科学通报 2026 年 第 71 卷 第 23 期: 5507 ~ 5516
人工智能-能源材料交叉前沿专题 进 展
《中国科学》杂志社
SCIENCE CHINA PRESS

机器学习在电催化剂的设计与发现中的应用

张驰, 王哲, 张葆坤, 刘宏, 于小雯*

山东大学晶体材料全国重点实验室, 济南 250100
* 联系人, E-mail: xiaowen.yu@sdu.edu.cn

2025-11-04 收稿, 2026-01-15 修回, 2026-01-16 接受, 2026-01-19 网络版发表
国家自然科学基金(52572243)和山东省泰山学者青年专家项目(tsqn202306025)资助

摘要 电催化技术可结合可再生能源, 实现清洁燃料的生产与高值化学品的合成, 在减缓全球变暖和缓解环境污染等方面具有重要意义, 因此被视为实现可持续发展的关键技术之一. 然而, 传统电催化剂的研究仍高度依赖实验试错法, 实验合成与理论模拟均面临效率低、周期长等瓶颈, 制约了新材料的开发进程. 近年来, 随着数据科学与人工智能的迅速发展, 机器学习在加速电催化研究中展现出巨大潜力, 尤其在高效探索庞大材料化学空间、设计与发现新型电催化剂方面表现突出. 本综述系统总结了机器学习技术在不同类型电催化剂中的最新应用, 重点阐述其在性能预测、高通量筛选及机制解析中的作用. 同时, 本文还探讨了该领域亟待解决的关键科学问题与未来发展方向, 以期推动电催化研究范式的转变, 促进高效且可持续的能源体系构建.

关键词 人工智能, 机器学习, 电催化剂, 性能预测, 高通量筛选, 机制解析

随着全球能源需求持续攀升与环境问题日益严峻, 开发高效清洁能源转换技术已成为实现可持续发展的重要课题[1]. 作为能源转换体系的核心环节, 催化技术在燃料电池[2]、水分解[3]及二氧化碳还原[4]等领域展现出巨大潜力. 尽管材料科学和电化学领域已取得显著进展, 但电催化剂的开发仍然主要依赖耗时的试错式研究. 无论是实验合成与性能评估, 还是数值模拟与理论计算, 这些过程往往依赖研究者的经验与主观判断, 导致研究效率低下且结果不尽理想. 此外, 传统研究方法难以有效处理具有多变量和高维特征的复杂体系, 设计空间受限, 从而使催化性能的优化成本高昂、进展缓慢[5]. 因此, 亟需发展更加高效的策略, 以在更广阔的化学空间中系统探索潜在电催化剂及最优结构组合.

机器学习(machine learning, ML)的迅猛发展正深刻变革着人类社会的各个领域[6]. 作为一种数据驱动的方法, ML能够从大规模数据中学习潜在规律与内在模式, 从而建立预测材料特性的有效模型. 其基本流程通常包括数据准备、模型构建与实际应用三个关键阶段(图1)[5]. 凭借其处理多层次输入变量的能力—涵盖从原子尺度描述符到宏观工程参数, ML为复杂催化体系的全局性能优化提供了强有力的工具[7], 其主要体现在三个方面: 性能预测、高通量筛选与机制解析. 在性能预测方面, ML不仅加速了材料筛选与设计, 更通过多尺度建模、描述符创新和动态模拟, 实现了从原子级结构到宏观性能的系统性关联[8]. 在高通量筛选方面, ML展现出在广阔化学空间中高效搜索与优化的能力, 可显著加速新材料的发现, 从而在近乎无限的材料组合中实现目标催化剂的快速锁定[9]. 在机制解析方面, ML推动了催化研究从静态描述向动态模拟的转变. 借助卷积神经网络(convolutional neural network, CNN)和图神经网络(graph neural network, GNN)等深度学习模型, 研究者可以在实际反应条件下解析覆盖度效应、活性位点协同机制及反应路径的动态演化. 这种“数据驱动”与“物理模型”相融合的多尺度研究框架, 为设计

引用格式: 张驰, 王哲, 张葆坤, 等. 机器学习在电催化剂的设计与发现中的应用. 科学通报, 2026, 71: 5507–5516
Zhang C, Wang Z, Zhang B, et al. Applications of machine learning in the design and discovery of electrocatalysts (in Chinese). Chin Sci Bull, 2026, 71: 5507–5516, doi: 10.1360/CSB-2025-5742

