

<https://doi.org/10.1038/s41524-025-01607-4>

具有最小化特征参数的机器学习模型，用于多型氢进化催化剂预测

Chao Wang, Bing Wang, Changhao Wang, Aojian Li, Zhipeng Chang & Ruzhi Wang  

庞大的化学成分空间带来了使用传统方法在催化剂开发中的挑战。机器学习（ML）提供了新的机会，但是当前的ML模型通常仅限于筛选单个催化剂类型。在这项工作中，我们开发了一种有效的ML模型，以预测各种催化剂的氢进化反应（HER）活性。通过最大程度地减少功能，我们引入了与关键能源相关的功能 $\phi = \text{nd}0^2/\psi 0$ ，这与她的自由能量相关。仅使用十个功能，非常随机的树模型可实现 $R^2 = 0.922$ 。我们预测了物料项目数据库中的132个新催化剂，其中一些表现出了她的表现。ML模型用于预测所消耗的时间是传统密度功能理论（DFT）方法所需的时间的200,000。该模型提供了一种有效的方法，可以使用少量关键功能来发现催化剂高性能，并为开发其他催化剂提供见解。

随着对环境污染和化石燃料耗尽的越来越关注，寻找清洁和可持续的能源已成为紧迫的¹⁻³。由于其高能密度和零碳排放⁴，氢（ H_2 ）被认为是最有前途的替代源之一。目前，由可再生能源提供动力的水电解生产是一种非常有希望的技术^{5,6}。水电解产生的氢会受到氢进化反应（HER）的控制。但是，电化学反应表现出缓慢的动力学，导致水电解⁷⁻⁹的高电势。因此，必须开发有效的催化剂来增强电化学反应并减少过电势。贵金属（例如PT，IR和RU）及其衍生物在水电解方面表现出极好的电导率，并且在贵金属表面上的氢原子的吸附自由能接近零，因此，高贵金属被视为她^{10,11}的最有效的催化剂。但是，高贵的金属材料具有高成本和有限的可用性等缺点，这限制了其大规模的商业应用。因此，从生产成本和效率¹²⁻¹⁴的角度来看，低成本，高效率的电催化剂材料的设计和开发至关重要。

已经开发了各种类型的氢进化催化剂（HEC），显示了某些催化活性，例如合金，碳化物，氮化物，氧化物，氧化物，磷化物，硫化物，硫酸盐和perovskites¹⁵⁻¹⁸。但是，基于传统实验方法的开发催化剂面临几个问题，包括长期的开发周期和显着的随机性。

此外，使用密度功能理论（DFT）开发有效催化剂的高通量计算方法需要大量的计算资源¹⁹。因此，使用经验实验和DFT计算从巨大的组成空间中开发出优质的催化剂仍然是一个重要的挑战^{20,21}。

机器学习（ML）是一种强大的统计方法，它基于输入数据构建模型，并通过计算算法提供目标值。ML可用于分析输入特征和目标性能之间的复杂关系。此外，ML可以评估每个输入特征的重要性，并预测许多未知催化剂的催化活性。ML的强大功能已应用于出色的催化剂²²⁻²⁷的快速筛选。强大的ML算法可以帮助揭示催化剂的物理化学特性与其活性之间的关系，从而加快其电催化剂^{28,29}的效果。例如，Chandra Veer Singh等。开发了一种神经网络模型，用于通过分解配体和协调效果³⁰设计高渗透合金（HEA）催化剂，实现了 $\text{MAE} \$ = 0.09 \text{ eV}$ 和 $\text{RMSE} \$ = 0.12 \$ 0.12 \$ 0.12 \text{ eV}$ 的测试集预测准确性。Lin等。使用147个功能开发了ML模型，以快速预测二进制合金HEAS³¹的活动，其测试集预测准确性为 $R^2 = 0.921$ 和 $\text{RMSE} = 0.224 \text{ eV}$ 。S. Kim等。使用20个功能来构建用于过渡金属单原子的CATBOOST回归模型³²，其测试集预测准确性为 $R^2 = 0.88$ 和 $\text{RMSE} = 0.18$ 。Mu等。使用了13个功能来发展

在石墨烯³³上支持的双原子催化剂的随机森林回归模型，达到了测试集预测准确性 $R^2 = 0.871$ 和 $MSE = 0.150$ 。但是，当前用于探索有效HEC的ML方法仅适用于单一类型的HEC的设计，它们遭受了诸如使用众多功能和低模型准确性之类的问题。这种限制源于HEC多类型的特征的显著变化。因此，迫切需要开发使用更少功能，具有更高准确性并可以预测各种HEC活动的ML模型。

