MATERIALS SCIENCE

Leveraging data mining, active learning, and domain adaptation for efficient discovery of advanced oxygen evolution electrocatalysts

Rui Ding[1,2]†, Jianguo Liu[3]†, Kang Hua[3], Xuebin Wang[4], Xiaoben Zhang[1,2], Minhua Shao[5,6], Yuxin Chen[7]*, Junhong Chen[1,2]*

Developing advanced catalysts for acidic oxygen evolution reaction (OER) is crucial for sustainable hydrogen production. This study presents a multistage machine learning (ML) approach to streamline the discovery and optimization of complex multimetallic catalysts. Our method integrates data mining, active learning, and domain adaptation throughout the materials discovery process. Unlike traditional trial-and-error methods, this approach systematically narrows the exploration space using domain knowledge with minimized reliance on subjective intuition. Then, the active learning module efficiently refines element composition and synthesis conditions through iterative experimental feedback. The process culminated in the discovery of a promising Ru-Mn-Ca-Pr oxide catalyst. Our workflow also enhances theoretical simulations with domain adaptation strategy, providing deeper mechanistic insights aligned with experimental findings. By leveraging diverse data sources and multiple ML strategies, we demonstrate an efficient pathway for electrocatalyst discovery and optimization. This comprehensive, data-driven approach represents a paradigm shift and potentially benchmark in electrocatalysts research.

INTRODUCTION

The development of advanced electrocatalysts for the acidic oxygen evolution reaction (OER) in proton exchange membrane (PEM) water electrolysis is imperative for enabling sustainable hydrogen production and achieving carbon neutrality targets (*1*). Despite their promise due to their electronic properties (*2*, *3*), RuO_2- and IrO_2-based materials still face inherent trade-offs between activity and stability under harsh acidic conditions (*4*). Prior efforts in the field via nearly exhaustive, brute-force search (fig. S1A) approaches have extensively explored a wide range of doping (*5*–*10*) and morphological (*11*–*13*) strategies, yet a comprehensive understanding of the optimal electrocatalytic systems remains elusive. The allure of multimetallic or "high-entropy" alloy/oxide materials, with their multi-principal element composition, lies in their potential for synergistic catalytic effects (*14*–*16*). Recent advances have proven such solutions as promising to break the limitations of traditional catalysts (*17*–*19*). However, the exploration of such complex multimetallic systems for acidic OER remains in its infancy, grappling with the immense inherent complexity and the formidable task of navigating the vast compositional and parameter spaces inherent in these systems.

Traditionally, the development of electrocatalysts has focused on empirical, trial-and-error methodologies, heavily dependent on limited prior knowledge and heuristic exploration (*20*). The Edisonian approaches, while grounded in chemical intuition, are insufficient for the nuanced optimization required for our targeted systems, given their expansive array of constituent elements and synthesis parameters (fig. S1B). Moreover, the intrinsic limitations of trial-and-error approaches stem from an over-reliance on subjective intuition with the lack of systematic integration, leading to a narrow and potentially suboptimal exploration of the materials landscape.

To address these challenges and unlock the untapped potential of multimetallic oxides for the acidic OER, we developed a transformative, multistage machine learning ML-driven approach. Our workflow embodied the synergistic integration of data mining, active learning, and domain adaptation at different discovery stages. Each later ML stage leverages and builds on the output of the previous stage, ensuring a seamless and comprehensive exploration of the vast parameter space. By harnessing the collective power of these ML techniques, our methodology has minimized subjective biases and maximized data-driven decision-making in a rationally hierarchical way. As illustrated in Fig. 1, we began by harnessing the breadth of available domain knowledge and conducting data mining. This step successfully distilled key parameters, established foundational pattern understandings, and systematically narrowed the initial exploration space. Subsequently, an active learning strategy was used and synergistically coupled with high-throughput experimental feedback. This iterative active learning–driven process navigated an efficient and refined search within the vast parameter space of quaternary element compositions and synthesis conditions. With this approach, called "DASH," the overpotentials at 10 mA cm^{-2} (η_{10}) observed in the best samples from each of the five experimental batches systematically decreased from 209 to 154 mV, reflecting the dynamic and continuous improvement characteristic of active learning. This dynamic optimization led toward the discovery of a promising Ru-Mn-Ca-Pr oxide catalyst out of an enormous candidate chemical and engineering parameter space. Like directly "dashing" to the endpoint in a winding maze, this is likely unachievable through subjective expertise and intuition. With the help of data mining, the active learning approach could systematically optimize

[1]Pritzker School of Molecular Engineering, University of Chicago, 5640 S Ellis Ave., Chicago, IL 60637, USA. [2]Chemical Sciences and Engineering Division, Physical Sciences and Engineering Directorate, Argonne National Laboratory, 9700 S Cass Ave., Lemont, IL 60439, USA. [3]Institute of Energy Power Innovation, North China Electric Power University, 2 Beinong Road, Beijing 102206, P. R. China. [4]National Laboratory of Solid State Microstructures, College of Engineering and Applied Sciences, Nanjing University, 22 Hankou Road, Nanjing 210093, P. R. China. [5]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong 999077, China. [6]Energy Institute, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong 999077, China. [7]Department of Computer Science, University of Chicago, 5730 S Ellis Ave., Chicago, IL 60637, USA.
*Corresponding author. Email: junhongchen@uchicago.edu (J.C.); chenyuxin@uchicago.edu (Y.C.)
†These authors contributed equally to this work.

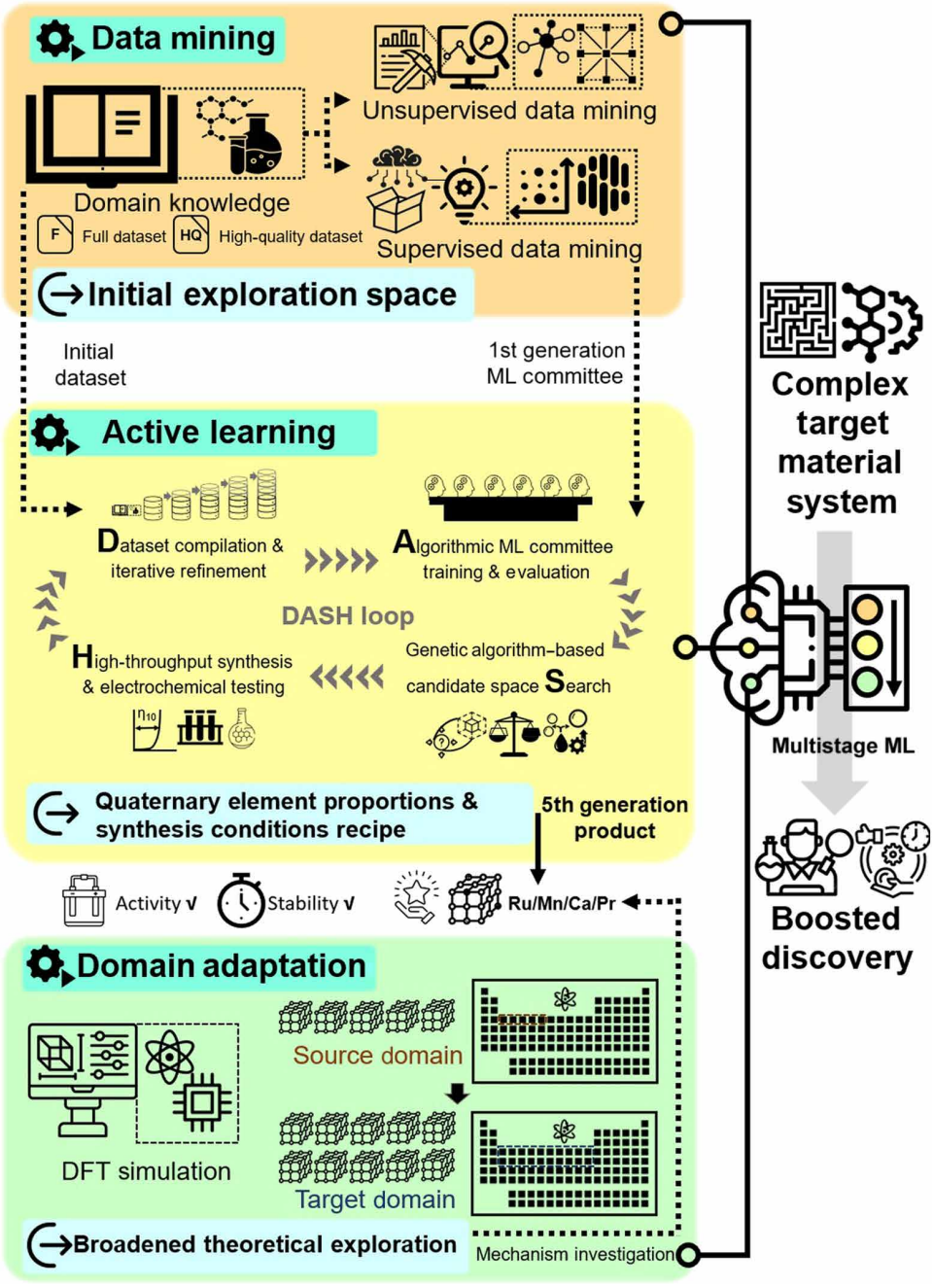

Fig. 1. Schematic illustration of the comprehensive, multistage ML workflow. This workflow guides the entire discovery process of OER electrocatalyst materials with ML. This workflow incorporates different specialized ML expert modules that are derived from various models and data sources.

catalyst composition and synthesis conditions from a refined parameter space, reducing the number of necessary experiments and effectively navigating the wide range of experimental options. In the final stage after material characterization, domain adaptation used within density functional theory (DFT) simulations effectively narrowed the theoretical design space to configurations that are potentially valid. Through enhanced ML surrogate modeling, this approach allowed us to allocate limited resources to critical DFT simulations, enabling a broader and in-depth investigation and thus providing valuable atomic scale insights.

Our multifaceted, integrated ML-powered methodology transcends the mere acceleration of optimal candidate discovery, heralding a transformative shift in electrocatalysis research. By seamlessly weaving diverse ML techniques with diverse knowledge sources throughout the materials discovery process, this work eclipses the current stage-specific, single-expert-system ML applications in OER electrocatalysts (*21*–*25*). Specifically, Hong *et al.* (*23*) used a statistical approach to create predictive models for OER activity using a dataset that is focused on perovskites and is based on publications only. Even though this work provided insightful information, its

limited material space and absence of clear, product-oriented design guidance prevented it from being used practically in synthesis and optimization. In a similar vein, Back and colleagues (*25*) used ML in high-throughput DFT calculations to identify active sites on IrO_2 surfaces and guide theoretical design. However, they only focused on a small number of IrO_2 polymorphs, and their method only used theoretical predictions with large discrepancies between calculated and experimental activity or stability. Additionally, Kim *et al.* (*21*) successfully reduced 77,946 possibilities to 110 experimental trials by optimizing multimetallic quaternary alloy compositions for water splitting using active learning. While impactful, this study did not provide theoretical insights into the underlying mechanisms. Furthermore, their approach, designed for a relatively low-dimensional design space, may lack the sophistication or robustness needed to handle more complex scenarios, such as higher-dimensional parameters like synthesis conditions and a broader spectrum of material properties.

As presented in fig. S1, none of these single-stage ML approaches might be capable of handling an enormous design space with over 3 billion possibilities. Our integrated ML approach, on the other hand, connects data mining, active learning, and theoretical modeling in a coherent framework to holistically leverage knowledge from all available sources to address the more challenging but realistic electrocatalyst discovery task. Using in-depth domain knowledge, the first data mining stage logically defines the exploration space, providing a strong basis for further experimental endeavors. Iterative experimentation based on active learning and theoretical validation enhanced by domain adaptation are then seamlessly linked to these data mining insights through our multistage approach, resulting in a synergistic and interconnected exploration strategy. Demonstrating the efficacy of this comprehensive ML-driven strategy in acidic OER, we have chartered an unprecedented blueprint for the future of electrocatalyst research.

RESULTS

Data mining on domain knowledge for initial exploration space

Facing the astronomical number of possible candidate configurations, we need to rationally decide on the initial exploration space. In the past, this was often accomplished by reading numerous domain publications in combination with researchers' own subjective experience to decide on the initial trial schemes. Serving a similar but more rational role, we have undertaken a comprehensive data mining approach centered around the digitalization and distillation of domain knowledge as a starting point of the data science workflow. This endeavor culminates in the digitalization of an extensive literature body (tables S1 and S2): 534 publications with relatively complete experimental records dating from November 2009 to January 2023. Four complementary datasets were extracted for subsequent data mining, segmented on the basis of (i) different publication qualities: an initial full dataset and a high-quality subset; and (ii) different fitting targets: η_{10} and decay rate, representing OER activity and stability, respectively (fig. S2 and Supplementary Note 1. Note: Supplementary Notes 1 to 8, archived at Dryad: https://doi.org/10.5061/dryad.nk98sf83g, provide independent secondary details for readers interested). Specifically, high-quality literature was defined using the following criteria: (i) publications in journals with an impact factor exceeding 10; (ii) an average daily citation rate of at least 0.025 per Web of Science records up to 20 January 2023; or (iii) publications within the 365 days leading up to 20 January 2023. Applying these criteria resulted in a curated dataset comprising 1358 entries related to activity and 345 entries addressing stability tests, from the original full 1847/453 entries.

Unsupervised data mining

The initial stage of our research involved straightforward yet effective unsupervised techniques to identify fundamental patterns in the datasets extracted from the literature. This included the Bibliometric Interconnected Network Graph (*26*, *27*) and Apriori association rule mining (*28*, *29*). We use the former (figs. S3 and S4) to visualize the frequency of the occurrence of common metal elements alone and in pairs in the aforementioned datasets, as well as their relationship as itemset with η_{10} and decay rate. Concise summaries of statistical patterns of these complicated element-related graphs are provided in Fig. 2 (A and B) (for activity data) and fig. S5 (A and B) (for stability data). The latter (Fig. 2, C and D, and fig. S5, C and D) further extends the association rules from element types to more empirical synthesis parameters, resulting in statistical suggestions based on domain-wide knowledge. With the support of unsupervised data mining of the extensive domain expertise (Supplementary Note 2), we could derive key rational objective decisions on defining the initial exploration.

From the element network analysis, we found that, among the knowledge base, the most frequently used elements are Ir, Ru, Sr, Mn, Ni, and Co. In addition, among them, Ru is the most competitive element. Ru shows promising activity, with qualification rates notably better than Ir and other metals (Fig. 2A), thus aligning with domain consensus (*30*). Furthermore, opposite of consensus (*31*), Ru-based candidates demonstrate evenly matched stability expectations compared to Ir (fig. S5A). Such advantages also existed with Ru's coexistence with other dopants. (For consistency in terminology, all metal elements in the precursor are referred to as "dopants," although the first "dopant" is the primary metal. These are arranged in descending order of their proportions, from the first to the fourth.). Therefore, we believe that (i) Ru versus Ir: Intuitive and impactful insights could be drawn by the distribution of different element's locations in Fig. 2B and fig. S5B for qualified rates (percentage of entries in the databases that reached $\eta_{10} < 250$ mV or decay rate < 1 mV hour^{-1}). If considering 20% as a qualified rate threshold, then 92%/67% (Fig. 2B for activity) and 58%/50% (fig. S5B for stability) of these frequent doping elements would be below it when coexisting with Ir, whereas, for Ru, corresponding ratios would only be 17%/8% and 42%/42%. Such a statistical result suggests that Ru is an indispensable element to explore with other elements as dopants because it offers better potential than Ir to achieve both good activity and stability through synergistic effects.