© 2026《中国科学》杂志社

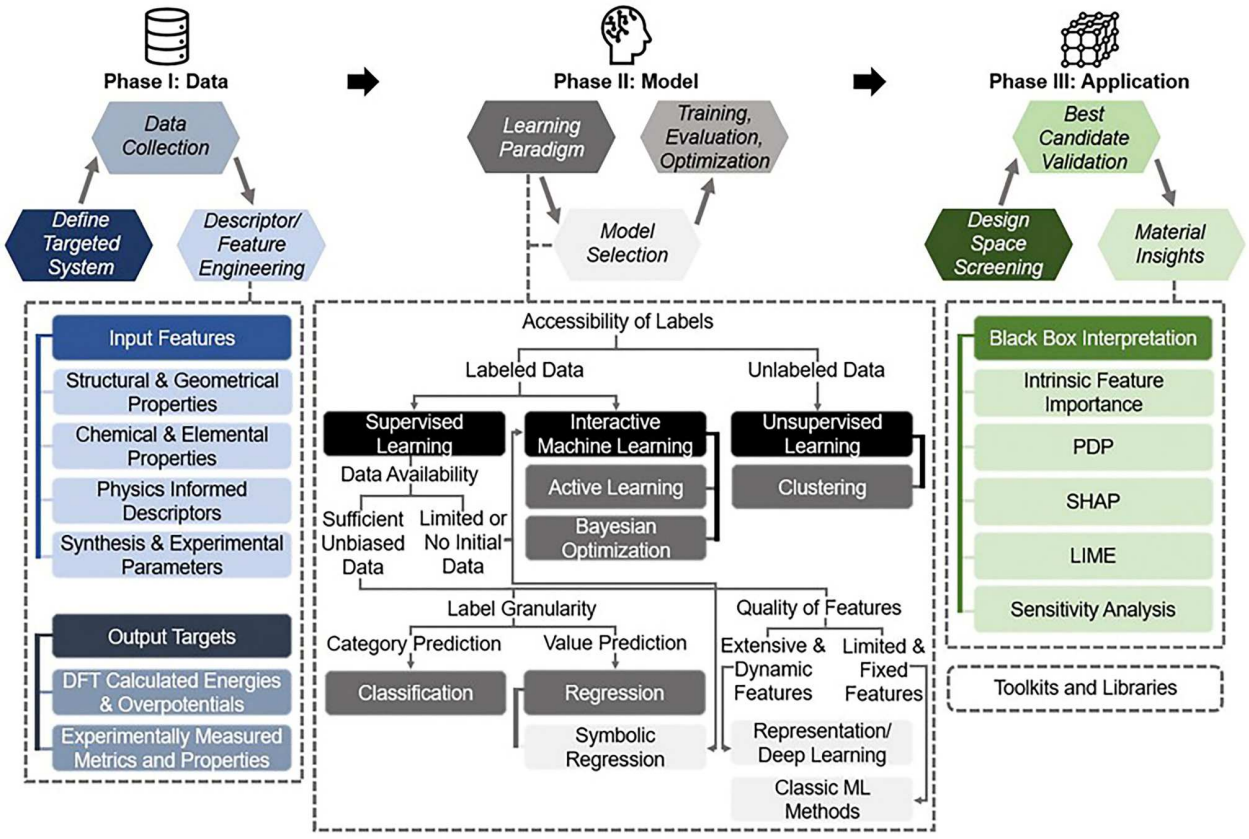

图 1 (网络版彩色)ML三个关键阶段: 数据准备、模型构建与实际应用[5], Copyright © 2024, Royal Society of Chemistry

Figure 1 (Color online) ML key stages: data preparation, model construction, and practical application[5], Copyright © 2024, Royal Society of Chemistry

不同功能的催化体系提供了指导[10]. 这三个维度构成了一个层层递进、相互反馈的完整体系, 共同推动电催化剂的设计从“经验试错”向“理性设计”的范式转变, 为探索广阔的催化材料空间开辟了高效路径.

本综述系统回顾了近五年来人工智能技术在电催化剂的设计领域所引发的范式转变, 重点聚焦于合金、金属氧化物及单原子催化剂等典型体系. 特别是从性能预测、高通量筛选与机制解析三个维度, 系统阐述了ML在不同催化材料的最新进展, 揭示了其在理论计算与实验验证之间构建高效协同桥梁的关键作用, 从而加速了电催化剂的闭环智能设计进程. 最后, 本文展望了该领域所面临的关键挑战与未来机遇, 为电催化研究范式的持续革新与高效、可持续能源体系的构建提供了新的思路与方法.

1 合金催化剂体系

合金材料凭借其多样化的组成与结构特征以及金属间的协同效应[11], 在催化、气体吸附、电子学等多个领域展现出广泛的应用潜力. 因此, 深入理解合金的组成–结构–性能关系对于指导高性能材料设计具有重要意义. 然而, 由于合金体系的高度复杂性, 将算法与量子力学相结合进行合金催化剂设计仍面临计算成本高、时间消耗大等挑战[12]. 在此背景下, ML与合金催化剂研究的深度融合, 正从“结构–性能”预测逐步拓展至“机理–设计”闭环, 推动着电催化材料开发向动态、精准、可解释的方向发展.

首先, 在催化性能预测方面, ML能够有效识别合金表面活性位点的分布规律, 从而实现对宏观催化活性的精准预测. Zhou等人[13]采用密度泛函理论(density functional theory, DFT)与ML相结合的高通量计算策略, 系统解析了Pt–Ru无序固溶体合金的活性位点特性. 作者基于DFT构建了大量随机合金表面模型, 精确计算氢吸附自由能(ΔG_{H^*}), 并以此为数据集训练CNN模型, 成功实现了对约2400个中空位点ΔG_{H^*}的高效预测(图2

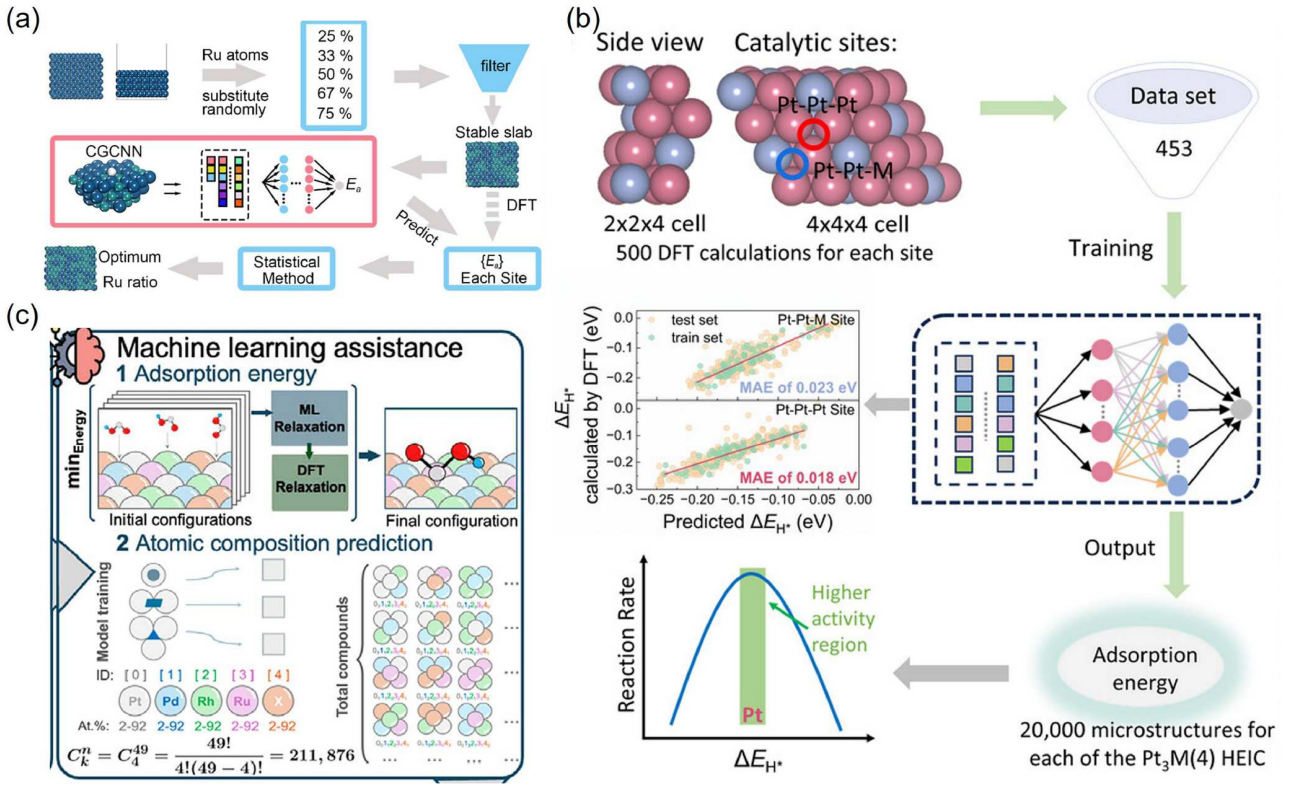

图 2 (网络版彩色)(a) ML用于合金催化剂活性中心分析及总活性预测的流程示意图[13], Copyright © 2024, American Chemical Society. (b) ML辅助设计Pt_3M HEICs用于HER的流程示意图[17], Copyright © 2025, John Wiley and Sons. (c) ML辅助理论预测与设计五元HEA示意图[18], Copyright © 2025, American Chemical Society