在这项工作中，我们开发了高精度ML模型来设计高度活跃的HEC。我们从Catalysis-Hub中获得了10,855 HEC的原子结构功能和氢吸附自由能 ΔG_H 数据，用于培训和预测³⁴。该数据集包括各种类型的HEC，例如纯金属，金属金属间化合物，轻质金属间的金属间化合物，非金属化合物和钙钛矿。仅使用23个基于原子结构和催化剂活性位点的电子信息，而无需进行其他DFT计算，我们建立了六个ML模型：随机森林回归（RFR），梯度增强回归（GBR），极端梯度增长（极端梯度促进回归（GBR），XGBR），XGBR回归（XGBR），决策树回归（DTR），轻度梯度启用了lgbmri机器调节（LGBBRIING机器）。ETR模型的R平方（ R^2 ）得分为0.921，预测 ΔG_H ，表现优于其他ML模型。通过功能重要性分析和功能工程，我们重新指出并确定了更相关的功能，将功能的数量从23降低到10，并将 R^2 得分提高到0.922。此外，两个深度学习（DL）模型，晶体图卷积神经网络（CGCNN）和轨道图概述神经网络（OGCNN）之间的比较，ETR模型表明，在准确性方面表现出这些DL模型的表现，表明在实现高预测性能中具有特征选择的重要作用。最后，我们预测了132种不同的HEC的性能，并使用DFT方法进一步验证了ML模型的预测准确性。这项工作提供了

可解释的见解，用于加速高性能HEC的组成设计。

结果

实施工作流的示意图如图1所示。该过程包括数据收集，功能提取，ML模型训练和测试，功能工程和ML优化，催化剂预测和筛选。数据是ML的基础。为了使用ML模型生成预测，可靠和能力的数据集至关重要。

数据收集

我们从Catalysis-Hub数据库³⁵中获得了11,068个她的自由能和相应的吸附结构。该数据库中的数据来自已发表的文献，经过同行评审，并经过验证，以确保数据准确性。该数据集包括各种类型的HEC，例如纯金属，金属金属间化合物，光金属间的金属间化合物，非金属化合物和perovskites，数据集中的所有数据均来自DFT计算。如图S2所示，过渡金属金属间化合物，光金属间的金属间化合物和非金属化合物一起占数据集的90%以上，与当前的催化剂开发中的研究重点保持一致。在数据集中，氢吸附位点包括顶部站点，桥接位点和空位位点，空缺位点吸附是最普遍的。在这里，我们介绍了图2中数据集的三个代表性结构的顶部和侧视图：Tair，Lair和Ptco分别在顶部站点，桥接位点和空心位置吸附的氢原子。通常，当氢原子吸附在表面上时，氢原子与表面原子之间的距离通常落在1.5至2.5Å的范围内。该范围表明氢原子与表面原子之间存在舒适的相互作用，以促进吸附。在我们的数据集中，表面上氢吸附的距离也位于该范围内。如图2a所示，在Tair顶部的氢吸附的键长为1.892Å；在图2B中，桥接吸附的键长

