

Green Chemistry

Cutting-edge research for a greener sustainable future

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: B. Farris, J. J. Meckstroth and K. C. Leonard, *Green Chem.*, 2026, DOI: 10.1039/D6GC01753C.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Green Foundation

View Article Online
DOI: 10.1039/D6GC01753C

1. This work advances green chemistry by establishing a sustainable machine-learning framework that relies exclusively on existing or pre-trained models for data collection, thereby minimizing computational energy use while predicting key reaction parameters relevant to green catalysis. Our approach also uncovers mechanistic insights that enable the rational design of CO₂-reduction catalysts with improved selectivity and greater energy efficiency.
2. The shallow-learning random forest models developed here can be deployed entirely on local hardware, eliminating dependence on resource-intensive data centers and demonstrating both computational and environmental sustainability. Beyond methodological benefits, the study provides new understanding of pathways leading to high-value CO₂-reduction products.
3. Future work aimed at quantifying the energy savings from combining pre-trained models with shallow-learning techniques could further enhance the overall greenness of this approach.



Cite this: DOI: 00.0000/xxxxxxxxxx

Interpretable, Low-Compute Machine Learning Integrating Experimental and Catalytic Descriptors for Sustainable CO₂ Electroreduction

Brianna R. Farris,^{†a,b} Joshua J. Meckstroth,^{†a,b} and Kevin C. Leonard^{*a,b}

Received Date

Accepted Date

DOI: 00.0000/xxxxxxxxxx

Applying machine learning (ML) sustainably to green chemistry is challenging because reaction complexity often drives the use of large, energy-intensive models. Here, we combine pre-trained models for information extraction with low-compute, interpretable shallow-learning models to deliver mechanistic insight while minimizing computational cost. Using the electrocatalytic CO₂ reduction reaction (CO₂RR) as a model green chemistry reaction, we automatically extracted 3,880 experimentally reported reaction conditions from peer-reviewed literature with a pre-trained large language model and augmented these data with relaxation energies of key (CO₂RR) intermediates obtained via community-sourced density functional theory (DFT) and ML surrogates for DFT. Training 98 random-forest binary classifiers across diverse feature sets, we find that models integrating both experimental and computational descriptors consistently achieve the best performance. Because these models can be run locally—without data-center resources—they offer a computationally and environmentally sustainable route to discovery. Furthermore, interpretable ML analysis revealed mechanistic trends, such as CH₃OH formation needing catalysts with weak adsorptions of O* and H₂O* for selective production, while C₂H₄ production required catalysts that combine moderate adsorption of CO* with moderate to strong adsorption of O* and H₂O*. The model also identified that similar catalytic properties produce C₂H₄ and CH₄, but the applied voltage is the major driving force with more negative voltages favoring C₂H₄ production. These findings underscore the value of integrating experimental and theoretical insights into ML frameworks and demonstrate how pre-trained and interpretable ML can uncover fundamental principles governing catalytic selectivity for sustainable production of fuels and chemicals.

Green Foundation

1. This work advances green chemistry by establishing a sustainable machine-learning framework that relies exclusively on existing or pre-trained models for data collection, thereby minimizing computational energy use while predicting key reaction parameters relevant to green catalysis. Our approach also uncovers mechanistic insights that enable the rational design of CO₂-reduction catalysts with improved selectivity and greater energy efficiency.
2. The shallow-learning random forest models developed here can be deployed entirely on local hardware, eliminating dependence on resource-intensive data centers and demonstrating both computational and environmental sustainability. Beyond methodological benefits, the study provides new understanding of pathways leading to high-value CO₂-reduction products.
3. Future work aimed at quantifying the energy savings from combining pre-trained models with shallow-learning techniques could further enhance the overall greenness of this approach.

^a Department of Chemical & Petroleum Engineering, 4132 Learned Hall 1530 W 15th St, The University of Kansas, Lawrence, KS USA.

^b Center for Environmentally Beneficial Catalysis, 1501 Wakarusa Dr. LSRL Building A, Suite 110, The University of Kansas, Lawrence, KS USA.

[†] These authors contributed equally.

* E-mail: kcleonard@ku.edu



1 Introduction

Machine learning (ML) has accelerated discovery across biology, chemistry, and physics, enabling rapid hypothesis generation, property prediction, and mechanism elucidation.^{1–19} Yet green-chemistry applications are unusually complex, often motivating large ML models with growing energy demands.^{20–23} For example, in heterogeneous electrocatalysis, performance is governed not only by atomic-scale catalyst properties but also by the interfacial microenvironment and operating conditions, and capturing these coupled effects in a predictive, energy-efficient, and interpretable framework remains challenging. Meeting this challenge is essential for ML-guided catalyst design that can steer experiments toward improved selectivity and efficiency. Here, we demonstrate a lightweight ML workflow that (i) curates reaction data from the literature using pre-trained, extractive large language models (LLMs), (ii) augments these data with key catalytic descriptors obtained from low-compute sources, (iii) trains shallow, interpretable random-forest models, and (iv) interprets the resulting structure–condition–selectivity relationships for green electrocatalyst design. This work builds on our prior study focused solely on literature-derived reaction conditions,^{4,24} and extends it by integrating catalytic descriptors sourced from community databases (Catalysis-Hub) and from pre-trained surrogates (Open Catalyst Project), thereby saving on the order of ten-thousand CPU-core hours of density functional theory (DFT) calculations while improving model performance and mechanistic interpretability.

As the model system, we investigated the electrocatalytic carbon-dioxide reduction reaction (CO₂RR), which is widely studied for converting CO₂ into value-added fuels and chemicals, yet the key parameters controlling selectivity remain under active debate.^{25–31} The (CO₂RR) is often cited as a renewable pathway for valorizing carbon emissions from industrial point sources (e.g., power and ethanol plants),^{32–34} but practical deployment is hindered by kinetic and selectivity challenges arising from the thermodynamic stability of CO₂ and the complexity of the electrochemical interface.^{35,36} Using predictive and interpretable shallow ML, we show that combining experimental reaction conditions with catalytic descriptors yields superior predictive performance and richer mechanistic insight compared with using either data type alone. Moreover, catalytic descriptors predicted by pre-trained ML surrogates from the Open Catalyst Project provide performance comparable to DFT-derived descriptors from Catalysis-Hub,³⁷ suggesting that low-compute surrogates can effectively support interpretable ML analyses of catalytic reactions. Collectively, these results illustrate how pre-trained and interpretable ML can reveal trends governing electrocatalytic selectivity while mitigating computational energy consumption.

2 Methods

2.1 Automated Dataset Creation Methods

Data for the CO₂RR to train the machine learning models was obtained from three separate sources – the archival literature, the Open Catalyst Project, and Catalysis Hub. This dataset was subdivided into four Feature Sets, which we label I, II, III, and IV.

Automatic literature extraction from our previous work²⁴ yielded experimental conditions and catalyst compositions. Those experimental conditions became Feature Set I after a post-processing treatment. Elemental catalyst compositions were set aside as Feature Set II, but two copies were made which would become Sets III and IV. Set III was formed by converting catalyst compositions into Atom Simulation Environment (ASE) bulk structures using the Materials Project database, then simulating bulk-adsorbate relaxations using the Open Catalyst Project EquiformerV2 model. Those adsorbates were constructed as ASE Atoms objects using 3D structure files retrieved from PubChem. Set IV was comprised of adsorption energies (for select adsorbates) retrieved from the Catalysis Hub database, which was queried using the same aforementioned catalyst compositions. In this way, as shown in Fig. 1, each individual experiment from the literature extraction was transformed into a row of data with attributes (the features) from I, II, III, and IV as well as the reaction product associated with the experiment (the label). The entire transformation process was performed in Python using the following notable libraries: Pandas, Numpy, Matplotlib, ASE, fairchem, and MP API.

