

Supplementary Note 4 (Details of Supervised Data Mining) for

Leveraging Data Mining, Active Learning, and Domain Adaptation in a Multi-Stage, Machine Learning-Driven Approach for the Efficient Discovery of Advanced Acidic Oxygen Evolution Electrocatalysts

Rui Ding *et al.*

*Corresponding author. Email: junhongchen@uchicago.edu, chenyuxin@uchicago.edu

Supplementary Note Discussion SND 4-1

Blackbox Interpretation Methods Explained

As mentioned in the main text, we have introduced three different types of deep black box interpretation. Here we provide brief introduction and elaborate more on the analysis plots we presented as vivid visualized results. These methods are commonly used ML-related material studies(1-4).

1. SHAP (SHapley Additive exPlanations):

SHAP(5) uses a game-theoretic approach to decompose machine learning model predictions into contributions from each feature. Its foundation in Shapley values, derived from cooperative game theory, assigns a fair "payout" to each feature based on its contribution to the prediction. The uniqueness of SHAP lies in its ability to provide both global interpretability (across the entire model) and local interpretability (for individual predictions).

The formula for SHAP values is rooted in the concept of marginal contributions. The key idea is to compute the difference in the prediction with and without a feature, averaged over all possible combinations of other features:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_x(S \cup \{j\}) - f_x(S)]$$

SHAP Analysis Plots:

- a) Summary Plot: This plot provides a comprehensive view of feature impact by displaying the distribution of SHAP values for each feature. It helps identify which features most strongly influence the model's output. Positive SHAP values indicate a higher prediction, while negative values suggest a lower prediction. The plot's color coding often represents another feature, providing insights into potential feature interactions.
- b) Clustering Bar Plots: These plots highlight the grouped importance of features by aggregating their SHAP values. Features with similar effects are clustered together, showing their combined influence on the model's decisions. The connection lines between features indicate correlation, with a clustering cutoff typically set at 0.5, suggesting a strong enough association to consider these features collectively.
- c) Overview Heatmap: It displays the global importance of features across all predictions, with each cell's intensity indicating the SHAP value magnitude for a feature across different samples. This plot is useful for spotting patterns and outliers in feature contributions across the dataset.
- d) Clustering Cohort Plots: By finding a decisive splitting ridge, these plots divide the samples into distinct groups to compare their SHAP feature importance. This division helps understand how different subsets of data behave under the model, highlighting the variance in feature importance across these groups.
- e) Dependence Plot: It shows the effect of a single feature on the prediction while also revealing interactions with another feature, if selected. The plot helps in understanding not just the isolated impact of a feature but also how its influence changes when another feature varies, providing a deeper insight into model behavior.

2. Partial Dependence Plots (PDP):

PDPs(6) show the average effect of one or two features on the predicted outcome, assuming independence from other features. This assumption is a limitation but also simplifies interpretation. PDPs can reveal whether the relationship between the target and the feature is linear, monotonic, or more complex:

$$PDP(f) = \frac{1}{n} \sum_{i=1}^n \hat{f}(x_{i1}, x_{i2}, \dots, f, \dots, x_{ip})$$

PDPs offer a straightforward visualization of the relationship between features and the prediction, but without considering interactions. This makes them particularly useful for understanding univariate effects.

Like SHAP dependence plots, the PDP dependence plots illustrates the average effect of a feature on the model's predictions, with the plot's surface showing the prediction change as the feature value varies. If a second feature is included, the plot can also reveal interaction effects, depicted by varying color intensities or patterns. This color variation helps identify how the interaction between features affects the predicted outcome. By analyzing this plot, one can understand whether the relationship between the target and the feature is linear, nonlinear, or shows any other complex pattern, aiding in the interpretation of model behavior and feature importance. Similar to SHAP, PDP results are also the ensembled weighted averaged results from the top-3 models in each committee.

3. Friedman's H-statistics:

Friedman's H-statistics(7) measure the interaction strength among features in a predictive model. It calculates the proportion of model output variance explained by interactions between features, a crucial aspect in understanding complex models like those in catalyst research:

$$H^2 = \frac{\text{Var}[\mathbb{E}(\hat{y}|X_i, X_j)] - [\text{Var}(\mathbb{E}(\hat{y}|X_i)) + \text{Var}(\mathbb{E}(\hat{y}|X_j))]}{\text{Var}(\hat{y})}$$

H-statistics Heatmap Matrix Plot:

This plot highlights the interactions between pairs of input features based on 2nd order of Friedman's H-statistics decomposition of partial dependence, showing which feature combinations significantly affect the model's predictions.

The heatmap created from Friedman's H-statistics displays the strength of interactions between pairs of features. Each cell in the matrix represents the interaction strength between two features, quantified by the H-statistic value. Higher values indicate stronger interactions and more significant influence on the model's output variance. This heatmap is crucial for identifying which feature pairs have nonlinear interactions that significantly affect the model's predictions. By analyzing this plot, one can prioritize feature pairs for further investigation or modeling efforts, especially in complex systems like catalyst research where interactions play a critical role. Finally, similar to SHAP and PDP, H matrix results are also the ensembled weighted averaged results from the top-3 models in each committee.

Integrating these methods provides a multifaceted view of model behavior. SHAP serves as the main method since it gives detailed, interaction-aware insights into predictions. As supplementary methods, PDPs offer broader, simpler interpretations of feature effects, and Friedman's H-statistics unveil crucial input feature interactions, (which would help us to quickly lock which pair of input features to be visualized with priority since we have a high dimensional input feature space, traverse through all the possible two-feature interaction analysis would cost C_{54}^2 or C_{55}^2 attempts which is too many and may not be all worthy). Together, they enable a comprehensive and nuanced understanding, vital for ensuring scientific integrity and guiding informed decisions in materials science and catalysis. And finally, all the results we present are, like mentioned in the main text, ensembled results from the analysis values (of SHAP etc.) on the top-3 best ML models with R^2 as weight coefficient. Readers could inspect the raw original SHAP/PDP analysis and interaction plots for different ML models before ensembled further in the online repository: **Fig. OR17-OR68** (including zoomed version of H-statistic matrix of between only elemental or synthesis features, **Fig. OR35, 36, 61, 62** corresponding to **Fig. SN 4-1** and **Fig. SN**

4-2). Here we present the detailed results in **Fig. SN 4-1** to **Fig. SN 4-18** for readers' query and retrieve, along with more comprehensiveness and robustness as explained in the main text.

Supplementary Note Discussion SND 4-2

Extensive Discussion of Supervised Data Mining Results

To maintain conciseness, we have highlighted the key insights derived from the detailed analysis of the SHAP matrix in the main text, focusing on perspective points 1 to 4 that guided our initial exploration. In this supplementary note, we present a more extensive set of results and discussions that merit the researcher's attention.

Despite observable disparities in regression metrics, the underlying patterns within ML committees trained on both the full and high-quality domain knowledge datasets exhibit high-level of consistency. This is evidenced by various analyses, including SHAP summary, cohort, clustering bar, and overview heat map plots, as well as by the Friedman's H-statistic matrix(3, 7, 8) and PDP interaction plots. Hence, rather focusing on each Fig. and subgraphs for detailed discussion, we would directly subjectively summarize the results into viewpoints separately.

Scripts in this part for activity or stability could be retrieved in directory “Domain Knowledge Based Initial ML Committee and Blackbox Interpretation/Regression/Activity/Blackbox Interpretation/”

Besides regression committee results, we have also comprehensively conducted level set estimation (LSE)(9) with the same workflow and datasets for further strongly proving self-consistency and data balance with details in **Supplementary Note Discussion SND 4-3 and Fig. SN 4-37** in this note. This approach, applied to the same dataset but with discretized output for classification, serves to corroborate the high accuracy of our regression models. Crucially, it demonstrates the consistency between SHAP analysis results from both models, affirming the balanced nature of our data and the self-consistency of our work.

Blackbox Interpretation of the ML Committee based on Domain Knowledge: OER Activity

1. **Feature Interaction Analysis:** In the Methods section, we outlined our approach to analyzing over 50 features by categorizing them into two groups: atomic element properties, representing intrinsic chemical descriptors, and synthesis/testing parameters, which stem from engineering practices. In **Figs. SN 4-1 and Fig. SN 4-2**, we illustrated the interaction strengths using H statistics heatmaps from two committeees, averaging the degree of interactions across categories. We observed that, on average, interactions between elemental features and synthesis parameters are less pronounced than those within the same category (either elemental or synthesis). This observation aligns with findings from our prior research(8), suggesting that the synergy between element choices and synthesis parameters is minimal. Moreover, it is evident that synthesis and testing parameters, in both dataset cases, exhibit a higher average internal interaction based on H-statistics (0.19 and 0.20). Consequently, the OER activity appears to be influenced independently by elemental features and synthesis parameters. This supports our decision to focus our analysis on dependence plots within the same feature category, examining the impacts of either elemental features or synthesis parameters separately. **This ML analysis outcome is in line with established electrocatalysis research and domain consensus.** It is generally accepted that while synthesis parameters regulate physical characteristics like morphology and surface area, which affect the accessibility of active sites, elemental features largely determine the electronic structure and intrinsic catalytic activity(10). Crucially, the independence between synthesis conditions and elemental composition is indicative of a larger trend in the field, where physical attributes are altered by synthesis techniques without substantially changing electronic properties, and intrinsic activity is defined by elemental choices(11). This understanding is supported by the increased internal interaction between the synthesis and testing parameters, since physical characteristics that interact with testing conditions to determine observed performance are directly impacted by synthesis conditions. By confirming that elemental and synthesis features function as separate drivers of OER activity, this division of influence supports our emphasis on dependence plots within isolated feature categories.

Another method for examine the extent of non-linear dependencies and synergistic effects between feature pairs on the target outputs of the ML models is by presenting SHAP cohort plots like **Fig. SN 4-3b**. Observations from these clustering plots that tend to base on SHAP value distinguishing the clustering tendency among input features (with clustering cutoff set to 0.5 as a moderate value), suggest that elemental features, particularly the fundamental intrinsic atomic properties features, tend to cluster together with minimal interaction with synthesis and testing parameters. Notably, the descriptors of the primary metal element rarely cluster with those of other elements, as indicated by **Fig. SN 4-4b** and **Fig. SN 4-10b**. Such a finding that the primary metal element's descriptors, like radius and electronegativity, show little clustering with those of other constituent elements is consistent with what is known about metal oxides. The basic electronic structure and bonding properties of complex oxide systems are frequently determined by the primary metal element, whereas dopant elements cause localized perturbations without

substantially changing the overall electronic framework. This distinction is emphasized in research where dopants primarily affect secondary attributes and the primary metal's properties mostly determine the behavior of the material. For example, studies on doped perovskite oxides show that although dopants can alter catalytic activity and surface characteristics, the primary metal element still largely controls the core electronic structure. This underscores the limited interaction between primary metal descriptors and those of dopant elements, supporting the clustering patterns observed(12). However, an intriguing global perspective emerges from **Fig. SN 4-3b**, revealing a notable clustering tendency with one-hot features, particularly regarding the use of carbon as a support material. This was initially included in our dataset to distinguish whether researchers incorporated additional support materials like TiO_x or carbon nanoparticles. In the case of synthesis parameters, aside from a tendency to cluster between annealing temperature and duration, other synthesis parameters appear relatively independent. This is consistent with well-established electrocatalysis findings. It is well known that carbon supports might help increase the dispersion of active ingredients, which raises the electrochemically active surface area. Furthermore, carbon supports can facilitate electron transfer between the electrode and the active sites. For instance, a previous study demonstrated that using carbon supports in OER electrocatalysts improves electronic conductivity and surface area, thereby possibly contributing to enhanced catalytic performance(13).

2. **Elemental Features Insights:** When examining Figs. such as **Fig. SN 4-4a** and **Fig. SN 4-7a**, which focus on the internal ordering of elemental feature categories, broader insights into the ML model's internal workings are revealed. For elemental features, in models trained with the full dataset, descriptors of the primary metal and the 2nd almost exclusively dominate the top rankings, surpassing any descriptors of other minor doping elements. Interestingly, element-related features, primarily atomic properties, were also particularly influential in the full and high-quality activity datasets, where the atomic radius of the primary metal element consistently emerged as the most critical factor, followed by electronegativity of the primary metal element (**Fig. SN 4-6b& Fig. SN 4-12b**). This trend is slightly attenuated in the high-quality dataset, where the second most influential descriptor of the 2nd element changed from number of d atoms to relative atomic mass (as shown in **Fig. SN 4-10a& Fig. SN 4-13a**). However, this does not significantly alter our conclusion. Namely, the ML model seems to infer a relatively weaker role for the third or fourth metal element in enhancing activity. At the domain level, it generally suggests that the primary metal element predominantly dictates activity, but specific thresholds and choices for secondary dopant metals, identified through methods like SHAP, are also crucial. Moreover, the atomic radius and electronegativity have proven to be decisive in determining the activity, and for the 2nd element, which might be the major dopant element, its doping amount, atomic mass, and d electron number follows closely. These analysis results are consistent with well-established electrocatalysis principles, especially with regard to metal oxides RuO_2 and IrO_2 . Catalytic activity is largely determined by the atomic radius and electronegativity of the primary metal. For example, research has demonstrated that the atomic radius affects the strength of the metal-oxygen bond, which in turn affects the activity of OER(14). The adsorption energies of reaction intermediates are also affected.

by electronegativity, which affects the electron density distribution(15, 16). The atomic mass and d-electron count of secondary dopant metals are important factors in altering the electronic structure and catalytic characteristics of the primary metal oxide, like band structures(17, 18). These observations support the focus of our machine learning model on the characteristics of the primary metal and the impact of secondary dopants. Further examination of cohort plots unveiled a critical threshold at 127.5 pm for the atomic radius of the primary metal, autonomously determined by SHAP. Beyond this threshold, elements with a larger atomic radius increase in importance. Dependence plots indicate that positive SHAP values under this 127.5 pm threshold are undesirable, while preferred elements like Rh, Pd, Re, Os, Ir, and Ru exhibit negative SHAP values, with Ru being the most favorable at around 138 pm (Fig. SN 4-15, Fig. SN 4-17). These analysis results are consistent with domain knowledge. Ru, with an atomic radius of approximately 138 pm, is recognized for its exceptional OER activity. Theoretical studies attribute this to Ru's optimal binding energies with oxygen intermediates, facilitating efficient catalytic cycles. Experimental evidence further supports RuO₂'s superior performance in acidic media, demonstrating high activity and stability(14, 15). Ir, with an atomic radius close to Ru, also exhibits commendable OER activity. IrO₂ is widely used as a benchmark catalyst due to its favorable combination of activity and stability(12, 19). The similar atomic radii of Ru and Ir contribute to their comparable catalytic behaviors. Other metals such as Rh, Pd, Re, and Os have also been investigated for OER electrocatalysis. For instance, Rh-based catalysts have shown promising activity in alkaline media(20). Pd-based materials, while more commonly associated with hydrogen evolution reactions, have also been explored for OER applications(21). Re and Os oxides have been studied for their electrocatalytic properties, with some reports indicating potential for OER, though their activities are generally lower compared to Ru and Ir oxides(18, 22).

3. **Synthesis and Testing Parameters:** Regarding synthesis and testing parameter categories, the feature ranking results are almost identical between models trained on the full and high-quality datasets, with only minor changes in order. Catalyst loading is understandably considered the most critical factor, followed closely by annealing temperature and time duration. Notably, in most studies of metal/metal oxide nanomaterials, annealing duration is often not a primary focus. This was evidenced in our previous study on Zeolite Imidazole Framework-8 derived acid oxygen reduction catalysts using data science methods(23). Hence, in the current body of published knowledge within our field, there has been no exploration of how annealing duration as a kinetic factor influences the material characteristics and performance, with most studies focusing on thermodynamics and annealing temperature. This echoes our previous work(23), suggesting that ML models have the capacity to unearth crucial variables often overlooked by researchers in the field based on traditional chemical and materials science expertise. Following catalyst loading, annealing then hydrothermal temperature and duration are significant factors, likely due to their close association with product morphology, size, and crystallinity. Relevant experimental research works also highlight and support the significance of hydrothermal parameters, annealing conditions, and catalyst loading in the synthesis of OER electrocatalysts, particularly

complex metal oxide systems. Hydrothermal synthesis conditions alter catalyst shape, which affects electrocatalytic activity, whereas annealing temperature and time have a significant impact on crystallinity and performance, as evidenced by studies on NiCo₂O₄(24) and CuGaO₂(25) systems as OER electrocatalysts. As for annealing, experimental research on IrO₂ thin films showed that the annealing temperature has a substantial impact on the electrochemically active surface area and OER activity. Higher temperatures result in lower activity because they alter surface characteristics and crystallinity(26). Likewise, according to studies on Cu-Co oxide nanosheets, annealing at 250 °C improved the catalyst's surface area and electron conduction pathways, leading to a higher OER activity than at other annealing temperatures(27). These results are in good agreement with our ML findings, demonstrating that the essential variables identified are in line with known experimental evidence and domain consensus.

Lastly, we observed that other one-hot encoded synthesis parameters, typically conditions of mixing precursors are generally assigned lower importance, which also demonstrates that more nuanced and process-specific factors often overshadow broad categorical variables in influencing the synthesis outcome. Further, through the cohort plot analysis of different catalyst loadings (as automatically selected in the cohort plots, along with corresponding thresholds like the one discussed at 127.5 pm in main text), we find from **Fig. SN 4-8b** and **Fig. SN 4-12b** that 0.2 mg cm⁻² is chosen as a significant demarcation point. Below this value, the influence of catalyst loading becomes markedly diminished, accompanied by a decrease in annealing temperature's impact. This may be because for OER catalysts in acidic conditions, there is an optimal catalyst loading threshold beyond which the increased surface area does not proportionally enhance the catalytic activity, possibly due to mass transport limitations or agglomeration effects. Hence, we could further from this indicate a minimum catalyst loading value of 0.2 mg cm⁻² on electrodes, based on data mining results. We also observed that beyond a preferred catalyst loading of 0.5 mg cm⁻², SHAP values diminish less significantly (**Fig. SN 4-15**), highlighting an optimal threshold above which additional surface area contributions become marginal. Additionally, materials like TiO_x or carbon particles, initially considered as support materials, did not exhibit focused distribution impacts and were subsequently excluded from further analysis. These results could also be confirmed by experimental research works. For OER performance in acidic environments, catalyst loading optimization is essential. Research on RuO₂/CoMo₂O_x, for example, indicates that low Ru loading can produce a high efficiency, but further increases result in decreasing returns because of mass transport limitations(28). As previously discussed, carbon supports improve conductivity and stability without substantially changing the electronic structure of Ru/Ir oxides to increase intrinsic activity(29). Likewise, TiO_x is not considered having direct synergistic effect on the intrinsic electronic characteristics and catalytic activity of the active metal oxides in most circumstances, but mainly enhancing the durability of Ru/Ir-based catalysts by inhibiting dissolution and aggregation(30).