Delving deeper into the Apriori association rule mining, we obtained several other key insights that help determine the initial exploration space, especially for empirical parameters. (ii) Hydrothermal: For the hydrothermal process (which broadly includes precursor mixing), room temperature is sufficient from an activity perspective, and the corresponding time range suggested was 12 to 24 hours. However, for stability, a temperature range of 50° to 100°C was preferred, and the corresponding time range was 6 to 12 hours. To balance these factors, we hence define the hydrothermal parameter candidate space for temperature and time to be 25° to 60°C and 6 to 24 hours, respectively. (iii) Annealing: For the annealing process, the

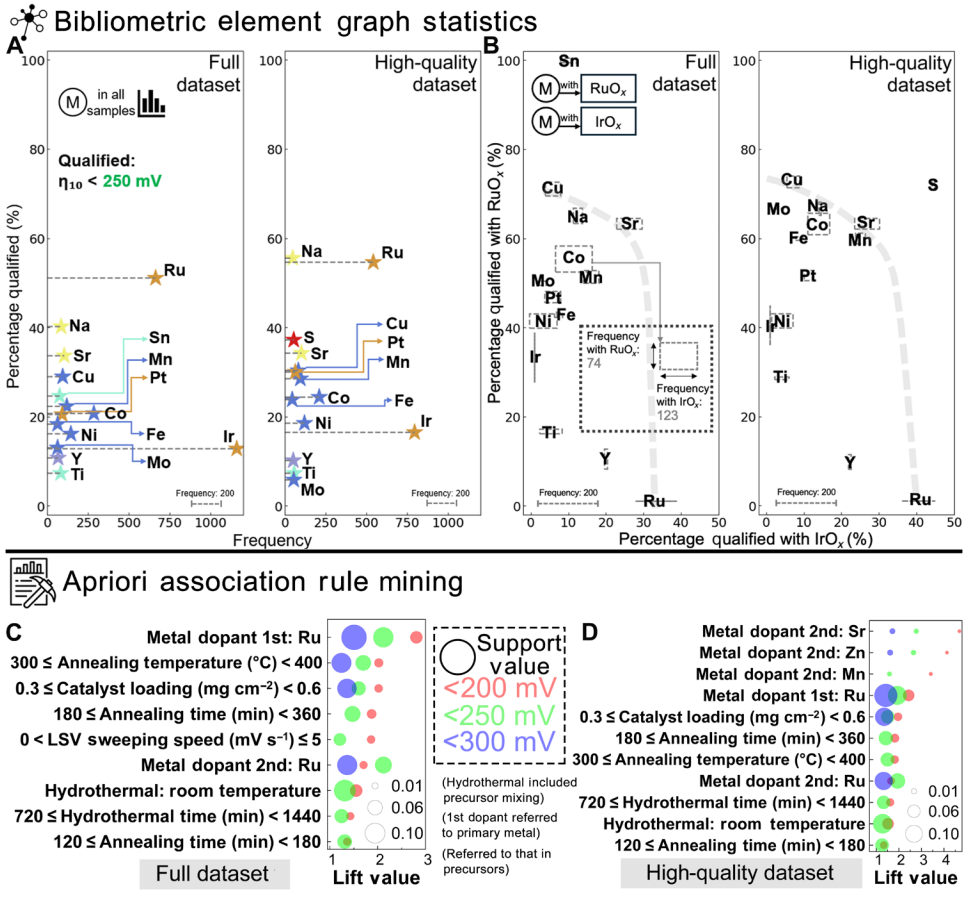

Fig. 2. Key results derived from unsupervised data mining focused on OER activity. (**A**) Frequencies and corresponding qualified percentages of different most frequently reported chemical elements in the full dataset and the high-quality dataset, with a qualified overpotential boundary set at 250 mV. (**B**) Frequencies and corresponding qualified percentages of different frequently reported chemical elements coexisting with either Ru or Ir in metal oxides, in the full dataset and the high-quality dataset with the same overpotential boundary. The dashed line represents the predicted frontier based on the promising elements in the figure that exhibit superior performance (i.e., higher frequencies) and simultaneously have larger *x*- and *y*-coordinate values [(A) and (B) are derived from the original graph in fig. S3]. (**C**) Results of high lift values in Apriori association rule mining based on the full dataset with a frequent itemset length of two for activity-related insights. (**D**) Results of high lift values in Apriori association rule mining based only on the high-quality dataset with the frequent itemset length of two for activity-related insights.

activity side suggests a temperature range of 300° to 400°C, while the stability side prefers 500° to 600°C. For the annealing time, the activity side requests 3 to 6 hours, while the stability side requests more than 6 hours. (iv) Dopants: The Apriori analysis also revealed that Sr, Zn, and Mn are recognized as good secondary elements, potentially serving as major dopants. They have high lift values (ratio of observed to expected co-occurrence), indicating a strong association with good activity. Such an insight is consistent with what we can observe in Fig. 2B and fig. S5B. Sr and Mn are seemingly close to the sketched frontier, suggesting their promising contributions to both activity and stability, when coexisting in either Ru- or Ir-based systems. Prior experimental studies (*32*–*35*) showed that Sr and Mn could enhance the electronic structure and oxygen vacancy formation in Ru and Ir oxide systems. Their coexistence lowers energy barriers for OER intermediates and mitigates active species dissolution, improving catalytic activity and stability in acidic environments. Another vital observation is the stability-focused itemsets suggested that a third type of metal should be present, with a proportion of 5 to 15% in the precursor. This emphasizes the potential of multi-metal synergy for improving catalyst stability.

Drawing on the insights from unsupervised data mining, we crafted from scratch a set of essential design rules to locate our initial exploration spaces to ensure balance between activity and stability. However, through dimensionality reduction techniques such as principal components analysis (*36*) and *t*-distributed stochastic neighbor embedding (Supplementary Note 3) (*37*), we delineated the intricate, high-dimensional nature of the datasets. This further revealed the limitations of unsupervised methods and underscored the need for more advanced ML strategies to accurately parse and interpret the complex patterns within our data. By leveraging ML models to discern data patterns, we can uncover more accurate qualitative patterns and laws governing parameter influence, enabling us to precisely define the most promising regions for continued investigation.

Supervised data mining

By leveraging supervised learning techniques to identify subtle patterns in our dataset, we complement our findings from unsupervised approaches, leading to more nuanced and comprehensive understandings. On the basis of the same datasets, we adopted a committee ensemble approach by training and evaluating (Supplementary

Discussion S1) various ML algorithms. The optimized models were then integrated into a committee, and the influence of each member ML algorithm was determined by its regression performance. Specifically, models capable of generating more accurate predictions quantified by coefficient of determination (R^2) yield greater influence on the overall decision-making process of the committee, whereas those with negligible or negative R^2 values have minimal or no impact on the overall decision-making process. The committee-based query method used here is a common strategy in batch active learning (*38*). Unlike ensemble algorithms such as random forest, which uses decision trees as basic learners, our approach aims to integrate models with completely different architectures and complexities on a higher scale, ranging from support vector regression (SVR), gradient boosting, light gradient-boosting machine (LightGBM), CatBoost, *k*-nearest neighbors regression (KNR), AdaBoost, decision tree, to extreme gradient boosting (XGBoost) to artificial neural networks, in addition to using random forest as one of the committee members. These different algorithmic architectures are capable of learning via different paths on the same dataset, enabling a more comprehensive and robust evaluation of the corresponding uncertainty.

Figures S6 to S13 illustrate the regression metrics of the four committees by two datasets and two prediction targets: full dataset, activity; full dataset, stability; high-quality dataset, activity; and high-quality dataset, stability. A low mean average error (MAE) and a high R^2 value notably highlight some of the committee members, especially the promising prediction accuracies of the boosting-based algorithms. The gradient boosting regressor could achieve an R^2 value of 0.84 and a MAE value of 29.76 mV for η_{10} on a full dataset (fig. S8). For a high-quality dataset (fig. S9), XGBoost stood out with the lowest MAE of 27.21 mV. On the basis of fig. S2B, we could infer that the best members of the two activity-predicting committees could limit their quantitative prediction error to less than 10% of the recognized level, illustrating its microscopic insights. Similarly, for the two stability-predicting committees (figs. S10 and S11), we could observe that a high R^2 value such as 0.86 could be achieved by XGBoost on a full dataset. In addition, on a high-quality dataset, CatBoost could reach an impressive R^2 value of 0.89, demonstrating its excellent forecast consistency and macroscopic grasp of long-term stability and complexity. These best members of the two stability-predicting committees have acceptably overcome the inherent selection bias and data noise caused by the selective reporting and missing details in the literature.

Hence, we transitioned to an in-depth analysis of the models' decision-making processes, systematically leveraging and integrating the interpretative tools Shapley additive explanations (SHAP) (*39*), assisted by Friedman's *H* statistics (*40*), and partial dependence plot (PDP) (*41*). On the basis of the SHAP matrices assembled by the top performers in the committees, as illustrated in Fig. 3 (A and B), we can discern patterns that illustrate how features influence the activity and stability of acidic OER catalysts. The pipeline used is depicted in fig. S14 and unfolds as follows: Initially, we identify the most impactful input features through the weighted average SHAP values, which are then coupled with *H* statistics to pinpoint the most nonlinearly correlated pairs of informative features for two-dimensional dependence plots. By thoroughly analyzing the results (Supplementary Note 4), we vividly capture and present the key insights from this data-mining stage in Fig. 3 (two committees trained on the high-quality datasets) and fig. S15 (two committees trained on the full datasets). Beyond narrowing the design space, this step also assesses whether the ML models' predictive abilities are based on meaningful patterns rather than spurious correlations. It ensures that the models align with established field consensus while still allowing for the discovery of previously unidentified insights.

We begin with (i) activity versus stability factors: Feature importance analyses are provided via cohort bar plots. For activity, element-related features, primarily atomic properties, were particularly influential. Both full and high-quality activity datasets consistently pointed to the atomic radius and electronegativity of the primary metal element as the most critical factors (Fig. 3C and fig. S15C) with great advantages. This aligns with electrocatalytic principles, especially in metal oxides like RuO_2 and IrO_2, where these properties govern the metal-oxygen bond strength and electron density distribution, affecting the overall activity (*42*–*44*). For the second metal, descriptors like d-band electron count and atomic mass, proven effective in modifying the oxide system's electronic structure, alter band structures and binding energies, affecting the catalytic performance (*45*, *46*). While the choice of the primary metal element appears crucial for activity, the atomic radius of the second and, intriguingly, third elements ranked highly for stability despite their lower proportions (Fig. 3D and fig. S15D). Moreover, for stability, parameters related to synthesis thermodynamics and kinetics, such as temperature and time duration, prevailed. These findings prompted further investigation into how distinct feature categories affect OER activity and stability using *H* statistics based on variance decomposition as shown in fig. S16. We found that the average nonlinear interaction between synthesis parameters and elemental features was insufficient to outweigh internal interactions within each category, especially compared to synthesis parameters. Generally, in consensus, the catalyst's morphology, surface area, and crystallinity, which maximize active site accessibility, are largely changed by synthesis parameters (*47*, *48*). In the meantime, intrinsic electronic characteristics defined by elemental features, such as oxygen binding energies and d-band centers, primarily affect activity by adjusting absorption energies for reaction intermediates (*43*, *44*, *46*). This comparatively independent role distribution is consistent with the well-established understanding that elemental composition determines the electronic structure and basic microscopic catalytic behavior, whereas synthesis parameters affect macroscopic physical characteristics (*49*, *50*). (ii) Primary and secondary element choices: Further examination of the cohort plots unveiled a critical threshold at 127.5 pm for the atomic radius of the primary metal, a value autonomously identified by SHAP. Beyond this threshold, elements with a larger atomic radius exhibit a substantial increase in importance. Dependence plots in Fig. 3E and fig. S15E further indicate positive SHAP values under 127.5 pm as undesirable for decreasing η_{10}. A cluster of negative SHAP values around 138 pm identifies preferred elements such as Rh, Pd, Re, Os, Ir, and Ru, with Ru standing out as the most favorable, corresponding to the peak. This is expected as Ru and Ir with similar atomic radius have been theoretically proven having optimal binding energies with oxygen-containing intermediate species (O, OH, and OOH) as benchmarks (*42*, *43*, *47*, *51*). As for secondary metal possibly as major dopant, a wide array of elements as electronic property modifiers was acceptable while completely avoid doping is unfavorable. For stability, Fig. 3F and fig. S15F indicate a preference for a higher proportion of the second metal (greater than ~35%) and the selection from larger atomic radius elements (greater than ~136 pm), predominantly rare earth metals like La, Ce, and Nd and post-transition metals such as In, Sn, and Pb, along with alkaline earth

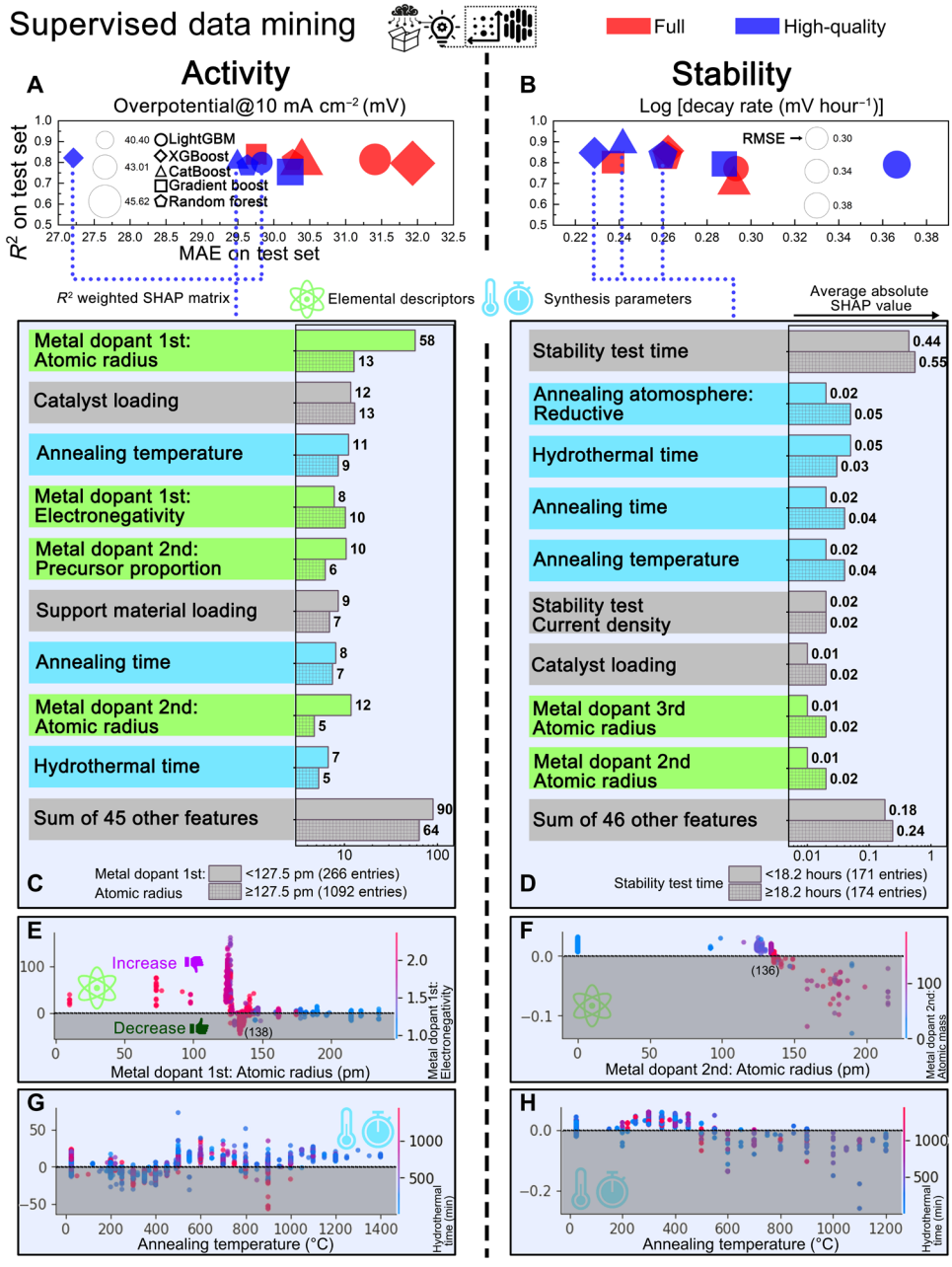

Fig. 3. Key results derived from supervised data mining on the high-quality domain-knowledge dataset. (**A** and **B**) The top-five ML models, as identified from committees trained on various datasets, are evaluated on the basis of R^2 and MAE metrics on the test set. RMSE, root mean square error. (**C** and **D**) SHAP cohort bar plots that highlight the important features with light green and blue frames highlighting the element-related features: atomic properties and synthesis condition parameters, respectively. (**E** to **H**) Selected SHAP two-dimensional interaction plots that feature an interaction of the primary studied feature on the *x* axis with a second feature, which is indicated by the color bars. The second features, also chosen from the top features with a similar type, are those with a high degree of interaction in the Friedman's *H* statistic interaction matrix. Dashed lines at $y = 0$ in each dependence plots split-gray areas that indicate the preferred value ranges.

