ScieRxiv
Research Highlight

Ariadne's thread: Machine learning navigates the labyrinth of cross-material catalyst discovery

Xin Wu[1,3], Xinzhang Lin[1,2,*], Harun Tüysüz[1,2,*]

[1]IMDEA Materials Institute, Getafe 28906, Madrid, Spain.
[2]Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr 45470, Germany.
[3]Universidad Politécnica de Madrid, Madrid 28040, Spain.

Correspondence to: Dr. Xinzhang Lin, Prof. Harun Tüysüz, IMDEA Materials Institute, Getafe 28906, Madrid, Spain. E-mail: xinzhang.lin@imdeamaterials.org; harun.tuysuz@imdea.org.

Received: 12 July 2026 | Approved: 07 August 2026 | Online: 07 August 2026

Keywords: Machine learning, catalyst discovery, single-atom catalyst, perovskite oxide, oxygen evolution reaction

MAIN TEXT

Catalysts are indispensable pillars of human societal development, underpinning a broad range of applications, from large-scale chemical manufacturing to sustainable energy technologies. However, the discovery of high-performance catalysts resembles a journey through a labyrinth, where each turn holds a slim chance of a breakthrough but a much greater possibility of ending in a blind alley. Traditionally, catalyst discovery has relied on laborious and time-consuming trial-and-error experiments involving repeated synthesis and performance evaluation. Theoretical approaches have been developed to guide catalyst design, but they remain challenging due to the intrinsic complexity of catalysts and catalytic reactions, demanding accurate computational models and meaningful descriptors to obtain reliable predictions[1]. Machine learning has emerged as

a powerful tool for catalyst discovery within specific material families[2-4], but achieving robust generalization across chemically diverse catalyst classes within a unified framework remains a huge challenge[5]. A major obstacle lies in the fragmented nature of experimental datasets, which often differ in descriptors, synthetic protocols and evaluation conditions. Bridging these disparate data spaces would allow machine learning models to extract transferable chemical knowledge and uncover promising catalysts beyond conventional design intuition.

This challenge is particularly evident in the oxygen evolution reaction (OER), the kinetically sluggish four-electron process that limits the efficiency of water electrolysis for green hydrogen production[6-9], Among alkaline OER catalysts, multimetallic perovskite oxides[9] and single-atom catalysts (SACs)[6-7] are particularly promising material platforms, yet they belong to distinct material families with little overlap in design strategies. Recently, Moon *et al.* reported in *Nature Materials* that deep machine learning can bridge this gap[10]. By integrating experimental datasets from SACs on carbon and bulk perovskite oxides into a unified framework, they developed a crossbreeding neural network (CBNN) capable of learning shared chemical descriptors across both catalyst families. The trained model successfully predicted OER activity in a previously unexplored catalyst class: SACs on perovskite oxides [Figure 1A], closely matching experimental activity trends and identifying a high-performing multimetallic catalyst beyond the original training space.

The conceptual advance of the CBNN lies in how it enables transferable learning across fundamentally different catalyst families. Two parallel network branches process surface and bulk structural information before integrating them into a unified predictive framework [Figure 1B]. A key innovation is the co- descriptor selection pipeline, which combines statistical independence testing, correlation analysis and natural language processing to identify five descriptors shared by both material families: oxidation state, ionic radius, valence d-electron count, Pauling electronegativity and metal coordination number. Rather than forcing disparate datasets into a common descriptor space, this strategy distilled a shared chemical language that enabled reliable cross-material prediction. Using this shared descriptor set, the CBNN accurately reproduced the experimentally observed activity ranking of 14 monometallic SACs supported on two representative perovskite oxides [Figure 1C]. This study demonstrates that chemically

meaningful knowledge can be transferred across material boundaries, allowing machine learning to uncover catalyst design principles beyond the scope of any individual experimental dataset.