2.2 Interpretable Machine Learning Methods

Relaxation energies from the Open Catalyst Project implementation were plotted against adsorption energies gathered from Catalysis Hub^{38–44} to establish a benchmark using CO gas on (111)-faceted metallic and bimetallic surfaces (Supporting Information Figure S1). Expected discrepancies of varying magnitudes were observed between OCP energies and Catalysis Hub energies. As detailed in the OC20 dataset methodology, adsorption energies are calculated using a gas-phase reference (E_{gas}) derived from a linear combination of N₂, H₂O, CO, and H₂, which differs from the absolute single-molecule DFT energy references frequently employed in the datasets hosted on Catalysis-Hub. Other caveats include the OCP framework utilizing non-spin-polarized calculations to maintain large-scale throughput, excluding entropic contributions to stability (relaxation energies are enthalpies), and approximating surfaces as being very small and defect-free. The integration of both Open Catalyst Project (OCP) relaxations and Catalysis Hub DFT data was not intended to achieve numerical parity between the two frameworks. Rather, these distinct computational sources were utilized to construct a denser and more diverse feature set. By incorporating descriptors from both ML-surrogate relaxations and traditional DFT, we aimed to capture complementary representations of the CO₂RR landscape, acknowledging that each method may reflect nuanced, model-specific trends in reaction mechanisms.

Random forest classifiers, each composed of 100 estimators (trees), were implemented in Python using the Sci-Kit Learn package. A hyperparameters search revealed that the cross-entropy loss function should be used and that $\log_2(n)$ total features should be visible to each tree. Each time a Feature Set combination was used for classification, it was stratified and partitioned into 80% training, 10% testing, and 10% validation sets. Thirty forests, each with a different number of layers (depths) from 1 to 30, were trained on the training set then evaluated on the validation set.



Training and validation scores were plotted versus tree depth in order to select the best-fitting depth where the model was neither overfit nor underfit. Models over 10 layers deep were not commonly selected due to overfitting. Finally, models of that chosen depth were trained on the combined training and validation sets and scored on the testing set in a 10-fold cross validation. Binary classification of reaction products was done using random forests trained on different combinations of Feature Sets I, II, III, and/or IV as input features. User-defined parameters were written into the Python code which allowed control over whether the “structure” and “electrolyte” features of Set I should be label encoded or one-hot encoded using built-in functions from Scikit-learn. Label encoding of structure and electrolyte (the scheme for which can be found in the supporting information) was enabled when evaluating a model’s accuracy, and one-hot encoding was enabled when generating SHAP and feature importance charts. Random forests were trained and cross-validated using built-in functions from Scikit-learn. Additionally, confusion matrices for random forests were produced using the library seaborn. All source code and results are provided in Supporting Information.

3 Results and Discussion

3.1 Automated Dataset Creation

Three overarching tasks were performed to utilize machine-learning models to uncover novel insights into the CO₂RR: (1) Automated Dataset Generation (2) Machine Learning on Various Feature Set Permutations and (3) Interpretable Machine Learning with Binary Classification. For the automated dataset generation, four different Feature Sets were created as outlined in the flowchart Figure 1. As shown in the green box in Figure 1, we created dataset of CO₂RR experimental conditions paired with the structure and elemental composition of the electrocatalysts used for the reaction, and the resulting CO₂RR product formed (e.g., CO, HCOOH, C₂H₄) as described in previous work from our group.²⁴ This dataset was automatically compiled using an extractive large language model which processed thousands of manuscripts from the archival literature into a single data table. After postprocessing from the LLM extraction, this dataset consisted of 3,880 individual data points. This dataset was then split into two Feature Sets. Feature Set I contains only the experimental conditions including: voltage, use of a gas diffusion electrode (GDE), electrolyte concentration, electrolyte pH, electrolyte species, and catalyst structure. Feature Set II contains only the elemental composition of the catalysts used in the re-

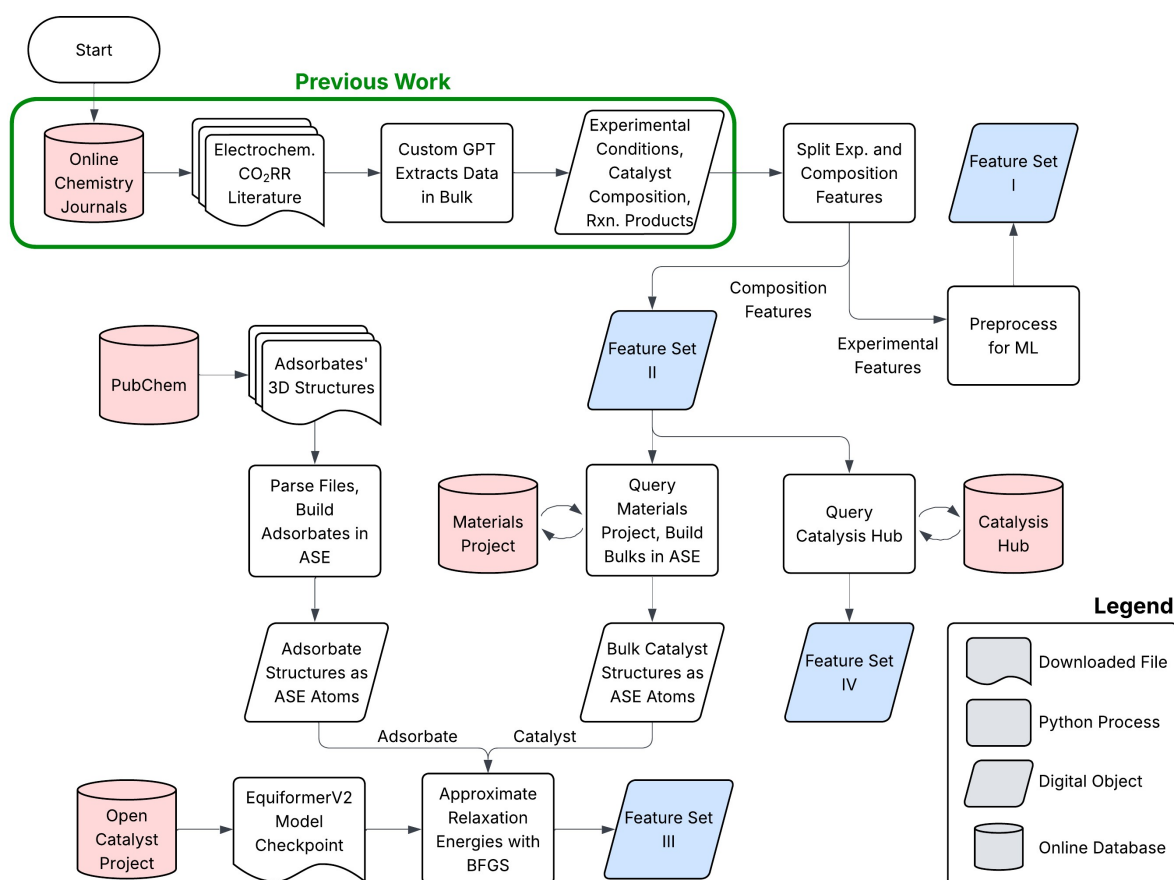
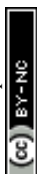


Fig. 1 Methodology for constructing the four main Feature Sets: arrows represent transfer of information, and Feature Sets are marked in blue.