4. **Feature Importance Ranking:** As we have discussed via cohort plots, when observing the importance ranking of both elemental and synthesis and testing parameter features together, it becomes apparent that besides catalyst loading

and parameters related to annealing and hydrothermal processes, descriptors related to elemental features hold a distinct advantage. The SHAP summary plots like **Fig. SN 4-3a** also support this with more features included. This likely aligns with our existing domain knowledge, suggesting that in this material system, the composition and ratio of elements play a more decisive role in OER catalytic activity compared to other factors that might influence nanomaterial crystallinity morphology. This can be explained by the role that elemental features play in determining electronic structure and the availability of active sites, since the d-band center theory emphasizes how electronic configuration directly affects reaction intermediate adsorption energies, which in turn affects overall catalytic efficiency(31, 32). Generally, the common sense from catalyst researchers would be elemental composition often plays a slightly more critical role in determining activity than synthesis parameters. For example, in Fischer–Tropsch synthesis, promoters like Cu, K, and Ni in iron-based catalysts significantly impact selectivity and conversion rates compared to reaction conditions(33). Nevertheless, thermodynamic and kinetic parameters in synthesis, such as annealing and hydrothermal conditions, should not be ignored. Performance may be impacted by changes in phase purity, particle size distribution, or defect density brought about by changes in synthesis conditions, but these effects are often less significant than the electrical and geometric parameters determined by elemental composition(34, 35). But optimization costs differ due to the high energy consumption associated with precise temperature control, so it is generally considered to spend more in extensively exploring these variables.

5. **Dependence Plots:** For the most direct insights, two-dimensional PDP plots generally exhibit trends akin to SHAP interaction plots, albeit with some variations in optimum areas. For instance, optimal catalyst loading appears to be around 0.5 mg cm^{-2} , and the annealing temperature is refined to approximately 400°C as we have discussed in the main text. Corresponding experimental studies also firmly support our observation regarding the optimal annealing temperature range for $\text{RuO}_2/\text{IrO}_2$ -based oxide systems in acidic OER conditions. For instance, annealing temperatures between 400°C and 500°C have been repeatedly demonstrated to produce nanoparticles with desirable catalytic properties, such as optimal size, enhanced activity, and improved stability. In one case, calcination at 500°C was shown to yield highly efficient $\text{RuO}_2@\text{IrO}_2$ core-shell structures, contributing to significant durability over extended periods(36). Another study established that annealing at 500°C is effective for controlling nanoparticle size, which directly influences catalytic performance(37). Yet another study showed hydrothermally prepared IrO_x nanocatalyst layers achieved peak activity and stability when treated at 400°C (38). These findings on Ru/Ir systems substantially align well with our interpretation from the two-dimensional PDP and SHAP interaction plots, particularly regarding the annealing temperature's critical role in fine-tuning the catalyst's performance.
Another intriguing aspect is the interaction between the proportions of the main and secondary metals in the precursor. As **Fig. SN 4-15** and **Fig. SN 4-16** indicate, it seems unfavorable to completely avoid doping, i.e., maintaining only a single metal like Ru or Ir. Such a scenario would result in an excessively high SHAP value, indicating a detrimental impact on overpotential. Therefore, not only doping is encouraged for stability purpose

as we stated in perspective 2 in the main text, doping with at least one element seems also imperative for enhancing activity. Moreover, **Fig. SN 4-17** and **Fig. SN 4-18** also indicate that for the 2nd metal, the choices of elements are quite abundant if the atomic radius is over ~50 pm. These results again support us and emphasize the significance of doping and synergy between different metal elements would benefit OER catalysis. As we have stated in introduction, a wide array of metal elements: Cu(39), Pb(40), Co(41), Mo(42), Ce(42), Au(43), Mn(44), Ni(45), Ta(46), and Gd(47) have been found to benefit OER performances in previous studies. At the theoretical level, doping transition metal oxides into RuO₂ and IrO₂ matrices modifies the electronic structure by adjusting the d-band center as mentioned previously(48), which optimizes the binding energies of OER intermediates and enhances catalytic activity. It can also create oxygen vacancies that serve as additional active sites, facilitating the reaction process. A typical example is an element that has a large difference in atomic radius, like Sr(49). Additionally, doping might induce lattice strain, which further influences the electronic structure and contributes to better OER performance(50). Hence, introducing at least one dopant into the oxide system is a very effective way to enhance OER activity. And it again supports our initial intention to explore the multi-metallic oxide system due to its unlimited potential.

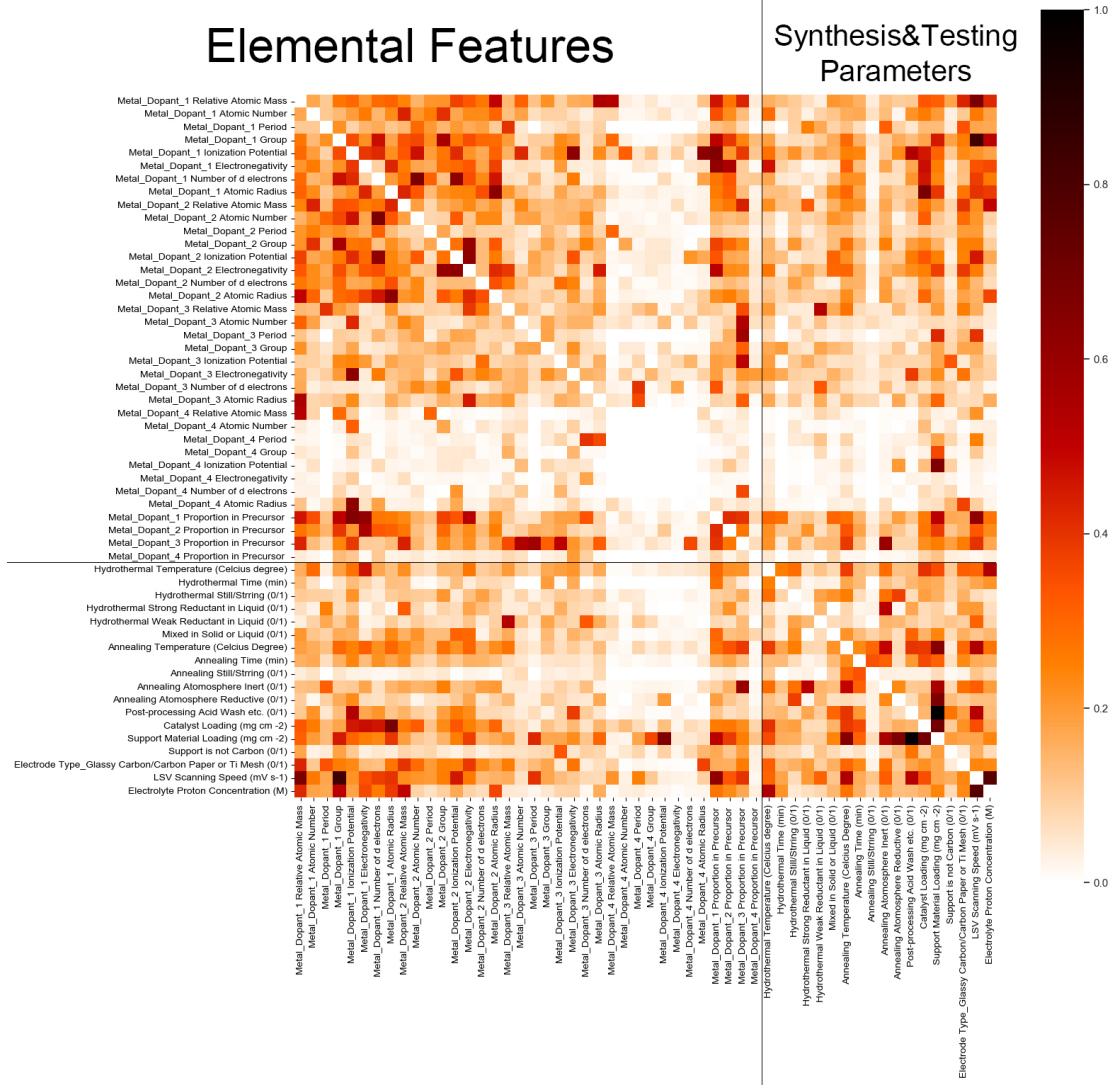


Fig. SN 4-1 Weighted averaged Friedman's H statistic matrix heatmap of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity. The averaged values of interaction between different regions are elemental-elemental: 0.13, elemental-synthesis & test: 0.12, synthesis & test-synthesis & test: 0.19. If only do not consider the fourth metal, the values would be elemental-elemental: 0.18, elemental-synthesis & test: 0.15, synthesis & test-synthesis & test: 0.19.

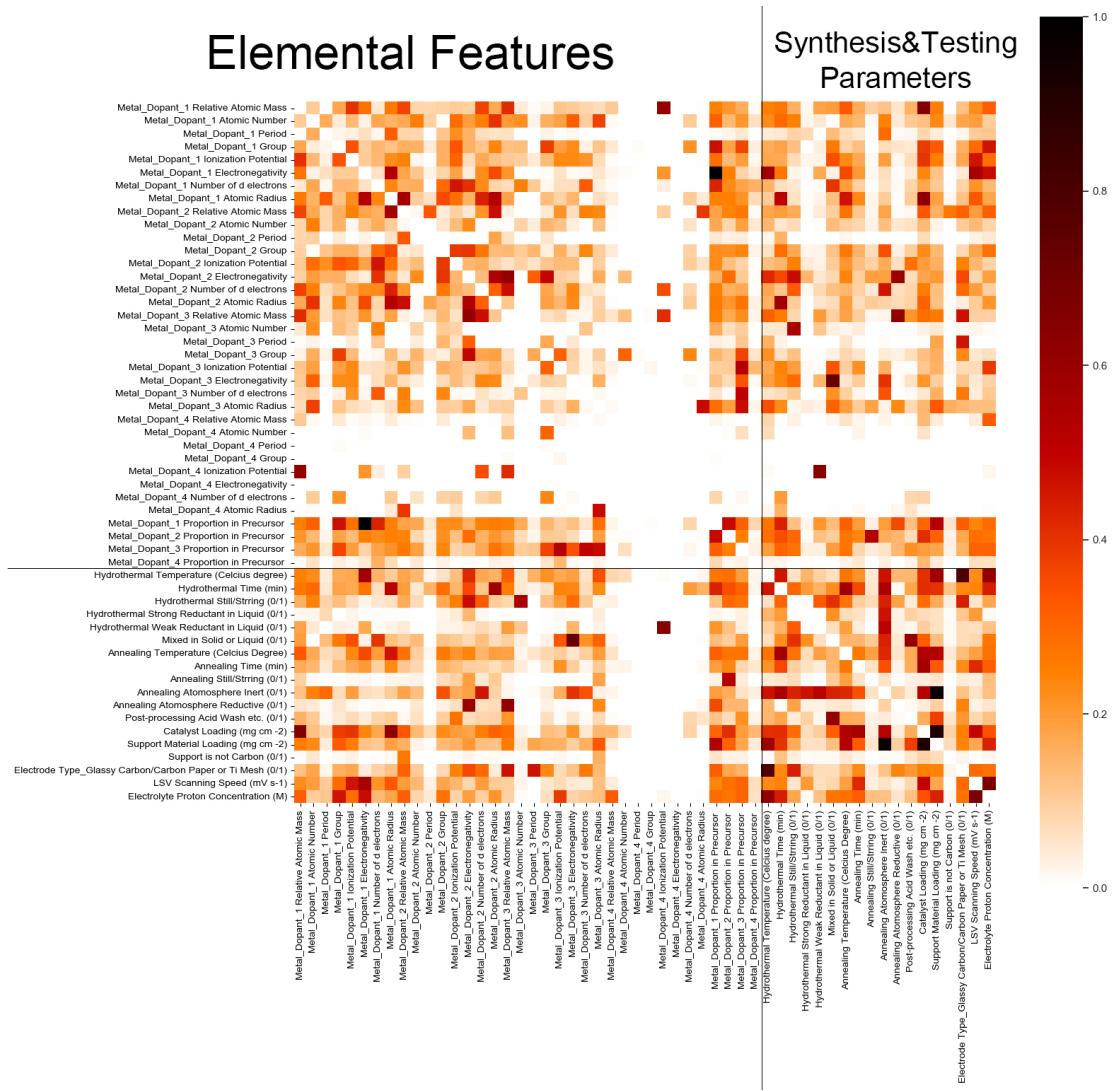


Fig. SN 4-2 Weighted averaged Friedman's H statistic matrix heatmap of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity. The averaged values of interaction between different regions are elemental-elemental: 0.09, elemental-synthesis & test: 0.11, synthesis & test-synthesis & test: 0.20. If only do not consider the fourth metal, the values would be elemental-elemental: 0.15, elemental-synthesis & test: 0.14, synthesis & test-synthesis & test: 0.20.

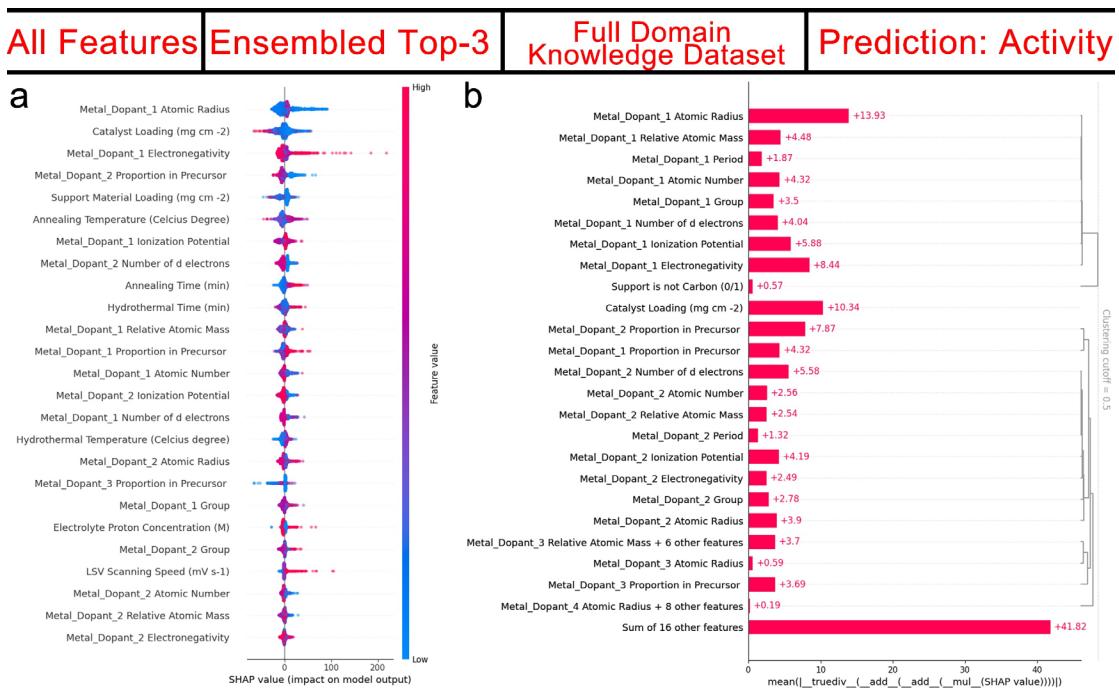


Fig. SN 4-3 (a) SHAP summary plot and (b) SHAP clustering bar plot included all features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity.

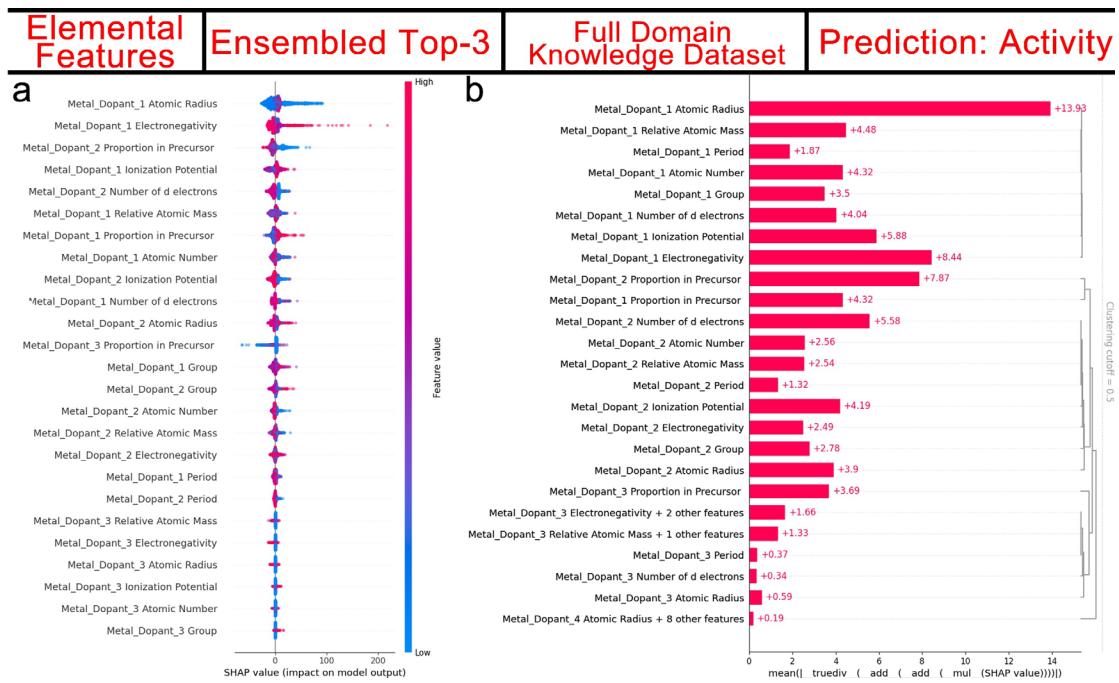


Fig. SN 4-4 (a) SHAP summary plot and (b) SHAP clustering bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity.