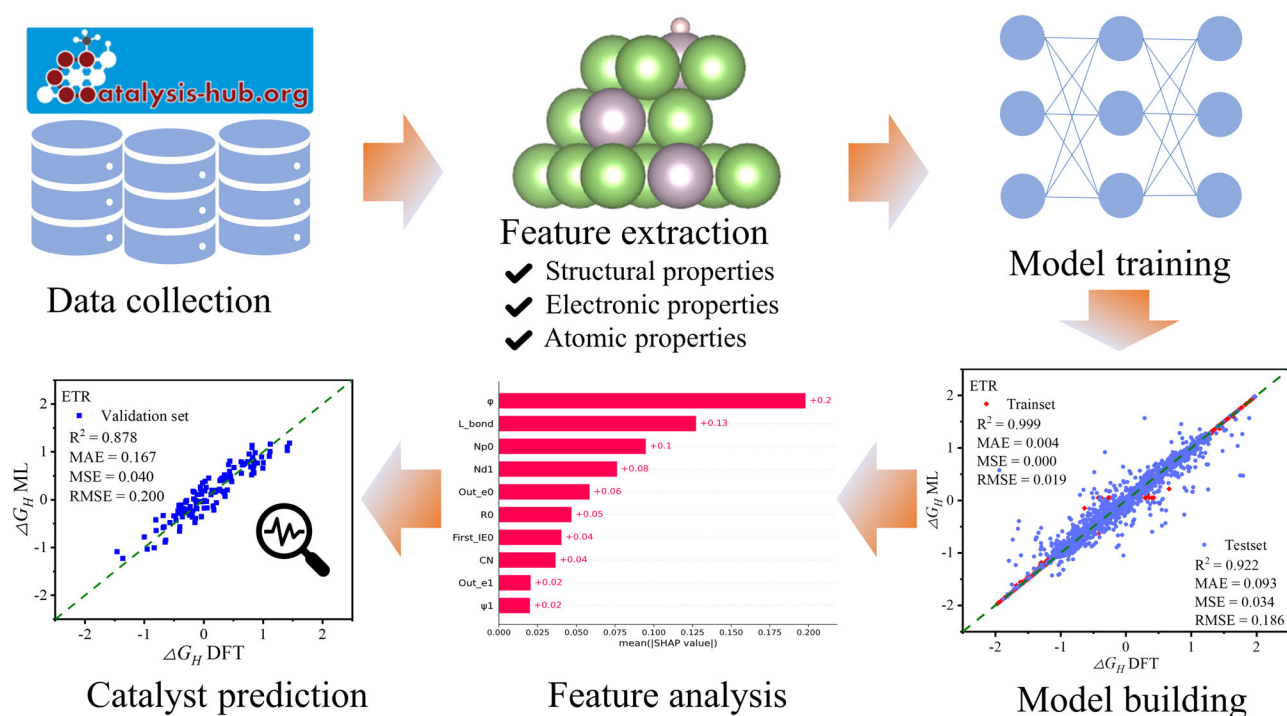
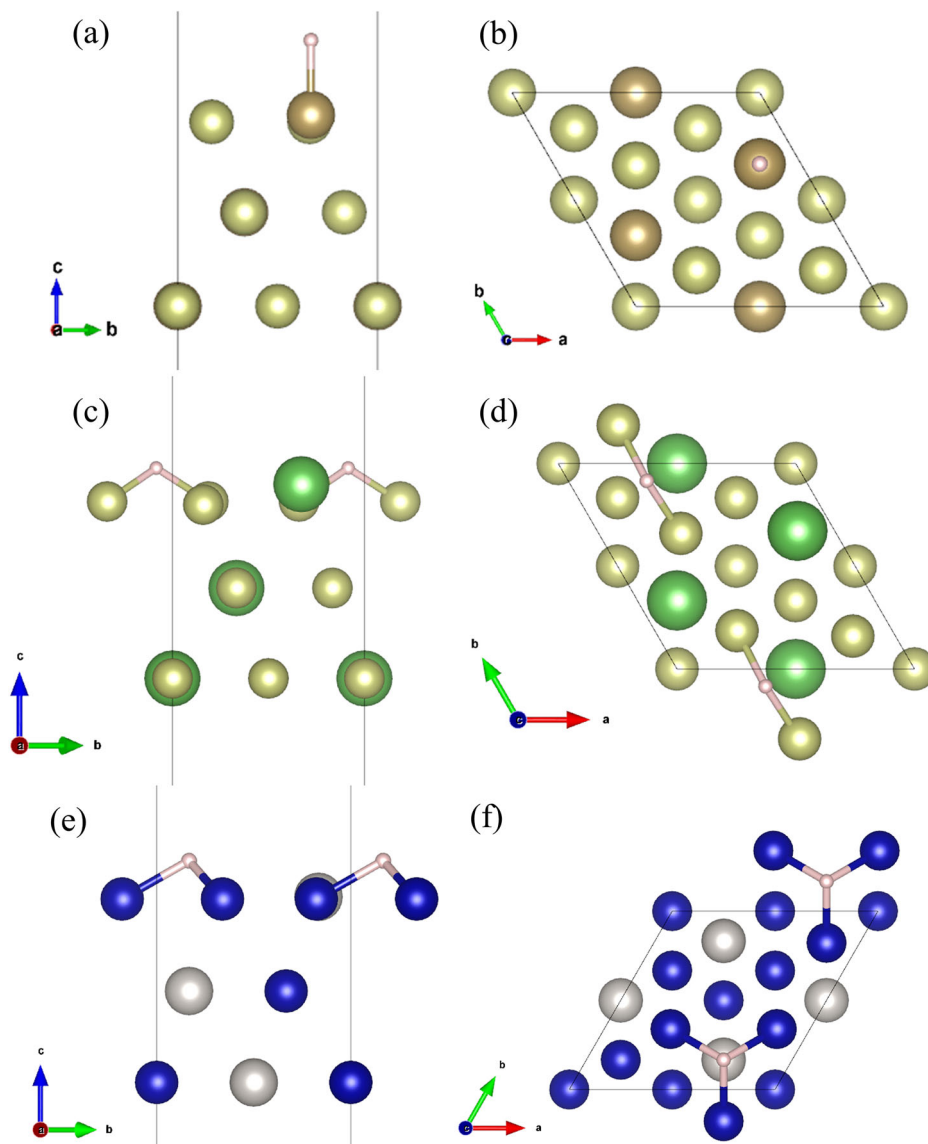