metals like Ba and Sr. Elements with a larger atomic radius contribute to enhanced lattice stability by reducing strain and facilitating a flexible matrix that accommodates structural deformations during catalytic reactions (*52*, *53*). This property is crucial in harsh conditions like high temperatures or acidic environments, where phase stability prolongs catalyst life. Hence, by all means, it is inferred from the domain perspective that, while Ru is essential as the primary metal for ensuring activity, extensive doping with a diverse range of elements is beneficial for both better activity and stability. Other broad domain-level insights include (iii) testing parameters: Despite its high-importance ranking, the nearly always reported stability testing time was included in the ML model to bolster data integrity, as it serves as the denominator when calculating time-averaged decay. Intriguingly, the cohort plots identified ~18 to 19 hours as a pivotal point, further suggesting that annealing parameters assume greater importance over extended durations. Consistent with

chemical intuition, adequate incubation and heat treatment help complex oxide samples form a more stable phase against degradation (*54*–*57*). Catalyst loading also plays a notable role, shedding light on macroscale marginal effects. As depicted in fig. S15G, beyond a preferred loading of 0.5 mg cm^{-2}, the reduction in SHAP values becomes less marked as restrictions of mass transport would emerge (*58*). Additionally, supporting materials like TiO_x or carbon particles did not exhibit a focused distribution in the analysis, indicating their negligible impact and thus leading to our decision to exclude them from further consideration. As indicated by previous studies, these supports would benefit the dispersion and facilitate the electron transfer (*59*, *60*). However, it is generally believed that they would have limited effects on the intrinsic electronic characteristics of active oxides but are more decisive in benefiting stability by inhibiting dissolution and aggregation (*61*). (iv) Synthesis parameters: As both derived from ML insights and validated by domain expertise in previous discussion, the thermodynamic and kinetic parameters (temperature and time) in the sample synthesis, such as annealing and hydrothermal conditions, are suggested to be nearly as crucial as the choice of metal elements for both activity and stability. Although they could be optimized directly like the proportion of metal elements, the exploratory costs in experimental synthesis differ, especially because the annealing step is time-consuming, and fine-tuning the optimal temperature can lead to substantial energy consumption. Thus, a prudent approach is to confirm them as constants through domain-knowledge data mining. Figure 3 (G and H) and fig. S15H visualize the pattern distributions of annealing temperatures in different committees. Figure 3G reveals a notable negative peak in the wavy SHAP value distribution trend around 200° to 500°C. Intriguingly, Fig. 3H and fig. S15H from stability perspective display a similar yet inverse peak, with the intersection point with the SHAP value zero line around 450°C. Given the initial coarse-grained results from unsupervised data mining, we believe that setting the annealing temperature directly to 400° and 500°C for exploration will most likely balance activity and stability. This temperature interval has been validated in numerous studies, consistently shown to yield RuO_2/IrO_2-based oxides with balanced structural and compositional features conducive to overall promising performances (*62*–*64*). Last, the interaction features of the annealing and hydrothermal time did not exhibit a concentrated color distribution. Hence, considering the need to control costs, we fixed the annealing time to 6 hours on the basis of unsupervised data-mining outcomes, while the hydrothermal time is included as one of the variables for optimization in the subsequent module.

By far, we have extensively narrowed the exploration space based on domain-knowledge data mining. The key parameters identified, and the refined data representations derived from supervised learning have provided a solid foundation for defining the exploration boundaries used in the subsequent active learning stage.

Active learning–guided experimental exploration

Our data mining results provide qualitative ranges rather than precise guidance for optimizing complex multielement systems, especially concerning the selection and proportions of various metal elements. A substantial budget for real-world experimentation would be needed for the corresponding investigation (discussed in fig. S1). To mitigate this challenge, this phase uses an active learning strategy that balances exploration and exploitation to address optimization in nuanced recipes, including the optimal numerical values for elements and proportions and for undetermined synthesis conditions.

Figure S17 both displays our experimental process for synthesizing multimetallic oxide samples (Supplementary Discussion S2) and lists the specific parameter determined by the ML committee in each round by a balanced query strategy. Table S3 further records the set ranges, constants, and constraints (we required Ru to be one of the four elements in the selection group) for the corresponding experimental variables, namely, the initial exploration space defined through previous data mining on domain knowledge. For cost efficiency, we focused on observing η_{10} in synthesized samples, refining through activity-focused ML committees and experimental feedback.

Our comprehensive OER experimental outcomes, conducted in 0.5 M H_2SO_4, are summarized in Fig. 4. Figure 4A (referred to as "batch 0" in the dataset) displays the linear sweep voltammetry polarization curves with Ru as the sole metal precursor. Figure 4 (B to F) presents the top samples and their corresponding curves from each experimental batch. Additionally, Fig. 4G uses violin plots to showcase the distribution of the total 258 samples in five iterations, exclusively guided by the ML committee without human decision interference. Over iterations, the failure rates (low activity products with no η_{10} available) also decreased from 47% in the first batch to 0% in the fourth and fifth batches (64 trials in each batch). A comprehensive pattern analysis of the encodings of the synthesis recipes of different batches is provided in fig. S18, which further revealed that the ML committee was able to evolve itself and navigate out of the failure "dead zone." Meanwhile, a more important and discernible trend emerges, as the best samples exhibiting progressively lower η_{10} values, underscoring the active learning workflow's efficacy in refining this complex material system. Initially, in the first batch, our domain knowledge–based ML committee showed negligible enhancements compared to the baseline RuO_2-400 and RuO_2-500 in Fig. 4A. However, by the fifth iteration, as depicted in Fig. 4F, the final samples demonstrated substantial enhancements. These top-performing samples, labeled A to D, with sample B achieving an exceptionally low η_{10} of 154 mV, surpassed 99.5% of the samples in our domain-knowledge datasets. Notably, our synthesis approach, characterized by its simplicity, holds potential for lower costs and scalability. Through fig. S19, table S4, and Supplementary Discussion S3, a comparison with other recent state-of-the-art reports driven by human domain experts, highlights our samples' unique advantages.

While η_{10} is an important metric, it is not the sole criterion for practical applications. Examining the Tafel slopes of the samples in Fig. 4 (A to F), as presented in fig. S20, reveals that the top-ranked samples do not always excel in kinetic performance. Specifically, a reversed ranking from η_{10} could be found in the final batch. Sample B, despite having the lowest η_{10}, exhibits the highest Tafel slope (80.9 mV $decade^{-1}$), indicative of its slower electrochemical kinetics compared with sample C (69.0 mV $decade^{-1}$). Further evaluation of the samples' electrochemically active surface area (ECSA) is presented in fig. S21, where the ECSA values were calculated on the basis of the measured double-layer capacitance (C_{dl}). B exhibits the largest ECSA (2357 cm^2), and C has the lowest ECSA among the four (923.3 cm^2). However, normalizing the OER current to ECSA at η = 200 mV, C demonstrates the highest specific current density (ECSA of 0.0186 mA cm^{-2}). We also conducted nitrogen adsorption-desorption isotherms of A to D (fig. S22). The fitting results indicate a consistent trend: sample B exhibits the highest Brunauer-Emmett-Teller (BET) surface area of 94.35 $m^2 g^{-1}$, greater than that of A at 88.88 $m^2 g^{-1}$, D at 74.47 $m^2 g^{-1}$, and C at 39.92 $m^2 g^{-1}$. On the basis of these results, we conclude that A/B/D might have advantages in surface area,

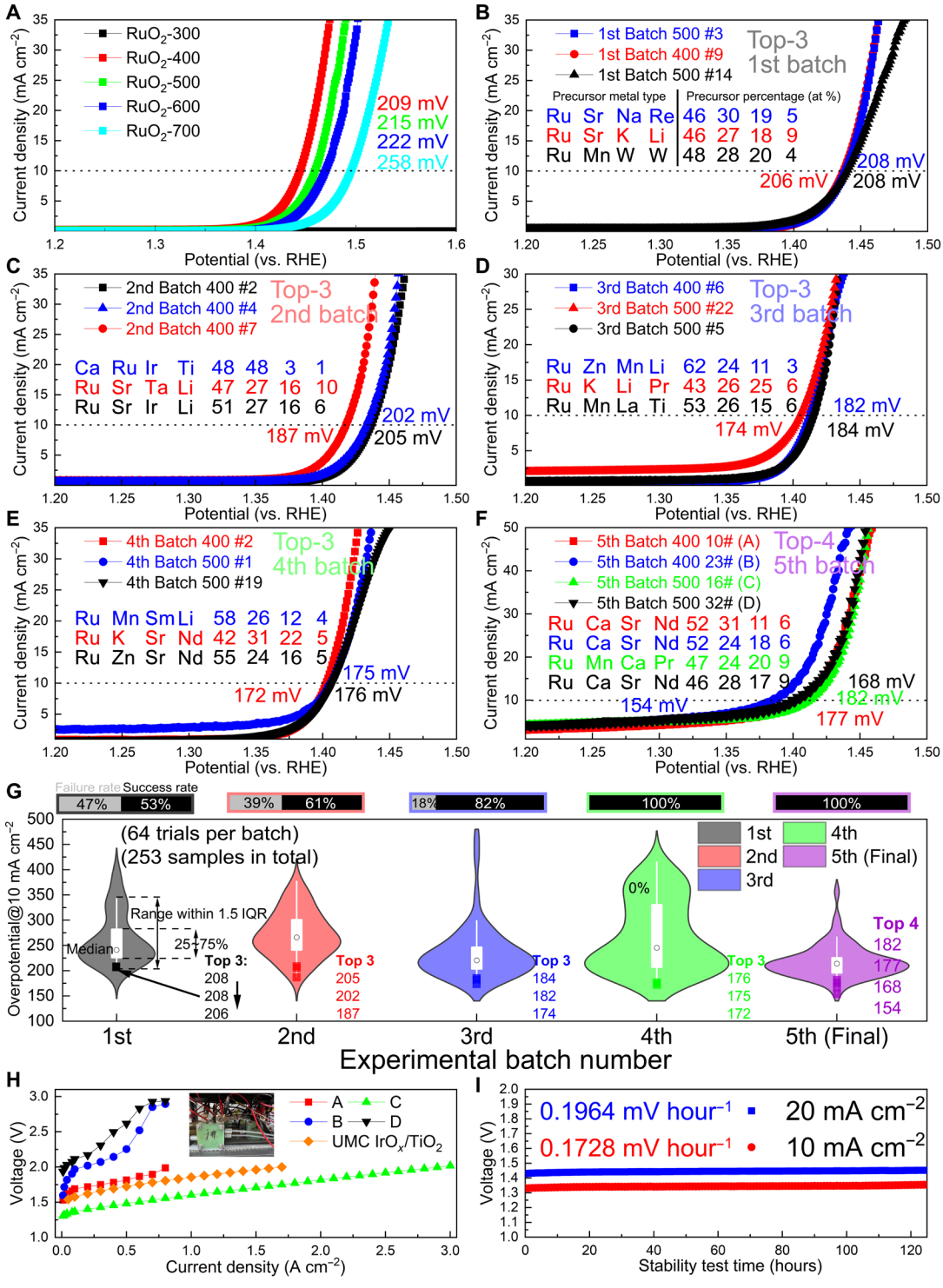

Fig. 4. Electrochemical performance evaluation results of the main/final electrocatalyst products synthesized under active learning guidance. (**A**) Linear sweep voltammetry (LSV) curves of batch 0, Ru-only samples obtained by different temperatures. All LSV curves are represented using line-scatter plots with legends. Due to inferior activity, the curve of RuO_2-300 nearly overlaps with the *x* axis. (**B** to **F**) LSV curves of the best-performing top samples in the first to the fifth (final) ML committee-guided experiment batch. at %, atomic %; RHE, Reversible Hydrogen Electrode. (**G**) The violin plots showing the distribution pattern of η_{10} values of the samples tested and synthesized under the guidance given by the DASH workflow in each experimental batch, along with corresponding failure and success rates in the shape of progress bars. IQR, interquartile range. (**H**) The polarization curves in a PEM electrolyzer single cell with membrane electrode assembly (MEA) samples loaded with different samples of A to D obtained in the final batch. (**I**) The Voltage-Time (*V-t*) curves of sample C loaded on MEA in a PEM electrolyzer single cell with a constant current density of 10/20 mA cm^{-2}.

which could enhance exposure and result in a higher density of active sites and, thus, lower their overpotentials under low-current density half-cell tests. However, C exhibits a better reaction kinetic rate, supported by its best Tafel slope (smallest) and specific current density (highest) normalized by ECSA. However, a balance is still required for maintaining enough surface area. As shown in figs. S21 and S22, the baseline RuO_2-500, although demonstrating a relatively satisfactory specific current density (ECSA of 0.0162 mA cm^{-2}), indicative of good intrinsic activity, has a notably low BET surface area of 45.58 m^2 g^{-1} and an ultralow ECSA of only 310 cm^2. This limited surface area restricts active site accessibility and mass transport, leading to a higher η_{10} of 215 mV (Fig. 4A) compared to those of samples A to D.

This insight led to additional testing in real-world scenarios, specifically focusing on single-cell PEM electrolyzers to assess performance at higher current densities, reaching levels of 1 A cm^{-2}. As depicted in Fig. 4H, sample C, despite having a comparatively higher η_{10} and superior kinetics among the final four, displayed remarkable performance under high-current density conditions in the single cell. A commercially purchased IrO_x sample with an Ir loading of 0.5 mg cm^{-2} was outperformed by sample C. While sample C achieved a notable current density of 3 A cm^{-2} at 2 V, samples A/B/D were deemed subpar. Electrochemical impedance spectroscopy conducted on the corresponding MEA samples (fig. S23), verified that sample C, as part of the MEA component, exhibited comparatively lower contact, charge transfer, and mass transfer resistances at both lower (0.1 A cm^{-2}) and higher (1 A cm^{-2}) current densities. This outcome suggests that, despite its smaller surface area, sample C demonstrates superior comprehensive behavior under high-current density conditions. Conversely, samples A/B/D experienced higher impedance in practical applications despite their initial advantage in η_{10}. Therefore, to provide a more comprehensive evaluation, we further conducted accelerated stability tests under different scenarios. Initially, we performed half-cell constant current tests, which are quicker due to harsher conditions. Consistent with predictions by the stability ML committee in the supervised data-mining module that sample C is the most competitive in long-term performance (fig. S24), real experimental results consistently revealed that sample C outperformed all others (fig. S25). In practical-level electrolyzer cell tests, MEAs loaded with sample C demonstrated considerable stability, as illustrated in Fig. 4I. Over a prolonged testing duration of 125 hours, the average decay rates were only 0.1728 and 0.1964 mV $hour^{-1}$ at 10 and 20 mA cm^{-2}, respectively. Compared with recent studies, C's overall performance in PEM electrolyzer, a basic component in the industrial stack, is proven competitive as shown in fig. S19B.

In summary, while the ML committee's proposed element recipes for samples A/B/D (all containing Ru, Ca, Sr, and Nd) had maximized η_{10} reduction after five iterations in half-cell tests, they underperformed in practical single-cell tests due to kinetic or stability issues. In contrast, sample C showed enhanced kinetics, charge/mass transfer characteristics, and electrochemical stability, which are crucial for device-level applications. Hence, while recipes for A/B/D may result in a higher surface area, C demonstrates superior performance overall and the potential for practical PEM electrolyzer applications. This outcome highlights the necessity of assessing electrocatalyst performance across diverse test conditions and metrics. It also emphasizes that expert input and thorough analysis are essential to complement ML-driven results for the multi-objective optimization that is crucial for practical scenarios, thus ensuring the continuous refinement and applicability of ML strategies.

