K. Wenderich, B. A. M. Nieuweweme, G. Mul, B. T. Mei, *ACS Sustainable Chem. Eng.* **2021**, *9*, 7803–7812, https://doi.org/10.1021/acssuschemeng.1c01244.
[66] J. Xu, Y. Wang, C. Dong, X. Yang, K. Zhang, J. Shen, *J. Environ. Chem. Eng.* **2024**, *12*, 112779, https://doi.org/10.1016/j.jece.2024.112779.
[67] Z. Chen, S. Geng, Y. Wang, Y. Wang, S. Song, *Appl. Catal., B* **2022**, *317*, 121756, https://doi.org/10.1016/j.apcatb.2022.121756.
[68] G. Zhao, I. Muhammad, K. Suzuki, M. Hirayama, R. Kanno, *Mater. Trans.* **2016**, *57*, 1370–1373.
[69] M. C. Biesinger, L. W. M. Lau, A. R. Gerson, R. S. C. Smart, *Appl. Surf. Sci.* **2010**, *257*, 887–898, https://doi.org/10.1016/j.apsusc.2010.07.086.
[70] D. Zhang, F. She, J. Chen, L. Wei, H. Li, *J. Am. Chem. Soc.* **2025**, *147*, 6076–6086, https://doi.org/10.1021/jacs.4c16733.

Manuscript received: August 16, 2025
Revised manuscript received: December 01, 2025
Manuscript accepted: December 02, 2025
Version of record online: ■■, ■■