action. This Feature Set is a one-hot encoded dataset of 40 catalyst elements each one denoted either the presence or absence of that element in the catalyst's composition. Two additional Feature Sets were created to incorporate surface adsorption energies of the key CO₂RR intermediates for each catalyst obtained from the automated literature extracted dataset. Feature Set III contained machine-learning based predictions of bulk-adsorbate surface adsorption energies (called relaxation energies) sourced from the Open Catalyst Project.^{45,46} Feature Set III was created using a customized automated framework to obtain relaxation energies for eight different adsorbates for each of the 3,880 rows in the original dataset. Following the left-hand side of Figure 1, the catalyst composition from Feature Set II was used to obtain the bulk structures for the catalysts from The Materials Project Materials Explorer API. These structures were then converted into ASE Atoms surfaces with a 111, 211, or 110 Miller Index. In this study, we used a subset of surface facets (111, 211, and 110) to represent a wide range of catalytic surfaces. We chose this subset because performing OCP on a large range of high-index facets would be computationally expensive. By choosing this facet subset, we capture primary morphologies of catalytic sites: close-packed terraces (111), step/edge sites (211), and more open surfaces (110). Since we do not observe a decrease in performance on our testing sets, we do not anticipate that the model is overfitting to specific non-physical correlations due to the generalization of catalytic surfaces. Moreover, since different facets of the same material are often highly correlated via scaling relations, the model is forced to find the most generalizable descriptors holding true across various experimental conditions. Concurrently, the key intermediates for the CO₂RR (CO*, CH*, C*, O*, H₂O*, and COOH*) were manually downloaded from PubChem and then converted into ASE Atoms. Using the pretrained EquiformerV2 OC20 model checkpoint from the Open Catalyst Project database, we obtained 1,472 bulk-adsorbate ASE pairs (184 unique bulks, 8 unique adsorbate/miller index configurations) and obtained the relaxation energies using the trained model and BFGS optimization. This processes allowed us to obtain Feature Set III, the OCP-predicted relaxation energy for each intermediate on each catalysts in our CO₂RR database.

The fourth and final Feature Set (IV) was obtained directly from the Catalysis Hub – an open source database comprised of published binding energies derived from DFT calculations across thousands of articles. To automatically scrape this dataset, custom python code looped over the Catalysis Hub OpenQL API, querying for any surface binding energies associated with the 198 unique bulks matched with 11 unique adsorbates. In this way, thousands of computational values could be retrieved in minutes on a personal device with minimal power draw. If a search yielded multiple energies, the lowest one was selected, and if the search yielded no energies, then the energy was set to zero. All available bulk miller indices were included in the search so that the lowest reported energy could be retrieved each time. Performing this zero-computation process gave the fourth and final Feature Set for the 3,880 reactions extracted from the literature.

3.2 Reduced Computational Cost

Our method for using machine learning to perform the relaxations and binary classifications is exceptionally lightweight. On a personal desktop (utilizing an Intel i5-14400F CPU and NVIDIA GeForce RTX 4060 GPU), the 1,472 EquiformerV2 relaxations required 51 hours (wall-clock) to complete. Compare this to using direct DFT calculation, which could require days to compute a single relaxation on faster computers. Using a pre-trained ML model significantly reduced the required computational time and cost, hence why we consider ML as a more sustainable method for acquiring DFT calculations. Furthermore, hours of compute time were saved by our employment of interpretable shallow learning models as opposed to complex neural networks like tuned multi-layer perceptron networks without sacrificing valuable insights.

It is a known concern that current DFT-based workflows calculating adsorption energies from the ground-up for each experiment demand high energy usage. Our framework employs pre-trained models and public datasets to save CPU-core hours on the order of 10⁴ compared to conventional DFT approaches. However, it is important to contextualize these results within the life-cycle energy footprint of machine learning workflows. While the GPT models used for LLM extraction and the EquiformerV2 OCP model used in this study required significant computational resources to train initially, these training costs have already been paid and are shared across innumerable downstream applications and users. Thus, the burden of energy expenditure for creating those resources becomes distributed across a continuously expanding pool of end users. From this perspective, the present framework dramatically reduces the additional computational burden of catalytic screening. However, a quantitative life-cycle assessment of machine learning approaches in computational chemistry remains an important area for future work, especially when considering their role in advancing sustainability goals in green chemistry.

3.3 Machine Learning on Feature Set Permutations

To show the interplay between the experimental parameters, ML-predicted relaxation energies, and calculated DFT adsorption energies affect the prediction and understanding of the CO₂RR, the machine learning models were performed on 14 permutations of these features sets as described in Table 1. As described in the Methods section, Random Forest Classifiers, each composed of 100 estimators (trees), were trained and tuned with the training and validation sets, and the prediction was compared to the ground truth on the testing set. The CO₂RR has many products with varying selectivity maximum insights from machine learning predictions can be obtained when organizing them into binary classification tasks. The six product pairs we investigated were (1) CO vs HCOOH, (2) C₂H₄ vs CH₃OH, (3) C₂H₄ vs CO, (4) C₁ vs C₂₊, (5) C₂H₄ vs CH₄, and (6) CH₃OH vs CH₄. A seventh investigation of CH₄ vs CO was performed but excluded due to highly imbalanced classes; However, these results from this can be found in the Supporting Information. In total, 98 random forest models were trained and tested to cover the 6 binary classifications and 14 Feature Set combinations using 10-fold cross



Table 1 Feature Set Combinations Overview: descriptions of Feature Sets and grouping logic.

Feature Set Combination	Description	Investigation
I	Experimental conditions (reaction parameters)	Effectiveness of individual Feature Sets
II	Catalyst composition on elemental basis	
III	Relaxation energies approximated with Open Catalyst Project (OCP)	
IV	DFT adsorption energies sourced from Catalysis Hub	
I + II	Experimental conditions and catalyst composition	Effectiveness of supplementing the experimental conditions (Set I)
I + III	Experimental conditions and OCP energies	
I + IV	Experimental conditions and Catalysis Hub energies	
I + II + III	Experimental conditions, catalyst compositions, and OCP energies	Effectiveness of supplementing the previous work (Sets I + II)
I + II + IV	Experimental conditions, catalyst compositions, and Catalysis Hub energies	
I + II + III + IV	All Feature Sets	
II + III	Catalyst compositions and OCP energies	Effectiveness of using catalyst-based information only
II + IV	Catalyst compositions and Catalysis Hub energies	
II + III + IV	Catalyst compositions and all adsorption energies	
III + IV	Open Catalyst Project and Catalysis Hub adsorption energies	Effectiveness of using relaxation energies only

validation. Random forest models were selected over other shallow learning models because they have a low number of tunable hyperparameters and they are usually more efficient out-of-the-box. Furthermore, random forests randomly select a subset of features at every split, forcing the model to have a broader scope over possible relevant chemical descriptors. To demonstrate this, a small-scale comparison between random forest classifiers and gradient boost classifiers was performed using CO and HCOOH labels (see Supporting Information Figure S2). XGBoost demonstrated only marginally higher cross-validation accuracies and F1-scores for some of the trials. However, XGBoost is less interpretable than Random Forest and is more computationally expensive to tune.^{47–49} Significant trends emerge when comparing the average accuracies and F1-scores for majority and minority classes across all six binary classification tasks and fourteen Feature Set combinations (Table 2). First, the highest prediction accuracies and F1-scores occur when the Feature Set includes both experimental conditions and at least one catalytic descriptor (Table 2). This underscores that machine learning models require a combination of actual reaction conditions and catalyst-specific information to achieve the best predictive performance.