Material characterization

After identifying the optimal sample C through the DASH loop, we sought to gain deeper insights behind the enhanced performance. A comprehensive suite of characterization techniques from a materials science perspective was conducted (Supplementary Note 5, results of 11 samples other than C). A total of 12 samples were included: samples A to D, RuO_2-400/500, and their respective states before the acid wash (identified with the suffix "Pre"). High-resolution transmission electron microscopy (TEM) results of the 12 samples are provided in Fig. 5 (A and B) and figs. S26 to S28. In general, the samples obtained consist of small, single-crystalline RuO_2 particles, ~5 to 10 nm in diameter, interconnected to form a polycrystalline structure. As indicated by Fig. 5C and fig. S26A, sample C's lattice spacing of the (110) lattice plane expanded to 3.25 Å when compared to the standard lattice spacing of the pure Ru samples both before (RuO_2-400-Pre, 3.12 Å) and after (RuO_2-400, 3.09 Å) the acid wash process. Subsequent energy-dispersive x-ray analysis via TEM in Fig. 5D further strengthened our belief that the second to fourth metal elements are uniformly distributed with non-aggregated signals detected, suggesting their existence as dopants in the major RuO_2 lattice. To further validate our hypothesis, we examined x-ray diffraction (XRD) spectra for more crystallographic information. Having been acid washed, sample C seemed to be dominated by the RuO_2 phase (Fig. 5E) with the impurities in C-Pre well removed, and no traces of $CaRuO_3$ and $SrRuO_3$ could be found like that in A/B/D. Such perovskites are known by consensus to dissolute cations and cause the irreversible collapse of polymetallic oxide structures (*65*–*68*), thus possibly contributing to the inferior stability behavior as observed in experiments. Additionally, we observed noticeable leftward shifts in the XRD peaks for the main phases, signifying slight expansions in the lattice parameters in all A to D samples, especially for (110) and (101) facets compared with RuO_2-400/500 (*69*). This is further supported by the comparison of the average interplanar spacing from TEM images across 12 samples (fig. S29).

To further prove the doping of metal elements into RuO_2 lattice, we obtained x-ray photoelectron spectroscopy (XPS) spectra for the surface chemical states, with a focus on sample C. Combining the XPS semi- and inductively coupled plasma quantitative analysis summarized in fig. S30, we confirmed that Mn was the major dopant. Hence, we first examined the XPS Mn-2p spectra in Fig. 5F. Compared with that before acid wash, the Mn^{2+} 3/2 peak of sample C shifts toward a lower binding energy, which indicates Mn cation's difference with that in Mn oxides and as a dopant into the Ru oxide lattice as reported (*70*). This was further supported by electron paramagnetic resonance comparisons of magnetic properties (fig. S31). We further investigated the Ru 3d and 3p XPS spectra in Fig. 5G. Different from A/B/D, the Ru 3d 3/2 peak of sample C exhibited a slight shift toward a higher binding energy compared with RuO_2, consistent with previous reports indicating Mn's doping effect on Ru's local electronic structures (*71*, *72*). As for Ru's 3p peaks, no obvious shift in peaks could be found, but the results are hidden in peak fitting. On the basis of previous reports, we calculated the $Ru^{>4+}$ species ratio, which has been proven to optimize the OER reaction pathway and enhance the charge transport by both experimental and theoretical approaches (*70*). As shown in Fig. 5H, we could find that sample C has the highest ratio (35.24%) among A to D. The result again supports C's superiority in terms of intrinsic activity and charge transfer characteristic, which could

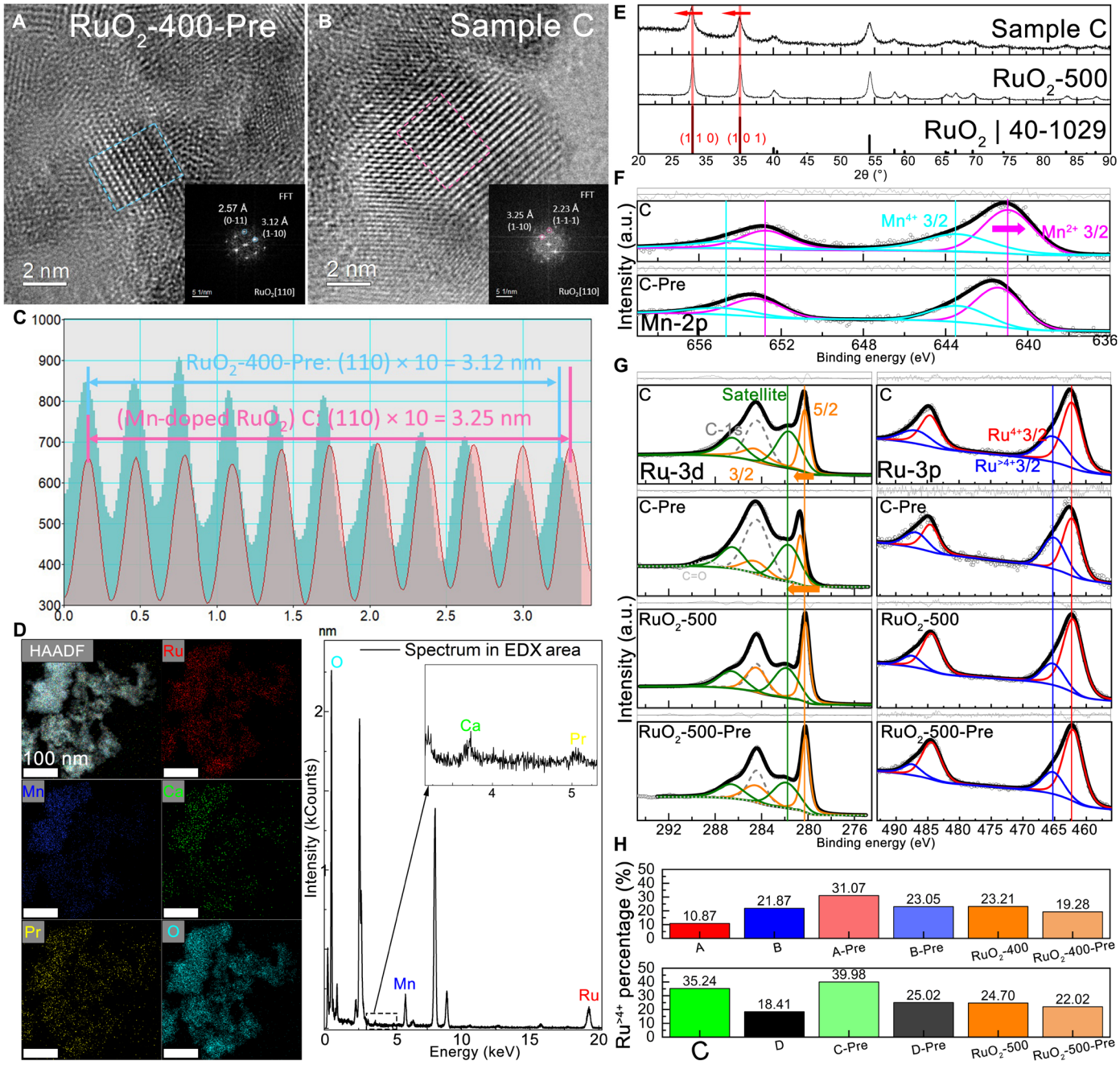

Fig. 5. Material characterization results of the optimal sample C identified through DASH loop. (**A** and **B**) High-resolution TEM images of RuO_2-400-Pre and sample C, respectively, with cyan and pink frames selected for performing fast Fourier transform (FFT) to analyze the diffraction pattern. (**C**) Corresponding statistical analysis of the average interplanar spacing. (**D**) Energy-dispersive x-ray (EDX) elemental distribution mapping spectrum for various elements in sample C. HAADF, high-angle annular dark-field; kCounts, kilocounts; keV, kiloelectron volts. (**E**) X-ray diffraction XRD spectra of sample C compared with RuO_2-500. (**F**) XPS Mn-2p spectra of sample C and its pre-acid wash state: "C-Pre." (**G**) XPS Ru-3d and 3p spectra of sample C, RuO_2-500, and their corresponding states before acid wash. a.u., arbitrary units. (**H**) Statistical analysis of the XPS-derived $Ru^{>4+}$ species percentage in the samples.

overcome its disadvantage in surface area and become the most practical sample. These thorough characterization results laid the groundwork for choosing realistic input configurations for atomic-level simulations, which helped us better understand the identified optimal catalysts and validate our experimental findings in the next phase.

Domain adaptation–assisted DFT theoretical simulation

To further understand the origin of the superiority of sample C's activity and stability, DFT simulation in the next stage is performed. However, because numerous elements are involved with varying doping ratios, we need to perform a large number of expensive DFT relaxations to find the stable structure of each sample before

conducting surface catalysis simulations. Therefore, in this last module, a domain-adaptation strategy based on commonly reported DFT ML surrogate modeling (*73*–*75*) was illustrated.

The characterization results suggested that the C is RuO_2 doped with multiple metal elements, and the same for A/B/D's surfaces after Sr and Ca were dissolved during acid wash. Therefore, to simplify and unify the qualitative discussion, we selected RuO_2 (110) as the subject of multi-metal doping research. Figure 6A presents the schematic workflow for the domain adaptation strategy. We aimed to achieve effective prediction of stable configurations by the ML committee when dopant contents were provided. As we are interested in handling a wide range of both the types and amounts of dopants, similar to previous experimental modules, the DFT relaxation calculations for sampling this vast candidate space from scratch would be prohibitively expensive for ML surrogate modeling. Hence, we divided the entire set of metal elements of interest into two domains: The source (S) domain candidates include common earth-abundant metals like Fe, Co, and Ni, and the relatively rare elements such as La and Ce are categorized into the target (T) domain. We provided abundant data entries in the S domain (3973) through high-throughput DFT relaxation calculations (Supplementary Note 6), while only about one-fifth (848) of the data entries were provided with the T domain concerned. Therefore, in Fig. 6B, Committee S achieved satisfactory predictions for slab energy values across its ML committee members that had R^2 values close to or over 0.99. In comparison, Committee T exhibited inferior performance due to its broader candidate space to be sampled and fewer data entries available as expected. Next, instead of directly expanding the dataset size, we used domain adaptation to avoid potential exponential growth in the supplemental DFT calculations needed to cover additional elements in the T domain. This involved fine-tuning the members of Committee S on Dataset T for a limited number of epochs. Figure 6B along with fig. S32 demonstrates the corresponding results, showing that the fine-tuned Committee S-T substantially improved its regression metrics on Dataset T. Compared to Committee T, Committee S-T was able to use the same small-volume dataset T to adaptively fine-tune the general pattern learned from the large-volume dataset S, further including situations when elements in the T domain were added. Moreover, similar patterns (figs. S33 to S35) could be observed when OER-related intermediate species, such as O, OH, and OOH, are present on the slab surface, further supporting the broad effectiveness of this strategy. In Supplementary Note 7, we further address the potential "forgetting" issues (*76*) of Committee S-T through detailed cross-domain evaluations. Post-adaptation, Committee S-T retains notable predictive power on domain S, with R^2 values primarily in the range of 0.8 to 0.9, demonstrating that the model successfully preserves key patterns from the source domain despite domain adaptation. This performance contrasts sharply to the significantly negative R^2 values and orders-of-magnitude higher mean square error/MAE observed when Committee S or Committee T was directly applied to the other domain, highlighting the inability of single-domain models to generalize effectively. As Committee S and Committee T were exclusively trained on single domains, this clearly reflect the fundamental gap between two domains. Committee S-T's robustness could partly be attributed to the ensemble structure and diversity of Committee S-T, which helps balance adaptation to domain T while retaining meaningful predictive power on domain S. Given the higher importance of domain T in our study, Committee S-T represents a cost-effective and reliable surrogate model for exploring the far vaster candidate space central to this work.

Next, we provided the quantitative composition results to Committee S-T and identified stable doping configurations, specifically the lowest slab energy, using the method described in DASH as illustrated in Fig. 6C. Without extensive calculations to determine the locations of the dopant atoms, four representative structures for samples A to D were selected using our ML surrogate on the basis of domain adaptation. Subsequently, we conducted OER simulations on these slabs. Initially, we comprehensively investigated the theoretical OER activity by simulating reaction pathways on the surface covered by various species (*25*, *77*). First, we investigated the adsorption evolution mechanism (AEM) that is recognized as the most common reaction pathway proposed by Man *et al.* (*30*). Figure 6D indicates that the theoretical OER overpotentials for sample C, computed from the energy barriers of the rate-determining steps (fig. S36), were overall the best compared to those for the other samples. Although slightly higher than sample B and undoped RuO_2 in the O-covered scenario, the OER overpotentials for C were the lowest in the other two scenarios at 403 and 376 mV, with the latter considered the commonly studied descriptor of OER activity (*30*). We also examined the lattice oxygen mechanism (LOM) pathway, which has been reported to explain the circumvention of the theoretical 370 mV overpotential limit of the AEM (*78*–*83*) by oxide electrocatalysts, albeit at the expense of stability. Our findings suggest that, depending on the surface coverage conditions, all samples could exhibit lower theoretical overpotentials via the LOM pathway (fig. S37). To ensure a more comprehensive analysis, both previously calculated AEM and LOM pathways under default vacuum conditions were further evaluated within an implicit solvent field (fig. S38), yielding consistent trends. Additionally, explicit solvent field effects, simulated through ab initio molecular dynamics (AIMD), were investigated under no surface coverage conditions (fig. S39). Although numerical discrepancies between simulated and experimental overpotential values are common—likely attributable to model simplifications (*84*), scaling limitations (*85*), and unmodeled environmental effects (*51*) in such complex materials systems designed for OER (*86*–*91*)—the qualitative agreement across various simulation conditions supports our prior experimental observation. Specifically, sample C demonstrates optimal intrinsic single-site activity due to its refined electronic structure, reaffirming its superiority over the other samples. Further examination of the density of states in fig. S40 revealed that C's Ru 4d and O 2p band centers were closer to the Fermi energy level, contributing to its optimized reaction pathway as reported (*92*–*94*). This might explain our experimental observation that C has the best electrochemical kinetics among A to D. Although undoped RuO_2 is proven similarly competitive in microscopic reaction kinetic by both simulation and experimentation, its drawback in low surface area makes it impossible to be the best candidate for application.