Figure 2 (Color online) (a) Workflow of ML for active-site analysis and overall catalytic activity prediction in alloy catalysts[13], Copyright © 2024, American Chemical Society. (b) Schematic workflow of ML-assisted design of Pt_3M HEICs for the HER[17], Copyright © 2025, John Wiley and Sons. (c) Schematic illustration of ML-assisted theoretical prediction and rational design of five-component HEAs[18], Copyright © 2025, American Chemical Society

(a)). 模型结果表明, 当Ru含量为20%~30%时, 合金整体的活性最优. 实验验证显示, 当Ru含量为21%时, 其析氢反应(hydrogen evolution reaction, HER)过电位仅为21 mV. 该研究从微观活性位点能量出发, 建立了其与宏观催化性能之间的定量关联, 为催化活性预测提供了可靠研究范式. 在此基础上, Mao等人[14]进一步将研究对象由二维表面位点拓展至三维纳米团簇结构. 通过结合ML与高通量DFT计算, 作者发现活性位点周围的电荷分布与氢吸附的强度密切相关, 并由此提出一种表面电荷组成的描述符来评估合金团簇的HER活性. 最终, 作者发现过渡金属掺杂可显著提升Cu基团簇的HER活性, 并且其高活性位点数量大幅增加. 其中, $Cu_{49}Ni_6$核壳型合金团簇表现出最佳的HER性能. 这些研究展示了ML在催化性能预测中由成分优化到结构设计、由位点尺度到纳米尺度逐层拓展的强大潜力.

其次, 在催化剂高通量筛选方面, 面对高熵合金(high-entropy alloys, HEAs)所涉及的高度复杂组分空间, ML展现出在庞大化学空间中进行高效搜索与筛选的独特优势. HEAs因其独特的成分复杂性与结构多样性而成为一类极具潜力的新型材料[15]. 典型的HEA通常由五种或更多主要元素组成, 每种元素的原子分数占总量的5%~35%. 这种高维组分复杂性赋予HEA一系列独特的物理化学特性, 使其在电化学能量转换领域展现出卓越的催化性能[16], 但同时也显著增加了材料设计与筛选的难度. 在此背景下, Wang等人[17]在固定晶体结构约束下引入ML, 实现了HEA催化剂的高效组分筛选(图2(b)). 作者以$L1_2$型Pt基高熵金属间化合物(high-entropy intermetallic compounds, HEICs)为研究对象, 基于Pm-3m空间群构建了15种Pt_3M型HEICs, 并结合453组DFT计算数据(不同表面ΔG_{H^*})训练深度神经网络模型, 从而对数以万计不同表面构型的ΔG_{H^*}进行快速预测. 结果表明, Pt_3(CrMnFeCo)组分在理论活性

范围内对应的微观结构数量最多. 进一步的数据挖掘分析揭示, 表层Co、Cr、Fe原子主要调控Pt–Pt–M中空位点, 而次表层Ni与Co原子则调节Pt–Pt–Pt相互作用. 实验结果进一步验证了Pt_3(CrMnFeCo) HEIC在宽pH范围内均表现出优异的HER催化活性与稳定性. 该研究代表了一种在结构固定条件下, 通过ML实现组分组合优化的"约束空间内精准搜索"策略. 此外, Nam等人[18]进一步放宽结构与成分约束, 提出了一种ML辅助的DFT自下而上研究策略, 以加速五元HEA的理性设计(图2(c)). 作者以四元合金PtPdRhRu为基础, 从7种候选元素中筛选最优"第五元素". 通过训练高性能的CatBoost模型, 对全部可能的成分组合进行快速预测, 并进一步利用以局部原子环境为描述符的ML模型精准预测关键中间体的吸附能, 确定了PtPdRhRuMo合金的最优原子比例. 实验结果表明, 该合金在甲醇氧化反应中的质量活性高达9.89 A mg^{-1}, 并表现出优异的长期稳定性. 该工作通过逐步引入元素选择与比例优化, 展示了ML在更高维HEA化学空间中实现高效筛选与实验验证的能力. 进一步地, Pan等人[19]将研究策略由计算主导拓展为实验–计算–ML协同的闭环筛选模式, 实现了HEA催化剂性能的快速迭代优化. 通过将高通量实验与理论计算紧密结合, 作者在大规模元素组合与成分空间中系统筛选出FeNiCuCoPt HEA. 实验结果显示, 该材料在0.85 V条件下的氧还原反应(oxygen reduction reaction, ORR)质量活性高达1.32 A mg^{-1}, 并兼具优异的稳定性. 该研究体现了闭环集成设计方法在复杂HEA体系中的显著优势, 即通过实验–计算反馈机制显著提升了化学空间探索的效率与可靠性.

最后, 在催化机理解析方面, ML已逐渐发展为揭示电催化反应本质机制的重要工具. 在此方面, Gu等人[20]提出了一种端到端的多尺度模拟框架, 系统阐明了锯齿状Pt纳米线在碱性条件下高效催化HER的内在机理. 作者首先基于DFT计算获得ΔG_{H^*}数据, 并以此训练ML模型, 从而实现对纳米线表面不同活性位点吸附能的快速预测. 在此基础上, 研究进一步引入覆盖度效应与线性自由能关系, 并结合动力学蒙特卡罗方法, 对实际HER动力学过程进行了模拟. 研究结果揭示了一种"自适应双功能机制": 强结合位点主要负责质子吸附, 而弱结合位点有利于氢分子活化, 其中ΔG_{H^*}位于−0.2~0.0 eV区间的活性位点表现出最优的HER催化性能. 该研究展示了ML在催化机制解析中的关键作用, 与传统的单一尺度计算相比, 该方法能够更真实地反映实际反应环境下的覆盖度效应与位点协同机制, 实现了从静态吸附能预测到动态过程模拟的跨越. 这不仅适用于HER, 也为其他涉及复杂表面过程和多中间体的电催化体系提供了可推广的机理研究工具.

2 金属氧化物催化剂体系

尖晶石型氧化物(AB_2O_4)与钙钛矿型氧化物(ABO_3)由于其A位与B位均可灵活调控催化活性中心, 从而能够引入多种候选元素, 成为设计高性能电催化剂的重要结构体系[21]. 近年来, 随着人工智能技术的迅速发展, 基于ML模型预测材料性能并指导新型氧化物设计的方法日趋成熟[22]. 与此同时, 许多过渡金属氧化物由于其结构–活性关系明确、储量丰富且成本低廉, 在电催化领域受到了广泛关注. 鉴于氧化物材料的化学空间近乎无限, 传统试错方法在筛选高性能催化剂时面临巨大挑战. 在此背景下, ML为从海量材料数据与文献中高效挖掘潜在候选材料提供了理想工具[23]. 目前, ML正推动金属氧化物催化剂体系实现从性能预测、高通量筛选到反应机制解析的闭环研究范式, 加速了该领域从经验探索到理性设计的转变.