Conversely, catalyst descriptors alone are insufficient for reliable predictions, specifically for predicting the minority class (Table 2). Second, incorporating adsorption energies of key intermediates provides greater predictive power than using only elemental composition of the catalysts. This finding highlights the importance of capturing interaction dynamics between intermediates and bimetallic surfaces, which offer valuable mechanistic insights for model training. Finally, no significant difference was observed between using adsorption energies derived from DFT calculations (Catalysis Hub) and those predicted by machine learning models (Open Catalyst Project). This result demonstrates the viability of ML-predicted adsorption energies as interpretable features, enabling data-driven models to deliver actionable insights into elec-

Table 2 Table of the average accuracy and average F1-score of the minority class across all six binary classification questions.

Feature Set Combination	Average Minority Class F1	Average Majority Class F1	Average Cross Val. Score
I	31.4	83.3	76.2
II	27.9	85.8	77.7
III	40.5	84.8	77.2
IV	47.7	85.4	77.0
I + II	49.7	86.9	80.8
I + III	56.4	86.7	81.2
I + IV	55.9	86.1	80.5
I + II + III	56.8	86.9	81.9
I + II + IV	54.7	86.1	80.4
I + II + III + IV	54.0	86.5	80.5
II + III	34.1	85.3	77.6
II + IV	31.9	84.8	77.6
II + III + IV	34.1	85.5	77.9
III + IV	47.3	84.7	77.9

trocatalytic performance metrics.

3.4 Interpretable Machine Learning

To fully leverage the interpretability of our machine-learning models, we conducted a direct analysis of each binary classification task to identify the most influential features governing the selectivity of the CO₂RR reaction.

3.4.1 CO versus HCOOH Classification

The CO-HCOOH data subset was constructed from the full dataset, comprising the 583 reactions yielding HCOOH and the 1,427 reactions yielding CO. Random forest models were trained on this subset using each Feature Set permutation outlined in Table 1. Figure 2a or Figures S3 and S4 in the Supporting Information show the cross-validation accuracies (colors) and F1-scores



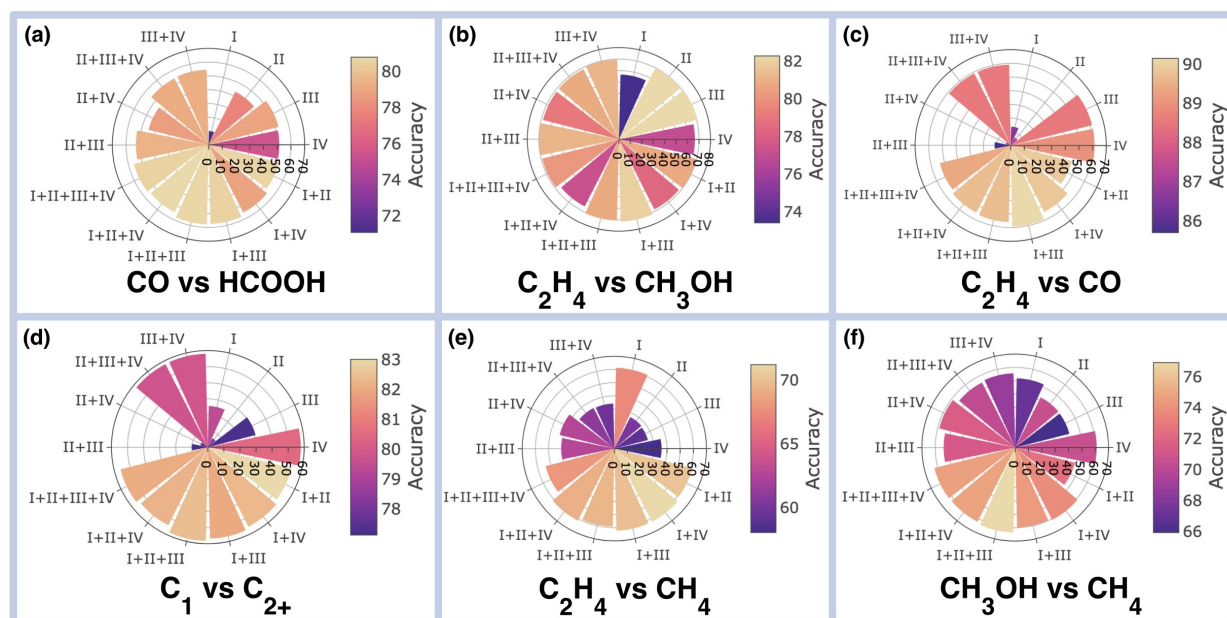


Fig. 2 Random Forest Classification Performance Plots: Cross validation accuracies (color gradient) and minority class F1-scores (bar lengths) for the six binary classification questions – CO vs HCOOH (a), C₂H₄ vs CH₃OH (b), C₂H₄ vs CO (c), C₁ vs C₂₊ (d), C₂H₄ vs CH₄ (e), and CH₃OH vs CH₄ (f); longer, brighter bars indicate better forest performance for a given Feature Set and binary question.

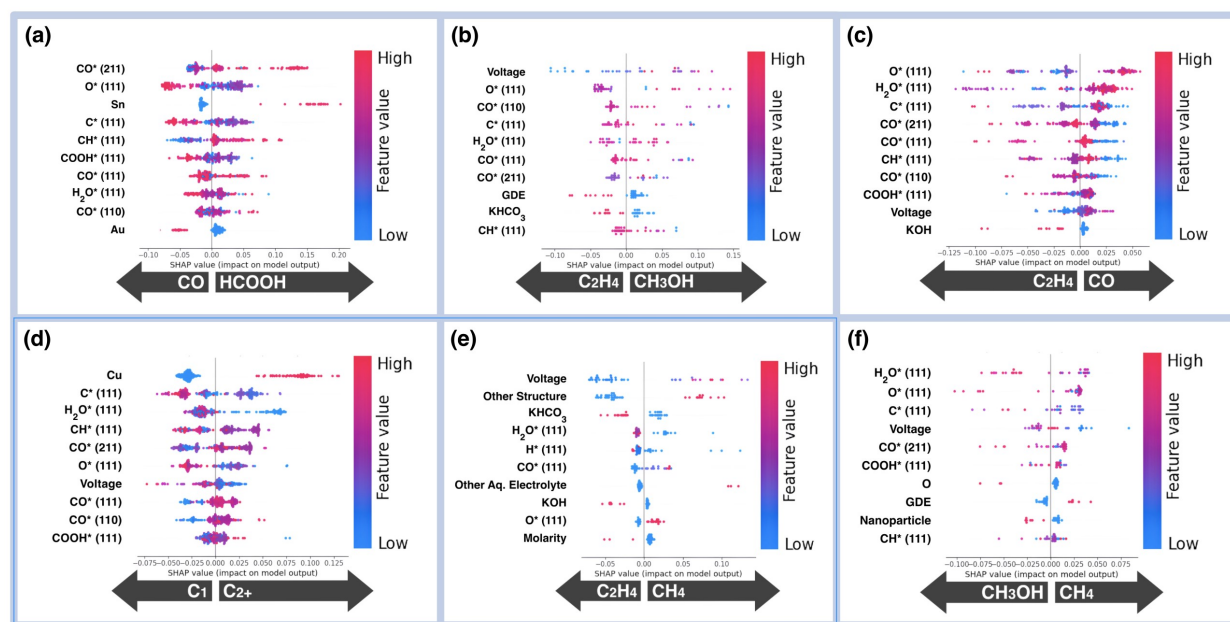


Fig. 3 SHAP Beeswarm Plots: SHAP charts for the six binary classification questions – CO vs HCOOH (a), C₂H₄ vs CH₃OH (b), C₂H₄ vs CO (c), C₁ vs C₂₊ (d), C₂H₄ vs CH₄ (e), and CH₃OH vs CH₄ (f); each chart originates from the random forest trained on the feature set giving the highest performance quantified by F1-score and cross-validation accuracy.