图1 | 的工作流程的催化活动预测。该过程包括数据收集，功能提取，模型培训，模型构建，功能分析和模型预测。数据收集：为她收集原子结构和 ΔG_H 。特征提取：从HE的原子结构中提取结构特征，电子特征和原子特征。模型训练：通过高参数调整提高ML模型的精度。模型构建：使用在训练集上拟合的ML模型对测试集进行预测。功能分析：分析特征的重要性和相关性，并使用功能工程来减少功能集，同时引入关键功能以增强模型精度。模型预测：使用ML模型预测潜在的HEC。

图2 |结构示意图说明了三个不同原子位点的水力吸附。它分别显示了(a, c, e)在 t_{air} , l_{air} 和 pt_{co} 上的氢吸附的前视图, 分别在(b, d, f)中分别显示了在 t_{air} , l_{air} 和 pt_{co} 上的氢吸附的最高视图。



巢穴的位置为1.795Å和1.796Å;在图2C中,PTCO 空位位点的氢吸附的键长均为1.758Å。另外,某些催化剂在Y SC等氢原子中表现出相对较大的吸附距离,其中空位位点的氢吸附键长为2.262Å,2.268Å和2.269Å。为了确保在我们的特征提取脚本中,我们将表面原子和氢原子之间的截止距离设置为2.4Å,以此为活性中心原子,我们将截止距离为2.4。

数据集中HEC的自由能的分布如图3A所示,范围为 $[-12.4, 22.1]$ EV,插图代表了某些区域的大量信息。

根据Nørskov的工作⁵,当 Δg_H 的绝对值为零时,她的催化活动是最佳的。值得注意的是,95.5%的数据范围内 $[-2, 2]$ ev。我们将氢吸附的自由能范围缩小到 $[-2, 2]$ EV,并去除了不合理的氢吸附结构。图3B显示了氢吸附数据集中自由能的调整分布。调整后的数据点的总数为10,855,涉及42个元素,如图S3所示,其中包括大多数过渡金属。此外,该数据集中氢原子的表面覆盖范围为 $\leq 25\%$ 。根据J.K. Nørskov对她的纯金属自由能³⁶的计算,她的自由能在低点计算

覆盖范围与实验结果很好。因此,预计使用该数据集计算出的自由能可以准确反映实际催化剂的催化活性。

功能提取和ML模型构建

特征提取是ML的必不可少的组件,设计适当,全面的特征是结构ML模型中最关键的阶段。因此,至关重要是建立主要基于吸附结构和电子培养物的催化反应描述符。催化剂的表面结构在研究她的催化作用中起着重要作用。本研究使用的特征提取脚本利用原子模拟环境(ASE)的Python模块自动识别吸附的氢原子和材料表面结构,从而提取相关特征。根据先前的研究,我们为催化剂及其最近的邻居³⁷⁻⁴³设计了电子和元素特征属性。

我们收集的功能包括活跃位点原子的特性,在该原子中发生氢吸附以及活动位点附近的周围原子。特定特征如下: D波段电子计数, P波段电子计数, S波段电子计数, 价电子计数, 电负性, 第一个电离能, 原子半径, 原子半径, 活性位点和几何氢之间的吸附键长, 以及氢原子以及该原子的几何键