We also assessed the theoretical stability (shown in fig. S41 and summarized in Fig. 6E). Although not the best (−1.20 eV), C and all other samples demonstrated favorable water adsorption energies (ΔG^*_{H2O}), which can be regarded as a straightforward yet effective descriptor of oxide's stability in water (*95*–*97*). Consistent with experimental observations, they are stable and unlikely to dissolve. Further calculations of the surface Ru dissolution potentials (U_{diss}) (*73*, *98*, *99*) and surface O vacancy formation energy (ΔG_{VO}) indicated that C was the best, with values of 3.34 V and 3.84 eV,

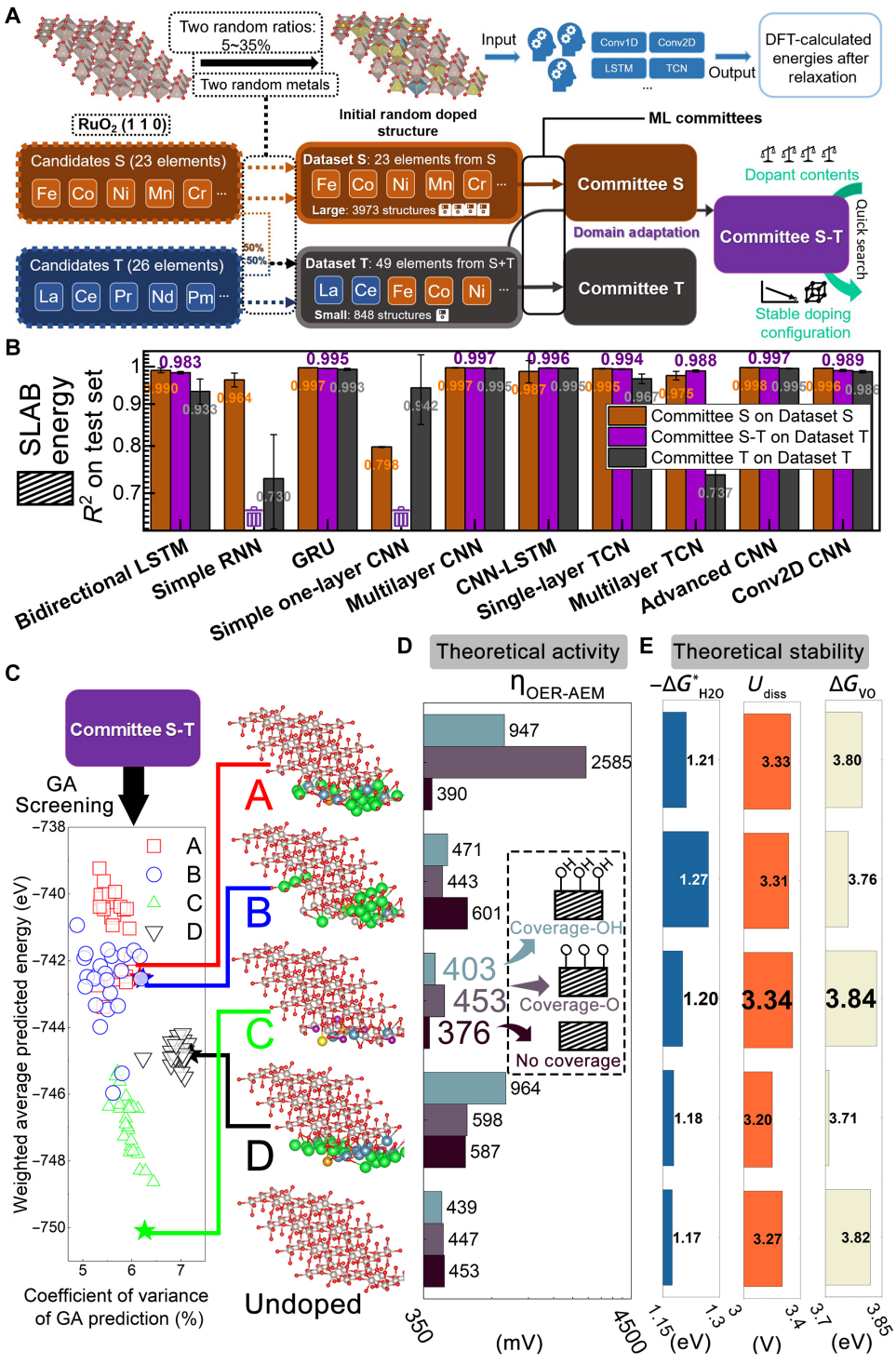

Fig. 6. OER mechanism simulation for the final batch samples enhanced by domain adaptation ML strategies in DFT methods. (**A**) Schematic of the domain adaptation workflow to obtain Committee S-T. (**B**) A summary of the committees' R^2 values on the test sets. The trash bin icons were placed to indicate that, for Committee S-T, corresponding members from Committee S and Committee T were deprecated because of inferior performances. LSTM, long short-term memory; GRU, gated recurrent unit; TCN, temporal convolutional network. (**C**) Results of the GA screening using committee models S-T for identifying the lowest slab energy of compositions corresponding to experimentally obtained samples A to D, and the corresponding determined doping configuration structures with the lowest slab energy. (**D**) Summary of the theoretical OER overpotential of the samples under different coverage scenarios. (**E**) Summary of the theoretical stability descriptors of the samples.

respectively. The former suggests that C is the most electrochemically stable. The latter, which is commonly examined in OER-related studies (*100*–*102*), indicates the stability of Ru-based materials in acidic electrolytes, as O vacancies could lead to over-oxidation and the generation of soluble Ru species. Results of the above simulations for stability performed in the implicit solvent field are also consistent (fig. S42). Last, although partial switching to the LOM pathway can contribute to a reduced overpotential, as supported by previous activity simulations, it introduces a substantial trade-off: the potential dissolution of lattice Ru species, which could be detrimental to catalyst stability. To address this concern (*103*), we investigated the free-energy barriers for the leaching of RuO_4 species ($\Delta G_{Diss\text{-}RuO4}$) under various environmental conditions (fig. S43). The results revealed consistent trends across different metrics, with an overall ranking of stability (from most to least stable) as follows: C > A > undoped RuO_2 > B > D. These findings further emphasize the superior stability of sample C under conditions that challenge its long-term structural integrity. In summary, our comprehensive DFT simulations of both theoretical activity and stability for samples A to D correlate well with the experimental results.

Last, for readers interested in further explorations based on Committee S-T, Supplementary Note 8 provides additional insights into broader candidate spaces and innovative dopants. To this end, the combination of data-driven parameter selection, adaptive experimental validation, and theoretically supported optimization highlights the interconnected efficiency of our multistage approach, as elaborated in the discussion below.

DISCUSSION

This comprehensive study represents a paradigm shift in the development of advanced electrocatalysts for acidic OER. Our research unfolds a multistage, ML-driven approach that embodies the organic fusion of data mining, active learning, and domain adaptation, leveraging these methodologies synergistically across different stages of material discovery based on varied data sources. This integration not only streamlines the journey from conceptualization to experimental validation and theoretical investigation but also ensures a flexible and reliable data-driven exploration of complex multimetallic systems. Unlike traditional methods that rely on chemical intuition and trial and error, our approach integrates a rational, hierarchical, and data-driven decision-making process across a vast compositional space, using different ML modules at various stages of the pipeline. The synergistic integration of these modules—from initial data mining to guide exploration space definition, through active learning for limited-budget experimental exploration and exploitation, to domain adaptation for theoretical validation—creates a cohesive framework where each component informs and strengthens the subsequent stages. Furthermore, while current ML studies in the field often remain confined to single stages of research based on a single modality of data, our work extends the application of ML to encompass the entire process of materials discovery, thus addressing challenges that single-stage models cannot. Through this comprehensive framework, we have successfully identified an Ru-Mn-Ca-Pr catalyst out of billions of possibilities that exhibits exceptional activity and stability and has been validated extensively through experimental and theoretical approaches. As autonomous laboratories and high-throughput testing capabilities continue to advance, this framework could be extended to incorporate multi-objective optimization strategies and device-level engineering parameters, potentially enabling even more comprehensive catalyst development workflows. In addition, the modular design of DASH suggests its adaptability to other catalytic systems. Researchers can tailor specific workflow components, such as feature engineering, ML algorithms, and active learning strategies, to align with the unique requirements of diverse materials discovery challenges. Therefore, this methodology potentially establishes a benchmark for the field of electrocatalysis, demonstrating the potential to revolutionize the discovery and optimization of OER catalysts through a comprehensive, data-driven strategy.

MATERIALS AND METHODS

Domain-knowledge dataset

A systematic literature review was conducted to compile a dataset of experimental studies on metal-oxide type electrocatalysts for OER in acidic conditions. Studies involving theoretical DFT or molecular dynamics simulations were excluded because of the variability in outcome metrics compared to experimental results. The literature search was performed using the Web of Science, resulting in a comprehensive collection of relevant publications. The curated dataset includes input features such as transition metal elements, proportions of precursors, conditions of hydrothermal mixing (precursor mixing), annealing process conditions, post-treatment conditions, and testing conditions. Output fitting targets for catalyst activity and stability, specifically η_{10} and time-averaged voltage decay rate, were recorded in the data entries. To manage variability in experimental conditions and testing standards, we standardized key experimental parameters across studies. For instance, catalyst loadings, precursor compositions, annealing temperatures, and testing conditions were normalized to common units where applicable. For minor missing values in the dataset, such as precursor mixing duration (missing in 11.6% of cases), the median value from the dataset was used for imputation, a method validated by previous studies (*104*) in the field of ML (*105*–*107*). These strategies aim to minimize biases and enhance the dataset's consistency for reliable model training.

Criteria were established to differentiate a "high-quality dataset" from the initial dataset, focusing on records from high-impact journals with high citation counts and more recent publication dates to ensure data reliability. Specifically, publication that fulfilled one of the three conditions: (i) published in journals with an impact factor greater than 10; (ii) having an average daily citation rate of at least 0.025 (as of 20 January 2023); (iii) published within 1 year leading up to the cutoff date (20 January 2023). The high-quality dataset post-screening included 1358 entries (fig. S2B) related to electrocatalytic activity and 345 entries addressing stability under constant current density tests (1847/453 in the full dataset). Four versions of datasets capturing domain knowledge were thus established: high-quality activity, full activity, high-quality stability, and full stability. This carefully curated dataset forms the foundation for our subsequent unsupervised and supervised learning analyses.

In both unsupervised and supervised data mining, insights from the high-quality datasets and the full initial datasets were integrated. This strategy was also applied in the active learning module called DASH flow. Two ML committees, one initiated from the high-quality dataset and the other from the full dataset, were updated iteratively with the data supplements derived from experimental observations. Candidate observations were determined on the basis

of recommendations from both committees, maintained at a 1:1 ratio in all genetic algorithm (GA) search result batches, to ensure diversity and reduce potential biases in the learning process. Details regarding reproducibility are documented in Supplementary Note 1.

Unsupervised data mining

The Bibliometric Interconnected Network Graph and the Apriori association rule mining are applied, building upon the foundational work referenced in previous study (*27*). These methods aim to explore relationships among itemsets categorized as "qualified" or "disqualified." Thresholds for OER activity have been established at 200, 250, and 300 mV for overpotential; for OER stability, thresholds are set at −1, 0, and +1 log (millivolts per hour) for decay rate, consistent with those used in the level set estimation.

A technique similar to that used by VOSviewer bibliometric software (*26*) facilitates visualization of the distribution of key elements within the dataset via an interconnected network graph. In such graphs like figs. S3 and S4, nodes symbolize different elements, where the color of the node edges differentiates element groups and node size denotes occurrence frequency. The intensity of a node's inner color indicates the proportion of high-quality samples associated with that element (e.g., η_{10} of OER < 250 mV,); here, its values are referred to as "quantification rates." The width of the lines between the nodes represents the frequency of element pairs co-occurring, and the depth of line color reflects the likelihood of these pairs being part of high-quality entries. Qualification rates, defined as the probability that the presence of an element or a pair of elements in the system contributes to reducing overpotentials or decay rates, are determined. These values for different chemical elements are further visualized as scatter plots in Fig. 2 (A and B) and fig. S5 (A and B). Network graphs are produced using the Python package, NetworkX.

Similarly, Apriori association rule mining is applied to extend the analysis to additional features for pattern discovery. The same thresholds previously mentioned are used for itemset comparison across different datasets. This method incorporates continuous variables such as synthesis and testing parameters into the analysis. For interpreting results, particularly in bubble figures, two key metrics are emphasized: "lift" and "support." A high lift value, calculated as the ratio of the observed support to the expected support if the items were independent, indicates a strong association with a target itemset linked to favorable performance. Conversely, a high support value, representing the frequency of occurrence of an itemset within the dataset, shows its significance in the analysis. In visualized figures, itemsets with the highest lift values are displayed, indicating the parameters most likely associated with high performance. Details on reproducibility are provided in Supplementary Note 2.

Supervised data mining

A committee ensemble strategy was used in the supervised data mining module to enhance the robustness of the ML workflow and minimize potential biases or errors. This ensemble, comprising 11 models including SVR, KNR, LightGBM, CatBoost, XGBoost, gradient boosting, random forest, decision tree, AdaBoost, and multilayer perceptron with one or two hidden layers, integrates insights from diverse architectures. It is noted that LightGBM, CatBoost, XGBoost, gradient boosting, random forest, and AdaBoost are generally recognized as "ensemble models" in computer science, indicating their composition from base learners such as single decision trees. In this study, each of these ensemble models is considered a member of a higher-level ensemble, the "committee," alongside other models like SVR or KNR. Each committee member was subject to fivefold cross-validation to fine-tune hyperparameters, aiming to optimize performance.

Interpretation of the internal decision-making processes of these models was primarily performed using SHAP analysis, supplemented by PDP and Friedman's *H* statistics. For a broader perspective, the concept of insight assembly was expanded, as indicated by dashed lines in Fig. 3. Rather than relying solely on the SHAP value matrix of the best-performing committee member model, the weighted average SHAP values of the top-three performing models (e.g., LightGBM, XGBoost, and CatBoost for the committee trained on the high-quality dataset) based on R^2 values were used. This approach enhances the SHAP matrix to support parameter insights in SHAP analysis plots. Additionally, because of high-input dimensionality in supervised ML models, the SHAP results were segmented by predefined feature categories into two domains: "element" and "synthesis and testing" for clearer analysis. This segmentation facilitates intuitive inspection and clearer insights. Alongside this, Friedman's *H* statistics were used to identify high-order, nonlinear correlations within influential features. Target feature pairs, exhibiting both high importance in SHAP and elevated second-order *H* statistics, indicative of notable nonlinear dependencies, were analyzed. The use of PDPs together with SHAP dependence plots for these feature pairs allows a comprehensive understanding of the patterns and relationships that influence model predictions. Details regarding reproducibility and discussion on physical chemical rationales can be found in Supplementary Note 4.

Active learning–guided experimental exploration

Building upon the insights gained from our data mining stage, an iterative experimental optimization process guided by active learning was implemented. The DASH loop was developed to synthesize Ru-based multimetallic oxides with up to four metal elements, using an iterative active learning cycle driven by ML committees for predicting activity. For each iteration, the two ML committees, trained on the current datasets (high-quality and full datasets, with supplement from each iteration), estimate the expected values and uncertainties for unexplored points within the exploration space, as indicated in table S3, derived from earlier domain-knowledge data mining. The robustness and fairness of these estimations are ensured through a weighted and averaged approach based on the R^2 value of each model in the committee from the test set, consistent with methods used in prior data-mining modules. The experimentally validated samples were further characterized, providing critical input for the subsequent domain adaptation–enhanced DFT simulations. For inferior models with R^2 values lower than 0, its corresponding weight coefficient would be set to 0 to nullify its contribution in the committee decision. In the first iteration, the two ML committees based on high-quality and full datasets for activity in previous supervised data-mining modules are directly applied for a batch parameter suggestion GA search. Similarly, after the first iteration's experimental observations, the corresponding data point records would be supplemented with the existing datasets for the next iteration.

The GA search identifies a batch of preferred observation points within the exploration space. Parameters determined during the GA search include types of metal elements, corresponding precursor proportions, and hydrothermal conditions, as detailed in table S3. A balanced batch query strategy accounts for both the expected R^2 weighted low average (overpotential) and high variance values in

predictions (prediction discrepancies among committee members). The GA search is conducted at random, generating a batch of suggestions and maintaining a 1:1 ratio between suggestions with lower weighted averages and those with higher variances. Furthermore, the pyrolysis temperature settings alternate between 400° and 500°C, also in a 1:1 ratio. The ML committee alternates between models based on the full dataset and high-quality–only dataset, in a 1:1 ratio, leading to $2 \times 2 \times 2 = 8$ types of parameter suggestions for each batch for experimental synthesis and validation, all in equal proportion to minimize bias. This method incorporates diversity in batch candidate generation, in line with the community consensus regarding the integration of various decision-making systems and expert systems within the active learning framework, as supported by recent studies (*108*, *109*). The hyperparameters for the GA search are set to a population number of 3000, a maximum of 50 iterations, and a variation probability of 0.01. Each parameter suggestion type is repeated eight times, resulting in a batch of 64 suggestions for experiments. Following the synthesis of the recommended formulations indicated by the 13-dimensional vector from the GA search, the corresponding experimental synthesis and sample evaluations are undertaken. These observations are then digitized and incorporated into the dataset to inform updates. Subsequently, the next iteration of the ML committee is trained on this enriched dataset to commence the next round.