for each Feature Set permutation. Models trained exclusively on experimental conditions (Feature Set I) exhibited the lowest accuracy and F1-scores. Model performance improved substantially with the catalyst descriptor Feature Sets (Feature Sets II, III, and IV). Furthermore, Figure 2a demonstrates that the highest accuracies were achieved when experimental conditions were combined with at least one catalytic descriptor Feature Set (Feature Set I + Feature Set II, III, or IV). This result shows that the interpretable

ML-model can independently determine the well-known fact that the catalyst plays a very important role with determining the selectivity of between CO and HCOOH, however the experimental conditions can influence the overall reaction outcome. The Feature Set combination that provided the best performance was I + II + III, with an accuracy of 80.7% and a HCOOH F1-score of 57.7%. Interpretable SHAP analysis was then performed on this Feature Set with the results shown in Figure 3a. The SHAP anal-



ysis showed that Open Catalyst relaxation energies were ranked highly for classifying between CO and HCOOH. The interpretable machine learning analysis showed that relaxation energy for CO* on the 211 Miller index was the most important feature were low relaxation energy (high ΔG° favored HCOOH and high relaxation energy (low ΔG° favored CO as expected. Again, this demonstrated that this ML-based approach can independently determine that catalysts which strongly adsorb key intermediates typically favor CO production, whereas those with weaker adsorption favor HCOOH formation. Interestingly, the relaxation energy of the CO* on the 211 Miller index gave the best correlation over the other Miller Indices, indicating the importance of high index edge sites catalysis. In addition, the SHAP analysis also showed that stronger O* relaxation energies predict towards formic acid, which may indicate that CO₂ to HCOOH mechanisms involving multiple surface–oxygen bonds, which is also supported in the literature.⁵⁰

3.4.2 C₂H₄ versus CH₃OH Classification

The C₂H₄-CH₃OH data set was comprised of 218 reactions that produced C₂H₄ and 125 reactions that produced CH₃OH from the full dataset. This data subset was then again used to train random forest models on the Feature Set permutations in Table 1. The cross validation accuracies (colors) and F1-scores (bar heights) from each permutation can be found in Figure 2b or Figures S3 and S4 in the Supporting Information. Similar to the CO-HCOOH data subset the model trained only on experimental conditions (Feature Set I) had the lowest cross validation accuracy and F1-scores. Model performance increases with the introduction of catalytic descriptors (Feature Sets II, II and IV), reinforcing the importance of electrocatalytic interactions in selective production of C₂H₄ and CH₃OH. The Feature Set that had the overall best performance with a high F1-score of the minority class (CH₃OH) was Feature Set I + III, with a cross validation score of 82.0% and an CH₃OH F1-score of 71.6%. SHAP analysis was performed on this Feature Set and is shown in Figure 3b. The SHAP analysis showed that the most important feature for classifying between C₂H₄ and CH₃OH was voltage, with more negative voltages leaning producing C₂H₄. This is interesting as the electrocatalytic production of C₂H₄ requires 12 electron transfers while CH₃OH only requires 6 electron transfers. It also demonstrated that similar catalysts produce C₂H₄ and CH₃OH, but it is the experimental condition of applied voltage that drives the selectivity. After voltage, the next most important features are many Open Catalyst relaxation energies. Interestingly, O* on the 111 Miller index was ranked highly with weakly bonding O* favoring methanol and intermediate to high bonding favoring C₂H₄. This would suggest that CO₂ to CH₃OH pathways do not go through O* adsorption to the catalysts surface and thus favoring the O* remaining on the product to produce CH₃OH over C₂H₄. In addition, the SHAP analysis also showed that more intermediate adsorption of CO* on the 110, 111 and 211 planes favor the production of C₂H₄, corresponding to the known volcano-type relationship between CO* and C₂H₄ production. In addition, using KHCO₃ and GDEs favor the production of C₂H₄ over CH₃OH. Thus, to design a system that produces CH₃OH, one would use a catalysts with weak O*

adsorption, and apply less negative potential. To design a system that produces C₂H₄, one would want a catalysts with moderate to high adsorption of O*, intermediate adsorption of CO*, and apply a highly negative potential in KHCO₃ electrolytes.

3.4.3 C₂H₄ versus CO Classification

The C₂H₄-CO data subset was comprised of 1,427 reactions yielding CO and 218 reactions yielding C₂H₄. Random forest models were again trained on the Feature Set permutations in Table 1. The cross validation accuracies and F1-scores for each feature set permutation are shown in Figure 2c and Figures S3 and S4 in the Supporting Information. Models trained exclusively on Feature Sets I or Feature Set II had dramatically lower minority F1-scores (C₂H₄ F1-score). This finding highlights the need for adsorption energy data combined with experimental data or catalyst composition in the classification of C₂H₄ and CO. When Feature Set I was combined with the catalyst descriptor Feature Sets (Feature Sets II, III, and IV), the model performance improved. Again reaffirming that even when adsorption energies dictate reaction pathways, experimental conditions maintain an influence on the overall outcome.

The Feature Set with the best performance was I + III, with a cross validation accuracy of 90.2% and a minority F1-score of 59.3%. Consistent with previous results the SHAP analysis ranked the Open Catalysts relaxation energies highly. This interpretable machine learning analysis showed that to favor CO formation over C₂H₄, one would require a catalysts that has low O* adsorption, low H₂O* adsorption, strong CO* adsorption, and less negative potentials. To favor C₂H₄, one would require moderate to strong O* and H₂O* adsorption, moderate CO₂ adsorption, and very negative applied potentials. This is potentially due to how the formation of C₂H₄ can involve multiple surface-oxygen bonds,⁵¹ yielding O* adsorption as an important feature for designing C₂H₄ producing catalysts. Interestingly, the significance of the Other Structure feature in promoting methanol production suggests that designer structures (e.g., single-atom catalysts) may be particularly beneficial for converting CO₂ to methanol.

3.4.4 C₁ versus C₂₊ Classification

The next data subset is unique as it is the only data subset tested that is a combination of products. It was comprised of 1,956 single carbon (C₁) products and 563 multicarbon (C₂₊) products. Figure 2d contains the cross validation scores and minority class (C₂₊) F1-scores from the random forest models. When analyzing the combined Feature Sets it is seen that performance improves when Feature Set 1 is combined with any combination of the other Feature sets. This further highlights that reliable predictions hinge on uniting experimental conditions with electrocatalytic information. The best performing Feature Set combination was I + II + III with a cross validation accuracy of 82.4% and a minority class (C₂₊) F1-score of 58.2%. Interpretable SHAP analysis was performed on this Feature Set shown in Figure 3d. Unsurprisingly, the most important feature identified by SHAP was Cu as the catalyst. SHAP was able to identify a well known trend that catalyst containing Cu produce C₂₊ products. This is due to Cu catalysts ability to facilitate carbon-carbon coupling. Another interesting



trend that SHAP was able to discern was that intermediate C*, CH*, and CO* all favored multi-carbon products, and that strong bonds of CO*, regardless of Miller index, favored the formation of single carbon products. This shows the classic volcano relationship for producing multi-carbon products and verifies that strong binding energies of CO leading to CO poisoning of the active sites or leading to the limited ability to diffuse reducing the C-C coupling.⁵² Of the top ten SHAP features, adsorption of the COOH* plays the least important role, which is logical since it is an intermediate for the CO₂ to CO, CH₃OH, and HCOOH pathways, but some multicarbon products as well.^{50,53,54} SHAP was also able to identify that strong H₂O relaxation energies favored the formation of multi-carbon products, leading to another key descriptor for catalyst design.