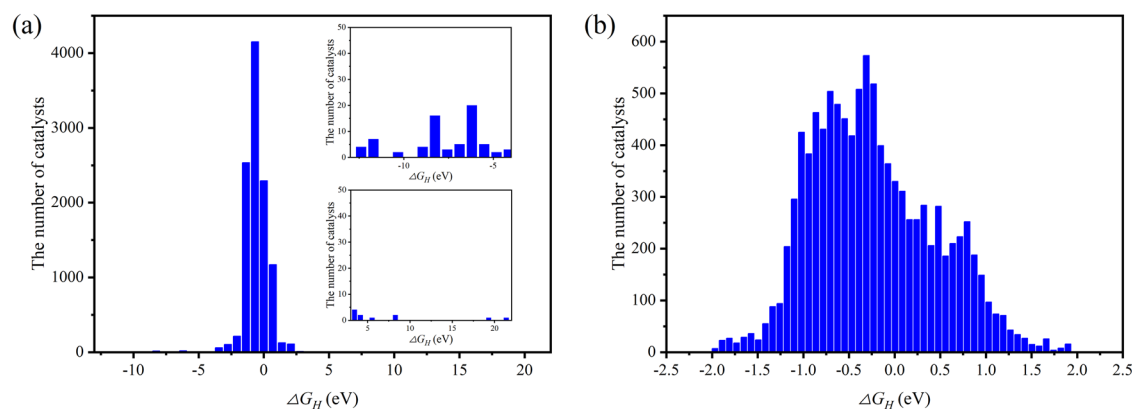


图3 | 氢进化自由能的分布图。在整个数据集中的氢进化自由能A的分布在 $[-2, 2]$ eV范围内。(a)中的插图代表某些区域的宏伟。

最接近吸附中心的原子的配位数的平均值。所有选定的特征都具有明确的物理和化学意义。

先前的研究表明，D波段和P波段电子数量直接影响D型波段和P波段中心的催化剂^{44–46}的位置。通常，较高的D电子计数与更正面的吸附能相关，而较高的P电子计数与更负吸附能相关。S-Electron计数的包含很大程度上是为了区分Cu, Ag, Au和Pt的HE(氢进化反应)与其他元素的性能，因为这些金属在其最外层的壳中只有一个电子。为了区分主要组元素和过渡金属的催化活性，与S-和P轨道相关的特征起着至关重要的作用。与过渡金属不同，主要组元素的化学性质主要取决于其S-和P轨道的电子结构，而D-轨道的贡献很小。S, P电子计数的引入特征有效地区分了主要组元素和过渡金属对表面催化的影响，从而促进了捕获主要组元素的化学行为和催化性能。价电子计数与选择s, p, d电子计数的选择有些重叠，最初被包括在模型中以探索特征适合性。电负性和第一个电离能分别代表了原子获得和丢失电子的能力，电子传递在催化反应中起着至关重要的作用。原子半径影响活性原子在催化剂表面，电子分布和状态表面密度上的排列。较小的原子半径通常会导致表面上较短的原子间差异，从而导致较高的表面电子密度，从而可以增强吸附的氢与催化剂之间的键合。这种情况使 ΔG_H 更加负面，阻碍氢的解吸，从而阻碍H₂产量。相反，较大的原子半径通常会增加表面上的原子间距离，从而降低了表面电子密度并削弱了氢的吸附，从而使 ΔG_H 更阳性。但是，如果吸附太弱，则氢不能稳定地吸附在催化剂表面上，这也不对H₂产生不利。

吸附键的长度是指氢原子与催化剂表面上的活性位点之间的距离。这直接影响氢原子与催化剂表面之间的相互作用强度。较短的吸附键长通常对应于由于氢原子与表面原子之间更强的相互作用而引起的更强的吸附。在这种情况下，氢原子更加紧密地与表面结合，从而产生了更负 ΔG_H 。另一方面，由于较弱的相互作用，较长的吸附键长对应于较弱的吸附，从而使 ΔG_H 更高。在先验知识的指导下，我们使用几何均值方法来研究特征。这种方法平均涉及的特征，并确保所有催化剂的处理特征保持在相同的数量级上，从而促进了影响氢进化性能的因素的探索。计算几何平均值的公式显示在等式中。1。每个

表S1中详细介绍了处理的功能。

$$X = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (1)$$

在获得催化剂的特征之后，由于树模型的稳健性和稳定性，我们在这项研究中采用了六种基于树的ML算法，即RFR, GBR, XGBR, DTR, DTR, LGBMR和ETR，表S2中提供了六个ML模型的详细参数。我们使用诸如R平方(r^2)，平均绝对误差(MAE)，均方误差(MSE)和根平方平方误差(RMSE)等指标评估了这六个模型的拟合性能，这些指标分别代表了准确性得分，平均分量误差，平均平方误差和平均误差。特定的预测结果如图4所示。

DTR算法在测试集上的 R^2 值为0.855，MAE，MSE和RMSE值分别为0.131、0.064和0.253。GBR算法在测试集中获得了 R^2 值为0.881的 R^2 值，其中包括MAE，MSE和RMSE值为0.112、0.055和0.235。RFR在测试集上的 R^2 值为0.912，MAE，MSE和RMSE值分别为0.110、0.040和0.200。XGBR在测试集中产生的 R^2 值为0.912，MAE，MSE和RMSE值为0.107、0.040和0.200。LGBMR在测试集中的 R^2 值为0.913，MAE，MSE和RMSE值为0.117、0.039和0.198。ETR在测试集上达到了最高的 R^2 值为0.921，相应的MAE，MSE和RMSE值为0.104、0.036和0.189。值得注意的是，所有六个基于树的型号均在测试集中获得了大于0.85的 R^2 值，这表明它们有效地描述所选功能之间的关系和 ΔG_H 之间的关系。这表明所选功能可以有效地建立催化剂功能与 ΔG_H 之间的关系。在六棵树模型中，ETR算法模型在测试集中达到了最高的 R^2 值。因此，我们选择了ETR模型进行后续预测和特征重要性分析。此外，我们还比较了两种基于非树的算法MLP模型和SVM模型，这两个模型的预测精度如图S4所示。对于MLP模型，训练集的拟合精度为 $R^2 = 0.847$ ，MAE，MSE和RMSE值分别为0.156、0.070和0.264。在测试集中，预测精度为 $R^2 = 0.815$ ，MAE，MSE和RMSE值分别为0.178、0.084和0.290。对于SVM模型，训练集的拟合精度为 $R^2 = 0.829$ ，MAE，MSE和RMSE值分别为0.148、0.078和0.279。在测试集中，预测精度为 $R^2 = 0.809$ ，MAE，MSE和RMSE值分别为0.164、0.086和0.294。表S2提供了MLP模型和SVM模型的详细参数。结果表明，MLP和SVM模型在训练和测试集上表现出相对较差的拟合性能，这主要是由于它们无法有效捕获复杂的非线性关系。相反，合奏树模型，结合了多个决策树或优化