Domain adaptation–assisted DFT theoretical simulation

Domain adaptation was applied in the DFT section to align the predictive capabilities of machine-learning models between two datasets: Dataset S, which includes transition metals commonly used as dopants, and Dataset T, which encompasses a wider array of rare earth elements. To begin, separate ML committees, Committee S and Committee T, were formed and trained rigorously on their respective datasets to achieve robust predictive accuracy for each specific domain. Training included a variety of deep-learning architectures such as bidirectional long short-term memory, gated recurrent unit, convolutional neural networks (CNNs), and temporal convolutional networks, which underwent systematic, 10-fold cross-validation to ensure model robustness.

Domain adaptation was carried out by fine-tuning the best-performing models from Committee S using Dataset T, adapting their learned features to the unique characteristics of the rare elements in Dataset T. Fine-tuning was carefully managed to limit the number of training epochs and used a protocol starting from the pretrained weights and biases of Committee S to minimize the risk of catastrophic forgetting. During this adaptation, simpler models that showed poor performance or instability, namely, the simple recurrent neural network (RNN) and one-layer CNN, were omitted from Committee S-T to ensure the reliability and stability of the adapted models. In the preprocessing stages, advanced normalization techniques were used to standardize input and output data. Different scaling techniques, including MinMaxScaler and RobustScaler, were applied to various parts of the data to maintain realistic conditions and variability in training. The performance of the adapted models in Committee S-T was then evaluated to confirm their improved predictive accuracy on Dataset T.

Following the establishment of Committee S-T, GA was again used as in the final step of the DASH loop to determine the doping configuration, specifically the position of dopant atoms leading to the lowest slab energy. This process was iterated 25 times to generate a batch of calculation results. The weighted (by R^2) average and variance of the committee members' predictions in Committee S-T were evaluated. The optimal structure was selected on the basis of the optimistic estimation strategy (*110*), specifically choosing the point in the repeated results that has the most negative value of committee expectation (*y*-axis value) * [1 + uncertainty percentage (*x*-axis value)].

Refined simulations were then performed on screened slabs to assess the theoretical OER activity and stability of the samples. DFT calculations assessed the slab energies and adsorption properties of OER intermediates [O, OH, and OOH for AEM; for LOM two configurations of OOH were considered; (*86*)] on doped RuO_2 surfaces under various surface-coverage scenarios based on a standard adsorption evolution mechanism. The assessment of stability involved calculations of Gibbs free energy changes for water adsorption on the surfaces (ΔG^*_{H2O}), the dissolution potentials of surface Ru (U_{diss}), and the formation energies of O vacancies (ΔG_{VO}). These calculations were conducted both in vacuum and implicit solvent field. Additionally, free energy barriers for RuO_4 species leaching ($\Delta G_{Diss\text{-}RuO4}$) were evaluated under LOM conditions. For explicit solvent field, the oxide slabs were further expanded with 100 H_2O molecules placed on the surface. AIMD was run for 2 ps for the relaxation of surface explicit H_2O layer. OER reaction pathway simulations under clean surface scenario (no coverage) were then conducted with both oxide slab and relaxed surface water layer fixed. Explicit solvation effects on $\Delta G_{Diss\text{-}RuO4}$ were further investigated using a 200-fs equilibrium sequence to extract distributions of the energies. Secondary reproducibility details are provided in Supplementary Note 6.

Supplementary Materials

This PDF file includes:

Figs. S1 to S43
Tables S1 to S4
Supplementary Discussion S1 to S3
References

REFERENCES AND NOTES

1. Y. Wang, Y. H. Pang, H. Xu, A. Martinez, K. S. Chen, PEM Fuel cell and electrolysis cell technologies and hydrogen infrastructure development - A review. *Energy Environ. Sci.* **15**, 2288–2328 (2022).
2. Z. C. Chen, L. Guo, L. Pan, T. Q. Yan, Z. X. He, Y. Li, C. X. Shi, Z. F. Huang, X. W. Zhang, J. J. Zou, Advances in oxygen evolution electrocatalysts for proton exchange membrane water electrolyzers. *Adv. Energy Mater.* **12**, 2103670 (2022).
3. T. Reier, H. N. Nong, D. Teschner, R. Schlögl, P. Strasser, Electrocatalytic oxygen evolution reaction in acidic environments–reaction mechanisms and catalysts. *Adv. Energy Mater.* **7**, 1601275 (2017).
4. L. An, C. Wei, M. Lu, H. W. Liu, Y. B. Chen, G. G. Scherer, A. C. Fisher, P. X. Xi, Z. C. J. Xu, C. H. Yan, Recent development of oxygen evolution electrocatalysts in acidic environment. *Adv. Mater.* **33**, e2006328 (2021).
5. Y. Y. Feng, S. Si, G. Deng, Z. X. Xu, Z. Pu, H. S. Hu, C. B. Wang, Copper-doped ruthenium oxide as highly efficient electrocatalysts for the evolution of oxygen in acidic media. *J. Alloys Compd.* **892**, 162113 (2022).
6. 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).
7. J. He, W. Q. Li, P. Xu, J. M. Sun, Tuning electron correlations of RuO_2 by co-doping of Mo and Ce for boosting electrocatalytic water oxidation in acidic media. *Appl. Catal. B Environ.* **298**, 120528 (2021).
8. T. Kwon, A. Yu, S. J. Kim, M. H. Kim, C. Lee, Y. Lee, Au-Ru alloy nanofibers as a highly stable and active bifunctional electrocatalyst for acidic water splitting. *Appl. Surf. Sci.* **563**, 150293 (2021).
9. W. Jin, H. B. Wu, W. Q. Cai, B. H. Jia, M. Batmunkh, Z. X. Wu, T. Y. Ma, Evolution of interfacial coupling interaction of Ni-Ru species for pH-universal water splitting. *Chem. Eng. J.* **426**, 130762 (2021).

10. Y. B. Wang, S. Hou, R. P. Ma, J. D. Jiang, Z. P. Shi, C. P. Liu, J. J. Ge, W. Xing, Modulating crystallinity and surface electronic structure of IrO_2 via gadolinium doping to promote acidic oxygen evolution. *ACS Sustain. Chem. Eng.* **9**, 10710–10716 (2021).
11. L. G. Li, P. T. Wang, Z. F. Cheng, Q. Shao, X. Q. Huang, One-dimensional iridium-based nanowires for efficient water electrooxidation and beyond. *Nano Res.* **15**, 1087–1093 (2022).
12. I. G. Kim, A. Lim, J. H. Jang, K. Y. Lee, I. W. Nah, S. Park, Leveraging metal alloy-hybrid support interaction to enhance oxygen evolution kinetics and stability in proton exchange membrane water electrolyzers. *J. Power Sources* **501**, 230002 (2021).
13. H. Hu, F. M. D. Kazim, Z. H. Ye, Y. H. Xie, Q. Zhang, K. G. Qu, J. X. Xu, W. W. Cai, S. L. Xiao, Z. H. Yang, Electronically delocalized Ir enables efficient and stable acidic water splitting. *J. Mater. Chem. A* **8**, 20168–20174 (2020).
14. Y. F. Sun, S. Dai, High-entropy materials for catalysis: A new frontier. *Sci. Adv.* **7**, eabg1600 (2021).
15. T. Loffler, A. Ludwig, J. Rossmeisl, W. Schuhmann, What makes high-entropy alloys exceptional electrocatalysts? *Angew. Chem. Int. Ed. Engl.* **60**, 26894–26903 (2021).
16. A. Amiri, R. Shahbazian-Yassar, Recent progress of high-entropy materials for energy storage and conversion. *J. Mater. Chem. A* **9**, 782–823 (2021).
17. C. Hu, K. Yue, J. Han, X. Liu, L. Liu, Q. Liu, Q. Kong, C. W. Pao, Z. Hu, K. Suenaga, D. Su, Q. Zhang, X. Wang, Y. Tan, X. Huang, Misoriented high-entropy iridium ruthenium oxide for acidic water splitting. *Sci. Adv.* **9**, eadf9144 (2023).
18. K. Miao, W. Jiang, Z. Chen, Y. Luo, D. Xiang, C. Wang, X. Kang, Hollow-structured and polyhedron-shaped high entropy oxide toward highly active and robust oxygen evolution reaction in a full pH range. *Adv. Mater.* **36**, e2308490 (2024).
19. H. Zhu, Z. Zhu, J. Hao, S. Sun, S. Lu, C. Wang, P. Ma, W. Dong, M. Du, High-entropy alloy stabilized active Ir for highly efficient acidic oxygen evolution. *Chem. Eng. J.* **431**, 133251 (2022).
20. R. Ding, Y. Chen, Z. Rui, K. Hua, Y. Wu, X. Li, X. Duan, J. Li, X. Wang, J. Liu, Machine learning utilized for the development of proton exchange membrane electrolyzers. *J. Power Sources* **556**, 232389 (2023).
21. M. Kim, Y. Kim, M. Y. Ha, E. Shin, S. J. Kwak, M. Park, I. D. Kim, W. B. Jung, W. B. Lee, Y. Kim, H. T. Jung, Exploring optimal water splitting bifunctional alloy catalyst by pareto active learning. *Adv. Mater.* **35**, e2211497 (2023).
22. B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, W. J. Yin, Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts. *Nat. Commun.* **11**, 3513 (2020).
23. W. T. Hong, R. E. Welsch, Y. Shao-Horn, Descriptors of oxygen-evolution activity for oxides: A statistical evaluation. *J. Phys. Chem. C* **120**, 78–86 (2015).
24. Y. M. Sun, H. B. Liao, J. R. Wang, B. Chen, S. N. Sun, S. J. H. Ong, S. B. Xi, C. Z. Diao, Y. H. Du, J. O. Wang, M. B. H. Breese, S. Z. Li, H. Zhang, Z. C. J. Xu, Covalency competition dominates the water oxidation structure-activity relationship on spinel oxides. *Nat. Catal.* **3**, 554–563 (2020).
25. S. Back, K. Tran, Z. W. Ulissi, Toward a design of active oxygen evolution catalysts: Insights from automated density functional theory calculations and machine learning. *ACS Catal.* **9**, 7651–7659 (2019).
26. N. J. van Eck, L. Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **84**, 523–538 (2010).
27. 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 Catal.* **13**, 13267–13281 (2023).
28. N. A. Tapan, M. E. Günay, R. Yildirim, Constructing global models from past publications to improve design and operating conditions for direct alcohol fuel cells. *Chem. Eng. Res. Des.* **105**, 162–170 (2016).
29. E. Can, R. Yildirim, Data mining in photocatalytic water splitting over perovskites literature for higher hydrogen production. *Appl. Catal. B Environ.* **242**, 267–283 (2019).
30. I. C. Man, H. Y. Su, F. Calle-Vallejo, H. A. Hansen, J. I. Martinez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Norskov, J. Rossmeisl, Universality in oxygen evolution electrocatalysis on oxide surfaces. *ChemCatChem* **3**, 1159–1165 (2011).
31. T. Reier, M. Oezaslan, P. Strasser, Electrocatalytic oxygen evolution reaction (OER) on Ru, Ir, and Pt catalysts: A comparative study of nanoparticles and bulk materials. *ACS Catal.* **2**, 1765–1772 (2012).
32. M. E. C. Pascuzzi, J. P. Hofmann, E. J. M. Hensen, Promoting oxygen evolution of IrO2 in acid electrolyte by Mn. *Electrochim. Acta* **366**, 137448 (2021).
33. C. Zheng, B. Huang, X. Liu, H. Wang, L. Guan, Mn-doped RuO 2 nanocrystals with abundant oxygen vacancies for enhanced oxygen evolution in acidic media. *Inorg. Chem. Front.* **11**, 1912–1922 (2024).
34. M. Ji, X. Yang, S. Chang, W. Chen, J. Wang, D. He, Y. Hu, Q. Deng, Y. Sun, B. Li, J. Xi, T. Yamada, J. Zhang, H. Xiao, C. Zhu, J. Li, Y. Li, RuO 2 clusters derived from bulk SrRuO 3: Robust catalyst for oxygen evolution reaction in acid. *Nano Res.* **15**, 1959–1965 (2022).
35. L. C. Seitz, C. F. Dickens, K. Nishio, Y. Hikita, J. Montoya, A. Doyle, C. Kirk, A. Vojvodic, H. Y. Hwang, J. K. Norskov, T. F. Jaramillo, A highly active and stable IrO x/SrIrO3 catalyst for the oxygen evolution reaction. *Science* **353**, 1011–1014 (2016).
36. S. Wold, K. Esbensen, P. Geladi, Principal component analysis. *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
37. L. van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
38. S. Kee, E. Del Castillo, G. Runger, Query-by-committee improvement with diversity and density in batch active learning. *Inf. Sci.* **454**, 401–418 (2018).
39. H. Aziz, B. de Keijzer, in *31st International Symposium on Theoretical Aspects of Computer Science (STACS)* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2014), pp. 99–111.
40. J. H. Friedman, B. E. Popescu, Predictive learning via rule ensembles. *Ann. Appl. Stat.* **2**, 916–954 (2008).
41. B. M. Greenwell, pdp: An R package for constructing partial dependence plots. *R J.* **9**, 421–436 (2017).
42. H. Sun, W. Jung, Recent advances in doped ruthenium oxides as high-efficiency electrocatalysts for the oxygen evolution reaction. *J. Mater. Chem. A* **9**, 15506–15521 (2021).
43. J. Bai, W. Zhou, J. Xu, P. Zhou, Y. Deng, M. Xiang, D. Xiang, Y. Su, RuO_2 catalysts for electrocatalytic oxygen evolution in acidic media: Mechanism, activity promotion strategy and research progress. *Molecules* **29**, 537 (2024).
44. J. Hao, Z. Zhuang, K. Cao, G. Gao, C. Wang, F. Lai, S. Lu, P. Ma, W. Dong, T. Liu, M. du, H. Zhu, Unraveling the electronegativity-dominated intermediate adsorption on high-entropy alloy electrocatalysts. *Nat. Commun.* **13**, 2662 (2022).
45. L. Zhao, S.-G. Park, B. Magyari-Köpe, Y. Nishi, Dopant selection rules for desired electronic structure and vacancy formation characteristics of TiO_2 resistive memory. *Appl. Phys. Lett.* **102**, 083506 (2013).
46. H.-W. Lv, H.-B. Zhao, X.-Y. Peng, Z.-G. Ye, Q.-B. Huang, X.-T. Yuan, D.-S. Li, Z. Jin, Rhenium-boosted electrocatalytic activity and durability of pyrolytic IrO_2 for acidic oxygen evolution. *Rare Metals* **43**, 6758–6764 (2024).
47. S. R. Ede, Z. Luo, Tuning the intrinsic catalytic activities of oxygen-evolution catalysts by doping: A comprehensive review. *J. Mater. Chem. A* **9**, 20131–20163 (2021).
48. G. H. Kim, Y. S. Park, J. Yang, M. J. Jang, J. Jeong, J.-H. Lee, H.-S. Park, Y. H. Park, S. M. Choi, J. Lee, Effects of annealing temperature on the oxygen evolution reaction activity of copper–cobalt oxide nanosheets. *Nanomaterials* **11**, 657 (2021).
49. L. G. M. Pettersson, A. Nilsson, A molecular perspective on the d-band model: Synergy between experiment and theory. *Top. Catal.* **57**, 2–13 (2014).
50. J. Hu, A. Al-Salihy, B. Zhang, S. Li, P. Xu, Mastering the D-band center of iron-series metal-based electrocatalysts for enhanced electrocatalytic water splitting. *Int. J. Mol. Sci.* **23**, 15405 (2022).
51. T. Naito, T. Shinagawa, T. Nishimoto, K. Takanabe, Recent advances in understanding oxygen evolution reaction mechanisms over iridium oxide. *Inorg. Chem. Front.* **8**, 2900–2917 (2021).
52. Z. Wu, L. Jiang, Q. Dong, Q. Gao, J. Cai, X. Cheng, Effect and mechanism of rare Earth and alkaline Earth metals on the high-temperature stability of activated alumina. *Front. Mat.* **10**, 1103590 (2023).
53. Y. Liu, H. Huang, L. Xue, J. Sun, X. Wang, P. Xiong, J. Zhu, Recent advances in the heteroatom doping of perovskite oxides for efficient electrocatalytic reactions. *Nanoscale* **13**, 19840–19856 (2021).
54. H. Sun, X. Xu, G. Chen, Z. Shao, Perovskite oxides as electrocatalysts for water electrolysis: From crystalline to amorphous. *Carbon Energy* **6**, e595 (2024).
55. 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. *Mater. Lett.* **83**, 84–87 (2012).
56. 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 $CoFe_2O_4$ nanoparticles. *Ionics* **30**, 6559–6574 (2024).
57. S. B. Attanayake, A. Chanda, T. Hulse, R. Das, M.-H. Phan, H. Srikanth, Competing magnetic interactions and field-induced metamagnetic transition in highly crystalline phase-tunable iron oxide nanorods. *Nanomaterials* **13**, 1340 (2023).
58. Y. Wang, H. Pan, Z. Liu, P. Kang, $RuO_2/CoMo_2O_x$ catalyst with low ruthenium loading for long-term acidic oxygen evolution. *Trans. Tianjin Univ.* **30**, 395–405 (2024).
59. X. Zhang, Y. Liu, X. Ma, X. Liu, R. Zhang, Y. Wang, Metal–support interaction of carbon–based electrocatalysts for oxygen evolution reaction. *Nanoenergy Adv.* **3**, 48–72 (2023).
60. D. Coertzen, R. J. Kriek, P. B. J. Levecque, A. Falch, Vulcan carbon as support for sputtered oxygen evolution electrocatalysts. *Electrocatalysis* **10**, 604–612 (2019).
61. 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 (2024).
62. X. Teng, J. Gao, Z. Yang, X. Liang, X. Wu, J. Yun, J. Zhang, RuO_2@IrO_2/C core-shell structure catalyst for efficient and durable acidic oxygen evolution. *Catalysts* **13**, 1456 (2023).
63. Z. J. Xu, T. D. Nguyen, G. G. Scherer, A facile synthesis of size-controllable IrO2 and RuO2 nanoparticles for the oxygen evolution reaction. *Electrocatalysis* **7**, 420–427 (2016).