Building on the accurate prediction of monometallic catalysts, the CBNN was subsequently applied to the more challenging task of discovering multimetallic SAC catalysts[8]. By exploring more than 8,000 compositions comprising seven metal species on a perovskite support, the model identified a multimetallic SAC containing W, Mo, Ru and Rh as the most active candidate. Experimental validation confirmed that the multimetallic SAC not only outperformed all 14 monometallic SACs on perovskite oxides but also surpassed every catalyst included in the original training datasets, including bulk perovskite oxides and SACs on carbon. Remarkably, this catalyst belonged to a previously untrained material class, demonstrating that the CBNN can extend catalyst discovery beyond the material groups included in the training datasets. Explainable machine learning further provided insight into the origin of this exceptional performance. Feature-importance analysis identified the relative contributions of the five shared co-descriptors to OER activity [Figure 1D], while activation mapping revealed increasingly favorable activation values as multiple metal species were combined, providing computational evidence that the superior activity of the multimetallic SAC arises from genuine synergistic interactions rather than the simple addition of individual elemental effects.

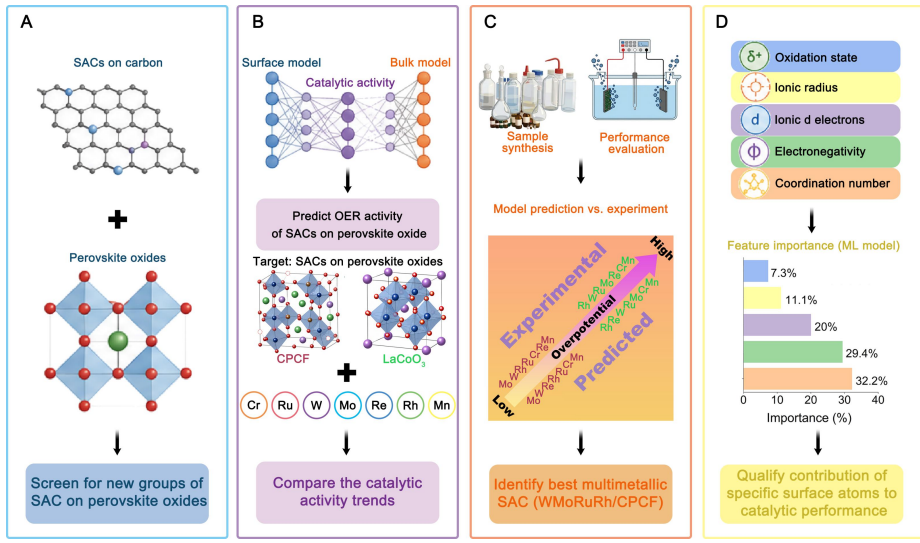

Figure 1. Cross-material catalyst discovery via deep machine learning. (A) Schematic

illustration of the catalyst screening strategy, in which structural information from SACs on carbon and bulk perovskite oxides is integrated to conduct the prediction of promising SACs on perovskite oxides; (B) Workflow of the machine learning model. Surface structures are represented by image-based descriptors, whereas bulk perovskite structures are encoded as graph-based descriptors. The combined structural information is employed to predict the OER activity of SACs on perovskite oxides and compare the catalytic activity trends of different SACs; (C) Experimental validation of the machine learning predictions through catalyst synthesis and electrochemical evaluation. The experimentally measured OER activity trends of the monometallic SACs show good agreement with the machine learning predictions, and the optimal multimetallic SAC supported on CPCF was subsequently identified; (D) Interpretation of the ML model by quantifying the relative importance of 5 key physicochemical descriptors. This figure is reproduced with permission from Moon *et al.*[10] Abbreviations: SAC: single-atom catalyst; OER: oxygen evolution reaction; CPCF: $Ca_{0.8}Pr_{0.2}Co_{0.8}Fe_{0.2}O_{3-\delta}$.