首先, 在催化性能预测方面, Li等人[24]提出了一种基于阳离子晶格环境预测钙钛矿氧化物在不同温度下氧空位浓度的方法(图(3a)). 研究构建包含235种钴基钙钛矿和200种铁基钙钛矿、总计超过5500个数据点的大规模数据集, 系统比较了36种ML算法, 最终选定梯度增强回归器作为预测模型. 该模型仅需输入阳离子晶格环境特征(如电负性、极化力、电荷等)和操作温度, 即可实现氧空位浓度的高精度预测(R^2>0.92). 进一步的DFT计算表明, 阳离子环境通过调控金属中心的电子态从而影响氧空位的形成能. 基于此, 研究设计并合成了两种新型钙钛矿材料: $SrCo_{0.8}Ta_{0.16}W_{0.04}O_{3-\delta}$和$SrCo_{0.8}Ta_{0.15}V_{0.05}O_{3-\delta}$. 这两种材料在500 ℃以下展现出优异的ORR活性, 该研究为开发宽温度范围高效电催化剂提供了可靠的预测框架. 此外, Moon等人[25]开发了一种基于主动学习(active learning, AL)的工作流程, 融合钙钛矿晶体结构信息与XRD等物理数据并结合GNN向量化策略, 通过迭代式数据增强, 仅依靠40组实验数据便实现了对数万候选材料的有效探索. 该模型成功再现了已知材料的析氧反应(oxygen evolution reaction, OER)活性趋势, 并预测验证了高性能催化剂$Ca_{0.8}Pr_{0.2}Co_{0.8}Fe_{0.2}O_{3-\delta}$, 其在10 mA cm^{-2}下的过电位仅为391 mV, 优于多数已报道材料. 这两项研究展示了催

化性能预测的两种方法，前者通过构建特征模型寻找关键描述符(氧空位浓度)来预测催化性能, 从而建立了可解释的“结构–描述符–性能”关联路径；而后者则利用数据融合与AL最大化有限数据的价值，直接对过电位进行预测与筛选，清楚地展示了ML推动催化性能预测从经验摸索向智能化设计的转变.

其次, 在催化剂高通量筛选方面, 针对酸性ORR中Ir基催化剂成本高、稳定性不足等关键问题，Esterhuizen等人[26]构建了一种融合贝叶斯优化(Bayesian optimization，BO)与DFT的智能筛选框架(图3(b)). 该框架首先从材料数据库中筛选Ir基混合氧化物结构, 并通过结构描述符进行特征提取; 随后, 基于高斯过程回归模型并结合期望改进采集函数，在有限计算预算下高效探索性能最优的材料组合. 通过对1014种$Ir_{0.5}M_{0.5}O_2$体系的系统筛选，模型成功预测Mo为最优掺杂元素. 实验结果表明，$Ir_{0.5}Mn_{0.5}O_2$催化剂在酸性ORR中的过电位降低约30 mV, Ir溶解率下降24%, 并展现出显著提升的结构稳定性. 该研究突出体现了BO方法在多目标高通量筛选中的高效性，能够在庞大搜索空间中快速锁定性能最优组合，为复杂电催化体系的智能设计提供了可扩展的新路径. 在此基础上，Nair等人[27]从模型可解释性与描述符自动发现的角度出发，开发了一种基于符号回归的AL工作流. 该方法能够在筛选过程中自动识别决定材料性能的关键物理描述符，并据此指导候选材料的迭代选择. 通过对包含1470种氧化物的数据库进行系统筛选，该方法成功发现了12种同时具备高稳定性与高OER活性的氧化物催化剂. 相较于传统黑箱式预测模型，符号回归方法在保持筛选效率的同时显著增强了模型的可解释性，有助于揭示催化性能背后的物理本质规律. 进一步地, 针对组成与合成条件

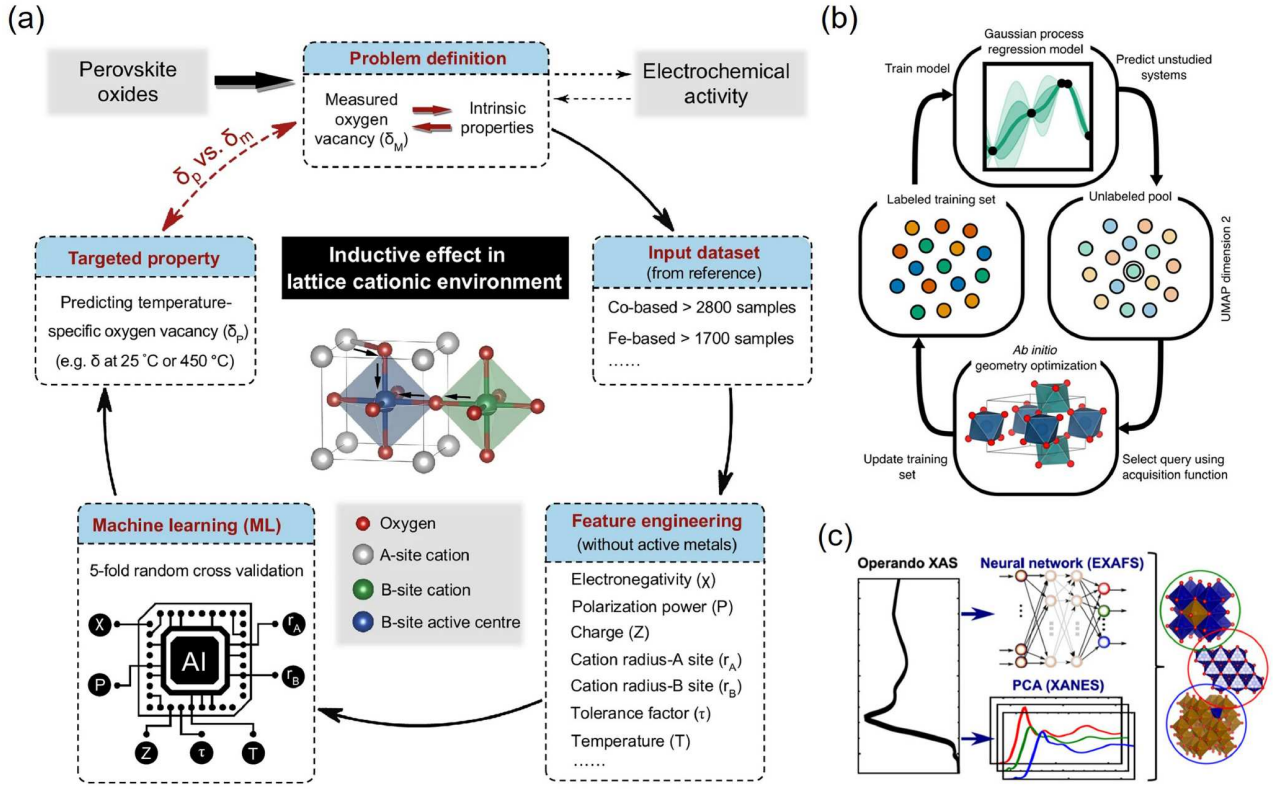

图 3 (网络版彩色)(a) 不同温度下$ABO_{3-\delta}$钙钛矿氧空位预测的工作流程示意图[24], Copyright © 2024, Nature Publishing Group. (b) BO算法结合DFT建模识别电化学稳定$Ir_{0.5}M_{0.5}O_2$催化剂的流程示意图[26], Copyright © 2024, American Chemical Society. (c) 利用快速X射线吸收光谱与ML分析OER过程中氧化物催化剂结构与化学转变的示意图[29], Copyright © 2023, American Chemical Society