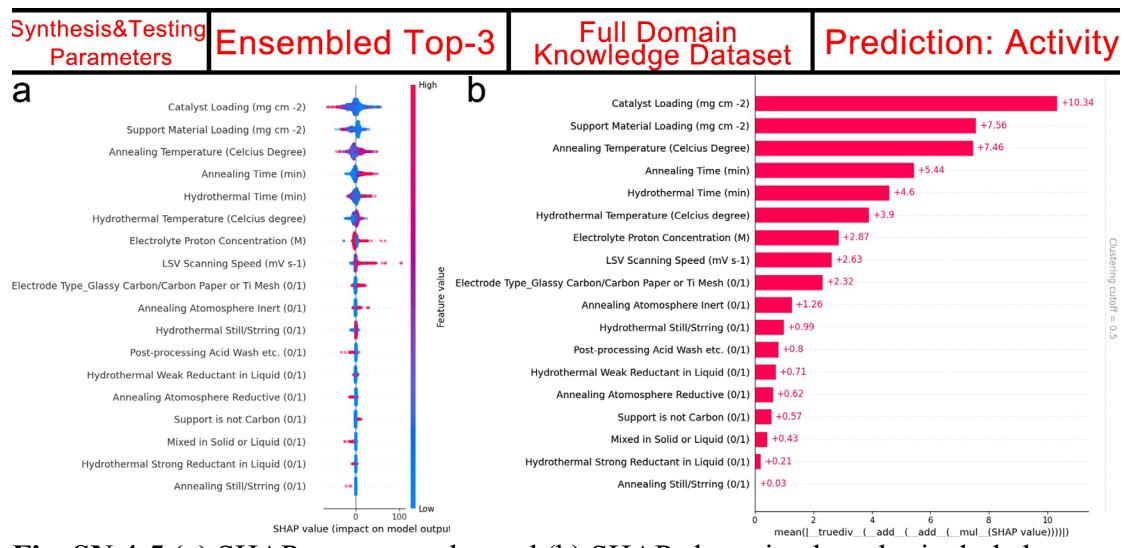


Fig. SN 4-5 (a) SHAP summary plot and **(b)** SHAP clustering bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity.

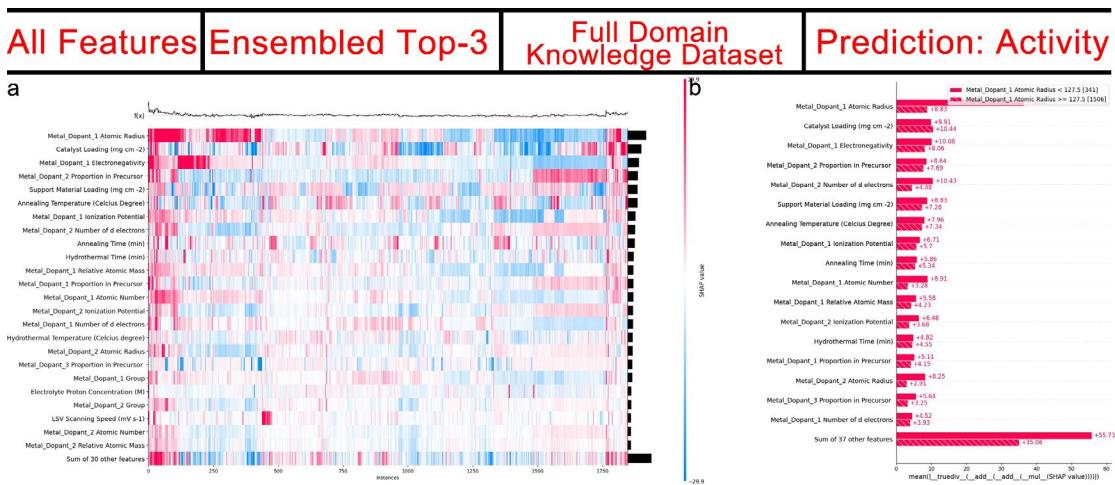
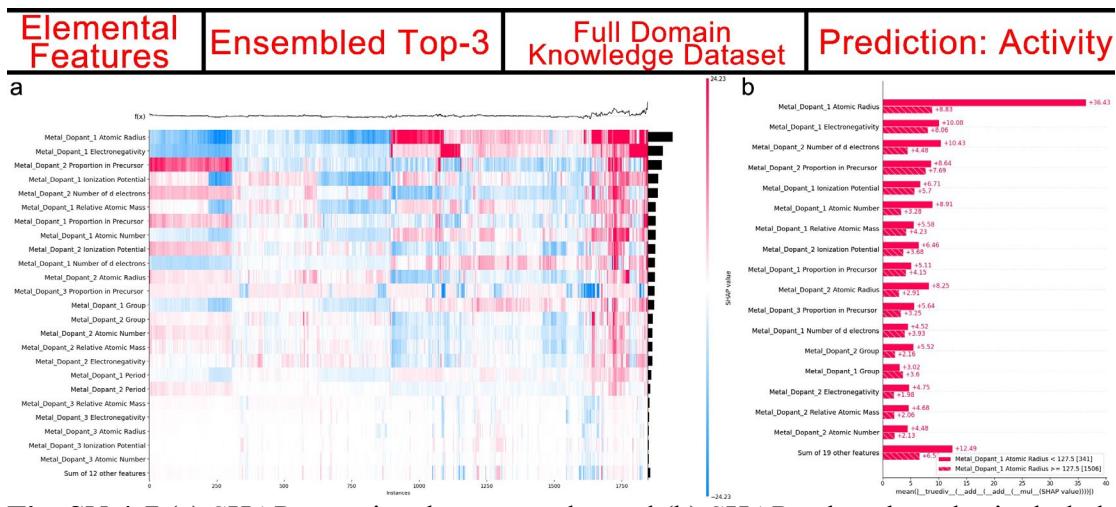


Fig. SN 4-6 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included all features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity.



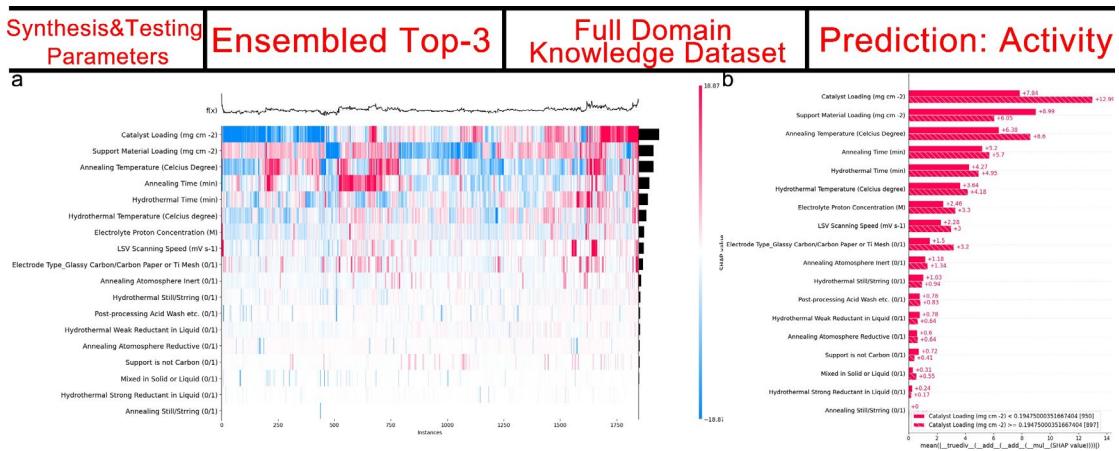


Fig. SN 4-8 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity.

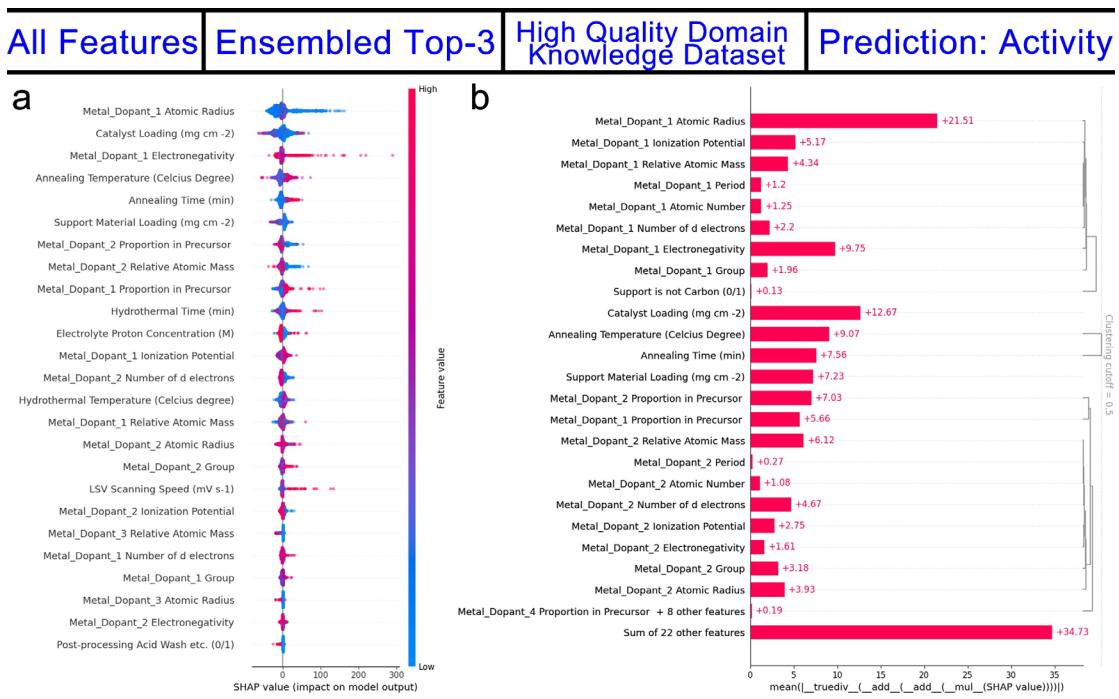


Fig. SN 4-9 (a) SHAP summary plot and (b) SHAP clustering bar plot included all features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity.

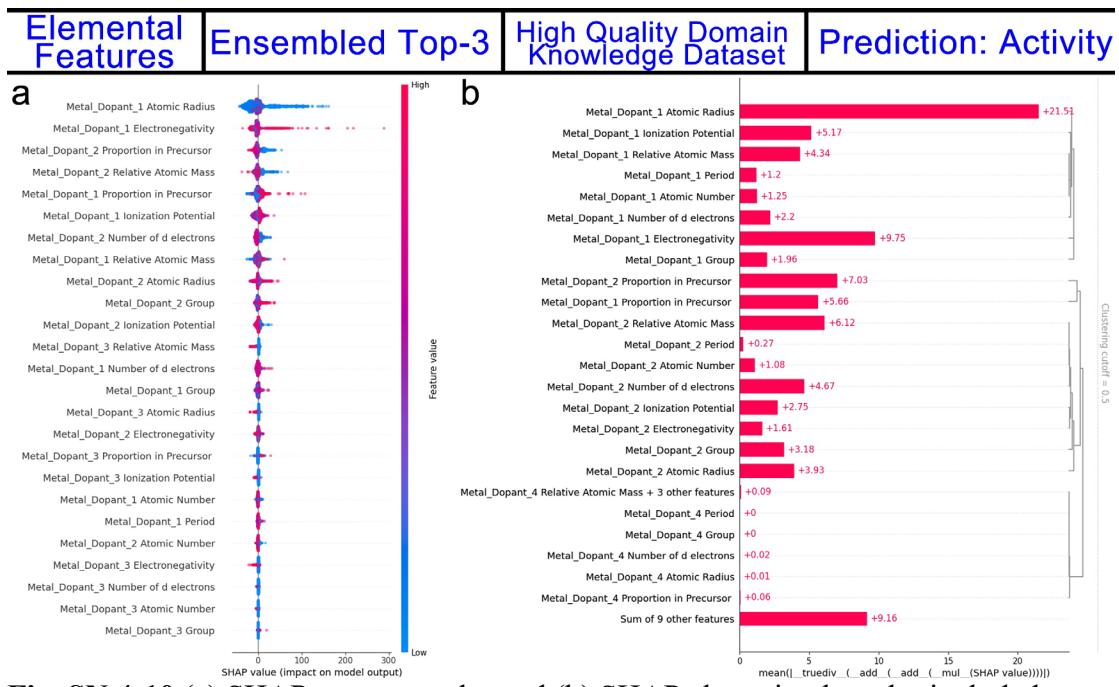


Fig. SN 4-10 (a) SHAP summary plot and (b) SHAP clustering bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity.

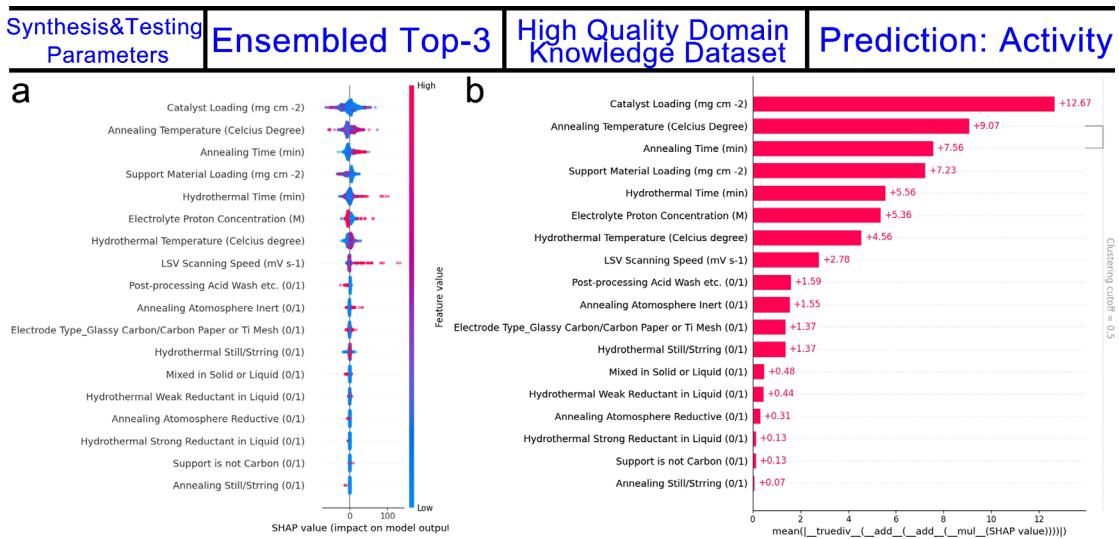


Fig. SN 4-11 (a) SHAP summary plot and (b) SHAP clustering bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity.

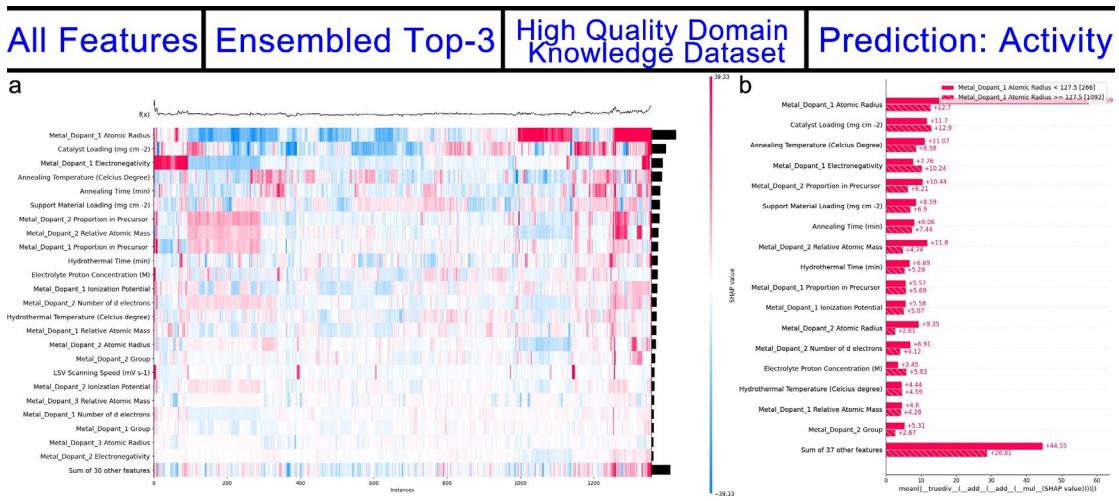


Fig. SN 4-12 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included all features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity.