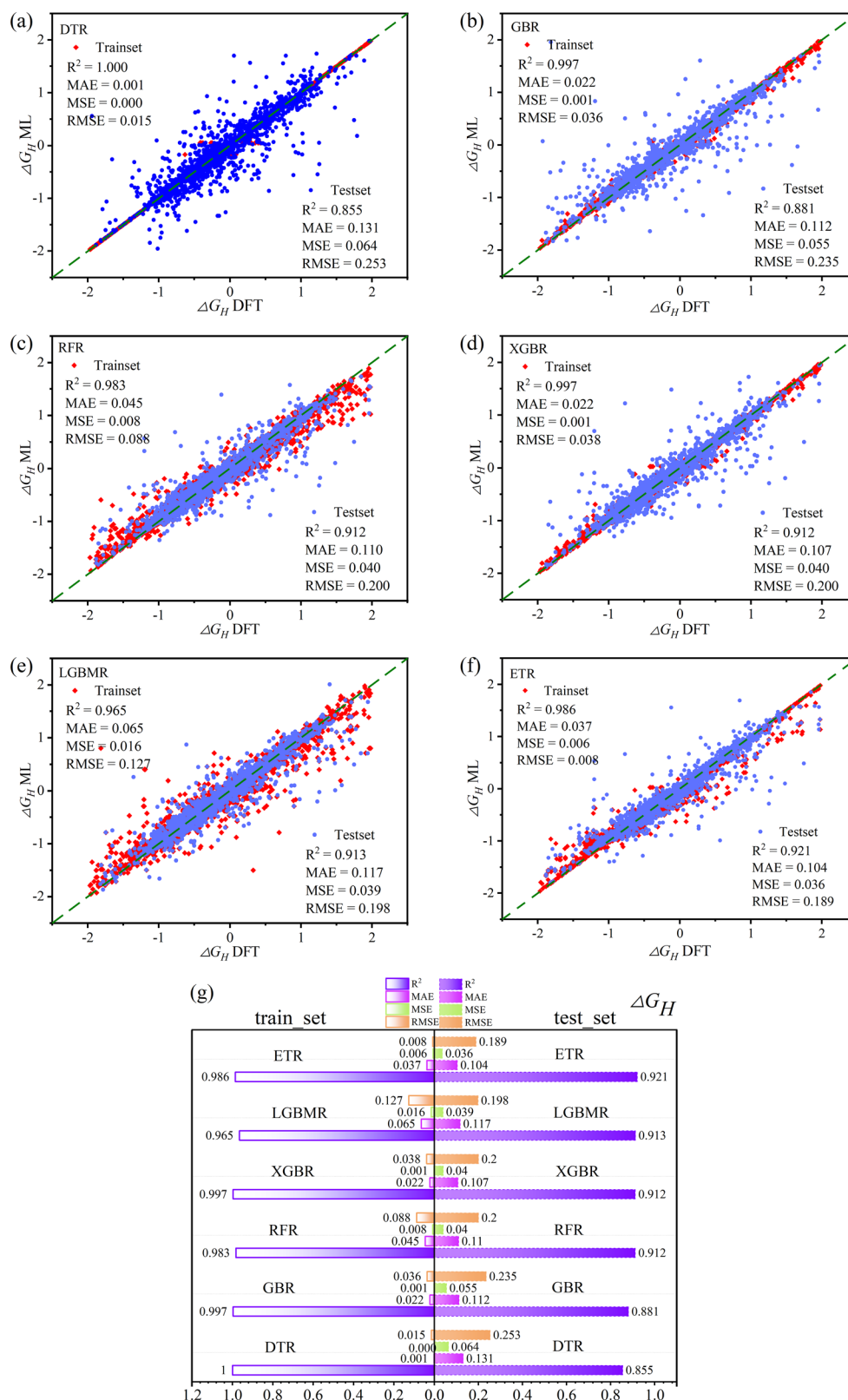


图4 |在培训和测试集中六个基于树的ML模型的预测性能。每种模型的A – FR²评分, MAE, MSE和RMSE: DTR, GBR, RFR, XGBR, LGBMR和ETR在训练和测试集中, G六个ML模型中预测准确性的总体比较。