3.4.5 C₂H₄ versus CH₄ Classification

The next data subset is composed of 218 C₂H₄-yielding reactions and 192 CH₄-yielding reactions. The cross validations accuracies and F1-scores for the random forest classifiers can be found in Figure 2e and Figures S3 and S4 in the Supporting Information for each Feature Set permutation. The Feature Sets containing only the catalytic descriptors (Feature Sets II, III, and IV) performed the worst out of all the permutations. However, adding Feature Set I with any one or two of the catalytic descriptor Feature sets II, III, or IV improved model performances. This demonstrates the importance of experimental values when classifying C₂H₄ and CH₄.

The best performing Feature Set combination was Feature Set I+IV. This combination had the highest cross validation of 71.2, minority class F1-score of 62.5, and majority class F1-score of 76.4. The SHAP analysis for this feature set can be found in Figure 3e. SHAP analysis also demonstrated the importance of with experimental parameters with voltage, structure, and electrolyte information being the top three features identified. This is interesting because it shows that similar catalytic properties can produce either C₂H₄ or CH₄— both molecules being fully hydrogenated with no oxygen atoms. However, driving selectivity toward C₂H₄ requires, more negative voltages to produce higher electron transfer products with KHCO₃ as the electrolyte. It should be noted that the structure being "other" means that the structure of the catalyst was not within the specified list of common structures therefore it was put into an "other" category. The influence of this "other" category on prediction suggests that unconventional catalyst structures may be more suitable for methane formation than ethylene.

3.4.6 CH₃OH versus CH₄ Classification

The CH₃OH-CH₄ data subset consisted of 125 reactions producing CH₃OH and 192 reactions producing CH₄. The cross validation accuracies and F1-scores for each feature permutation from the random forest models can be found in Figure 2f and Figures S3 and S4 in the Supporting Information. The Feature Set combination with the highest cross validation accuracy was I + II + III with 76.9. This Feature Set combination also had the highest F1-scores for both CH₄ and CH₃OH. SHAP analysis showed that the Open Catalyst relaxation energies were ranked high to

differentiate between CH₃OH and CH₄. The most important feature highlighted by SHAP was H₂O adsorption energy on the 111 Miller index with stronger bindings of H₂O leaning towards CH₄. Interestingly, the trend for H₂O on the 111 Miller index emerges in both the C₂H₄ vs CH₄ and CH₃OH vs CH₄ datasets, with SHAP assigning it greater importance in the CH₃OH vs CH₄ predictions. Surrounded by adsorption energies, voltage was ranked fourth: CH₄ predictions were promoted by slightly more negative voltages than CH₃OH. This supports the conclusion that, despite CH₄ production being more thermodynamically feasible in theory,⁵⁵ it's more common for it to require more negative overpotentials to produce in practice.^{56,57}

3.5 Model Confidence Evaluation

The results of an uncertainty quantification and model stability analysis across the 14 feature subsets for binary classification between CO and HCOOH are summarized in Supporting Information Table S5. Global model stability was high across all subsets, evident by the low standard deviation in 10-fold cross-validation accuracies, particularly for the comprehensive I+II+III+IV subset ($\pm 2.9\%$).

Notably, while Subset I displayed the lowest predictive variance when used by itself (9.5%), this was accompanied by the lowest accuracy, suggesting that the ensemble achieved consensus primarily by defaulting to the majority-class predictions rather than capturing meaningful chemical signal. As catalyst compositions were introduced, the initial increase in Mean Predictive Variance (e.g., 21.4% for Subset I+II) reflected the trees attempting to reconcile conflicting trends within the noisy experimental and compositional data. The subsequent inclusion of theoretical descriptors from the Open Catalyst Project (III) and Catalysis Hub (IV) stabilized these predictions, reducing the variance back down to 14.6% for the I+II+III+IV set. This reduction indicates that the physics-based features provided a necessary grounding that enhanced model confidence. By integrating physics-based descriptors, the model was able to maintain a high degree of consensus while significantly improving accuracy, confirming that the identified mechanistic drivers for CO vs HCOOH selectivity are robust and reliable despite the inherent variability in the underlying literature sources.

To evaluate the predictive power of our framework on out of sample data, we tested the trained I+II+III+IV Random Forest model against a report published within the last 6 months involving dendritic Bi-Ce alloys for CO₂RR.⁵⁸ The experimental parameters, including an applied voltage of -0.8437 V vs RHE, a molarity of 0.1 M, and a KHCO₃ electrolyte, were encoded and fed into the model trained on the historical 2018 to 2021 dataset. The model correctly identified HCOOH as the major product with a forest confidence of 82.5%. This prediction is in agreement with the experimental findings, where the authors achieved a peak Faradaic efficiency of 99.1% for formic acid production. The ability of a lightweight model to accurately categorize the selectivity of a novel catalyst from the most recent literature, despite the inherent noise in multi source experimental data, confirms that the physics based descriptors and LLM extracted features capture the



CO ₂ RR Product	Adsorption/Relaxation Energy Values that Favor Product Formation							Voltage
	CO* (low Miller)	CO* (high Miller)	O*	C*	CH*	COOH*	H ₂ O*	
CO	-	strong	weak	weak	moderate to strong	weak	weak	less negative
HCOOH	weak	weak	moderate to strong	moderate to strong	weak	moderate to strong	-	-
C ₂ H ₄	weak to moderate	moderate	moderate to strong	moderate	weak to moderate	-	moderate	more negative
CH ₃ OH	-	weak to moderate	weak	-	-	-	weak	-
CH ₄	-	moderate	-	moderate	-	-	moderate to strong	less negative
C ₂₊	weak to moderate	weak to moderate	moderate to strong	moderate to strong	moderate	-	strong	more negative

Fig. 4 Conclusions Drawn from SHAP: correlations between adsorption energies and CO₂RR major products as evident by SHAP analysis; blank cells indicate inconclusivity.

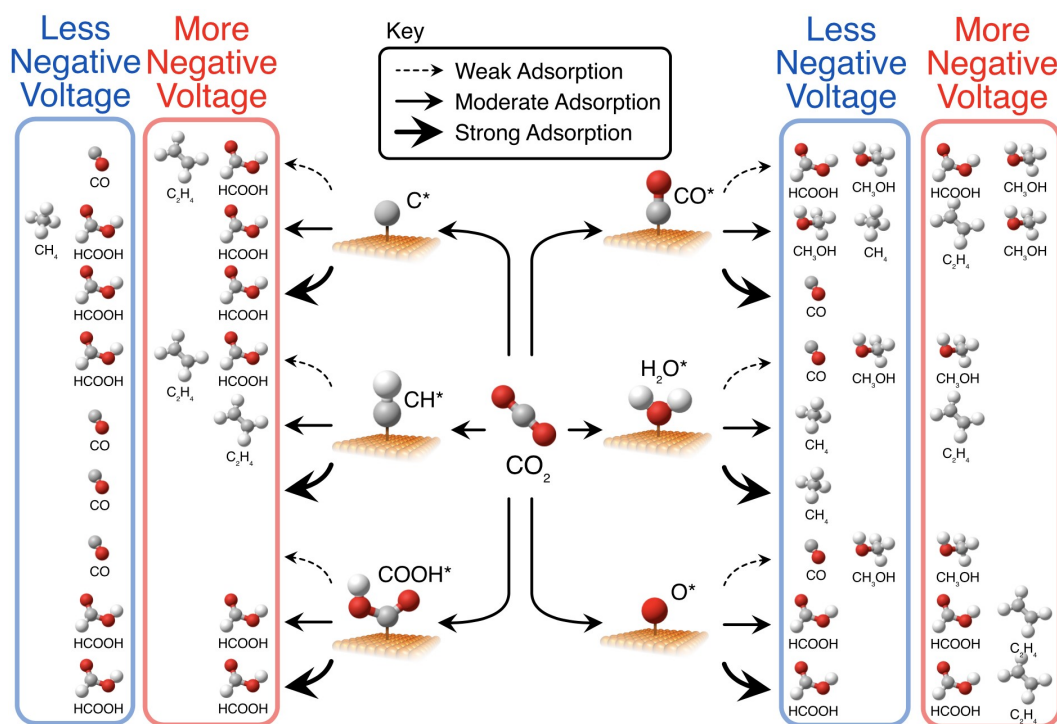


Fig. 5 Visual scheme of conclusions drawn from SHAP.

fundamental mechanistic drivers of the reaction.