64. 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. *Appl. Catal. B Environ.* **218**, 287–297 (2017).
65. I. Rodríguez-García, D. Galyamin, L. Pascual, P. Ferrer, M. A. Peña, D. Grinter, G. Held, M. Abdel Salam, M. Mokhtar, K. Narasimharao, M. Retuerto, S. Rojas, Enhanced stability of SrRuO3 mixed oxide via monovalent doping in $Sr_{1-x}K_xRuO_3$ for the oxygen evolution reaction. *J. Power Sources* **521**, 230950 (2022).
66. B.-J. Kim, D. F. Abbott, X. Cheng, E. Fabbri, M. Nachtegaal, F. Bozza, I. E. Castelli, D. Lebedev, R. Schäublin, C. Copéret, T. Graule, N. Marzari, T. J. Schmidt, Unraveling thermodynamics, stability, and oxygen evolution activity of strontium ruthenium perovskite oxide. *ACS Catal.* **7**, 3245–3256 (2017).
67. M. Retuerto, L. Pascual, F. Calle-Vallejo, P. Ferrer, D. Gianolio, A. G. Pereira, A. Garcia, J. Torrero, M. T. Fernandez-Diaz, P. Bencok, M. A. Pena, J. L. G. Fierro, S. Rojas, Na-doped ruthenium perovskite electrocatalysts with improved oxygen evolution activity and durability in acidic media. *Nat. Commun.* **10**, 2041 (2019).
68. M. Ji, X. Yang, S. Chang, W. Chen, J. Wang, D. He, Y. Hu, Q. Deng, Y. Sun, B. Li, J. Xi, T. Yamada, J. Zhang, H. Xiao, C. Zhu, J. Li, Y. Li, RuO_2 clusters derived from bulk $SrRuO_3$: Robust catalyst for oxygen evolution reaction in acid. *Nano Res.* **15**, 1959–1965 (2022).
69. S. Chen, H. Huang, P. Jiang, K. Yang, J. Diao, S. Gong, S. Liu, M. Huang, H. Wang, Q. Chen, Mn-Doped RuO_2 nanocrystals as highly active electrocatalysts for enhanced oxygen evolution in acidic media. *ACS Catal.* **10**, 1152–1160 (2020).
70. K. Wang, Y. Wang, B. Yang, Z. Li, X. Qin, Q. Zhang, L. Lei, M. Qiu, G. Wu, Y. Hou, Highly active ruthenium sites stabilized by modulating electron-feeding for sustainable acidic oxygen-evolution electrocatalysis. *Energy Environ. Sci.* **15**, 2356–2365 (2022).
71. W. Li, R. Liu, G. Yu, X. Chen, S. Yan, S. Ren, J. Chen, W. Chen, C. Wang, X. Lu, Rationally construction of Mn-doped RuO_2 nanofibers for high-activity and stable alkaline ampere-level current density overall water splitting. *Small* **20**, e2307164 (2024).
72. W. Xu, H. Huang, X. Wu, Y. Yuan, Y. Liu, Z. Wang, D. Zhang, Y. Qin, J. Lai, L. Wang, Mn-doped Ru/RuO_2 nanoclusters@CNT with strong metal-support interaction for efficient water splitting in acidic media. *Compos. Part B Eng.* **242**, 110013 (2022).
73. M. Umer, S. Umer, M. Zafari, M. Ha, R. Anand, A. Hajibabaei, A. Abbas, G. Lee, K. S. Kim, Machine learning assisted high-throughput screening of transition metal single atom based superb hydrogen evolution electrocatalysts. *J. Mater. Chem. A* **10**, 6679–6689 (2022).
74. G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, A. Fazzio, From DFT to machine learning: Recent approaches to materials science–A review. *J. Phys. Mater.* **2**, 032001 (2019).
75. S. Gao, H. Zhen, B. Wen, J. Ma, X. Zhang, Exploring the physical origin of the electrocatalytic performance of an amorphous alloy catalyst via machine learning accelerated DFT study. *Nanoscale* **14**, 2660–2667 (2022).
76. Z. Li, D. Hoiem, Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 2935–2947 (2018).
77. G. T. K. K. Gunasooriya, J. K. Nørskov, Analysis of acid-stable and active oxides for the oxygen evolution reaction. *ACS Energy Lett.* **5**, 3778–3787 (2020).
78. A. Grimaud, O. Diaz-Morales, B. Han, W. T. Hong, Y. L. Lee, L. Giordano, K. A. Stoerzinger, M. T. M. Koper, Y. Shao-Horn, Activating lattice oxygen redox reactions in metal oxides to catalyse oxygen evolution. *Nat. Chem.* **9**, 457–465 (2017).
79. S. Xin, Y. Tang, B. Jia, Z. Zhang, C. Li, R. Bao, C. Li, J. Yi, J. Wang, T. Ma, Coupling adsorbed evolution and lattice oxygen mechanism in Fe-$Co(OH)_2$/Fe_2O_3 heterostructure for enhanced electrochemical water oxidation. *Adv. Funct. Mater.* **33**, 2305243 (2023).
80. S. Zhu, L. Song, Z. Xu, F. Chen, H. Tao, X. Tang, Y. Wang, Theoretical and experimental aspects of electrocatalysts for oxygen evolution reaction. *Chemistry* **30**, e202303672 (2024).
81. X. Liu, Z. He, M. Ajmal, C. Shi, R. Gao, L. Pan, Z.-F. Huang, X. Zhang, J.-J. Zou, Recent advances in the comprehension and regulation of lattice oxygen oxidation mechanism in oxygen evolution reaction. *Trans. Tianjin Univ.* **29**, 247–253 (2023).
82. M. Chatenet, B. G. Pollet, D. R. Dekel, F. Dionigi, J. Deseure, P. Millet, R. D. Braatz, M. Z. Bazant, M. Eikerling, I. Staffell, P. Balcombe, Y. Shao-Horn, H. Schafer, Water electrolysis: From textbook knowledge to the latest scientific strategies and industrial developments. *Chem. Soc. Rev.* **51**, 4583–4762 (2022).
83. J. Song, C. Wei, Z. F. Huang, C. Liu, L. Zeng, X. Wang, Z. J. Xu, A review on fundamentals for designing oxygen evolution electrocatalysts. *Chem. Soc. Rev.* **49**, 2196–2214 (2020).
84. S. Divanis, A. M. Frandsen, T. Kutlusoy, J. Rossmeisl, Lifting the discrepancy between experimental results and the theoretical predictions for the catalytic activity of RuO_2(110) towards oxygen evolution reaction. *Phys. Chem. Chem. Phys.* **23**, 19141–19145 (2021).
85. S. Divanis, T. Kutlusoy, I. M. Ingmer Boye, I. C. Man, J. Rossmeisl, Oxygen evolution reaction: A perspective on a decade of atomic scale simulations. *Chem. Sci.* **11**, 2943–2950 (2020).
86. Y. Wang, R. Yang, Y. Ding, B. Zhang, H. Li, B. Bai, M. Li, Y. Cui, J. Xiao, Z. S. Wu, Unraveling oxygen vacancy site mechanism of Rh-doped RuO_2 catalyst for long-lasting acidic water oxidation. *Nat. Commun.* **14**, 1412 (2023).
87. S. Y. Hao, H. Y. Sheng, M. Liu, J. Z. Huang, G. K. Zheng, F. Zhang, X. N. Liu, Z. W. Su, J. J. Hu, Y. Qian, L. N. Zhou, Y. He, B. Song, L. C. Lei, X. W. Zhang, S. Jin, Torsion strained iridium oxide for efficient acidic water oxidation in proton exchange membrane electrolyzers. *Nat. Nanotechnol.* **16**, 1371–1377 (2021).
88. Z. P. Shi, Y. Wang, J. Li, X. Wang, Y. B. Wang, Y. Li, W. L. Xu, Z. Jiang, C. P. Liu, W. Xing, J. J. Ge, Confined Ir single sites with triggered lattice oxygen redox: Toward boosted and sustained water oxidation catalysis. *Aust. Dent. J.* **5**, 2164–2176 (2021).
89. A. M. Harzandi, S. Shadman, A. S. Nissimagoudar, D. Y. Kim, H. D. Lim, J. H. Lee, M. G. Kim, H. Y. Jeong, Y. Kim, K. S. Kim, Ruthenium core–shell engineering with nickel single atoms for selective oxygen evolution via nondestructive mechanism. *Adv. Energy Mater.* **11**, 2003448 (2021).
90. D. A. Kuznetsov, M. A. Naeem, P. V. Kumar, P. M. Abdala, A. Fedorov, C. R. Muller, Tailoring lattice oxygen binding in ruthenium pyrochlores to enhance oxygen evolution activity. *J. Am. Chem. Soc.* **142**, 7883–7888 (2020).
91. N. Yao, H. Jia, J. Zhu, Z. Shi, H. Cong, J. Ge, W. Luo, Atomically dispersed Ru oxide catalyst with lattice oxygen participation for efficient acidic water oxidation. *Chem* **9**, 1882–1896 (2023).
92. L. Giordano, K. Akkiraju, R. Jacobs, D. Vivona, D. Morgan, Y. Shao-Horn, Electronic structure-based descriptors for oxide properties and functions. *Acc. Chem. Res.* **55**, 298–308 (2022).
93. J. Kim, P. C. Shih, K. C. Tsao, Y. T. Pan, X. Yin, C. J. Sun, H. Yang, High-performance pyrochlore-type yttrium ruthenate electrocatalyst for oxygen evolution reaction in acidic media. *J. Am. Chem. Soc.* **139**, 12076–12083 (2017).
94. M. A. Hubert, A. M. Patel, A. Gallo, Y. Liu, E. Valle, M. Ben-Naim, J. Sanchez, D. Sokaras, R. Sinclair, J. K. Nørskov, L. A. King, M. Bajdich, T. F. Jaramillo, Acidic oxygen evolution reaction activity–stability relationships in Ru-based pyrochlores. *ACS Catal.* **10**, 12182–12196 (2020).
95. R. Ovcharenko, E. Voloshina, J. Sauer, Water adsorption and O-defect formation on Fe_2O_3 (0001) surfaces. *Phys. Chem. Chem. Phys.* **18**, 25560–25568 (2016).
96. S. Goverapet Srinivasan, R. Shivaramaiah, P. R. Kent, A. G. Stack, R. Riman, A. Anderko, A. Navrotsky, V. S. Bryantsev, A comparative study of surface energies and water adsorption on Ce-bastnasite, La-bastnasite, and calcite via density functional theory and water adsorption calorimetry. *Phys. Chem. Chem. Phys.* **19**, 7820–7832 (2017).
97. R. R. Rao, M. J. Kolb, J. Hwang, A. F. Pedersen, A. Mehta, H. You, K. A. Stoerzinger, Z. Feng, H. Zhou, H. Bluhm, L. Giordano, I. E. L. Stephens, Y. Shao-Horn, Surface orientation dependent water dissociation on rutile ruthenium dioxide. *J. Phys. Chem. C* **122**, 17802–17811 (2018).
98. M. Ha, D. Y. Kim, M. Umer, V. Gladkikh, C. W. Myung, K. S. Kim, Tuning metal single atoms embedded in NxCy moieties toward high-performance electrocatalysis. *Energy Environ. Sci.* **14**, 3455–3468 (2021).
99. J. Greeley, J. K. Nørskov, Electrochemical dissolution of surface alloys in acids: Thermodynamic trends from first-principles calculations. *Electrochim. Acta* **52**, 5829–5836 (2007).
100. S. Hao, M. Liu, J. Pan, X. Liu, X. Tan, N. Xu, Y. He, L. Lei, X. Zhang, Dopants fixation of Ruthenium for boosting acidic oxygen evolution stability and activity. *Nat. Commun.* **11**, 5368 (2020).
101. A. Li, S. Kong, C. Guo, H. Ooka, K. Adachi, D. Hashizume, Q. Jiang, H. Han, J. Xiao, R. Nakamura, Enhancing the stability of cobalt spinel oxide towards sustainable oxygen evolution in acid. *Nat. Catal.* **5**, 109–118 (2022).
102. H.-Y. Su, K. Sun, DFT study of the stability of oxygen vacancy in cubic ABO_3 perovskites. *J. Mater. Sci.* **50**, 1701–1709 (2014).
103. X. Ping, Y. Liu, L. Zheng, Y. Song, L. Guo, S. Chen, Z. Wei, Locking the lattice oxygen in RuO_2 to stabilize highly active Ru sites in acidic water oxidation. *Nat. Commun.* **15**, 2501 (2024).
104. R. Ding, R. Wang, Y. Q. Ding, W. J. Yin, Y. D. Liu, J. Li, J. G. Liu, Designing AI-aided analysis and prediction models for nonprecious metal electrocatalyst-based proton-exchange membrane fuel cells. *Angew. Chem. Int. Ed. Engl.* **59**, 19175–19183 (2020).
105. R. Rodríguez, M. Pastorini, L. Etcheverry, C. Chreties, M. Fossati, A. Castro, A. Gorgoglione, Water-quality data imputation with a high percentage of missing values: A machine learning approach. *Sustainability* **13**, 6318 (2021).
106. C. Li, X. Ren, G. Zhao, Machine-learning-based imputation method for filling missing values in ground meteorological observation data. *Algorithms* **16**, 422 (2023).
107. A. A. Dharmasaputro, N. M. Fauzan, M. Kallista, I. P. D. Wibawa, P. D. Kusuma, in *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)* (IEEE, 2022), pp. 140–145.
108. A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, E. D. Cubuk, Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
109. N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, G. Ceder, An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86–91 (2023).
110. H. Grosskreutz, S. Rüping, S. Wrobel, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Springer, 2008), pp. 440–456.