基于残差,在捕获这种复杂的非线性模式方面表现出色。此外,这些模型表现出对特征噪声,缺失值和高维数据的强大鲁棒性。

为了评估所选功能与 Δg_H 之间的相关性,我们利用Pearson相关系数 (PCCM) 表示他们

关系。图S5提出了一个 heatmap,描述了功能与功能之间的关系和 Δg_H 之间的关系。皮尔逊相关系数接近1表示两个变量之间的相关性很高。Pearson相关系数的检查和 Δg_H 之间的同系数表明,具有最高的Pearson相关性的三个功能

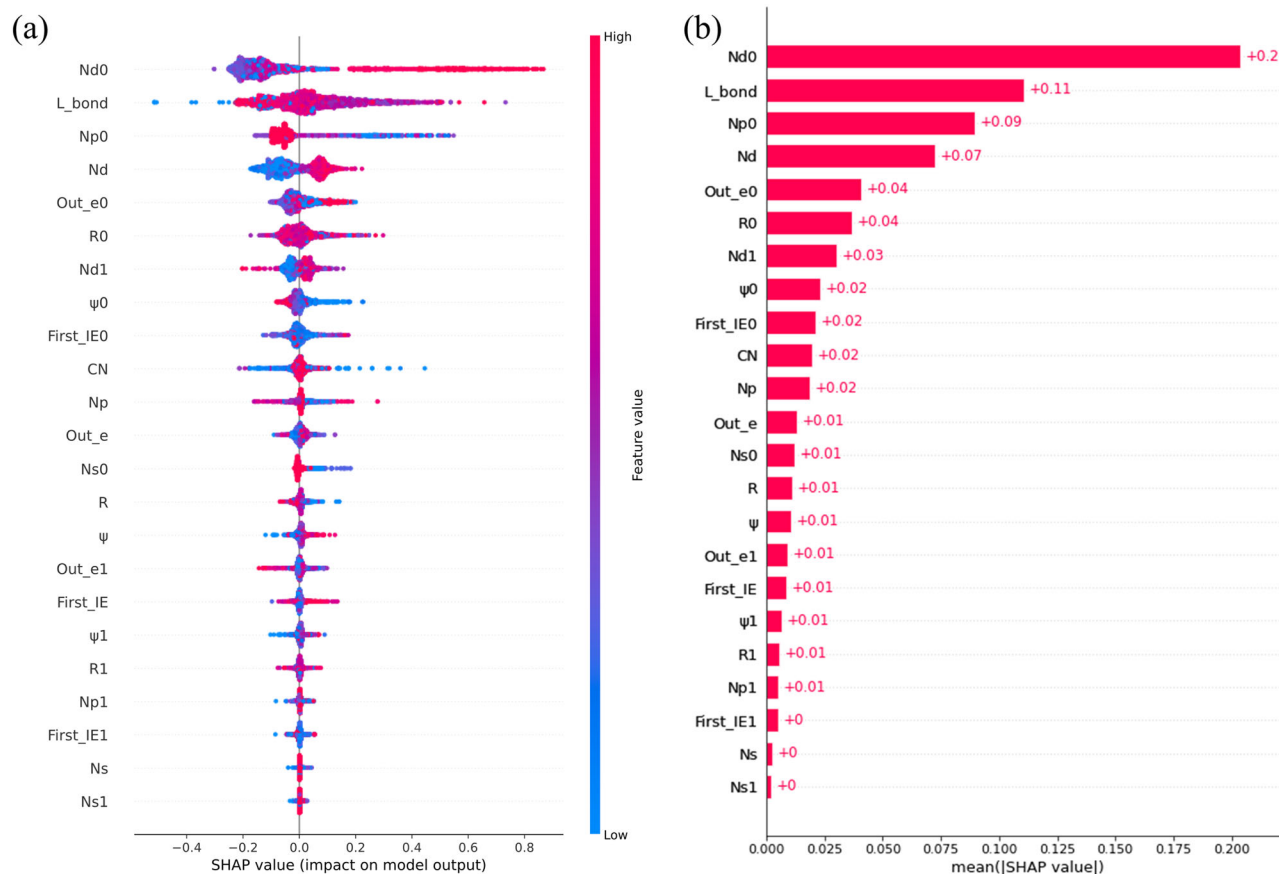


图5 | shap值分布分析用于评估所有功能的重要性。全球解释（平均特征重要性）和本地解释（塑造 Δg_H 的ETR模型的价值分布）。b对于 Δg_H ，每个输入功能的塑性功能在ETR模型上的重要性很重要。