Conclusions

We introduced a sustainable ML framework that unites pre-trained, LLM-extracted reaction conditions with low-compute catalytic descriptors—specifically, adsorption (relaxation) energies of key CO₂RR intermediates. The resulting hybrid dataset enabled training of lightweight, interpretable random forest models that recover mechanistic trends while maintaining a modest computational footprint.

Across tasks, integrating experimental parameters with catalytic descriptors yielded higher predictive performance and richer mechanistic insight than using either data type alone. No-

tably, descriptors predicted by pre-trained ML surrogates performed comparably to DFT-derived descriptors, indicating that computation intensive workflows can often be replaced with lower-compute alternatives without sacrificing accuracy.

By formulating the problem as a series of binary classification tasks, we obtained models amenable to interpretation and sensitivity analysis, clarifying structure–condition–selectivity relationships in CO₂RR. Summarized key findings that this hybrid LLM, catalytic descriptor machine-learning approach found are shown in Figure 4 and Figure 5. Here, the model independently verified that CO production requires strong CO* adsorption coupled with weak adsorptions of O*, C*, COOH*, and H₂O*. Interestingly, the model identified that CH₃OH formation needed weak adsorptions



of O^* and H_2O^* for selective production. Contrastingly, to design a system that produces C_2H_4 , catalysts that combine moderate adsorption of CO^* with moderate to strong adsorption of O^* and H_2O^* is required. The model also identified that similar catalytic properties produce C_2H_4 and CH_4 , but the applied voltage is the major driving force with more negative voltages favoring C_2H_4 production.

In addition, the model indicated that both $KHCO_3$ and KOH electrolytes were important for C_2H_4 production across many binary classification problems. This finding is consistent with the higher local pH in alkaline media, which suppresses proton availability, increases CO^* surface coverage, and promotes C–C coupling pathways over competing protonation steps. Moreover, adsorption of the CO^* intermediate on the higher Miller index (211) facet further emphasizes the role of high-index facets in facilitating C–C coupling. Together, these results underscore the importance of performing binary SHAP analyses across multiple product iterations to elucidate the underlying chemistry.

In summary, this work advances a human-interpretable, data-efficient route to AI-enabled electrocatalyst design. It leverages a rigorously curated, peer-reviewed literature dataset and sustainably derived descriptors to deliver interpretable, physics consistent insights rather than opaque predictions. This strategy ensures scientific reliability while enabling interpretable models that reveal mechanistic insights rather than opaque predictions. By identifying critical parameters and reducing computational overhead, the approach can streamline reaction screening and accelerate discovery of next-generation green electrocatalytic systems. Moreover, we show how pre-trained and low-compute interpretable ML can reveal novel insights about green chemical reactions while diminishing computational energy consumption.

Author contributions

B.R.F. and J.J.M. contributed equally to the data collection, model training, and manuscript writing. K.C.L. supervised the project and reviewed the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

All source code and results are provided in Supporting Information. Extracted datasets and code availability are detailed in the Methods section.

Acknowledgments

This work was funded by the U.S. National Science Foundation Research Traineeship (NRT) grant through Award DGE-1922649. We also acknowledge support by the U.S. Army DEVCOM ARL Army Research Office (ARO) Energy Sciences Competency, Electrochemistry Program award #W911NF-22-1-0293. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army or the U.S. Government.

Notes and references

- 1 D. A. Rosser, B. R. Farris and K. C. Leonard, *Digital Discovery*, 2024, **3**, 667–673.
- 2 K. C. Leonard and A. J. Bard, *Journal of the American Chemical Society*, 2013, **135**, 15885–15889.
- 3 K. C. Leonard, F. Hasan, H. F. Sneddon and F. You, *ACS Sustainable Chemistry and Engineering*, 2021, **9**, 6126–6129.
- 4 B. R. Farris, T. Niang-Trost, M. S. Branicky and K. C. Leonard, *ACS Sustainable Chemistry & Engineering*, 2022, **10**, 10934–10944.
- 5 H. Wang, K. Hu, W. Fan, M. Zhang, X. Xia and J. Cai, *International Journal of Hydrogen Energy*, 2025, **101**, 303–312.
- 6 H. Ji, D. Pu, L. Su, Q. Zhang, W. Yan, J. Kong, M. Zuo and Y. Zhang, *Food Research International*, 2025, **202**, 115707.
- 7 G. Yin, H. Zhu, S. Chen, T. Li, C. Wu, S. Jia, J. Shang, Z. Ren, T. Ding and Y. Li, *Molecules* 2025, Vol. 30, Page 759, 2025, **30**, 759.
- 8 C. Bozal-Ginesta, S. Pablo-García, C. Choi, A. Tarancón and A. Aspuru-Guzik, *Nature Reviews Chemistry* 2025 9:9, 2025, **9**, 601–616.
- 9 K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. Billinge, E. Holm, S. P. Ong and C. Wolverton, *npj Computational Materials*, 2022, **8**, 59.
- 10 A. L. Ferguson, J. Hachmann, T. F. Miller and J. Pfaendtner, *The Journal of Physical Chemistry B*, 2020, **124**, 9767–9772.
- 11 H. M. Cartwright, *Machine learning in chemistry : the impact of artificial intelligence*, 1st edn, 2020.
- 12 S. Peng and L. Rajjou, *Plant Cell Reports*, 2024, **43**, 208.
- 13 J. A. Pugar, C. M. Childs, C. Huang, K. W. Haider and N. R. Washburn, *Journal of Physical Chemistry B*, 2020, **124**, 9722–9733.
- 14 L. Dumortier and S. Mossa, *Journal of Physical Chemistry B*, 2020, **124**, 8918–8927.
- 15 J. G. Rittig, K. C. Felton, A. A. Lapkin and A. Mitsos, *Digital Discovery*, 2023, **2**, 1752–1767.
- 16 J. Willard, X. Jia, S. Xu, M. Steinbach and V. Kumar, *ACM Computing Surveys*, 2022, **55**, 1–37.
- 17 M. AlQuraishi, *Current Opinion in Chemical Biology*, 2021, **65**, 1–8.
- 18 Z. Zhang, H. Tang, M. Wang, B. Lyu, Z. Jiang and J. Jiang, *ACS Sustainable Chemistry & Engineering*, 2023, **11**, 8148–8160.
- 19 P. Kollenz, D. P. Herten and T. Backup, *Journal of Physical Chemistry B*, 2020, **124**, 6358–6368.
- 20 A. Dunn, J. Dagdelen, N. Walker, S. Lee, A. S. Rosen, G. Ceder, K. Persson and A. Jain, 2022.
- 21 M. Schilling-Wilhelmi, M. Ríos-García, S. Shabih, M. V. Gil, S. Miret, C. T. Koch, J. A. Márquez and K. M. Jablonka, 2024.
- 22 X. Chen, Y. Gao, L. Wang, W. Cui, J. Huang, Y. Du and B. Wang, *Scientific Data*, 2024, **11**, 347.
- 23 C. W. Kosonocky, C. O. Wilke, E. M. Marcotte Ac and A. D. Ellington, 2024.
- 24 B. R. Farris and K. C. Leonard, *JACS Au*, 2025, **5**, 5578–5589.