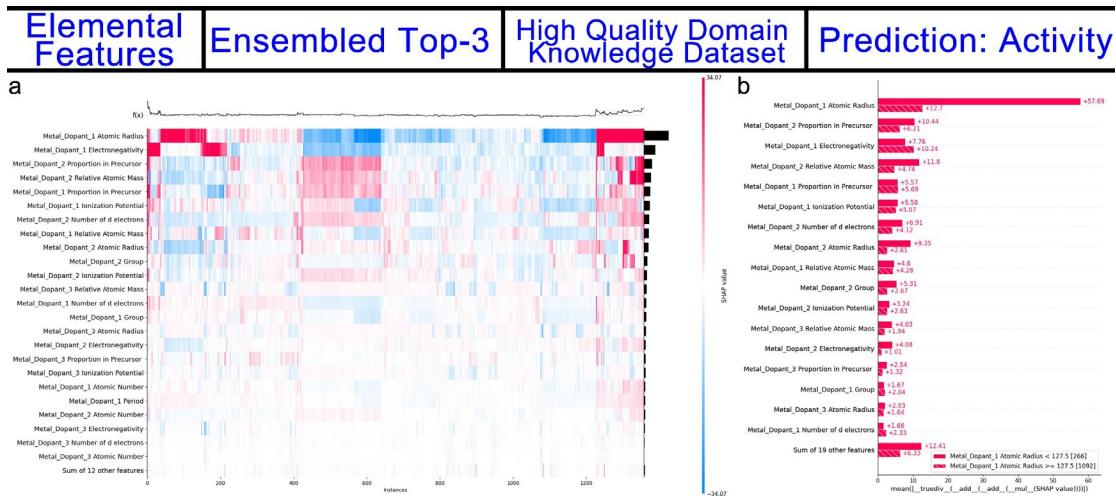
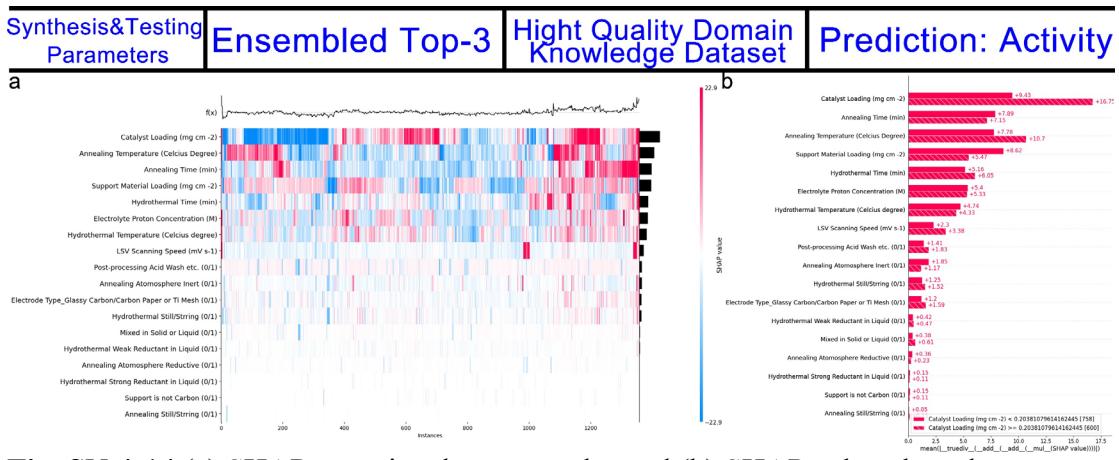


Fig. SN 4-13 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity.



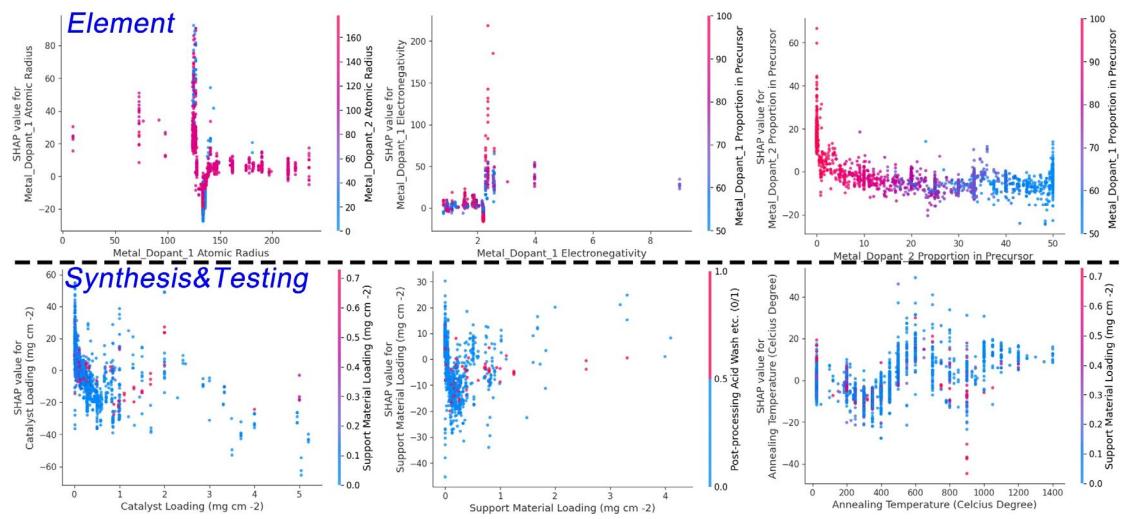


Fig. SN 4-15 SHAP two-dimensional interaction plots screened based on SHAP feature importance ranking's top 5 features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the activity. The interaction features are selected from same category high ranking features, also with high degree of input interaction strength as indicated by H statistics.

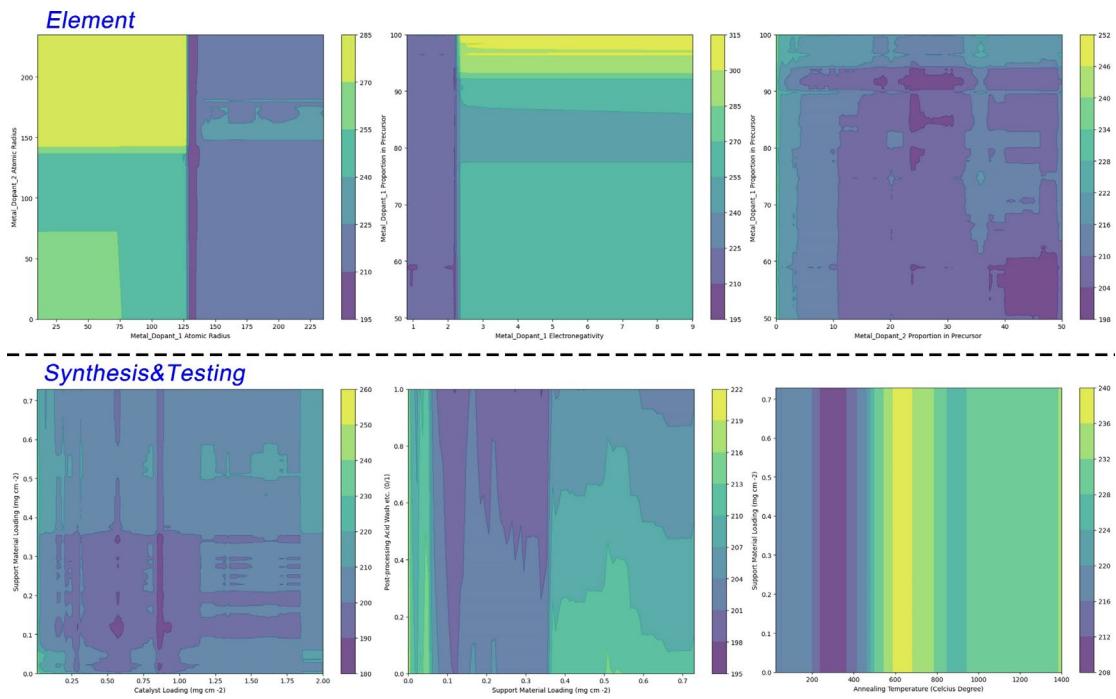


Fig. SN 4-16 PDP two-dimensional interaction plots analyzing the same input feature pairs as in Fig. SN 4-15.

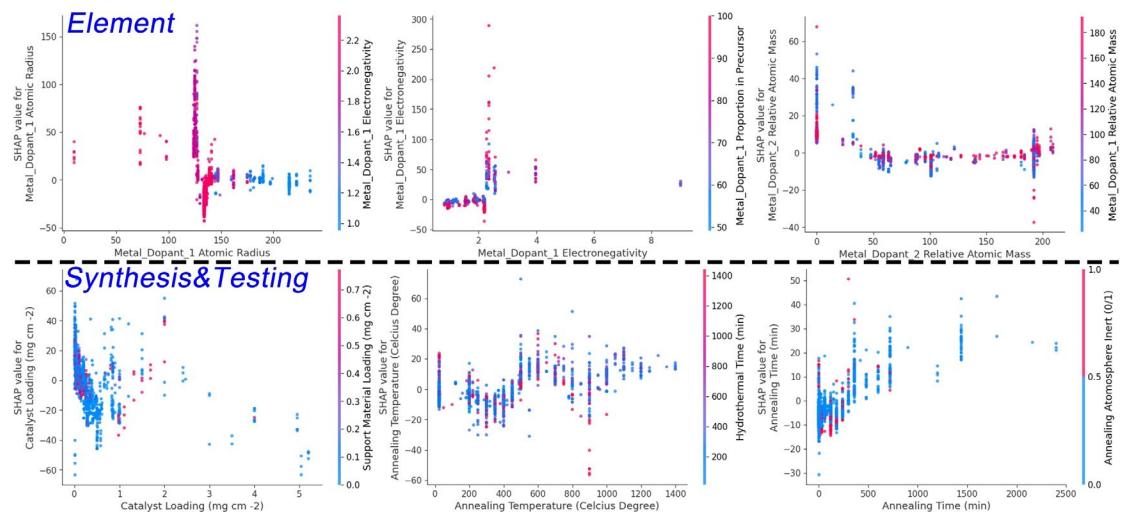


Fig. SN 4-17 SHAP two-dimensional interaction plots screened based on SHAP feature importance ranking's top 5 features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the activity. The interaction features are selected from same category high ranking features, also with high degree of input interaction strength as indicated by H statistics.

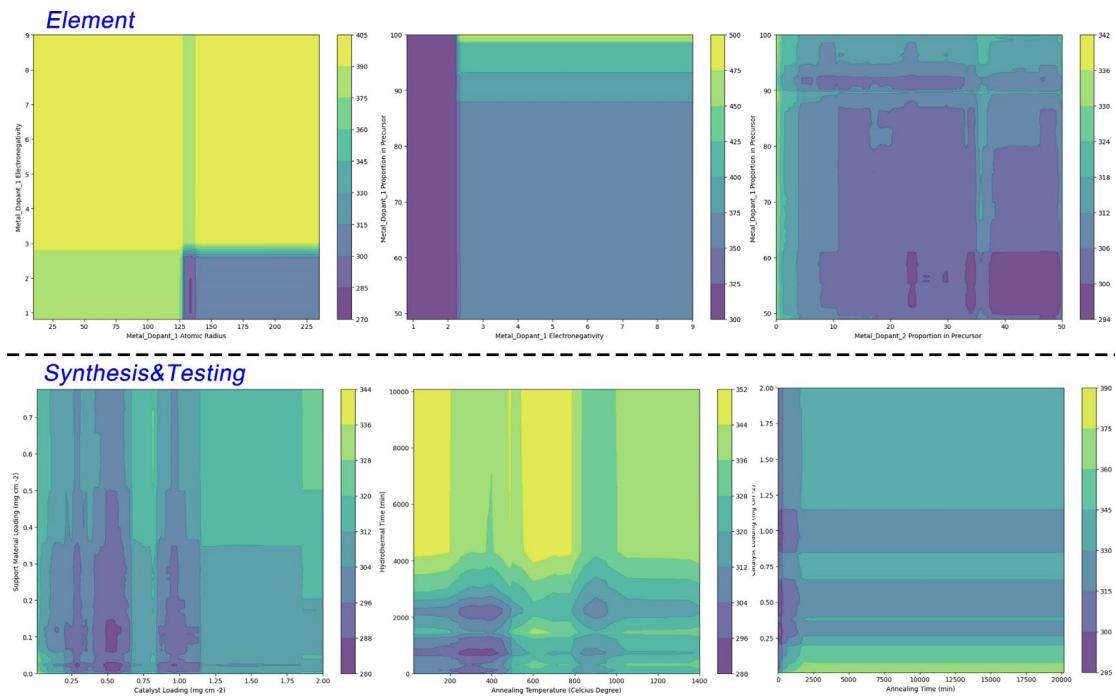


Fig. SN 4-18 PDP two-dimensional interaction plots analyzing the same input feature pairs as in **Fig. SN 4-17**.

Blackbox Interpretation of the ML Committee based on Domain Knowledge: OER Stability

It's imperative to recognize that, in practical applications for water electrolysis, the stability of these catalysts in acidic environments is as crucial, if not more so, than their catalytic activity. This heightened significance stems from the harsh operational conditions typical of acidic OER processes, where catalysts are subjected to high overpotentials and corrosive environments. For stability, the parameters related to synthesis thermodynamics and kinetics, like temperature and time duration, were found to be particularly influential, as visualized in Fig. SN 4-24 and Fig. SN 4-30.

Prolonged exposure to such conditions can lead to rapid degradation of catalytic materials, especially in the case of commonly used materials like RuO₂ and IrO₂, which, despite their high activity, often suffer from limited durability in acidic media. This degradation not only undermines the long-term efficiency and effectiveness of the electrolysis process but also escalates operational costs due to frequent catalyst replacement. Experimental electrocatalyst studies generally support such a rationale, demonstrating that synthesis parameters like temperature and duration play a crucial role in enhancing the stability of electrocatalysts in acidic conditions. The previous study we have discussed also showed that higher annealing temperatures improved the surface area and conductivity of Cu-Co oxide nanosheets, leading to better oxygen evolution reaction OER activity and durability(27). Another investigation of a typical binary alloy Pt-Cu system also validated that optimal annealing conditions significantly increased the stability of alloy catalysts by enhancing crystallinity and surface morphology(51). Similarly, research on rapid thermal annealing of Pt-Pd system thin film electrocatalysts revealed that higher temperatures resulted in improved structural integrity, reducing degradation during prolonged electrochemical reactions(52). Collectively, these findings highlight the critical influence of thermal treatment in reinforcing catalyst stability, especially in acidic environments.

As for metal elements, while the atomic radius of the primary metal element dictates activity as mentioned in the activity discussion section, the atomic radii of the second and third elements rank highly for stability (with even the third element surpassing the second element), suggesting their pivotal role despite their lower proportions. As we have discussed previously how atomic radius and electronegativity would be decisive in determining the electronic structure and hence intermediate binding energies, the intriguing insights here is that doping at least one and favorably two elements is important in boosting stability. Our findings have been confirmed by several recent experimental investigations. It appears that doping Ru and Ir-based catalysts for the acidic OER with one or more metal elements truly increases their stability and durability, especially for practical applications in proton exchange membrane water electrolysis. In practice, one significant issue with these catalysts is their propensity to deteriorate as a result of lattice oxygen participation, which speeds up metal dissolution and structural collapse. According to one study, IrO₂ can be made more durable in acidic OER conditions by shifting the Ir 5d band center closer to the Fermi level through double doping and strain engineering (50). Another strategy stabilizes the Ru active sites and stops degradation by altering the local Ru–O–Ir environment to suppress lattice oxygen involvement(53). It has also been demonstrated that tailored reaction pathways greatly increase stability for Ru-based oxides. The catalyst can sustain steady performance over extended operation by regulating the catalytic pathway and modifying the intermediate binding strengths in Sn-doped RuO₂(54). Ni-stabilized RuO₂, a non-Ir-based catalyst, further illustrates how adding nickel stabilizes the RuO₂ lattice and lowers degradation rates by keeping lattice oxygen

from taking part in the reaction(55). These investigations collectively imply that lattice oxygen involvement can be successfully reduced by altering the local environment surrounding Ru and Ir sites through doping, strain engineering, and customized reaction pathways, improving stability and durability. In all, while advancing the activity of OER catalysts is essential, ensuring their stability under acidic conditions becomes a pivotal challenge that must be addressed. Solving this issue is key to realizing sustainable and economically viable hydrogen production through water electrolysis, as it directly impacts the longevity and reliability of the system, thus playing a vital role in the transition towards carbon-neutral energy solutions.

Due to the above reasons, we have as mentioned in the main text, collected the datasets, and conducted the data mining processes as well. Here we apply the same ML committee training-black box interpretation workflow modules to gain insights where key points have been provided in the main text. The analysis highlights a preference for stability when the second metal is present in higher proportions, exceeding ~35%, and when selected from a range of larger atomic radius elements, generally above 136 pm, like alkali/alkaline or heavy metals. This is illustrated in **Fig. SN 4-33** and **Fig. SN 4-35**. Here, we would comprehensively discuss all the modeling and analysis results in detail. We have stored the original results for each model on the online repository as well as **Fig. OR69-OR120** for readers to check and retrieve. The following discussion are based on the ensembled SHAP/PDP/H-statistic values, which we have visualized in **Fig. SN 4-19- Fig. SN 4-36**. A separate discussion from different feature domain is as follows:

1. **Elemental Features Insights:** As previously mentioned, in contrast to the ranking order observed when studying activity, the 2nd and 3rd metal elements significantly impact stability performance, as evidenced by their high-ranking orders from SHAP summary plots. This suggests the pivotal role of doping elements in regulating long-term service performance. However, in contrast to activity analysis, the cohort plots for stability identified different features and boundaries. In high-quality models, the atomic mass of the primary metal element is highlighted, implying a preference for heavier metal elements (with atomic mass larger than 196) as the primary content for stability purposes (**Fig. SN 4-31a**). This seems to be favoring mainly Ir-based oxides, consistent with the domain knowledge. For models based on the full dataset, the SHAP-selected feature and boundary is the atomic radius of the second metal element at 154 pm. Comparing this with the periodic table, a unique message emerges. In our investigation of secondary metal elements as dopants, we identified a notable trend for elements with an atomic radius greater than 154 pm. These elements, including rare earth metals like La, Ce, Nd, and post-transition metals such as In, Sn, Pb, along with alkaline earth metals like Ba and Sr, demonstrate a significant increase in importance according to our cohort graph in **Fig. SN 4-25b**. The machine learning models, as per SHAP summary plot analysis (**Fig. SN 4-23a**, **Fig. SN 4-28a**), tend to favor these larger-radius elements, potentially due to the positive impact of alkaline earth metals on stability. This preference suggests exploring dopants from these categories, typically located in the s, f, and some p blocks of the periodic table. Conversely, elements with an atomic radius smaller than 154 pm, predominantly transition metals like Ti, Cr, Fe, and noble metals such as Pt and Cu, exhibit distinct properties and interactions. This dichotomy in elemental characteristics based on atomic radius is crucial for selecting

appropriate dopants to enhance stability in acidic OER catalysts. In general, we want to highlight the immense possibilities of doping that could lead to optimal results. But through ML insights, it appears that transition metals and noble metals may be more favorable for activity, while rare earth, post-transition, and alkaline earth metals are likely more crucial for enhancing stability.