Figure 3 (Color online) (a) Workflow for predicting oxygen vacancies in $ABO_{3-\delta}$ perovskite at different temperatures[24], Copyright © 2024, Nature Publishing Group. (b) Schematic workflow of BO combined with DFT modeling for identifying electrochemically stable $Ir_{0.5}M_{0.5}O_2$ catalyst[26], Copyright © 2024, American Chemical Society. (c) Schematic illustration of structural and chemical transformations of oxide catalysts during OER, revealed by rapid X-ray absorption spectroscopy coupled with ML analysis[29], Copyright © 2023, American Chemical Society

高度复杂的多金属氧化物体系，Ding等人[28]提出了一种融合系统性文献挖掘与机器学习的材料发现策略. 通过无监督与有监督学习方法，从已有文献中提炼元素特性与合成参数等关键信息，并在此基础上构建设计准则；随后引入AL模块，对候选体系进行迭代优化筛选. 该策略最终确定以Ru为核心元素，并成功发现了兼具高活性与高稳定性的Ru–Mn–Ca–Pr氧化物催化剂, 其OER理论过电位显著降低. 该研究展示了文献整合、计算预测与ML筛选协同作用下对复杂多元体系进行系统性探索的强大能力.

最后，在催化机制解析方面，ML与原位同步辐射光谱的融合，为揭示金属氧化物在OER中的动态机制提供了强大工具. Timoshenko等人[29]通过训练神经网络方法, 建立从X射线吸收精细结构光谱到径向分布函数的复杂非线性映射，实现了对金属氧化物四面体和八面体配位金属位点局部结构的分别和定量追踪(图3(c)). 研究者发现在OER条件下，活性位点并非均匀生成, 而是源于富Co相中Co^{3+}-O_6八面体团簇的可逆形成, 且其结构形态与催化剂组成密切相关，主成分分析进一步确认该过程的电化学信号与Co氧化态变化同步, 明确了其为活性中心. 同时, 通过独立追踪Co、Fe元素的光谱演化, 揭示了催化剂存在纳米尺度的相分离, 主成分分析也从变化微弱的Fe-K边光谱中捕捉到活性相中微量Fe的氧化态与Co同步可逆升高，证明Fe起到了关键的电子协同作用，表明适量的Fe掺入富Co相会优化活性位点的电子结构从而促进Co^{3+}-O_6活性物种的形成. 这一研究揭示了CoFe氧化物催化剂在OER过程中的动态过程以及ML与光谱技术在解析复杂催化体系动态行为中的强大能力.

3 单原子催化剂体系

传统的负载型金属催化剂通常由尺寸分布较宽、分散于载体表面的金属纳米颗粒构成. 为充分利用其活性位点, 需要尽可能减小并精确控制金属粒径. 当金属原子被完全孤立并单独锚定于载体表面时，即形成单原子催化剂(single-atom catalysts, SACs), 这一构型能够最大化原子利用率并优化反应选择性，被认为是实现高活性与高选择性催化反应的理想体系[30]. 单个金属原子可锚定于多种载体上，包括金属氧化物、金属以及石墨烯等材料. 近年来，ML结合DFT已被广泛用于加速SACs的发现与机制解析.

在催化性能预测方面：Xu等人[31]通过ML辅助的DFT计算，重新审视并优化了SACs的通用设计原则, 提出了改进的结构描述符ψ′. 该描述符以金属中心*d*轨道价电子数与局部配位环境的电负性为核心变量，可有效关联石墨烯基SACs在ORR、OER和HER中的催化活性. ψ′不仅能准确预测112种石墨烯负载SACs的吸附自由能，还可扩展至小/中/大环复合物负载的SACs体系. 基于火山图分析, 作者筛选出ORR的高性能候选Fe-吡啶-4N以及OER的高性能候选Co-吡咯-4N，计算预测与实验测试的催化活性趋势高度一致. 该研究通过构建普适性的描述符ψ′，实现了多类SACs在多反应体系中的性能快速预测，推动SACs理性设计由经验驱动向普适规律指导转变，并为面向多反应场景的材料筛选与结构优化提供了高效可靠的工具.

在催化剂高通量筛选方面, Xu等人[32]在AL框架内将渐进式学习策略与DFT计算相结合，以预测OER中间体吸附能(图4(a))，研究对261种过渡金属单原子掺杂金属氧化物开展高通量筛选，最终识别出9种性能突出的候选催化剂. 借助子群发现技术, 作者进一步提炼出5个关键电子结构描述符，并揭示Mn掺杂可调控RuO_2表面电子结构，适度削弱含氧中间体的吸附强度, 从而显著降低OER决速步骤($O^* \rightarrow OOH^*$)的反应能垒. 在理论指导下，研究成功合成Mn_{SA}-RuO_2 SAC并验证其优异的OER活性及长期稳定性. 该研究表明以“反应性能指标”为起点，借助ML反向解析关键结构因素，是实现复杂催化体系高效筛选与精准优化的有效路径. 在此基础上，Lin等人[33]将研究视角由“性能驱动的结构反演”进一步拓展至“结构本征属性的正向预测”. 针对石墨烯负载单原子体系，作者提出将ML与物理机理融合的筛选策略，构建基于金属*d*电子数与电负性的本征描述符，用于高效评估石墨烯基SACs的氮气还原反应性能(图4(b)). 该方法可对超过700种石墨烯基SACs进行高通量筛选，并兼容3*d*–5*d*过渡金属及C/N/P/B/O等多样配位环境，其解析表达式从分子轨道层面揭示了结构–活性之间的内在联系. 理论预测与实验测得的氨产率呈现典型火山型关系. 电化学测试表明, 理论预测的最佳催化剂Ru-N_3实现了最高的NH_3产率(134.2 μg h^{-1} mg^{-1}). 该研究体现了以“本征电子结构”为出发点构建物理可解释描述符，在实现大规模筛选的同时兼顾预测准确性与可推广性的优势.

