Research Highlight | Open Access

AI Agent

Tang *et al. AI Agent* **2026**, *2*, 35 **DOI:**10.20517/aiagent.2026.54

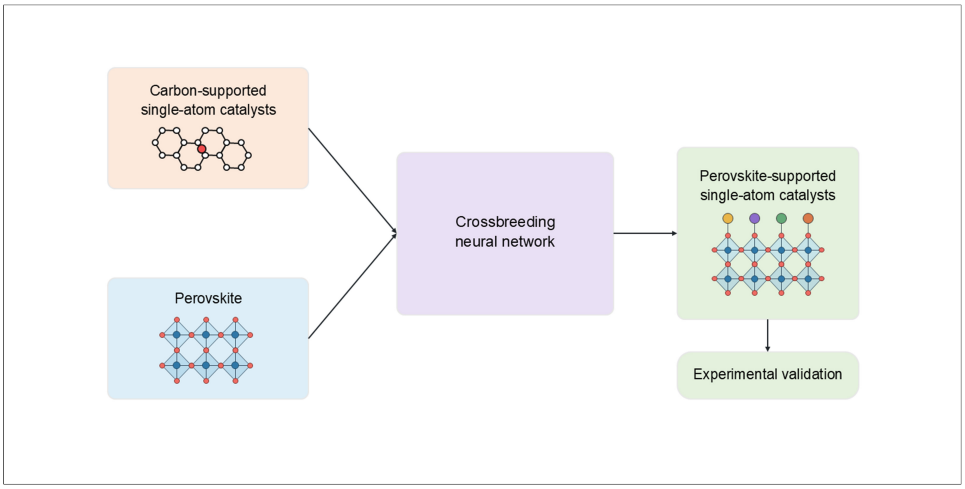
Beyond material silos: expanding catalyst discovery through cross-material deep learning

Yuanting Tang[1], Jihwan An[1,2]

Citation: Tang, Y.; An, J. Beyond material silos: expanding catalyst discovery through cross-material deep learning. *AI Agent* **2026**, *2*, 35. https://dx.doi.org/10.20517/aiagent.2026.54

Received: 28 Sep 2026 **First Decision:** 21 Oct 2026 **Revised:** 4 Nov 2026 **Accepted:** 10 Nov 2026 **Published:** 18 Nov 2026

Academic Editor: Hao Li **Copy Editor:** Tong Wang **Production Editor:** Tong Wang



Machine learning has accelerated catalyst development[1,2], yet differences in catalyst configurations, descriptors, and evaluation protocols, *etc.*, still confine data-driven discovery to individual material families, leaving the intersections between distinct material systems largely underexplored despite their potential to reveal unexpected catalytic behavior. For the oxygen evolution reaction (OER), carbon-supported single-atom catalysts (SACs) and perovskite oxides exemplify two major catalyst-design strategies based on surface-site engineering and bulk compositional tuning, respectively[3,4]. In a recent study, Moon *et al.* reported a crossbreeding neural network (CBNN) approach that integrates experimental data from these two families to predict and guide the synthesis of perovskite-supported SACs[5], a material class absent from the training set [Figure 1A]. Their work demonstrates a framework for cross-material machine learning that extends the predictive reach of existing experimental data across material boundaries, opening hybrid catalyst space to data-driven discovery.

The central methodological advance of the CBNN lies in connecting material-specific representations while preserving their distinct structural characteristics. As shown in

[1]Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea. [2]Graduate School of Semiconductor Technology, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea.