A deeper insight comparing the studies in the field for the rationales behind this is worth discussing. The choice of dopants is essential for adjusting the stability and activity of electrocatalysts, and various metal types may offer special advantages. As we have previously discussed, typical transition metals such as Fe, Cr are well known for their catalytic efficiency. This is primarily because of their partially filled d-orbitals, which improve electron transfer and speed up reaction kinetics, making them excellent for high activity in catalyst behavior. Because of their high electron mobility and low activation barriers, noble transitional metals—like Pt and Pd in particular—are advantageous for applications where catalytic activity is a top priority(56). Rare earth and alkaline earth metals, on the other hand, like Sr, Ba, Ce, and La, provide notable stability enhancements. These components can stop structural degradation at high temperatures by stabilizing the catalyst structure under operating conditions, frequently through processes like phase stabilization. Research on La- and Ce-doped oxides has shown that these elements can prevent phase changes, improving structural stability over long periods of time(57). By strengthening the perovskite matrix, alkaline earth metals such as Ba and Sr provide additional stability, which is essential for maintaining functionality in oxidative or acidic environments(58). These observations highlight the fact that rare earth and alkaline earth metals are essential for increasing durability, even though transition and noble metals are beneficial for increasing catalytic activity. In order to maximize both activity and stability for electrocatalysts, a careful selection of metal types is necessary, according to this nuanced understanding of dopant roles. This opens the door for more durable, multipurpose materials.

2. **Synthesis and Testing Parameters:** Focusing on synthesis and testing parameters, their significance aligns with expectations, given their role in calculating the average decay rate. As discussed previously, these parameters, specifically stability testing time, appear to have gained prominence in global ranking, alongside elemental features. Annealing temperature and hydrothermal time rank higher, underlining their importance. Consistent with our chemical intuition, longer durations, and higher temperatures for both hydrothermal processes (which also apply to precursor mixing when no hydrothermal treatment is used) and annealing processes seem beneficial for improved stability(27, 51, 52). This is relatively easy to agree with, because higher temperatures and longer annealing and hydrothermal times will help oxide samples form a more stable phase or enhance the crystallinity. This is typically experimentally validated especially in complex systems like perovskite oxides(59), or other common oxide systems(60-62). Cohort plots suggest that these factors become increasingly important when stability test durations exceed ~19 hours.

3. **Interaction Plots:** Examining feature interactions through SHAP clustering and H-statistics matrices offers further insights. Firstly, SHAP clustering analysis reveals that synthesis and testing parameters do not directly cluster

with any elemental features. Additionally, elemental feature clustering tends to occur within the same element rather than across different elements, except for a clustering tendency between the third and fourth metal elements, as shown in **Fig. SN 4-27b**. The Friedman's H-statistic matrix reveals similar patterns in overall interaction degrees between different feature types. Like previous insights that for activity committees, synthesis and testing features exhibit stronger interactions among themselves, while their interactions with elemental features are weaker. Interactions among elemental features are even less pronounced. This pattern indicates that for stability performance, chemical composition and synthesis parameters again tend to affect the output more independently, without strong synergistic effects.

Building on similar insights previously discussed for activity, we shall now delve deeper into the failure modes of OER electrocatalysts. The synthesis parameters and elemental composition have a significant impact on the long-term stability of OER catalysts, and each of these factors contributes differently to different aspects of catalyst durability. Research has shown that elemental composition and preparation methods have an independent impact on **structural degradation**, such as surface reconstruction or amorphization. For example, certain crystalline structures can be stabilized by optimizing the composition of perovskite oxides, and crystallinity can be further controlled by synthesis parameters such as hydrothermal conditions without depending on elemental changes(63). Similar to how corrosion-resistant metals can be chosen to control **chemical corrosion**, synthesis techniques such as annealing temperatures can also independently strengthen stability by producing more compact or well-adhered catalyst surfaces, which directly affects a material's resistance to dissolution without changing its elemental composition(64). Particularly in nickel-based catalysts (OER in alkaline specifically), elemental factors play a major role in controlling the formation of **surface passivation** layers like NiOx. However, synthesis parameters like annealing independently affect the extent and uniformity of these layers, stabilizing conductivity without changing elemental composition(65). Another crucial failure mode, **particle aggregation**, is highly influenced by synthesis decisions like hydrothermal time, which has a direct impact on particle size and dispersion while preserving stability regardless of the inherent characteristics of the selected elements(66). Lastly, the importance of synthesis parameters or methodology which controls morphology and ensures mechanical resilience independent of the catalyst's elemental composition, is highlighted by **mechanical degradation** from gas evolution reactions. This is important for resisting the adhesive force of generated oxygen bubbles and shedding of catalyst particles(67, 68).

All things considered, these results demonstrate the relatively independent contributions of elemental composition and synthesis parameters to various stability aspects in OER catalysts. Nevertheless, these factors still intrinsically interact with each other, and this is the reason why we choose not only to include elemental compositions but also synthesis parameters in the experimental-driven active learning exploration and exploitation phase for a global better outcome.

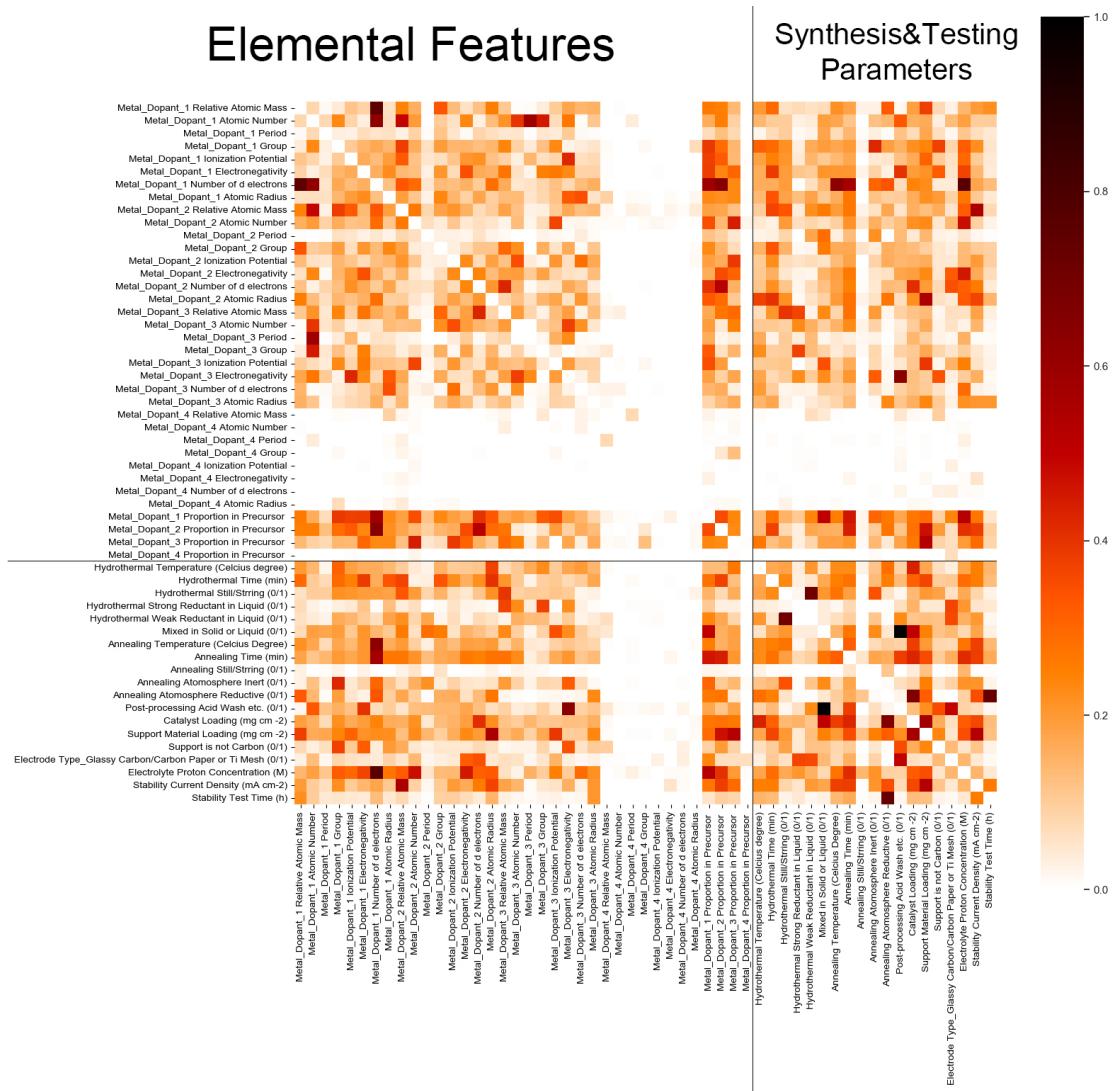


Fig. SN 4-19 Weighted averaged Friedman's H statistic matrix heatmap of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability. The averaged values of interaction between different regions are elemental-elemental: 0.08, elemental-synthesis & test: 0.11, synthesis & test-synthesis & test: 0.16. If only do not consider the fourth metal, the values would be elemental-elemental: 0.14, elemental-synthesis & test: 0.13, synthesis & test-synthesis & test: 0.16.

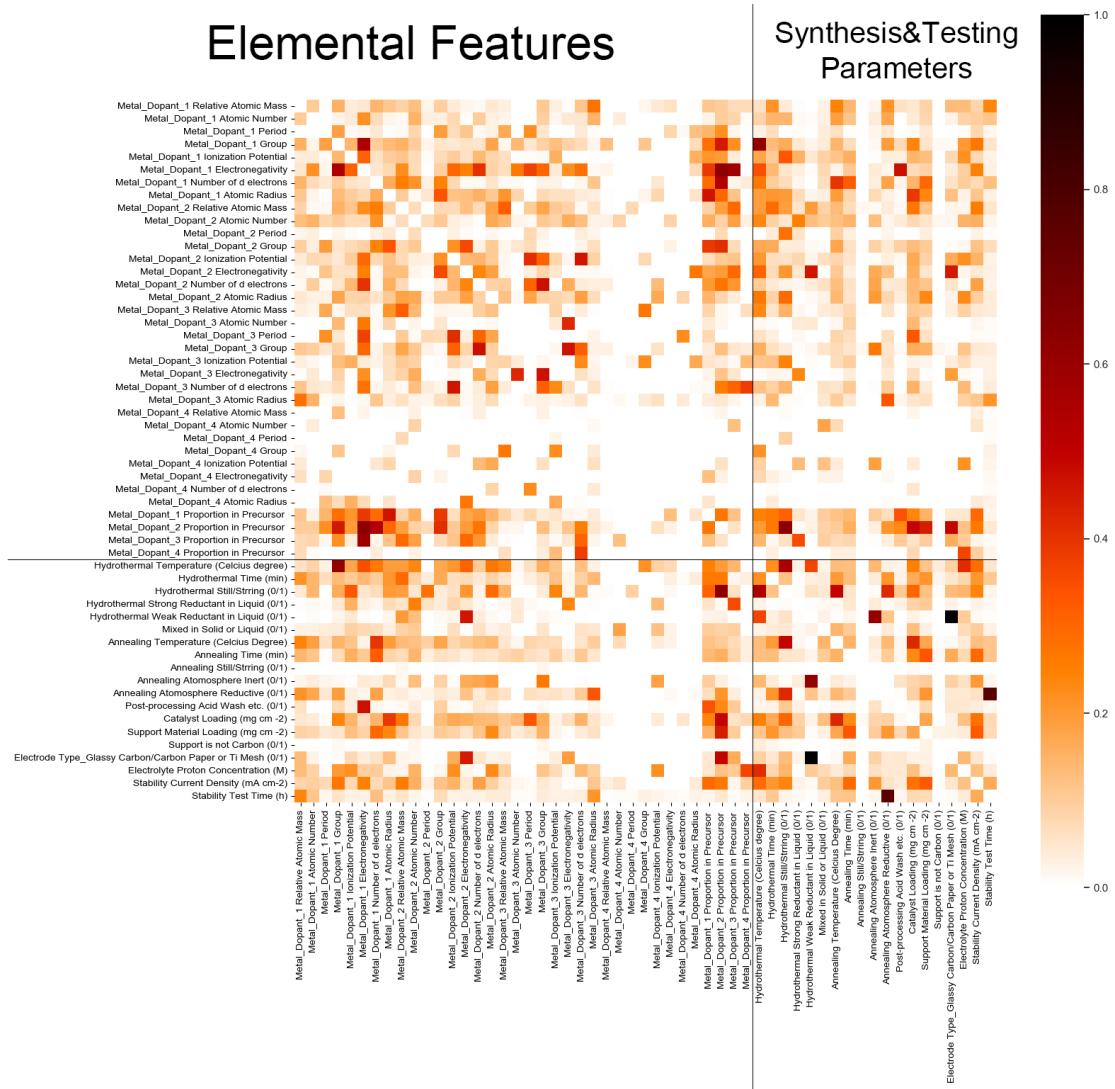


Fig. SN 4-20 Weighted averaged Friedman's H statistic matrix heatmap of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability. The averaged values of interaction between different regions are elemental-elemental: 0.06, elemental-synthesis & test: 0.07, synthesis & test-synthesis & test: 0.09. If only do not consider the fourth metal, the values would be elemental-elemental: 0.09, elemental-synthesis & test: 0.09, synthesis & test-synthesis & test: 0.09.

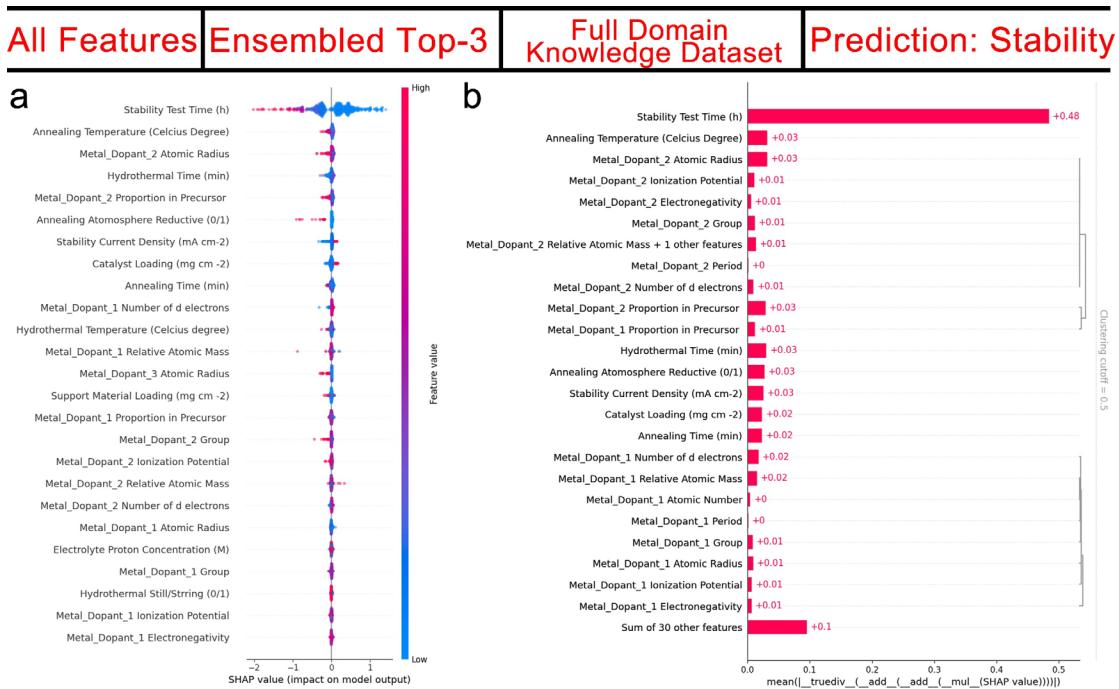


Fig. SN 4-21 (a) SHAP summary plot and (b) SHAP clustering bar plot included all features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability.

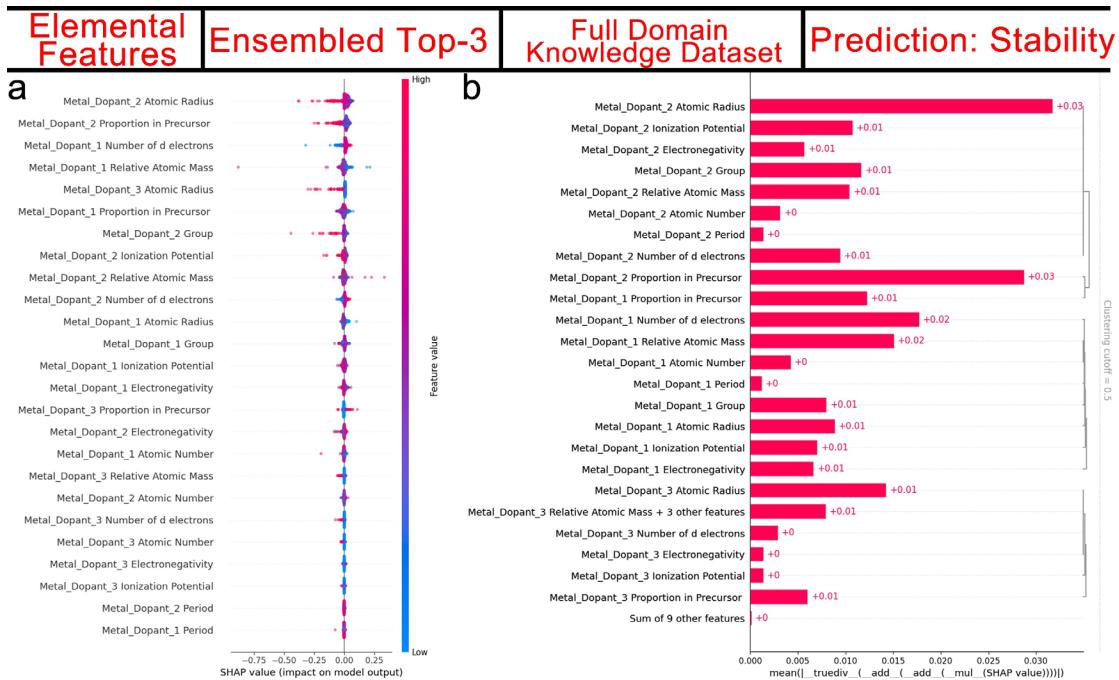


Fig. SN 4-22 (a) SHAP summary plot and (b) SHAP clustering bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability.