在催化机理解析层面，Zhang等人[34]利用ML对锚定在$MoSe_2$表面的Au单原子HER机制进行了系统研究(图4(c)). 作者基于高通量DFT计算，构建了涵盖多种

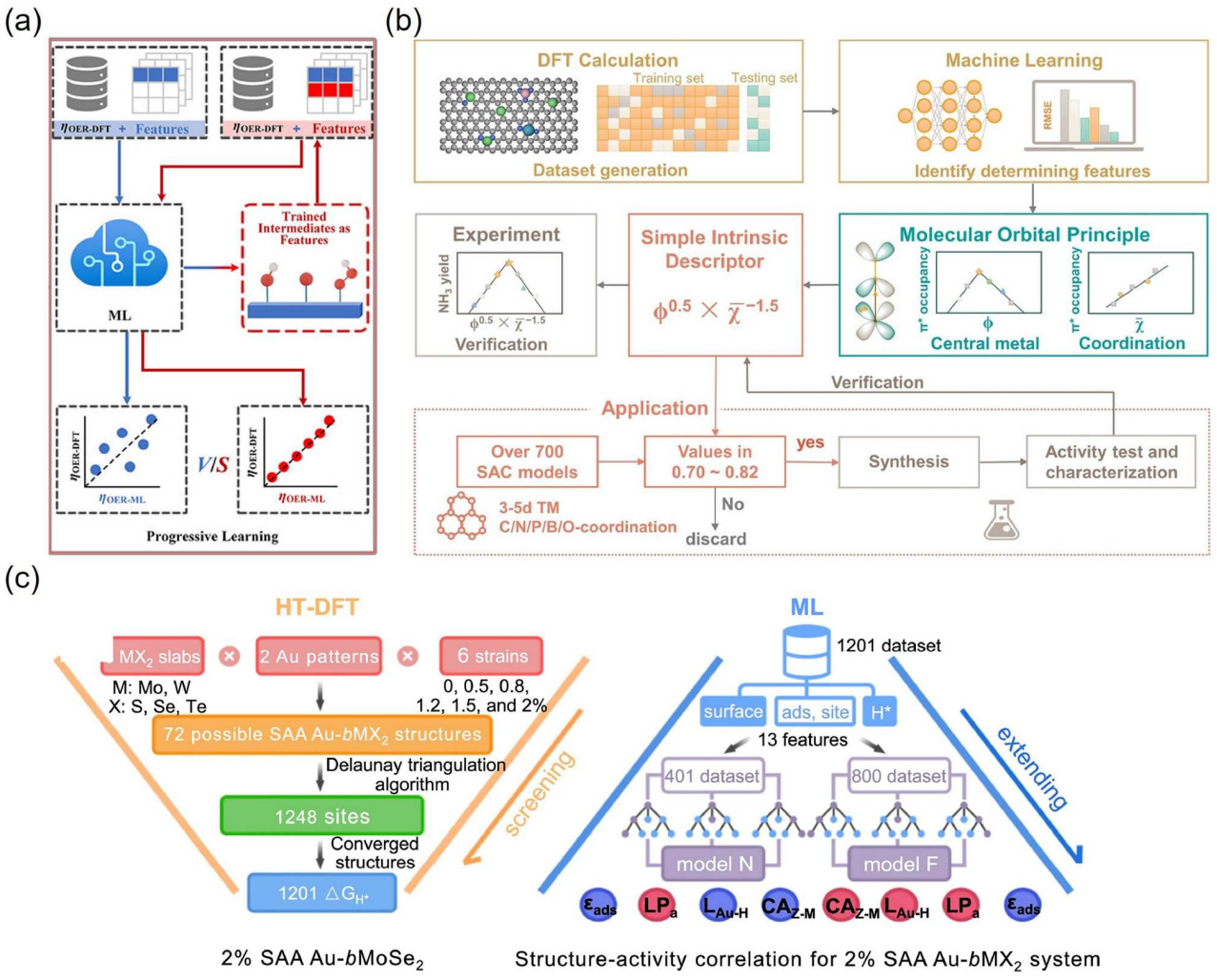

图 4 (网络版彩色)(a) ML筛选单原子金属氧化物催化剂的流程图[32], Copyright © 2025, John Wiley and Sons. (b) 基于压缩感知数据挖掘建立的二维描述符方程的预测能力[33], Copyright © 2023, John Wiley and Sons. (c) 结合高通量DFT计算和ML预测的数据驱动工作流程[34], Copyright © 2025, Nature Publishing Group

Figure 4 (Color online) (a) Flowchart of ML screening of single-atom metal oxide catalysts[32], Copyright © 2025, John Wiley and Sons. (b) Predictive performance of a two-dimensional descriptor equation established by compressed sensing data mining[33], Copyright © 2023, John Wiley and Sons. (c) Data-driven workflow developed by integrating high-throughput DFT calculations with ML prediction[34], Copyright © 2025, Nature Publishing Group

MX_2基底及不同应变水平的数据集，共获得1201个ΔG_{H^*}数据点. 研究发现，数据按吸附位点分类后呈现明显的双峰分布特征. 据此，作者将数据集预分为“近Au”与“远Au”两类，并分别训练相应的ML模型. 基于ML识别出的特征、原位拉曼光谱表征与晶体轨道哈密顿布居分析，作者进一步阐明了该体系的反应机制: Au与邻近Se原子协同构成“催化活性热点”. 该热点的空间尺寸显著大于传统SACs中孤立金属活性位点. 其中Au作为电子富集中心，可协同激活周围多个Se作为氢吸附位点，从而实现多位点协同催化并显著提升整体HER活性. 该研究通过对高通量计算数据进行ML分类建模, 并结合原位表征手段, 揭示了反应过程中真实活性位点的本质特征.

4 总结与展望

本综述系统梳理了ML在合金、金属氧化物与SACs设计与发现中的最新进展，重点涵盖性能预测、高通量筛选与机制解析三方面. 总体而言，ML可用于快速评估催化性能、在复杂材料空间中高效定位优选体系, 并进一步挖掘关键控制因素以深化机理理解; 三者相互衔接，构成从材料发现、理性设计到机理认知的闭环研究框架. 尽管三类体系在组成复杂度、活性

位点形态与反应机理上差异显著, 但其ML研究策略呈现高度共性. 在性能预测层面, 普遍强调建立“结构–描述符–性能”的可解释映射, 只是描述符侧重点不同: 合金更关注局部原子环境与组分协同, 氧化物侧重阳离子特征与氧空位调控, SACs突出金属中心*d*电子结构与金属–载体相互作用. ML的关键作用在于将这些不同物理背景下的核心特征统一映射到宏观性能, 实现高维化学空间中的快速、定量预测. 在高通量筛选层面, 三类体系广泛采用AL、BO等算法, 构建“预测–验证–反馈”的迭代闭环, 从而显著降低计算与实验成本. 尽管合金与氧化物常面向连续多组分空间的组合优化, 而SACs更偏向离散位点与配位环境的精细筛选, 但ML均通过持续更新模型认知来引导搜索向最优区域收敛. 在机制解析层面, ML由“预测工具”扩展为“知识发现工具”, 通过特征重要性分析、子群发现、符号回归与压缩感知等方法自动识别主导催化行为的关键物理化学参数(如*d*带中心、键长、价态或配位效应), 推动研究由静态吸附能关联走向对反应动态机理与协同机制的深入理解, 实现从材料筛选到构效关系与反应本质解析的跨越. 总体来看, ML的核心价值不依赖于材料类型, 而体现在其作为连接数据、物理与设计的枢纽: 以可解释描述符支撑预测、以智能算法驱动筛选、以数据驱动建模促进机理认知, 从而推动催化材料研究由经验试错向理性、可解释且可迁移的设计范式转变.