Correspondence to: Prof. Jihwan An, Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea. E-mail: jihwanan@postech.ac.kr

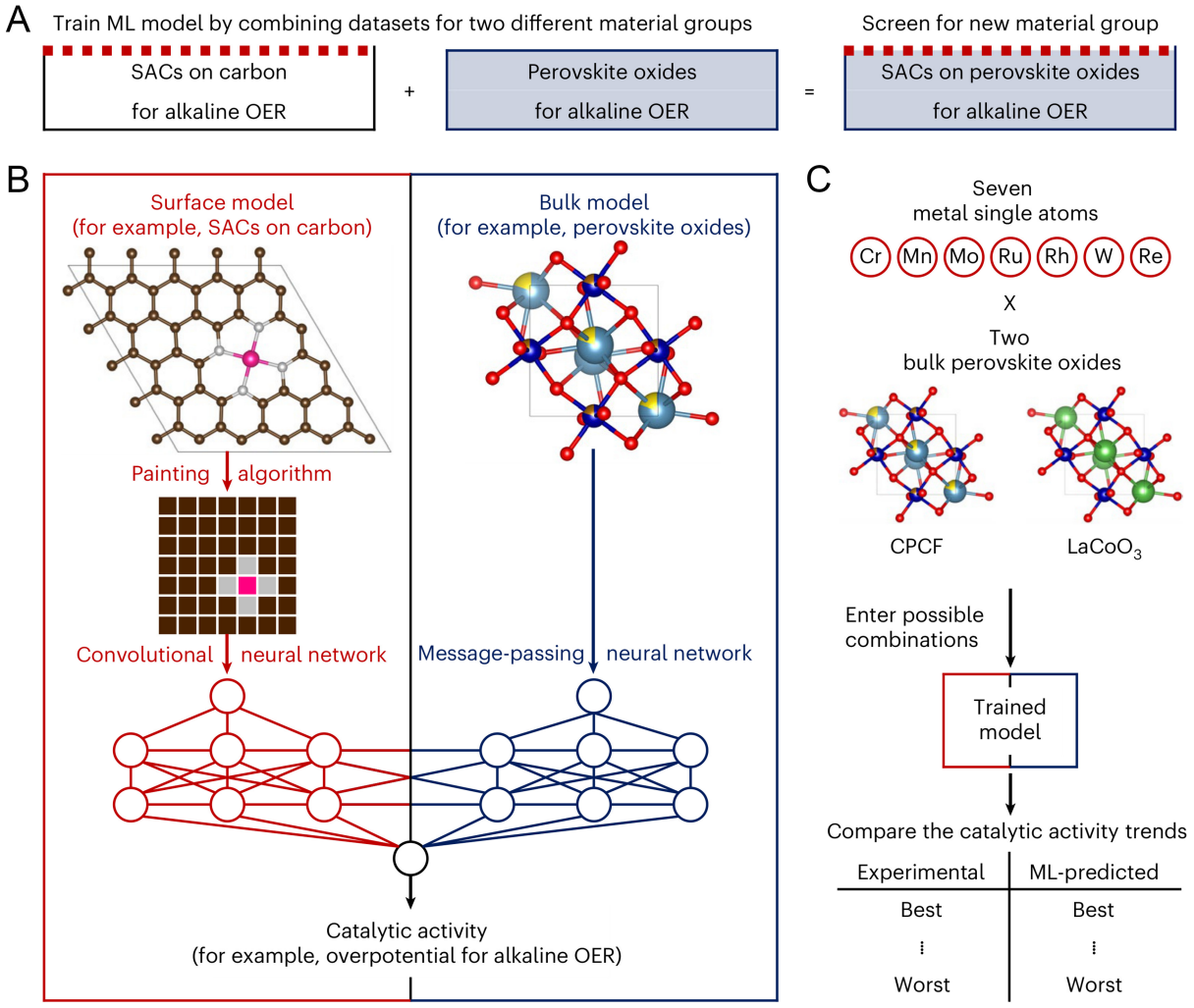


Figure 1. Overview of the CBNN for cross-material catalyst discovery. (A) Schematic illustration of integrating experimental datasets for carbon-supported SACs and perovskite oxides to predict the OER activity of perovskite-supported SACs; (B) Overview of the CBNN; (C) Screening of the relative OER activities of 14 monometallic SACs comprising seven metals and two perovskite supports. Reproduced with permission from Ref.[5], copyright 2026 Springer Nature. This figure is not included in the CC BY 4.0 license of the present article. ML: Machine learning; SACs: single-atom catalysts; OER: oxygen evolution reaction; CBNN: crossbreeding neural network.