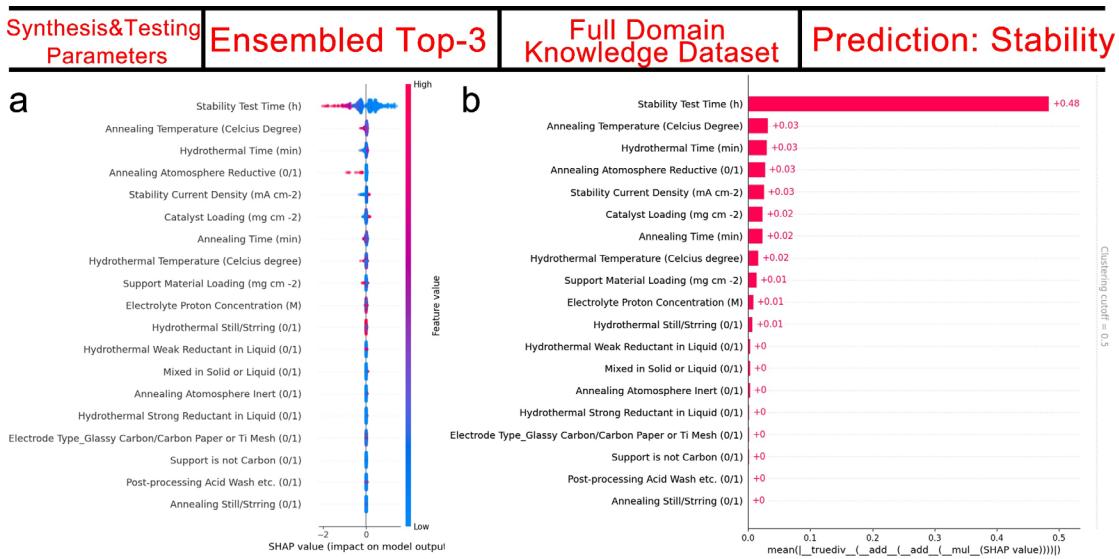


Fig. SN 4-23 (a) SHAP summary plot and (b) SHAP clustering bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability.

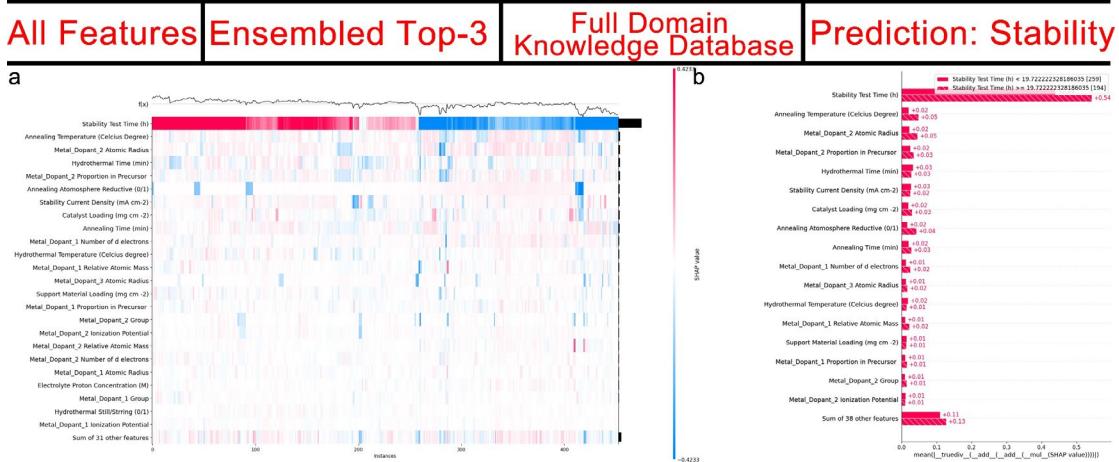


Fig. SN 4-24 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included all features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability.

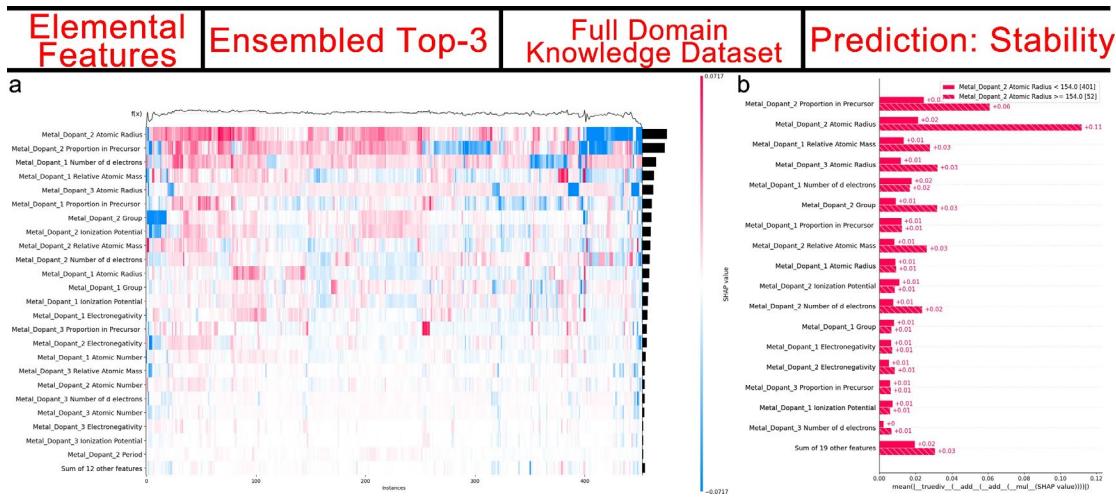


Fig. SN 4-25 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability.

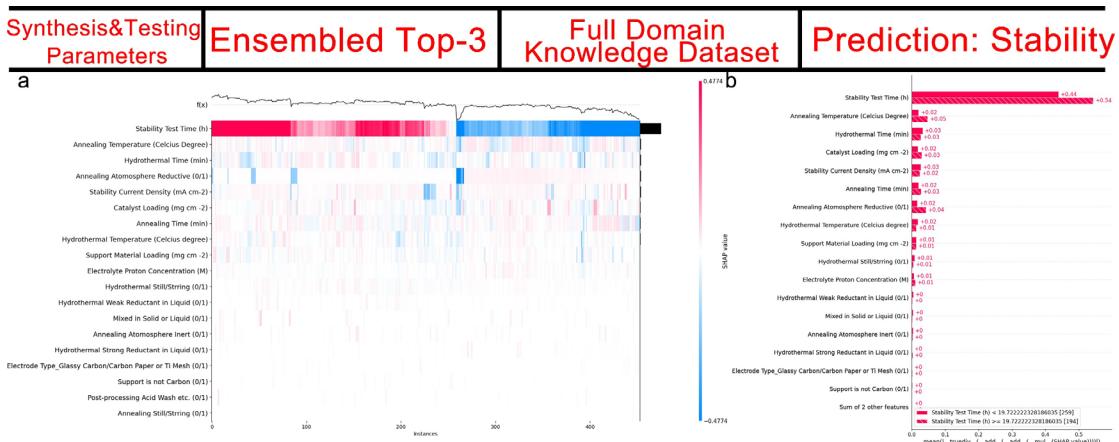


Fig. SN 4-26 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability.

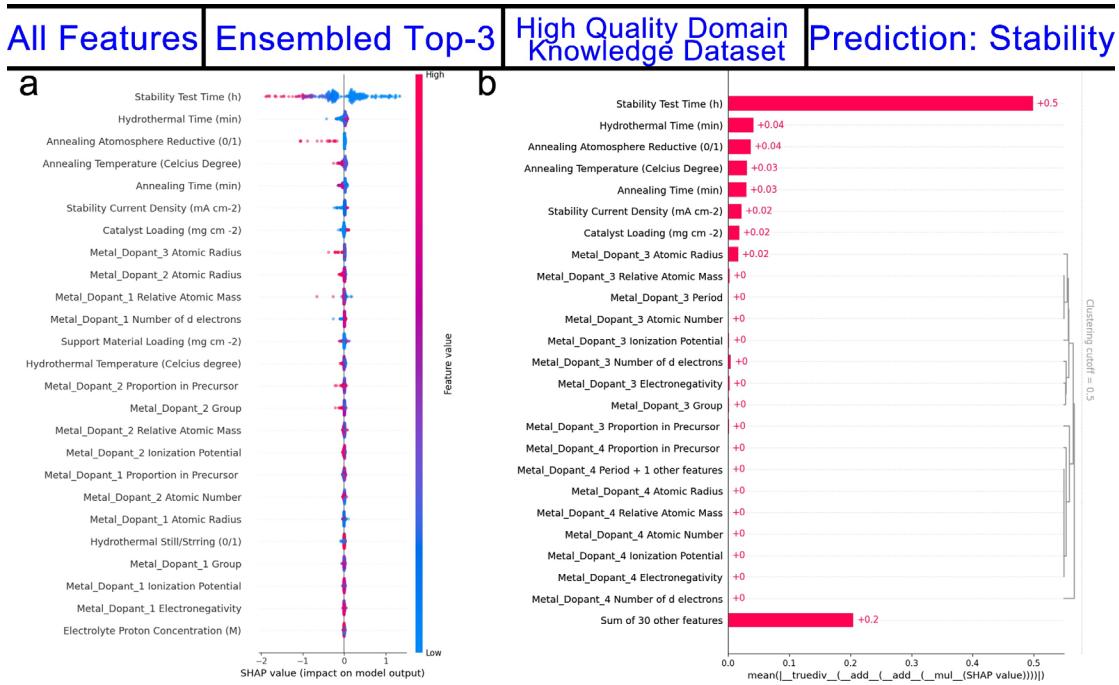


Fig. SN 4-27 (a) SHAP summary plot and **(b)** SHAP clustering bar plot included all features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability.

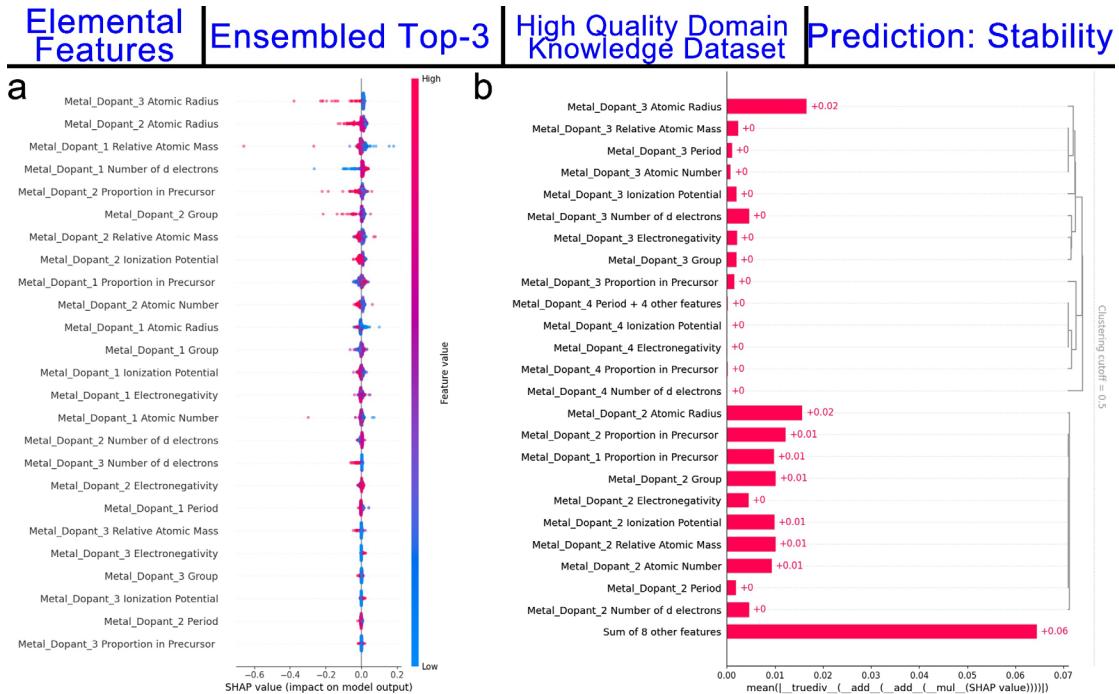


Fig. SN 4-28 (a) SHAP summary plot and (b) SHAP clustering bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability.

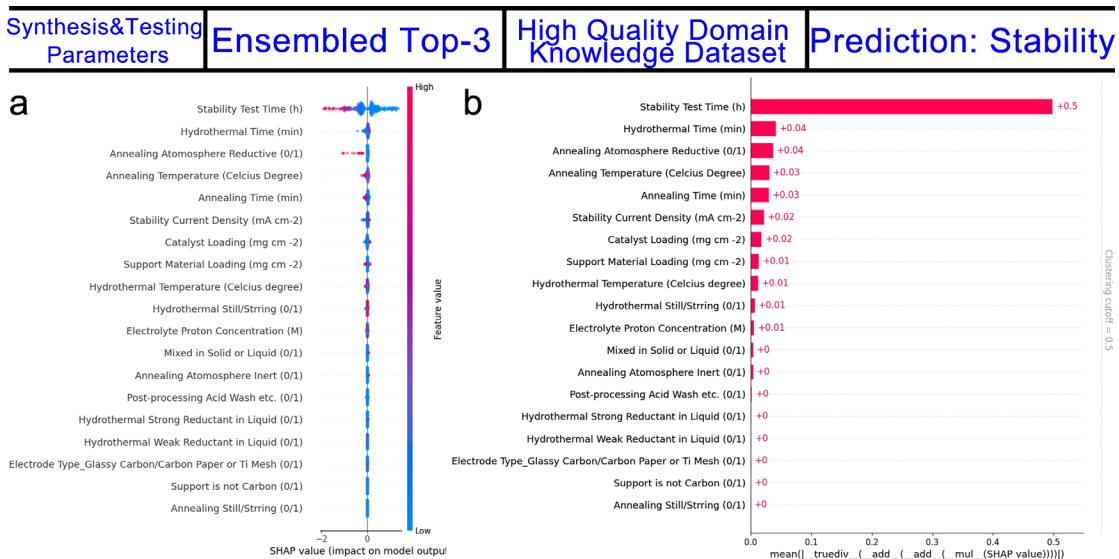


Fig. SN 4-29 (a) SHAP summary plot and **(b)** SHAP clustering bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability.

All Features | Ensembled Top-3 | High Quality Domain Knowledge Dataset | Prediction: Stability

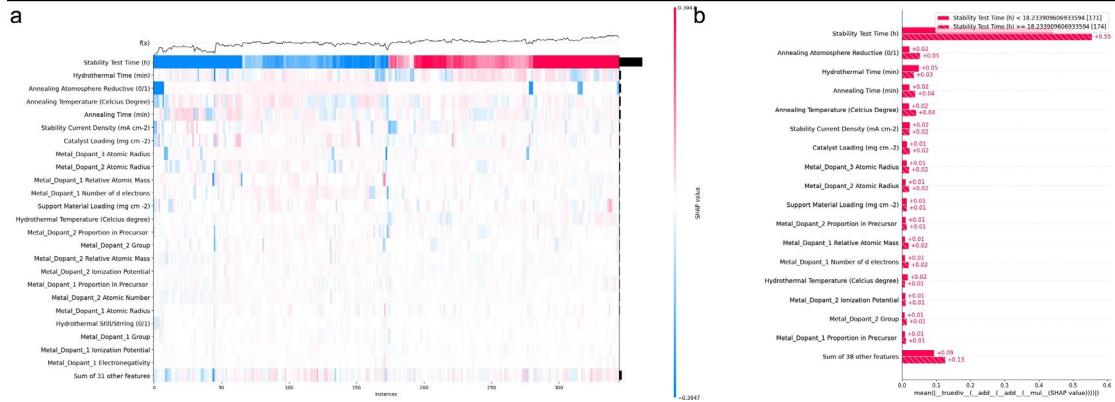


Fig. SN 4-30 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included all features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability.

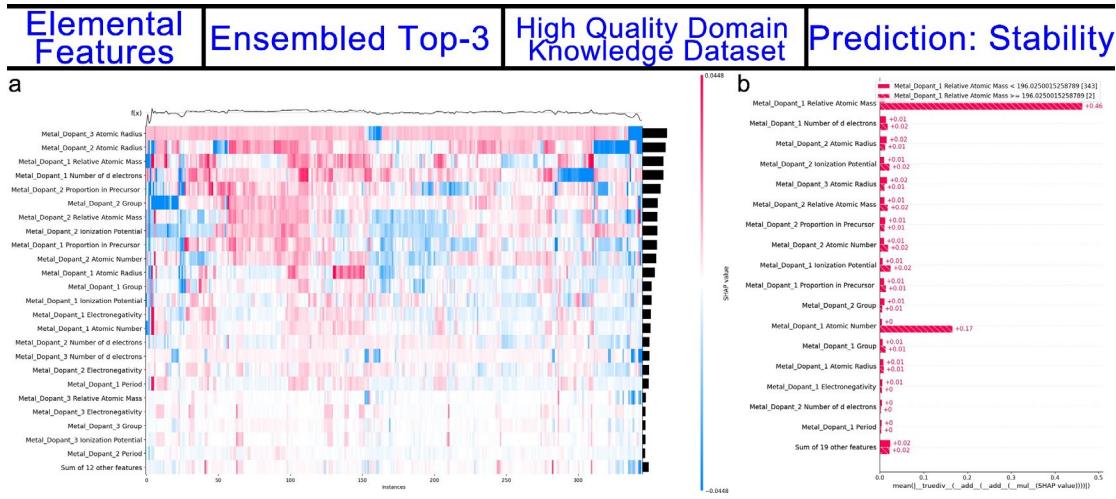


Fig. SN 4-31 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included elemental features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability.