111. X. Miao, Z. Peng, L. Shi, S. Zhou, Insulating high-entropy ruthenium oxide as a highly efficient oxygen-evolving electrocatalyst in acid. *ACS Catal.* **13**, 3983–3989 (2023).
112. J. Kwon, S. Sun, S. Choi, K. Lee, S. Jo, K. Park, Y. K. Kim, H. B. Park, H. Y. Park, J. H. Jang, H. Han, U. Paik, T. Song, Tailored electronic structure of Ir in high entropy alloy for highly active and durable bifunctional electrocatalyst for water splitting under an acidic environment. *Adv. Mater.* **35**, e2300091 (2023).
113. L. Yao, F. Zhang, S. Yang, H. Zhang, Y. Li, C. Yang, H. Yang, Q. Cheng, Sub-2 nm IrRuNiMoCo high-entropy alloy with iridium-rich medium-entropy oxide shell to boost acidic oxygen evolution. *Adv. Mater.* **36**, e2314049 (2024).
114. S. Jo, M. C. Kim, K. B. Lee, H. Choi, L. Zhang, J. I. Sohn, Nonprecious high-entropy chalcogenide glasses-based electrocatalysts for efficient and stable acidic oxygen evolution reaction in proton exchange membrane water electrolysis. *Adv. Energy Mater.* **13**, 2301420 (2023).
115. S.-Q. Chang, C.-C. Cheng, P.-Y. Cheng, C.-L. Huang, S.-Y. Lu, Pulse electrodeposited FeCoNiMnW high entropy alloys as efficient and stable bifunctional electrocatalysts for acidic water splitting. *Chem. Eng. J.* **446**, 137452 (2022).
116. D. Zhang, Y. Shi, X. Chen, J. Lai, B. Huang, L. Wang, High-entropy alloy metallene for highly efficient overall water splitting in acidic media. *Chin. J. Catal.* **45**, 174–183 (2023).
117. Y. Yu, H. Li, J. Liu, W. Xu, D. Zhang, J. Xiong, B. Li, A. O. Omelchuk, J. Lai, L. Wang, High entropy stabilizing lattice oxygen participation of Ru- based oxides in acidic water oxidation. *J. Mater. Chem. A* **10**, 21260–21265 (2022).
118. A. L. Maulana, P. C. Chen, Z. Shi, Y. Yang, C. Lizandara-Pueyo, F. Seeler, H. D. Abruna, D. Muller, K. Schierle-Arndt, P. Yang, Understanding the structural evolution of IrFeCoNiCu high-entropy alloy nanoparticles under the acidic oxygen evolution reaction. *Nano Lett.* **23**, 6637–6644 (2023).
119. J. Ni, Z. Shi, Y. Wang, J. Yang, H. Wu, P. Wang, K. Li, M. Xiao, C. Liu, W. Xing, Suppressing the lattice oxygen diffusion via high-entropy oxide construction towards stabilized acidic water oxidation. *Nano Res.* **17**, 1107–1113 (2024).
120. Z. Jin, J. Lv, H. Jia, W. Liu, H. Li, Z. Chen, X. Lin, G. Xie, X. Liu, S. Sun, H. J. Qiu, Nanoporous Al-Ni-Co-Ir-Mo high-entropy alloy for record-high water splitting activity in acidic environments. *Small* **15**, e1904180 (2019).
121. P. Keller, H.-H. Strehblow, XPS investigations of electrochemically formed passive layers on Fe/Cr-alloys in 0.5 M H_2SO_4. *Corros. Sci.* **46**, 1939–1952 (2004).
122. H. Su, M. A. Soldatov, V. Roldugin, Q. Liu, Platinum single-atom catalyst with self-adjustable valence state for large-current-density acidic water oxidation. *eScience* **2**, 102–109 (2022).
123. R. Ding, Y. Chen, Z. Rui, K. Hua, Y. Wu, X. Li, X. Duan, X. Wang, J. Li, J. Liu, Guiding the optimization of membrane electrode assembly in a proton exchange membrane water electrolyzer by machine learning modeling and black-box interpretation. *ACS Sustain. Chem. Eng.* **10**, 4561–4578 (2022).
124. A. Tan, Y. Zhang, X. Shi, C. Ju, P. Liu, T. Yang, J. Liu, The poisoning effects of Ti-ion from porous transport layers on the membrane electrode assembly of proton exchange membrane water electrolyzers. *Chem. Eng. J.* **471**, 144624 (2023).
125. C.-A. Tao, J.-F. Wang, Synthesis of metal organic frameworks by ball-milling. *Crystals* **11**, 15 (2021).
126. Y. Ye, H. Chen, Y. Ye, H. Zhang, J. Xu, L. Wang, L. Mo, Silica-supported copper (II) oxide cluster via ball milling method for catalytic combustion of ethyl acetate. *Catalysts* **12**, 497 (2022).
127. T. B. Ferriday, S. Nuggehalli Sampathkumar, P. H. Middleton, J. Van Herle, M. L. Kolhe, How acid washing nickel foam substrates improves the efficiency of the alkaline hydrogen evolution reaction. *Energies* **16**, 2083 (2023).
128. P. M. Bacirhonde, N. Y. Dzade, C. Chalony, J. Park, E.-S. Jeong, E. O. Afranie, S. Lee, C. S. Kim, D.-H. Kim, C. H. Park, Reduction of transition-metal columbite-tantalite as a highly efficient electrocatalyst for water splitting. *ACS Appl. Mater. Interfaces* **14**, 15090–15102 (2022).
129. B. Huang, H. Xu, N. Jiang, M. Wang, J. Huang, L. Guan, Tensile-strained RuO_2 loaded on antimony-tin oxide by fast quenching for proton-exchange membrane water electrolyzer. *Adv. Sci.* **9**, e2201654 (2022).
130. T. Niyitanga, H. Kim, Hematite-nickel oxide/carbon nanotube composite catalyst for oxygen evolution reaction. *Mater. Chem. Phys.* **275**, 125266 (2022).
131. W. Zhu, X. Song, F. Liao, H. Huang, Q. Shao, K. Feng, Y. Zhou, M. Ma, J. Wu, H. Yang, H. Yang, M. Wang, J. Shi, J. Zhong, T. Cheng, M. Shao, Y. Liu, Z. Kang, Stable and oxidative charged Ru enhance the acidic oxygen evolution reaction activity in two-dimensional ruthenium-iridium oxide. *Nat. Commun.* **14**, 5365 (2023).
132. X. Wu, C. Lin, W. Hu, C. Fu, L. Tan, H. Wang, F. Meharban, X. Pan, P. Fu, H. D. Um, Q. Xiao, X. Li, M. Yamauchi, W. Luo, Pluralistic electronic structure modulation of ruthenium oxide for enhanced acidic water electrolysis. *Small Struct.* **5**, 2300518 (2024).
133. Y. Wu, T. Xia, L. Yang, F. Guo, W. Jiang, J. Lang, Y. Ma, J. Feng, G. Che, C. Liu, Construction of an oxygen vacancy-enriched triple perovskite oxide electrocatalyst for efficient and stable oxygen evolution in acidic media. *Inorg. Chem. Front.* **11**, 6387–6395 (2024).
134. D. Zhang, M. Li, X. Yong, H. Song, G. I. N. Waterhouse, Y. Yi, B. Xue, D. Zhang, B. Liu, S. Lu, Construction of Zn-doped RuO2 nanowires for efficient and stable water oxidation in acidic media. *Nat. Commun.* **14**, 2517 (2023).
135. B. Huang, Y. Cui, X. Liu, C. Zheng, H. Wang, L. Guan, Dense-packed RuO_2 nanorods with in situ generated metal vacancies loaded on SnO_2 nanocubes for proton exchange membrane water electrolyzer with ultra-low noble metal loading. *Small* **19**, e2301516 (2023).
136. X. Wang, X. Wan, X. Qin, C. Chen, X. Qian, Y. Guo, Q. Xu, W.-B. Cai, H. Yang, K. Jiang, Electronic structure modulation of RuO_2 by TiO_2 enriched with oxygen vacancies to boost acidic O_2 evolution. *ACS Catal.* **12**, 9437–9445 (2022).
137. Y. Wang, X. Lei, B. Zhang, B. Bai, P. Das, T. Azam, J. Xiao, Z. S. Wu, Breaking the Ru—O—Ru symmetry of a RuO_2 catalyst for sustainable acidic water oxidation. *Angew. Chem.* **136**, e202316903 (2024).
138. H. Jin, X. Liu, P. An, C. Tang, H. Yu, Q. Zhang, H.-J. Peng, L. Gu, Y. Zheng, T. Song, K. Davey, U. Paik, J. Dong, S. Z. Qiao, Dynamic rhenium dopant boosts ruthenium oxide for durable oxygen evolution. *Nat. Commun.* **14**, 354 (2023).
139. X. Wu, W. Han, S. Hao, Y. He, L. Lei, X. Zhang, Synergistically optimizing the electrocatalytic performance of IrO 2 with double doping and bi-directional strains for acidic oxygen evolution reaction. *Cat. Sci. Technol.* **14**, 4599–4607 (2024).
140. D. Chen, R. Yu, K. Yu, R. Lu, H. Zhao, J. Jiao, Y. Yao, J. Zhu, J. Wu, S. Mu, Bicontinuous RuO_2 nanoreactors for acidic water oxidation. *Nat. Commun.* **15**, 3928 (2024).
141. B. An, X. Li, Y. Lin, F. Shang, H. He, H. Cai, X. Zeng, W. Wang, S. Yang, B. Wang, Strontium doping RuO2 electrocatalyst with abundant oxygen vacancies for boosting OER performance. *Inorganic Chem Front* 11, 8935–8944 (2024),.
142. H. Liu, Z. Zhang, J. Fang, M. Li, M. G. Sendeku, X. Wang, H. Wu, Y. Li, J. Ge, Z. Zhuang, D. Zhou, Y. Kuang, X. Sun, Eliminating over-oxidation of ruthenium oxides by niobium for highly stable electrocatalytic oxygen evolution in acidic media. *Aust. Dent. J.* **7**, 558–573 (2023).
143. H. Huang, H. Kim, A. Lee, S. Kim, W.-G. Lim, C.-Y. Park, S. Kim, S.-K. Kim, J. Lee, Structure engineering defective and mass transfer-enhanced RuO_2 nanosheets for proton exchange membrane water electrolyzer. *Nano Energy* **88**, 106276 (2021).
144. H. Lv, S. Wang, J. Li, C. Shao, W. Zhou, X. Shen, M. Xue, C. Zhang, Self-assembled RuO_2@ IrOx core-shell nanocomposite as high efficient anode catalyst for PEM water electrolyzer. *Appl. Surf. Sci.* **514**, 145943 (2020).
145. G. Zhao, W. Guo, M. Shan, Y. Fang, G. Wang, M. Gao, Y. Liu, H. Pan, W. Sun, Metallic Ru-Ru interaction in ruthenium oxide enabling durable proton exchange membrane water electrolysis. *Adv. Mater.* **36**, e2404213 (2024).
146. Y. Liu, M. Zhang, C. Zhang, H. Zhang, H. Wang, An IrRuO x catalyst supported by oxygen-vacant Ta oxide for the oxygen evolution reaction and proton exchange membrane water electrolysis. *Nanoscale* **16**, 9382–9391 (2024).
147. Y. Dong, Y. Li, Y. Lin, A. Chen, M. Deng, L. Zhang, Z. Tian, L. Chen, Atomically dispersed Ru doped on IrO x sub-nanoclusters for enhanced oxygen evolution reaction in acidic media. *J. Mater. Chem. A* **12**, 21905–21911 (2024).
148. F. Ospina-Acevedo, L. A. Albiter, K. O. Bailey, J. F. Godínez-Salomón, C. P. Rhodes, P. B. Balbuena, Catalytic Activity and Electrochemical Stability of Ru_{1-x} M_x O_2 (M= Zr, Nb, Ta): Computational and experimental study of the oxygen evolution reaction. *ACS Appl. Mater. Interfaces* **16**, 16373–16398 (2024).
149. W. Ren, K. Wang, D. Lu, C. Xu, High-stability RuO_2/MoO_3 electrocatalyst for the oxygen evolution reaction in proton-exchange-membrane water Electrolysis. *ACS Appl. Energy Mater* **6**, 12573–12578 (2023).
150. D. Galyamin, J. Torrero, I. Rodríguez, M. J. Kolb, P. Ferrer, L. Pascual, M. A. Salam, D. Gianolio, V. Celorrio, M. Mokhtar, D. Garcia Sanchez, A. S. Gago, K. A. Friedrich, M. A. Peña, J. A. Alonso, F. Calle-Vallejo, M. Retuerto, S. Rojas, Active and durable R_2MnRuO_7 pyrochlores with low Ru content for acidic oxygen evolution. *Nat. Commun.* **14**, 2010 (2023).
151. J. Wang, Y. Zhang, Y. Wang, J. Cho, T.-S. Chan, Y. Ha, S.-C. Haw, C.-W. Kao, Z. Wang, J. Lei, M. Ju, J. Tang, T. Liu, S. Zhao, Y. Dai, A. Baron-Wiechec, F. R. Chen, W. Wang, C. H. Choi, Z. Shao, M. Ni, Heterostructure boosts a noble-metal-free oxygen-evolving electrocatalyst in acid. *Energy Environ. Sci.* **17**, 5972–5983 (2024).
152. J. Y. Lim, G. Rahman, S. Y. Chae, K. Y. Lee, C. S. Kim, O. S. Joo, Highly stable RuO_2/SnO_2 nanocomposites as anode electrocatalysts in a PEM water electrolysis cell. *Int. J. Energy Res.* **38**, 875–883 (2014).
153. S. Kong, A. Li, J. Long, K. Adachi, D. Hashizume, Q. Jiang, K. Fushimi, H. Ooka, J. Xiao, R. Nakamura, Acid-stable manganese oxides for proton exchange membrane water electrolysis. *Nat. Catal.* **7**, 252–261 (2024).
154. J. S. Yoo, X. Rong, Y. Liu, A. M. Kolpak, Role of lattice oxygen participation in understanding trends in the oxygen evolution reaction on perovskites. *ACS Catal.* **8**, 4628–4636 (2018).
155. R. Ding, Y. Chen, P. Chen, R. Wang, J. Wang, Y. Ding, W. Yin, Y. Liu, J. Li, J. Liu, Machine learning-guided discovery of underlying decisive factors and new mechanisms for the design of nonprecious metal electrocatalysts. *ACS Catal.* **11**, 9798–9808 (2021).

Acknowledgments
Funding: This project is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences program, at the University of Chicago (R.D.); the National Key Research and Development Program of China (2021YFB4000100) (J.L.); the National Natural Science Foundation of China (U23B2075, 52272039, and 51972168) (X.W.); the Research Grant Council of Hong Kong Special Region (16308420) (M.S.); part of the computational resources used by this work are provided by the University of Chicago's Research Computing Center and the High-Performance Computing Center of Nanjing University, China; and Hong Kong Research Grant Council (C6011-20GF and JLFS/P-602/24) and Guangzhou Science and Technology Bureau (2024A03J0609) (M.S.). **Author contributions:** Conceptualization: R.D. Experiments: R.D. and K.H. Characterization: R.D. and K.H. Data analysis: R.D. and X.Z. Visualization: R.D. ML methods: R.D. and Y.C. ML modeling: R.D. DFT simulation: R.D. Drafting: R.D., Y.C., and J.C. Resources: J.L., X.W., M.S., Y.C., and J.C. Supervision: Y.C. and J.C. Manuscript reviewing: J.L., X.W., X.Z., M.S., Y.C., and J.C. Editing: Y.C. and J.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. In line with the principles of open access and knowledge-sharing in the ML community, all ML training and data mining scripts used in this study, datasets extracted from the literature for data mining and initial ML committee training, high-throughput experimental/DFT computational data, characterization results of the samples, and other supplementary data mining results and trivial detailed discussions (Supplementary Notes 1 to 8) are publicly accessible on the Dryad repository (https://doi.org/10.5061/dryad.nk98sf83g) and mirrored on GitHub (https://github.com/ruiding-uchicago/DASH) for interested readers to review in detail.

Submitted 21 July 2024
Accepted 28 February 2025
Published 4 April 2025
10.1126/sciadv.adr9038