Figure 1B, rather than forcing structurally distinct catalyst families into a uniform input format, the CBNN encodes the surface structures of carbon-supported SACs and bulk structures of perovskites through separate branches. Training on either source dataset (22 carbon-supported SACs or 41 perovskites) updates the corresponding branch while freezing the other. For hybrid catalysts, both branches are activated, and their outputs are concatenated and passed to a shared fully connected output layer trained on both datasets. The shared co-descriptors give the two representations a common chemical basis for combining their learned surface and bulk contributions. To identify suitable co-descriptors, the authors developed an automated selection pipeline integrating statistical independence testing, correlation analysis, and natural language processing (NLP). NLP supports descriptor selection by assessing textual similarity between candidate descriptor terms and "catalytic activity" or "overpotential" in the literature.

From ten candidate elemental features, the pipeline selected five key co-descriptors: oxidation state, ionic radius, valence *d*-electron count, Pauling electronegativity, and metal coordination number. These co-descriptors primarily characterize atom-level properties, including local chemical and coordination

environments. For the surface branch, the atomic arrangements of carbon-supported SACs are mapped onto pixels, and the descriptors are encoded in image channels. In the bulk branch, the descriptors form feature vectors for atomic nodes in perovskite graphs. For the source datasets, oxidation states are taken from reported data, while coordination numbers are taken from reported local coordination for carbon-supported sites or assigned from perovskite crystal structures. Ionic radii are obtained from Shannon's database for the corresponding oxidation states and coordination numbers, while valence *d*-electron counts follow the elemental electron configurations and oxidation states. Joint training with the five key co-descriptors improved the correlation coefficient from 0.904 to 0.931 compared with independent training. In contrast, when using the five lower-ranked descriptors, joint training decreased the correlation coefficient from 0.829 to 0.803[5]. These results highlight the critical role of co-descriptor selection in effective cross-material learning.

To infer co-descriptor values for hybrid catalysts, the authors used an auxiliary convolutional neural network, the Discriminator, trained independently of the CBNN. It was trained to distinguish surface images generated from experimentally determined descriptor values from images generated using other theoretically possible combinations of descriptor values. Its classification output guided the selection of surface images, which were then used as inputs to the CBNN for activity prediction[5]. Uncertainty in these inferred inputs should be considered when interpreting the predictions. The CBNN then ranked 14 monometallic SACs comprising seven metals supported on $Ca_{0.8}Pr_{0.2}Co_{0.8}Fe_{0.2}O_{3-\delta}$ (CPCF) or $LaCoO_3$ [Figure 1C]. All 14 catalysts were synthesized for experimental validation. Predicted and experimental activity trends agreed, with Mo/CPCF and Rh/$LaCoO_3$ as the most active catalysts on their respective supports and Mn as the least active metal on both supports. The authors further screened 8,008 CPCF-supported compositions, identifying W-Mo-Ru-Rh/CPCF[5]. Experimental validation showed that W-Mo-Ru-Rh/CPCF achieved an intrinsic OER overpotential of 349 mV at 10 $mA{\cdot}cm^{-2}_{oxide}$ and outperformed all 14 monometallic SACs on perovskite oxides tested in the study[5]. These results support transfer within the investigated space, rather than unrestricted generalizability[6]. Although both source studies used 0.1 M potassium hydroxide (KOH), the carbon-supported SAC dataset reports overpotentials at 5 $mA{\cdot}cm^{-2}$, whereas the bulk perovskite dataset reports them at 10 $mA{\cdot}cm^{-2}_{oxide}$. For experimental benchmarking, the original study compared W-Mo-Ru-Rh/CPCF with bulk perovskite oxides using intrinsic overpotential and with carbon-supported SACs using turnover frequency.