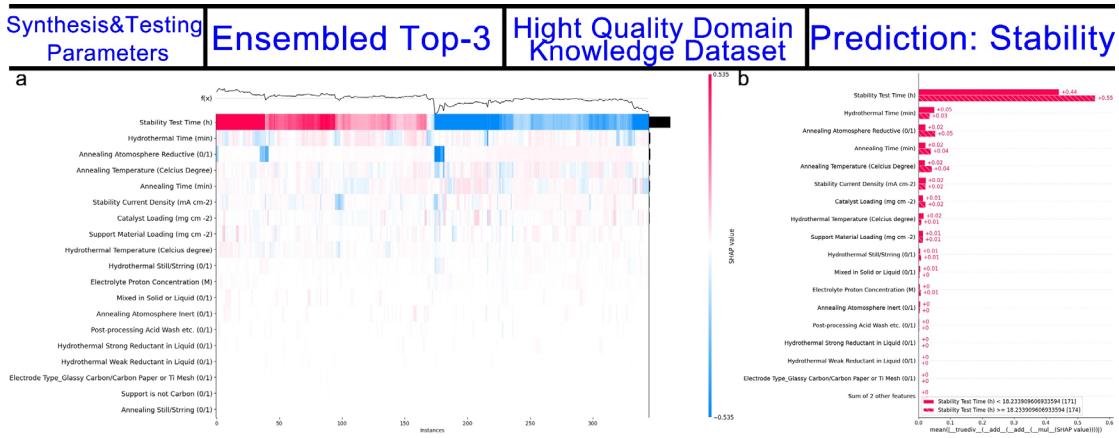


Fig. SN 4-32 (a) SHAP overview heat map plot and (b) SHAP cohort bar plot included synthesis & testing parameters of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability.

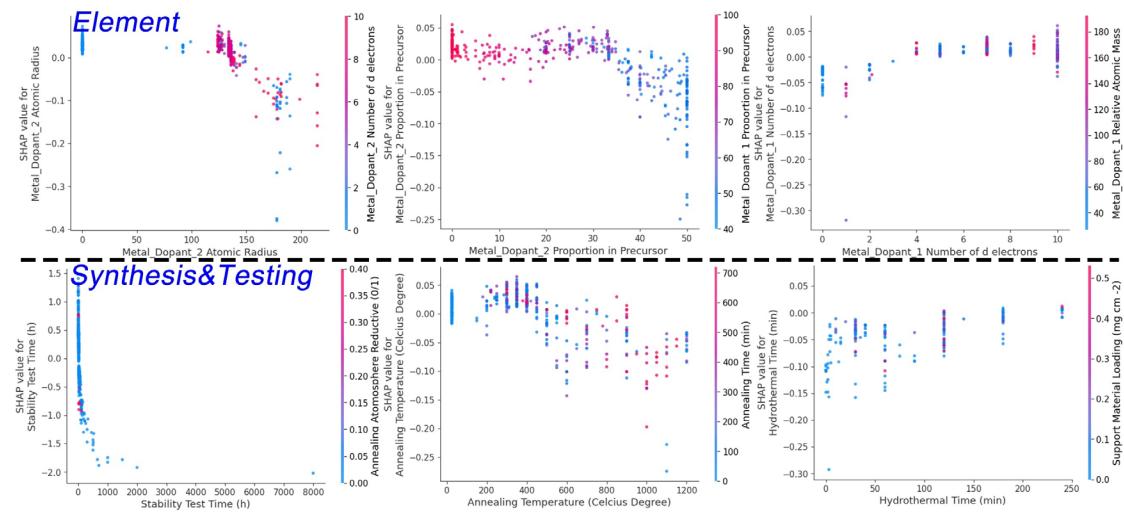


Fig. SN 4-33 SHAP two-dimensional interaction plots screened based on SHAP feature importance ranking's top 5 features of ensembled top-3 ML models' results, which are trained based on the full domain knowledge dataset to predict the stability. The interaction features are selected from same category high ranking features, also with high degree of input interaction strength as indicated by H statistics.

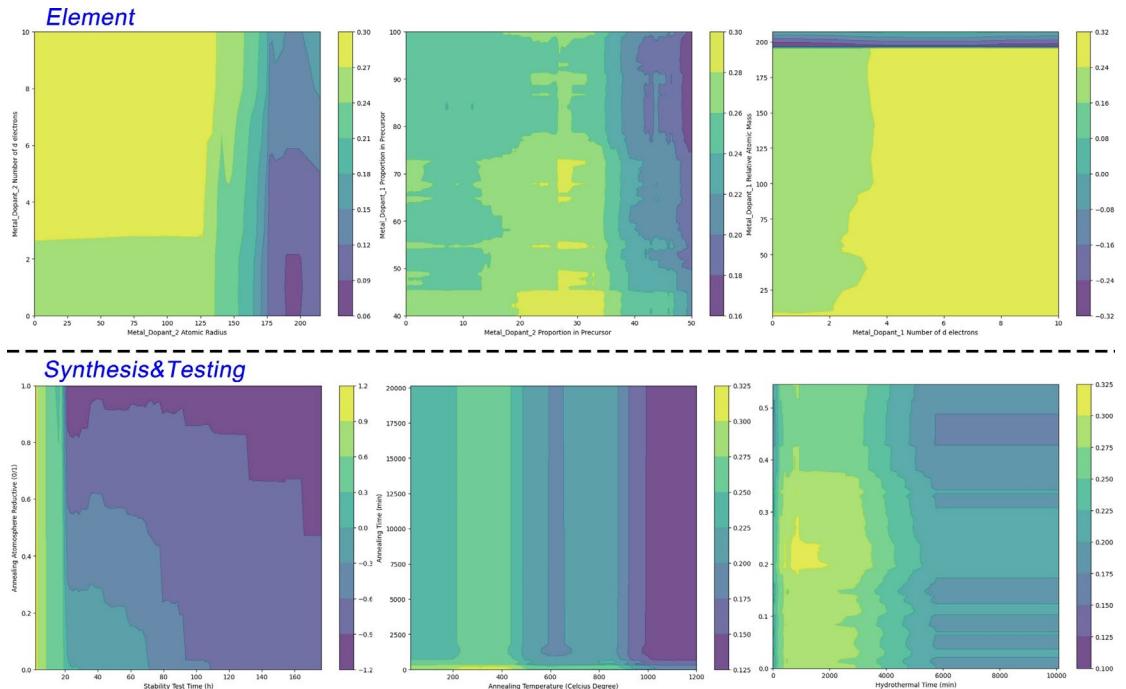


Fig. SN 4-34 PDP two-dimensional interaction plots analyzing the same input feature pairs as in Fig. SN 4-33.

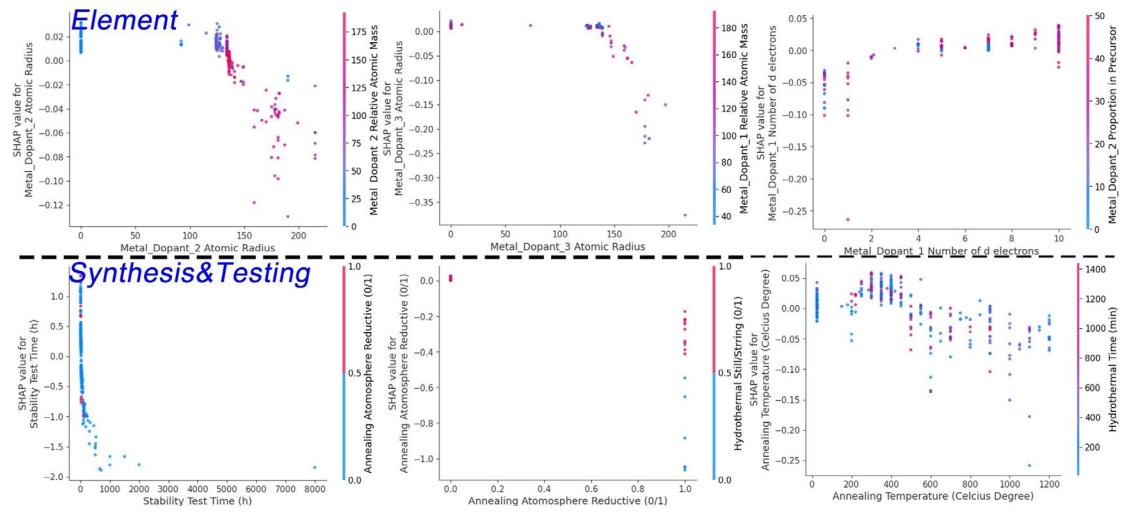


Fig. SN 4-35 SHAP two-dimensional interaction plots screened based on SHAP feature importance ranking's top 5 features of ensembled top-3 ML models' results, which are trained based on the high-quality domain knowledge dataset to predict the stability. The interaction features are selected from same category high ranking features, also with high degree of input interaction strength as indicated by H statistics.

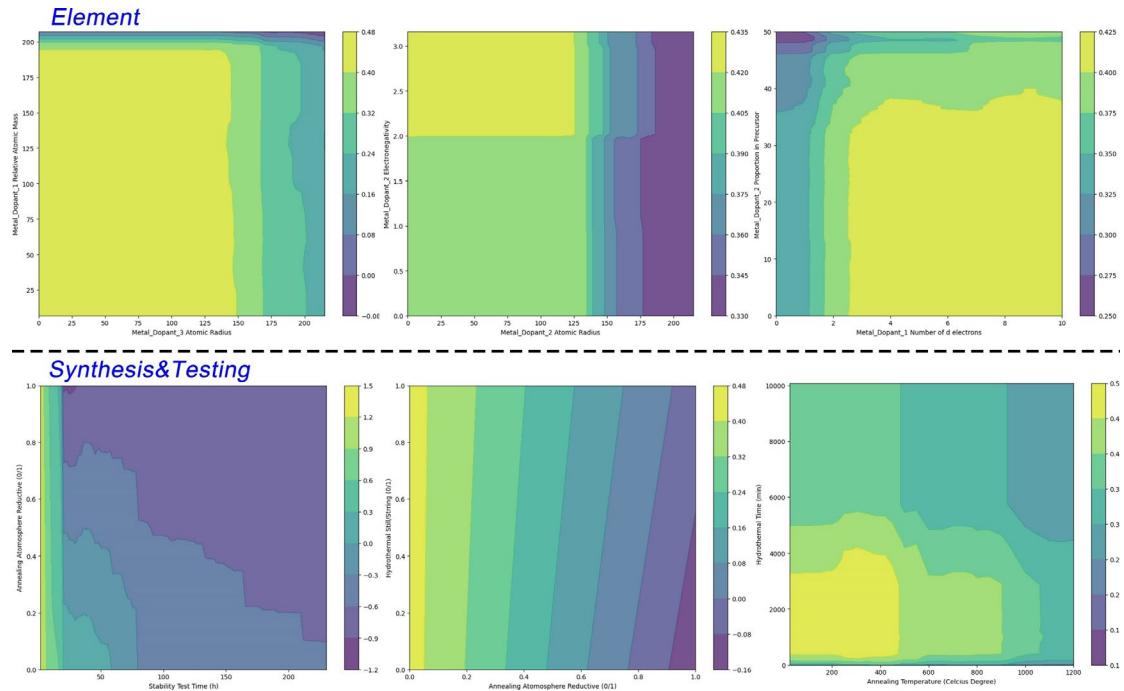


Fig. SN 4-36 PDP two-dimensional interaction plots analyzing the same input feature pairs as in Fig. SN 4-35.

Supplementary Note Discussion SND 4-3

Classification Models LSE

Level Set Estimation (LSE) is a pivotal method in ML, particularly beneficial for enhancing the precision of predictive models. At its core, LSE involves estimating the contours or “levels” of a function at which the function takes on specific values. In the context of our work, LSE is crucial for defining discrete output levels within a continuous dataset, enabling us to categorize data points into binary classes based on predetermined thresholds. This method is particularly useful in scenarios where a clear demarcation between different states or conditions is required, such as in classification tasks. The necessity of LSE alongside regression models stems from its ability to provide a different perspective on the dataset. While regression models excel in predicting continuous variables and capturing the nuanced variations within data, they may not always effectively delineate distinct categories or thresholds inherent in the data. LSE addresses this limitation by segmenting continuous data into distinct classes, enabling us to perform a two-pronged analysis: employing regression for understanding continuous trends and variations, and deploying LSE for more granular classification and to pinpoint threshold-based insights.

In our work, except for the previously focused regression modelling, we apply LSE to support and complement our regression models. By doing so, we achieve a more comprehensive analysis of our dataset, capable of capturing both the subtle continuous variations through regression and the distinct categorical divisions through LSE. This dual approach ensures a more robust and thorough understanding of our data, allowing us to derive insights that might be overlooked if only one method were employed. The use of LSE, therefore, not only enriches our analysis but also validates the versatility and applicability of our dataset across different types of predictive modeling tasks. In the core analysis, our focus primarily revolves around regression modeling for activity and stability. These regression models utilize the input features depicted in **Fig. SN 1-1** (in **Supplementary Note 1**) to predict η_{10} or Log (Decay Rate (mV h^{-1})). Complementing this, we have employed LSE which maintains the same input features and utilizes the identical dataset. However, unlike the continuous output in regression, LSE targets discrete outputs defined by various thresholds for binary classifications. We adhere to the six thresholds previously outlined in the unsupervised data mining section: 200/250/300 mV for activity and -1/0/+1 for stability.

This approach entails a comprehensive process, encompassing everything from ML committee training to the interpretation of black-box models. Through this, we achieve comparable results in forms akin to SHAP/PDP analyses, among others. The rationale behind implementing LSE is to validate that, aside from achieving high-quality regression ML models with test R^2 exceeding 0.8 or even nearing 0.9 in the main text; classification tasks on the same dataset can yield even more promising performance metrics. As **Fig. SN 4-37**, a pivotal summary, illustrates, out of the 12 committees and 132 hyperparameter-optimized models, the top-performing half demonstrate exceptionally high average levels of accuracy, F1 score, and Area Under the Curve (AUC) - often surpassing 0.95. This underscores the reliability of our domain knowledge dataset used for regression and its balanced nature. The refined granularity of our regression results is thus corroborated.

Additionally, **Fig. SN 4-37** reveals that ML models trained on high-quality datasets tend to exhibit superior performance metrics, reinforcing the conclusions drawn in the main text. The extensive repository of black-box interpretation results, ROC plots, and detailed data for all related LSE classification models, stored in online repository

Figs. OR121-OR564, are available for interested readers. These results show general high consistency with the regression analyses discussed in **Supplementary Note Discussion SND 4-2**, especially concerning metal elements, synthesis, and testing parameters. This further affirms the internal consistency of our work. However, notable distinctions that we could briefly summarize for readers include:

1. Classification models, due to their coarser granularity, tend to exhibit more discontinuous boundaries in SHAP and PDP interaction plots. This phenomenon is attributed to the binary nature of classification outputs, where the decision boundaries are more sharply defined as compared to the continuous spectrum of regression models.
2. The feature importance in classification models displays more dramatic differences, highlighting distinct predictive variables that significantly influence classification outcomes, as opposed to the more nuanced variations observed in regression.
3. In terms of black-box interpretation, classification models tend to emphasize threshold-based decision criteria, where features contribute to pushing the output towards one class or the other. This is in contrast to regression models, where the interpretation often revolves around how features influence the degree or magnitude of the response variable.
4. Another aspect where classification differs from regression in black-box interpretation is the prominence of interaction effects. In classification models, the interaction between features often plays a critical role in determining class boundaries, which can be more complex and nuanced than the additive or linear interactions typically seen in regression models. This is evidenced by generally higher value of H-statistics.

Regarding the metrics employed, it should be noted that this part of LSE study uses two distinct methods for calculating the AUC: one based on binary predictions and the other derived from predicted probability scores. The binary prediction-based AUC offers a direct measure of model performance at specific decision thresholds, essential in scenarios requiring definitive decisions. Conversely, the probability-based AUC provides a broader view of model performance across various thresholds, capturing the model's discriminatory power more effectively. Thus, both types of AUC contribute valuable insights for a comprehensive model evaluation. In the OR Figs. for visualization, the computed AUC shown along with ROC curves are probability-based AUC, but to represent each model's classification ability, the used AUC is binary-prediction-based AUC.

Different from regression models, where R^2 (ranging between 0 and 1) is used to weigh model results, we incorporate a combination of Accuracy, F1 score, and AUC to derive an average metric for classification models. This approach offers a more holistic evaluation, capturing different aspects of model performance that are crucial for classification tasks. Such a multi-metric approach is advantageous in classification models as it provides a more nuanced understanding of model efficacy, especially when delving into black-box interpretations for deeper insights.

In conclusion, through the LSE process, we have trained classification model committees using the same full/high-quality dataset with discrete output. The exceptional performance of these models not only validates the quality of our dataset but also reinforces the reliability of the insights presented in the discussed regression ML model committees.