- 25 Y. Xue, L. Zhang and G. Zheng, *Advanced Energy Materials*, 2025, e03560.
- 26 Y. Pang, Z. Ding, A. Ma, G. Fan and H. Xu, *Separation and Purification Technology*, 2025, **354**, 129422.
- 27 X. Cheng, H. Wang, X. Zhu, Y. Wang and Q. Fu, *Energy and AI*, 2025, **22**, 100613.
- 28 A. Halilu, Z. Amir, M. K. Hadj-Kali, S. K. Bhargava, M. G. Mohammed and M. A. Hashim, *Journal of Environmental Chemical Engineering*, 2025, **13**, 120019.
- 29 R. Zhu, H. Wang, K. Tang, X. Yang, X. Zhao, J. Yu and R. Hu, *Journal of Energy Chemistry*, 2026, **112**, 842–851.
- 30 F. Shen, S. Wu, M. Kurniawan, D. Ostheimer, J. Shi, T. Chen, A. Bund, T. Hannappel, J. Liu, P. Zhao and S. Miao, *Applied Surface Science*, 2025, **681**, 161459.
- 31 J. Leverett, G. Baghestani, T. Tran-Phu, J. A. Yuwono, P. Kumar, B. Johannessen, D. Simondson, H. Wen, S. L. Chang, A. Tricoli, A. N. Simonov, L. Dai, R. Amal, R. Daiyan and R. K. Hocking, *Angewandte Chemie - International Edition*, 2025, **64**, e202424087.
- 32 R. S. Brower, B. Wuille Bille, S. Chiu, J. T. Perryman, L. Yao, F. O. Agboola, C. A. Nagasaka, Y. Xie, R. Gomez-Caballero, A. Kumari, E. K. Neumann, A. N. Alexandrova, C. C. McCrory and J. M. Velázquez, *Advanced Energy Materials*, 2025, **15**, 2501286.
- 33 P. T. Benavides, U. R. Gracida-Alvarez, K. Richa, J. Port and T. R. Hawkins, *Bioresource Technology*, 2025, **430**, 132565.
- 34 P. Intarapong, S. Yongprapat, R. Saelim, S. Therdthianwong, M. Nithitanakul and A. Therdthianwong, *Journal of Industrial and Engineering Chemistry*, 2025, **145**, 773–782.
- 35 W. Xu, H. Shang, J. Guan, X. Yang, X. Jin, L. Tao and Z. Shao, *Advanced Functional Materials*, 2025, **35**, 2412812.
- 36 W. Xie, B. Li, L. Liu, H. Li, M. Yue, Q. Niu, S. Liang, X. Shao, H. Lee, J. Y. Lee, M. Shao, Q. Wang, D. O'Hare and H. He, *Chemical Society Reviews*, 2025, **54**, 898–959.
- 37 R. Mahajan, A. M. Aleman, C. F. Crago, S. Bhasker-Ranganath, M. E. Kreider, J. A. Z. Zeledon, J. Schröder, G. A. Kamat, M. A. Hubert, A. C. Nielander, T. F. Jaramillo, M. B. Stevens, J. Voss and K. T. Winther, *Journal of Chemical Physics*, 2025, **163**, 124704.
- 38 J. S. Hummelshøj, F. Abild-Pedersen, F. Studt, T. Bligaard and J. K. Nørskov, *Angewandte Chemie - International Edition*, 2012, **51**, 272–274.
- 39 R. B. Sandberg, M. H. Hansen, J. K. Nørskov, F. Abild-Pedersen and M. Bajdich, *ACS Catalysis*, 2018, **8**, 10555–10563.
- 40 J. K. Nørskov, *ACS National Meeting Book of Abstracts*, 2009, **120**, 4913–4917.
- 41 J. Schumann, A. J. Medford, J. S. Yoo, Z. J. Zhao, P. Bothra, A. Cao, F. Studt, F. Abild-Pedersen and J. K. Nørskov, *ACS Catalysis*, 2018, **8**, 3447–3453.
- 42 J. Li, J. H. Stenlid, M. T. Tang, H. J. Peng and F. Abild-Pedersen, *Journal of Materials Chemistry A*, 2022, **10**, 16171–16181.
- 43 S. M. Sharada, R. K. Karlsson, Y. Maimaiti, J. Voss and T. Bligaard, *Physical Review B*, 2019, **100**, 035439.
- 44 M. J. Hoffmann, A. J. Medford and T. Bligaard, *Journal of Physical Chemistry C*, 2016, **120**, 13087–13094.
- 45 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 and Z. Ulissi, *ACS Catalysis*, 2021, **11**, 6059–6072.
- 46 R. Tran, J. Lan, M. Shuaibi, B. M. Wood, S. Goyal, A. Das, J. Heras-Domingo, A. Kolluru, A. Rizvi, N. Shoghi, A. Sriram, F. Therrien, J. Abed, O. Voznyy, E. H. Sargent, Z. Ulissi and C. L. Zitnick, *ACS Catalysis*, 2023, **13**, 3066–3084.
- 47 M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim and A. Fernández-Delgado, *Journal of Machine Learning Research*, 2014, **15**, 3133–3181.
- 48 L. Breiman, *Machine Learning*, 2001, **45**, 5–32.
- 49 S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S. I. Lee, *Nature Machine Intelligence* 2020 2:1, 2020, **2**, 56–67.
- 50 D. Ewis, M. Arsalan, M. Khaled, D. Pant, M. M. Ba-Abbad, A. Amhamed and M. H. El-Naas, *Separation and Purification Technology*, 2023, **316**, 123811.
- 51 M. Zheng, X. Zhou, Y. Zhou and M. Li, *Applied Surface Science*, 2022, **572**, 151474.
- 52 G. Jiang, D. Han, Z. Han, J. Gao, X. Wang, Z. Weng and Q. Yang, *Trans. Tianjin Univ.*, 2022, **28**, 265–291.
- 53 S. Liang, L. Huang, Y. Gao, Q. Wang and B. Liu, *Advanced Science*, 2021, **8**, 2102886.
- 54 A. Mravak, S. Vajda and V. Bonačić-Koutecký, *The Journal of Physical Chemistry C*, 2022, **126**, 18306–18312.
- 55 J. Wu, S. Wang, J. Qi, D. Li, Z. Zhang, G. Liu and Y. Feng, *Materials Today Energy*, 2022, **28**, 101065.
- 56 P. Hirunsit, W. Soodsawang, J. Limtrakul, T. Khlong Nueng, A. Khlong Luang and P. Thani, *Journal of Physical Chemistry C*, 2015, **119**, 8238–8249.
- 57 X. Nie, W. Luo, M. J. Janik and A. Asthagiri, *Journal of Catalysis*, 2014, **312**, 108–122.
- 58 X. Zhang, M. Wu, Y. Xu, J. Ji, Y. Chen, W. Wang, N. Mitsuzaki and Z. Chen, *Langmuir*, 2025, **41**, 32382–32393.



Data availability

All source code and results are provided in Supporting Information. Extracted datasets and code availability are detailed in the Methods section.