To interpret these predictions, the authors used explainable machine learning to examine the role of key co-descriptors and local atomic environments. Sensitivity analysis quantified the relative influence of the five key co-descriptors on the predicted OER overpotential, while atom-resolved activation mapping visualized the model response around neighboring active sites. For CPCF-supported monometallic catalysts, the activation maps showed more favorable model responses for Mo and W than for Mn, in agreement with the experimental activity trends[5]. When W, Mo, Ru, and Rh were co-loaded, the average activation value decreased further relative to the monometallic cases, consistent with the superior performance of W-Mo-Ru-Rh/CPCF. These analyses help explain model behavior but do not independently establish catalytic mechanisms or multimetallic synergy. Complementary theoretical and experimental investigations are therefore needed to assess catalyst reconstruction, working-state coordination and durability under operating conditions[7].

Looking ahead, extending this approach requires consistent descriptor definitions, comparable activity labels and adequate coverage of target chemistry. The decline in performance with lower-ranked co-descriptors illustrates the risk of negative transfer and the need to benchmark joint training against independent training. Even when average predictive accuracy improves, performance can vary across materials space[6], so applicability limits and uncertainty in predictions and inferred descriptors require explicit assessment.

Beyond alkaline OER, the authors applied the CBNN architecture to acidic OER and the alkaline hydrogen evolution reaction, using reaction-specific datasets[5]. These extensions suggest broader applicability, but differences in reaction mechanisms, intermediates, and active-site requirements may alter descriptor-activity relationships and complicate knowledge transfer across reactions. To evaluate transfer in new settings, active learning and automated experimentation could support an iterative discovery framework in which predictions and uncertainty guide candidate selection and standardized activity and stability measurements feed back into model training.

Overall, the broader lesson of this study is that dataset boundaries should not become barriers to catalyst discovery. Rather than treating individual material families as isolated systems, cross-material approaches can draw on the different structure-activity information contained in existing datasets to investigate new combinations. The key is not the indiscriminate aggregation of heterogeneous datasets, but the construction of a chemically meaningful basis that allows otherwise dissimilar structural representations to contribute to a common predictive task. As experimental datasets become more comparable and more closely integrated with automated synthesis and testing, this principle could support a broader shift from optimizing catalysts within established material classes to designing across their boundaries.

DECLARATIONS

Authors' contributions

Original manuscript preparation: Tang, Y.
Review and editing of the manuscript: Tang, Y.; An, J.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2026-25480463) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A1A03039981).

Conflicts of interest

Both authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Zhong, M.; Tran, K.; Min, Y.; et al. Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature* **2020**, *581*, 178-83. DOI PubMed
2. Mok, D. H.; Li, H.; Zhang, G.; Lee, C.; Jiang, K.; Back, S. Data-driven discovery of electrocatalysts for CO_2 reduction using active motifs-based machine learning. *Nat. Commun.* **2023**, *14*, 7303. DOI PubMed PMC
3. Moon, J.; Beker, W.; Siek, M.; et al. Active learning guides discovery of a champion four-metal perovskite oxide for oxygen evolution electrocatalysis. *Nat. Mater.* **2024**, *23*, 108-15. DOI PubMed

4. Cao, L.; Luo, Q.; Chen, J.; et al. Dynamic oxygen adsorption on single-atomic Ruthenium catalyst with high performance for acidic oxygen evolution reaction. *Nat. Commun.* **2019**, *10*, 4849. DOI PubMed PMC
5. Moon, J.; Yoo, S.; Shim, J.; et al. Cross-material catalyst discovery via deep learning. *Nat. Mater.* **2026**, *25*, 1393-401. DOI PubMed
6. Sutton, C.; Boley, M.; Ghiringhelli, L. M.; Rupp, M.; Vreeken, J.; Scheffler, M. Identifying domains of applicability of machine learning models for materials science. *Nat. Commun.* **2020**, *11*, 4428. DOI PubMed PMC
7. Zhao, J. W.; Li, Y.; Luan, D.; Lou, X. W. D. Structural evolution and catalytic mechanisms of perovskite oxides in electrocatalysis. *Sci. Adv.* **2024**, *10*, eadq4696. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.


© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.