For this part's scripts, please check the online repository's directory “/Machine Learning Databases and Script/Domain Knowledge Based Initial ML Committee and Blackbox Interpretation/Classification (LSE)/*”

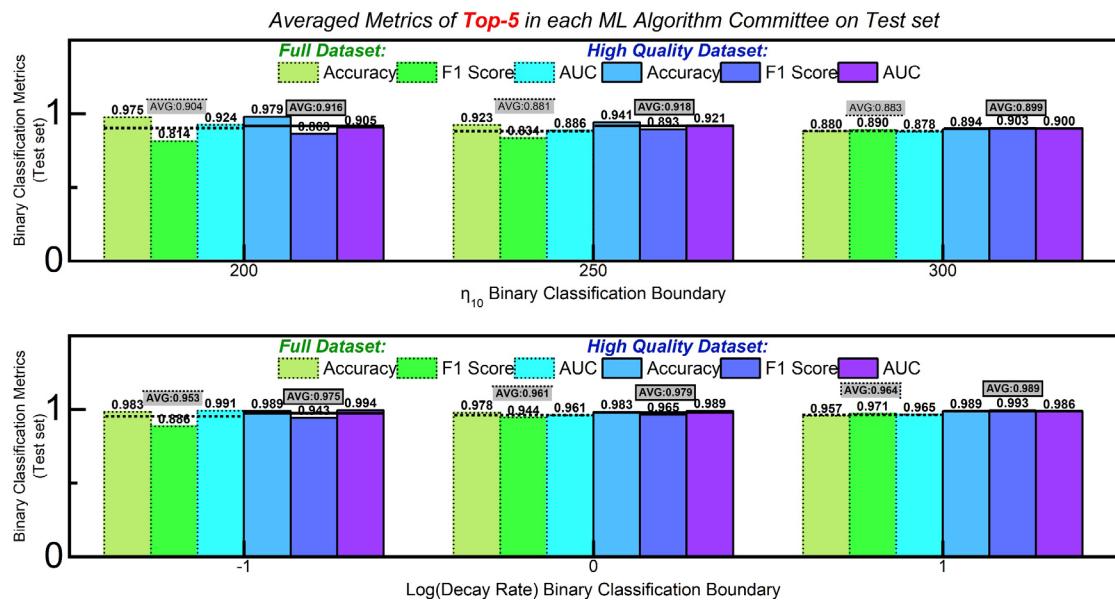


Fig. SN 4-37 A column summary plots showing the performances of classification models in the LSE processes. Each single column in the six groups represents the average value of corresponding binary classification metrics: accuracy, F1 score, AUC by the top-5 performing ML models in the corresponding committee.

Supplementary Note 4 References:

1. R. Ding *et al.*, Boosting the optimization of membrane electrode assembly in proton exchange membrane fuel cells guided by explainable artificial intelligence. *Energy and AI* **5**, 100098 (2021).
2. R. Ding *et al.*, Effectively Increasing Pt Utilization Efficiency of the Membrane Electrode Assembly in Proton Exchange Membrane Fuel Cells through Multiparameter Optimization Guided by Machine Learning. *ACS Applied Materials & Interfaces* **14**, 8010-8024 (2022).
3. R. Ding *et al.*, Guiding the Optimization of Membrane Electrode Assembly in a Proton Exchange Membrane Water Electrolyzer by Machine Learning Modeling and Black-Box Interpretation. *ACS Sustainable Chemistry & Engineering* **10**, 4561-4578 (2022).
4. R. Ding *et al.*, Inspecting design rules of metal-nitrogen-carbon catalysts for electrochemical CO₂ reduction reaction: From a data science perspective. *Nano Research* **16**, 264–280 (2022).
5. H. Aziz, B. de Keijzer, paper presented at the 31st International Symposium on Theoretical Aspects of Computer Science (STACS), Lyon, FRANCE, Mar 05-08 2014.
6. B. M. Greenwell, pdp: an R package for constructing partial dependence plots. *R Journal* **9**, 421-436 (2017).
7. J. H. Friedman, B. E. Popescu, Predictive learning via rule ensembles. *The Annals of Applied Statistics* **2**, 916-954 (2008).
8. R. Ding, X. Wang, A. Tan, J. Li, J. Liu, Unlocking New Insights for Electrocatalyst Design: A Unique Data Science Workflow Leveraging Internet-Sourced Big Data. *ACS Catalysis* **13**, 13267-13281 (2023).
9. C. Scott, M. Davenport, Regression level set estimation via cost-sensitive classification. *Ieee Transactions on Signal Processing* **55**, 2752-2757 (2007).
10. R. L. Doyle, M. E. G. Lyons, The oxygen evolution reaction: mechanistic concepts and catalyst design. *Photoelectrochemical Solar Fuel Production: From Basic Principles to Advanced Devices*, 41-104 %@ 3319296396 (2016).
11. C. Hu *et al.*, Advanced Catalyst Design Strategies and In-Situ Characterization Techniques for Enhancing Electrocatalytic Activity and Stability of Oxygen Evolution Reaction. *Electrochemical Energy Reviews* **7**, 19 %@ 2520-8489 (2024).
12. S. R. Ede, Z. Luo, Tuning the intrinsic catalytic activities of oxygen-evolution catalysts by doping: a comprehensive review. *Journal of Materials Chemistry A* **9**, 20131-20163 (2021).
13. D. Coertzen, R. J. Kriek, P. B. J. Levecque, A. Falch, Vulcan carbon as support for sputtered oxygen evolution electrocatalysts. *Electrocatalysis* **10**, 604-612 %@ 1868-2529 (2019).
14. H. Sun, W. Jung, Recent advances in doped ruthenium oxides as high-efficiency electrocatalysts for the oxygen evolution reaction. *Journal of Materials Chemistry A* **9**, 15506-15521 (2021).
15. J. Bai *et al.*, RuO₂ Catalysts for Electrocatalytic Oxygen Evolution in Acidic Media: Mechanism, Activity Promotion Strategy and Research Progress. *Molecules* **29**, 537 %@ 1420-3049 (2024).
16. J. Hao *et al.*, Unraveling the electronegativity-dominated intermediate adsorption on high-entropy alloy electrocatalysts. *Nature Communications* **13**, 2662 %@ 2041-1723 (2022).
17. L. Zhao, S.-G. Park, B. Magyari-Köpe, Y. Nishi, Dopant selection rules for desired electronic structure and vacancy formation characteristics of TiO₂ resistive memory. *Applied Physics Letters* **102**, (2013).
18. H.-W. Lv *et al.*, Rhenium-boosted electrocatalytic activity and durability of pyrolytic IrO₂ for acidic oxygen evolution. *Rare Metals*, 1-7 %@ 1001-0521 (2024).

19. T. Naito, T. Shinagawa, T. Nishimoto, K. Takanabe, Recent advances in understanding oxygen evolution reaction mechanisms over iridium oxide. *Inorganic Chemistry Frontiers* **8**, 2900-2917 (2021).
20. H. Guo *et al.*, Rational design of rhodium–iridium alloy nanoparticles as highly active catalysts for acidic oxygen evolution. *ACS nano* **13**, 13225-13234 %@ 11936-10851 (2019).
21. H. Li, G. Li, Novel palladium-based nanomaterials for multifunctional ORR/OER/HER electrocatalysis. *Journal of Materials Chemistry A* **11**, 9383-9400 (2023).
22. Y.-T. Kim *et al.*, Balancing activity, stability and conductivity of nanoporous core-shell iridium/iridium oxide oxygen evolution catalysts. *Nature communications* **8**, 1449 %@ 2041-1723 (2017).
23. R. Ding *et al.*, Machine Learning-Guided Discovery of Underlying Decisive Factors and New Mechanisms for the Design of Nonprecious Metal Electrocatalysts. *ACS Catalysis* **11**, 9798-9808 (2021).
24. C. L. I. Flores, G. Gupta, M. Mamlouk, M. D. L. Balela, Enhancing the physicochemical properties of nickel cobaltite catalyst for oxygen evolution reaction in anion exchange membrane water electrolyzers. *Materials for Renewable and Sustainable Energy*, 1-12 %@ 2194-1467 (2024).
25. H. Gao *et al.*, Hydrothermal synthesized delafossite CuGaO₂ as an electrocatalyst for water oxidation. *Frontiers of Optoelectronics* **15**, 8 %@ 2095-2759 (2022).
26. T. Tachikawa, A. Beniya, K. Shigetoh, S. Higashi, Relationship between OER activity and annealing temperature of sputter-deposited flat IrO₂ thin films. *Catalysis Letters* **150**, 1976-1984 %@ 1011-1372X (2020).
27. G. H. Kim *et al.*, Effects of Annealing Temperature on the Oxygen Evolution Reaction Activity of Copper–Cobalt Oxide Nanosheets. *Nanomaterials* **11**, 657 %@ 2079-4991 (2021).
28. Y. Wang, H. Pan, Z. Liu, P. Kang, RuO₂/CoMo₂O_x Catalyst with Low Ruthenium Loading for Long-Term Acidic Oxygen Evolution. *Transactions of Tianjin University* **30**, 395-405 %@ 1006-4982 (2024).
29. X. Zhang *et al.*, Metal–Support Interaction of Carbon–Based Electrocatalysts for Oxygen Evolution Reaction. *Nanoenergy Advances* **3**, 48-72 %@ 2673-2706X (2023).
30. L. Lahn, A. M. Mingers, A. Savan, A. Ludwig, O. Kasian, Low Ti Additions to Stabilize Ru-Ir Electrocatalysts for the Oxygen Evolution Reaction. *ChemElectroChem* **11**, e202300399 %@ 202302196-202300216 (2024).
31. L. G. M. Pettersson, A. Nilsson, A molecular perspective on the d-band model: Synergy between experiment and theory. *Topics in catalysis* **57**, 2-13 %@ 1022-5528 (2014).
32. J. Hu, A. Al-Salihi, B. Zhang, S. Li, P. Xu, Mastering the D-band center of iron-series metal-based electrocatalysts for enhanced electrocatalytic water splitting. *International Journal of Molecular Sciences* **23**, 15405 %@ 11422-10067 (2022).
33. J. Horáček, Fischer–Tropsch synthesis, the effect of promoters, catalyst support, and reaction conditions selection. *Monatshefte für Chemie-Chemical Monthly* **151**, 649-675 %@ 0026-9247 (2020).
34. Y. Tang, W. Cheng, Key parameters governing metallic nanoparticle electrocatalysis. *Nanoscale* **7**, 16151-16164 (2015).
35. X. Yan, L. Zhuang, Z. Zhu, X. Yao, Defect engineering and characterization of active sites for efficient electrocatalysis. *Nanoscale* **13**, 3327-3345 (2021).
36. X. Teng *et al.*, RuO₂@ IrO₂/C Core-Shell Structure Catalyst for Efficient and Durable Acidic Oxygen Evolution. *Catalysts* **13**, 1456 %@ 2073-4344 (2023).
37. Z. J. Xu, T. D. Nguyen, G. G. Scherer, A Facile Synthesis of Size-Controllable IrO₂ and RuO₂ Nanoparticles for the Oxygen Evolution Reaction. (2016).
38. G. C. da Silva, N. Perini, E. A. Ticianelli, Effect of temperature on the activities and stabilities of hydrothermally prepared IrO_x nanocatalyst layers for the oxygen

- evolution reaction. *Applied Catalysis B: Environmental* **218**, 287-297 %@ 0926-3373 (2017).
39. Y. Y. Feng *et al.*, Copper-doped ruthenium oxide as highly efficient electrocatalysts for the evolution of oxygen in acidic media. *Journal of Alloys and Compounds* **892**, 162113 (2022).
 40. R. Huang, Y. Z. Wen, H. S. Peng, B. Zhang, Improved kinetics of OER on Ru-Pb binary electrocatalyst by decoupling proton-electron transfer. *Chinese Journal of Catalysis* **43**, 130-138 (2022).
 41. G. H. Moon, Y. Wang, S. Kim, E. Budiyanto, H. Tuysuz, Preparation of practical high-performance electrodes for acidic and alkaline media water electrolysis. *Chemsuschem* **15**, e202102114 (2022).
 42. J. He, W. Q. Li, P. Xu, J. M. Sun, Tuning electron correlations of RuO₂ by co-doping of Mo and Ce for boosting electrocatalytic water oxidation in acidic media. *Applied Catalysis B: Environmental* **298**, 120528 (2021).
 43. T. Kwon *et al.*, Au-Ru alloy nanofibers as a highly stable and active bifunctional electrocatalyst for acidic water splitting. *Applied Surface Science* **563**, 150293 (2021).
 44. Z. P. Shi *et al.*, Confined Ir single sites with triggered lattice oxygen redox: toward boosted and sustained water oxidation catalysis. *Joule* **5**, 2164-2176 (2021).
 45. W. Jin *et al.*, Evolution of interfacial coupling interaction of Ni-Ru species for pH-universal water splitting. *Chemical Engineering Journal* **426**, 130762 (2021).
 46. S. Y. Hao *et al.*, Torsion strained iridium oxide for efficient acidic water oxidation in proton exchange membrane electrolyzers. *Nature Nanotechnology* **16**, 1371-1376 (2021).
 47. Y. B. Wang *et al.*, Modulating crystallinity and surface electronic structure of IrO₂ via gadolinium doping to promote acidic oxygen evolution. *ACS Sustainable Chemistry & Engineering* **9**, 10710-10716 (2021).
 48. Y. Ping, G. Galli, W. A. Goddard III, Electronic structure of IrO₂: the role of the metal d orbitals. *The Journal of Physical Chemistry C* **119**, 11570-11577 %@ 11932-17447 (2015).
 49. B. An *et al.*, Strontium doping RuO₂ electrocatalyst with abundant oxygen vacancies for boosting OER performance. *Inorganic Chemistry Frontiers*, (2024).
 50. X. Wu *et al.*, Synergistically optimizing the electrocatalytic performance of IrO₂ with double doping and bi-directional strains for acidic oxygen evolution reaction. *Catalysis Science & Technology* **14**, 4599-4607 (2024).
 51. B. Wu *et al.*, Effects of annealing temperature of PtCu/MWCNT catalysts on their electrocatalytic performance of electrooxidation of methanol. *Ionics*, 1-14 %@ 0947-7047 (2022).
 52. A. Falch, V. A. Lates, H. S. Kotzé, R. J. Kriek, The Effect of Rapid Thermal Annealing on Sputtered Pt and Pt 3 Pd 2 Thin Film Electrocatalysts for Aqueous SO₂ Electro-Oxidation. *Electrocatalysis* **7**, 33-41 %@ 1868-2529 (2016).
 53. Y. Wen *et al.*, Stabilizing highly active Ru sites by suppressing lattice oxygen participation in acidic water oxidation. *Journal of the American Chemical Society* **143**, 6482-6490 %@ 0002-7863 (2021).
 54. Z. Shi *et al.*, Customized reaction route for ruthenium oxide towards stabilized water oxidation in high-performance PEM electrolyzers. *Nature Communications* **14**, 843 %@ 2041-1723 (2023).
 55. Z.-Y. Wu *et al.*, Non-iridium-based electrocatalyst for durable acidic oxygen evolution reaction in proton exchange membrane water electrolysis. *Nature Materials* **22**, 100-108 %@ 1476-1122 (2023).
 56. S. Wang *et al.*, Recent advances in noble metal-based catalysts for CO oxidation. *RSC advances* **14**, 30566-30581 (2024).
 57. Z. Wu *et al.*, Effect and mechanism of rare Earth and alkaline Earth metals on the high-temperature stability of activated alumina. *Frontiers in Materials* **10**, 1103590 %@ 1102296-1108016 (2023).

58. Y. Liu *et al.*, Recent advances in the heteroatom doping of perovskite oxides for efficient electrocatalytic reactions. *Nanoscale* **13**, 19840-19856 (2021).
59. H. Sun, X. Xu, G. Chen, Z. Shao, Perovskite oxides as electrocatalysts for water electrolysis: From crystalline to amorphous. *Carbon Energy*, e595 %@ 2637-9368 (2024).
60. J. Sengupta, R. K. Sahoo, C. D. Mukherjee, Effect of annealing on the structural, topographical and optical properties of sol–gel derived ZnO and AZO thin films. *Materials Letters* **83**, 84-87 %@ 0167-0577X (2012).
61. D. Yadav, R. Gahlawat, R. Shukla, A comprehensive analysis of the impact of annealing temperature variation on the structural, optical, morphological, magnetic, and photocatalytic properties of CoFe₂O₄ nanoparticles. *Ionics* **30**, 6559-6574 %@ 0947-7047 (2024).
62. S. B. Attanayake *et al.*, Competing magnetic interactions and field-induced metamagnetic transition in highly crystalline phase-tunable iron oxide nanorods. *Nanomaterials* **13**, 1340 %@ 2079-4991 (2023).
63. S. Gong, T. Zhang, J. Meng, W. Sun, Y. Tian, Advances in the mechanism investigation for the oxygen evolution reaction: fundamental theory and monitoring techniques. *Materials Chemistry Frontiers*, (2024).
64. M. Risch, Upgrading the detection of electrocatalyst degradation during the oxygen evolution reaction. *Current Opinion in Electrochemistry* **38**, 101247 %@ 102451-109103 (2023).
65. G. Li, L. Anderson, Y. Chen, M. Pan, P.-Y. A. Chuang, New insights into evaluating catalyst activity and stability for oxygen evolution reactions in alkaline media. *Sustainable Energy & Fuels* **2**, 237-251 (2018).
66. A. Higareda, D. L. Hernández-Arellano, L. C. Ordoñez, R. Barbosa, N. Alonso-Vante, Advanced electrocatalysts for the oxygen evolution reaction: From single-to multielement materials. *Catalysts* **13**, 1346 %@ 2073-4344 (2023).
67. L. He, Y. Zhou, M. Wang, S. Li, Y. Lai, Recent Progress on Stability of Layered Double Hydroxide-Based Catalysts for Oxygen Evolution Reaction. *Nanomaterials* **14**, 1533 %@ 2079-4991 (2024).
68. Z. Li *et al.*, Seed-assisted formation of NiFe anode catalysts for anion exchange membrane water electrolysis at industrial-scale current density. *Nature Catalysis*, 1-9 %@ 2520-1158 (2024).