This work has taken an important step towards deep machine learning for catalyst discovery by demonstrating that experimental knowledge from chemically distinct catalyst families can be integrated into a unified predictive framework. Beyond alkaline OER, the CBNN model was also applied to acidic OER and alkaline hydrogen evolution reaction, suggesting that the framework could be extended to other electrocatalytic systems when structurally relevant training data are available. However, as with all data-driven approaches, the effectiveness of this framework depends critically on the quality and diversity of available experimental datasets, and its predictive robustness could be further improved by incorporating data from diverse synthesis methods and operating conditions. Overall, this work shifts machine learning from interpolation within isolated chemical families towards knowledge transfer across them, opening opportunities to discover catalysts beyond predefined chemical boundaries. Like Ariadne's thread through the labyrinth, the CBNN links previously isolated catalyst families, providing a unified framework for cross-material catalyst discovery.

DECLARATIONS

Authors' contributions

Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Wu X, Lin X, Tüysüz H.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the authors used AI and AI-assisted tools solely to improve the language, grammar, and readability of the text.

Financial support and sponsorship

This work was supported by the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101264623.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Orouji, N.; Bennett, J. A.; Canty, R. B.; et al. Autonomous catalysis research with human-AI-robot collaboration. *Nat. Catal.* **2025**, *8*, 1135-45.[DOI: 10.1038/s41929-025-01430-6]

2. Xin, H.; Kitchin, J. R.; López, N.; et al. Roadmap for transforming heterogeneous catalysis with artificial intelligence. *Nat. Catal.* **2026**, *9*, 102-11.[DOI: 10.1038/s41929-026-01479-x]

3. Zhang, M.; Wan, Y.; Lin, Y. Machine learning identifies efficient Ru-based electrocatalysts. *Chem. Synth.* **2025**, *5*, 11.[DOI: 10.20517/cs.2024.74]

4. Miyazaki, R.; Belthle, K. S.; Tuysuz, H.; et al. Materials genes of CO_2 hydrogenation on supported cobalt catalysts: An artificial intelligence approach integrating theoretical and experimental data. *J. Am. Chem. Soc.* **2024**, *146*, 5433-44.[DOI: 10.1021/jacs.3c12978]

5. Rivera-Arrieta, H. I.; Foppa, L. Rules describing CO_2 activation on single-atom alloys from DFT-meta-GGA calculations and artificial intelligence. *ACS Catal.* **2025**, *15*, 2916-26.[DOI: 10.1021/acscatal.4c06006]

6. Lin, X.; Xu, J.; Peng, Z. Atomically dispersed catalysts toward the oxygen evolution reaction in electrochemical water splitting: from catalyst design, performance to catalytic mechanism. *Next Sustain.* **2024**, *3*, 100023.[DOI: 10.1016/j.nxsust.2023.100023]

7. Kumar, A.; Gil-Sepulcre, M.; Fandré, J. P.; et al. Regulating local coordination sphere of Ir single atoms at the atomic interface for efficient oxygen evolution reaction. *J. Am. Chem. Soc.* **2024,** *146*, 32953-64.[DOI: 10.1021/jacs.4c08847]

8. Li, S.; Tang, D.; Jing, X. Metal-organic framework-based self-supported electrodes for oxygen evolution reaction. *Chem. Synth.* **2024**, *4*, 70.[DOI: 10.20517/cs.2024.26]

9. Sun, X.; Yuan, Y.; Liu, S.; et al. Recent advances in perovskite oxides for oxygen evolution reaction: Structures, mechanisms, and strategies for performance enhancement. *Adv. Funct. Mater.* **2025**, *35*, 2416705.[DOI: 10.1002/adfm.202416705]

10. Moon, J.; Yoo, S.; Shim, J.; et al. Cross-material catalyst discovery via deep learning. *Nat. Mater.* **2026**, Epub ahead of print[DOI: 10.1038/s41563-026-02622-6]