尽管ML在电催化研究中展现出突出的预测与筛选能力, 其进一步应用仍面临若干关键局限, 主要体现在以下四个方面: (1) 高质量数据稀缺与系统偏差. 高质量、标准化的数据集是模型准确性与可迁移性的基础, 但实验与计算数据获取成本高、格式不统一. 文献数据尤其受pH、温度、载量、电压及反应器构型等条件显著影响, 跨研究难以直接比较; 若简单合并, 实验差异易被模型误判为材料本征规律, 从而削弱预测可靠性. 改进方向包括: 完善实验条件数据标注, 将关键性能指标尽量本征化(如电位统一换算至标准氢电极、电流密度按电化学活性面积校正), 并将pH、温度等环境因素作为显式特征引入模型, 使其学习“材料–环境”耦合效应. (2) DFT数据与实验数据之间的鸿沟. DFT计算存在系统误差且多基于理想化结构与反应环境, 与真实实验条件下的数据分布不一致, 直接用于训练或融合可能产生偏差, 增加实验指导风险. 跨越这一鸿沟需要发展多保真建模、迁移学习等方法, 并构建“计算–实验”闭环, 以实验反馈校准理论与模型误差. (3) 物理可解释性不足. 许多ML模型, 尤其深度学习方法, 虽具备较高预测精度, 但往往呈现“黑箱”特征, 难以将结果与电子结构调控、吸附能演变及反应路径选择等物理化学机制建立明确关联. 不同模型各有局限: 传统模型依赖人工特征、难捕捉复杂非线性; CNN对输入规整性要求高且可解释性有限; GNN虽能自动学习结构信息, 但计算成本较高且机制可追溯性仍不足. 提升预测性能的同时增强物理一致性与可解释性, 是实现可验证理性设计的关键. (4) 复杂界面与结构重构难以刻画. 电催化过程发生在高度动态的电极–电解液界面, 涉及表面重构、价态演化、覆盖度变化及局部电场/溶剂效应等非平衡过程. 现有ML多基于静态结构与理想化数据, 难以反映“工作态”位点演化, 导致真实工况下预测不稳健. 未来需推进ML与第一性原理分子动力学、相场或连续介质模型等多尺度动态模拟的耦合, 以更真实地描述复杂界面行为.

总体而言, 随着现代数据科学与计算化学的快速发展, ML有望在电催化剂的理性设计与智能发现中发挥更为核心的作用. 自适应学习、可解释人工智能与生成模型的引入, 将进一步拓展模型的预测与创新能力. 同时, 跨学科融合, 尤其是化学、材料科学、计算科学与数据科学的深度结合将推动从“经验驱动”向“数据驱动”范式的转变. 可以预见, 随着这些新方法和理念的不断成熟, 人工智能将成为催化化学领域的基石技术, 加速高性能催化材料与绿色能源转化技术的协同创新.

参考文献

1 Tung C W, Zhang W, Lai T Y, et al. Spin crossover-driven diiron electrocatalyst boosts sustainable water oxidation. Nat Sustain, 2025, 8: 793–805

2 Lei J, Fan X X, Liu T, et al. Single-dispersed polyoxometalate clusters embedded on multilayer graphene as a bifunctional electrocatalyst for efficient Li-S batteries. Nat Commun, 2022, 13: 202

3 Wang J, Yang H, Li F, et al. Single-site Pt-doped RuO_2 hollow nanospheres with interstitial C for high-performance acidic overall water splitting. Sci Adv, 2022, 8: eabl9271

4 Hao S, Elgazzar A, Zhang S K, et al. Acid-humidified CO_2 gas input for stable electrochemical CO_2 reduction reaction. Science, 2025, 388:

eadr3834
5 Ding R, Chen J, Chen Y, et al. Unlocking the potential: machine learning applications in electrocatalyst design for electrochemical hydrogen energy transformation. Chem Soc Rev, 2024, 53: 11390–11461
6 Chen W, Jiang Y, Noble W S, et al. Error-controlled non-additive interaction discovery in machine learning models. Nat Mach Intell, 2025, 7: 1541–1554
7 Deringer V L, Caro M A, Csányi G. Machine learning interatomic potentials as emerging tools for materials science. Adv Mater, 2019, 31: 1902765
8 Li J, Wu N, Zhang J, et al. Machine learning-assisted low-dimensional electrocatalysts design for hydrogen evolution reaction. Nano-Micro Lett, 2023, 15: 227
9 Pu T, Ding J, Zhang F, et al. Dual atom catalysts for energy and environmental applications. Angew Chem Int Ed, 2023, 62: e202305964
10 Han Z, Gao R, Wang T, et al. Machine-learning-assisted design of a binary descriptor to decipher electronic and structural effects on sulfur reduction kinetics. Nat Catal, 2023, 6: 1073–1086
11 Zhou M, Li C, Fang J. Noble-metal based random alloy and intermetallic nanocrystals: syntheses and applications. Chem Rev, 2021, 121: 736–795
12 Mai H, Le T C, Chen D, et al. Machine learning for electrocatalyst and photocatalyst design and discovery. Chem Rev, 2022, 122: 13478–13515
13 Zhou Q, Shou H, Qiao S, et al. Analyzing the active site and predicting the overall activity of alloy catalysts. J Am Chem Soc, 2024, 146: 15167–15175
14 Mao X, Wang L, Xu Y, et al. Computational high-throughput screening of alloy nanoclusters for electrocatalytic hydrogen evolution. npj Comput Mater, 2021, 7: 46
15 Cai J, Zhu H. Surface-engineered nanostructured high-entropy alloys for advanced electrocatalysis. Commun Mater, 2025, 6: 118
16 Li M, Lin F, Zhang S, et al. High-entropy alloy electrocatalysts go to (sub-)nanoscale. Sci Adv, 2024, 10: eadn2877
17 Wang Z, Chen X, Lin T, et al. Machine learning-guided design of $L1_2$-type Pt-based high-entropy intermetallic compound for electrocatalytic hydrogen evolution. Adv Mater, 2026, 38: e10424
18 Nam H N, Nandan R, Fu L, et al. Modeling-making-modulating high-entropy alloy with activated water-dissociation centers for superior electrocatalysis. J Am Chem Soc, 2025, 147: 33545–33558
19 Pan Y, Shan X, Cai F, et al. Accelerating the discovery of oxygen reduction electrocatalysts: high-throughput screening of element combinations in Pt-based high-entropy alloys. Angew Chem Int Ed, 2024, 63: e202407116
20 Gu G H, Lim J, Wan C, et al. Autobifunctional mechanism of jagged Pt nanowires for hydrogen evolution kinetics via end-to-end simulation. J Am Chem Soc, 2021, 143: 5355–5363
21 Jeong I, Shim Y, Oh S, et al. A machine learning-enhanced framework for the accelerated development of spinel oxide electrocatalysts. Adv Energy Mater, 2024, 14: 2402342
22 Dhaliwal G, Nair P B, Singh C V. Machine learned interatomic potentials using random features. npj Comput Mater, 2022, 8: 7
23 Feng J, Dong Z, Ji Y, et al. Accelerating the discovery of metastable IrO_2 for the oxygen evolution reaction by the self-learning-input graph neural network. JACS Au, 2023, 3: 1131–1140
24 Li Z, Mao X, Feng D, et al. Prediction of perovskite oxygen vacancies for oxygen electrocatalysis at different temperatures. Nat Commun, 2024, 15: 9318
25 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: 108–115
26 Esterhuizen J A, Mathur A, Goldsmith B R, et al. High-performance iridium–molybdenum oxide electrocatalysts for water oxidation in acid: bayesian optimization discovery and experimental testing. J Am Chem Soc, 2024, 146: 5511–5522
27 Nair A S, Foppa L, Scheffler M. Materials-discovery workflow guided by symbolic regression for identifying acid-stable oxides for electrocatalysis. npj Comput Mater, 2025, 11: 150
28 Ding R, Liu J, Hua K, et al. Leveraging data mining, active learning, and domain adaptation for efficient discovery of advanced oxygen evolution electrocatalysts. Sci Adv, 2025, 11: eadr9038
29 Timoshenko J, Haase F T, Saddeler S, et al. Deciphering the structural and chemical transformations of oxide catalysts during oxygen evolution reaction using quick X-ray absorption spectroscopy and machine learning. J Am Chem Soc, 2023, 145: 4065–4080
30 Zhou L, Yang M, Liu Y, et al. Intrinsic metal-support interactions break the activity-stability dilemma in electrocatalysis. Nat Commun, 2025, 16: 8739
31 Xu H, Cheng D, Cao D, et al. Revisiting the universal principle for the rational design of single-atom electrocatalysts. Nat Catal, 2024, 7: 207–218
32 Xu L, Lu L, Xu N, et al. Progressive learning-guided discovery of single-atom metal oxide catalysts for acidic oxygen evolution reaction. Angew Chem Int Ed, 2025, 64: e202510965
33 Lin X, Wang Y, Chang X, et al. High-throughput screening of electrocatalysts for nitrogen reduction reactions accelerated by interpretable intrinsic descriptor. Angew Chem Int Ed, 2023, 62: e202300122
34 Zhang T, Ye Q, Liu Y, et al. Data-driven discovery of biaxially strained single atoms array for hydrogen production. Nat Commun, 2025, 16: 3644

Summary for “机器学习在电催化剂的设计与发现中的应用”

Applications of machine learning in the design and discovery of electrocatalysts

Chi Zhang, Zhe Wang, Baokun Zhang, Hong Liu & Xiaowen Yu*

State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, China
* Corresponding author, E-mail: xiaowen.yu@sdu.edu.cn

With the continuous growth of global energy demand and increasingly severe environmental issues, developing efficient and clean energy conversion technologies has become a critical task for sustainable development. Electrocatalysis, which can couple with renewable energy sources like solar and wind to produce clean fuels and high-value chemicals, plays a pivotal role in mitigating climate change and environmental pollution. It is widely regarded as a core technology supporting sustainable development goals. The fundamental challenge lies in discovering and designing high-performance, stable, and low-cost catalytic materials that can efficiently and selectively drive key reactions such as the oxygen evolution reaction, hydrogen evolution reaction, oxygen reduction reaction, and carbon dioxide reduction reaction. Despite significant progress in materials science and electrochemistry, the discovery and optimization of electrocatalysts still rely heavily on time-consuming and repetitive trial-and-error approaches. Both experimental workflows—involving the synthesis and characterization of numerous candidate materials—and theoretical simulations, such as computationally expensive density functional theory (DFT) calculations, are often constrained by empirical judgment and high resource costs. These methods struggle to navigate the vast, multi-dimensional composition-structure-property space of potential catalysts efficiently. In recent years, the rapid development of data science and artificial intelligence (AI) has created new opportunities for accelerating electrocatalysis research. Machine learning (ML), in particular, demonstrates strong potential for intelligently exploring the chemical and material design space. By learning complex, non-linear relationships from existing experimental or computational datasets, ML models can rapidly predict key properties like adsorption energies, overpotentials, and stability, bypassing the need for exhaustive calculations or experiments. This data-driven paradigm is fundamentally transforming traditional research workflows. This review systematically summarizes the AI-driven paradigm shift in electrocatalyst design over the past five years, with a focus on representative material systems, including alloys, metal oxides, and single-atom catalysts. We highlight recent progress in ML-assisted electrocatalysis from three key perspectives: performance prediction, high-throughput screening, and mechanism analysis. This review begins by outlining the general ML workflow, covering data acquisition, feature engineering, model selection, and validation strategies. We then discuss recent advances in ML-enabled high-throughput performance prediction and screening, which have accelerated the identification of promising catalyst candidates. The review also examines how ML aids in the microstructural optimization of known materials, for instance, by guiding doping strategies, surface modification, or defect engineering to enhance activity and durability. Furthermore, we highlight the emerging trend of integrating theoretical modeling with experimental validation through active learning and autonomous robotic platforms, paving the way for a closed-loop, intelligent design framework for electrocatalysts. This iterative cycle of ML prediction, automated synthesis, and high-throughput testing significantly shortens the innovation timeline. Finally, we address key challenges and future research directions, including the scarcity of high-quality, standardized data, the interpretability of complex ML models, and the need for closer integration between computational and experimental communities. Future advancements are expected to involve multi-task learning, generative models for inverse design, and more sophisticated digital platforms that unify data, models, and experiments. By addressing these issues, AI-powered electrocatalyst discovery promises to not only innovate research paradigms but also accelerate the transition towards an efficient and sustainable energy system. This review aims to provide new insights and methodologies for researchers working at the intersection of electrochemistry, materials science, and artificial intelligence.

artificial intelligence, machine learning, electrocatalyst, performance prediction, high-throughput screening, mechanism analysis

doi: 10.1360/CSB-2025-5742