系数，ND0，ND1和ND，均超过或等于0.4，表明D-Electron功能与 Δg_H 之间的显著关联。此发现与Nørskov的D波段中心理论^{47,48}一致。随着D电子的数量增加，D波段中心降低，导致更多的D电子占据抗BONT的轨道。这导致催化剂的不稳定增加，进一步削弱了H原子的吸附，并使 Δg_H 更高。为了更直观地确定每个功能对 Δg_H 的影响，我们采用了Shap方法来分析每个功能的重要性。形状值为特征重要性排名提供了新的见解。鉴于ETR模型在测试集（ $r^2 = 0.921$ ）上表现最佳，我们对ETR模型进行了SHAP评估。如图5a，b所示，所有特征的重要性都显示出来，其影响的程度由数据集中所有数据点的平均绝对形状值确定。这些功能是根据对模型输出的影响进行排名的，功能的重要性从上到下都降低。在图5A中，如果功能的形状值随着功能值的增加而增加，则表明与 Δg_H 的正相关。For instance, the feature Nd0 (The geometric mean of the d electron count of the adsorption center atom) exhibits a positive correlation with Δg_H , consistent with our aforementioned discussion: as the number of d-electron increases in the active center atom, more d-electrons occupy antibonding orbitals, resulting in weaker adsorption of H atoms and a more positive value of Δg_H . 一个值得注意的功能是功能 ψ 0（吸附中心原子的电负性的几何平均值），该功能表现出负面的，带有 Δg_H 的负相关性。更大的功能 ψ 0的价值表明具有更强的电子吸引了活动中心⁴⁹⁻⁵¹的能力。当吸附氢原子时，氢原子的电子转移到催化剂表面。较高的特征 ψ 0的值意味着从氢原子到催化剂表面的电子传输更大，从而导致催化剂对氢原子的吸附更强，因此， Δg_H 。

图5b显示，所有功能的平均绝对形状值均被用作数据库投票回归器的输入，而ND0对预测 Δg_H 值的预测具有最大的影响，其次是L_bond（随后是L_BOND（氢原子和供应中心之间的债券长度））。这两个特征的平均绝对外形值超过0.1 eV，这表明，这些特征平均有助于预测变化大于0.1 eV。

为了阐明功能ND0和 ψ 0对 Δg_H 的预测的影响，我们进一步分析了这些功能的外形值。图S6A，B分别说明了特征ND0和 ψ 0对形状值的影响。从视觉上看，随着功能ND0的值增加，Shap值的值会表现出向上的趋势（与 Δg_H 相关）。建议，对于功能 ψ 0，其价值的增加会导致外形值下降趋势（与 Δg_H 负相关）。SHAP特征分析结果验证了本研究中构建的电子和结构特征的有效性。但是，鉴于当前使用大量功能，需要优化，并且更相关的新功能正在等待发现。为了进一步优化ML模型，我们通过减少和结合现有功能来进行功能工程。

ML模型的功能工程和优化

功能工程对于ML是必不可少的，设计合适而全面的功能对于构建ML模型至关重要。因此，建立有效的催化特征对于增强模型准确性和探索因素影响反应至关重要。功能工程将单个电子和元素特性转化为复合特征，并为催化反应开发合适的特征。使用SHAP方法对特征重要性的分析揭示了活动中心D电子数量的几何平均值ND0与活性中心电负性的几何平均值 ψ 0之间的正相关性

Δg_H ，而观察到负相关。基于上述讨论，我们引入了一种与能源相关的新功能 φ ，定义如下：

$$\varphi = \frac{ND0^2}{\psi_0} \quad (2)$$

对该公式的分析表明， $ND0^2$ 和 $1/\psi_0$ 与 Δg_H 正相关，这表明新引入的功能 φ 也与 Δg 呈正相关。

基于D波段模型和Muf-Fin-Tin轨道理论^{52,53}，我们建立了吸附能与D-电子计数的平方之间的关系。根据这些理论，金属表面上的吸附能(E_{ad})与金属D轨道与吸附物的耦合强度密切相关。这种耦合强度由哈密顿元素 a_{ad} 表示，它受到金属d轨道的空间范围(r_d)的影响，而崇拜者之间的距离则影响了

金属表面(L)： $E_{ad}/a_{ad}^2/r_d^3$ 。在这里， r_d 是金属D-轨道的空间范围，它与活动中心原子的D-电子($ND0$)的数量直接相关。具体而言， $ND0$ 既影响金属的D波段中心的位置，又影响D轨道的空间范围。因此， R_d 可以通过金属的电子结构间接反映。

从上面的关系中，我们观察到吸附能量 E_{ad} 与空间范围 r_d 成正比，这受D-电子数量的影响。特别是随着D-电子计数的增加，空间范围 r_d 的D轨道增强，从而增强了金属和吸附物之间的耦合强度。因此，吸附能与D-电子计数(E_{ad}/a_{ad}^2)具有二次关系。这种二次关系反映了金属的D-轨道电子之间的协同作用，以及电子云的重叠如何影响通知者和金属表面之间电子相互作用的强度。随着D-电子计数的增加，金属的D-轨道电子云膨胀，从而增强了电子云的重叠。这种重叠不仅是加性的，而且导致由于电子-电子相互作用(例如电子排斥和轨道重叠效应)而导致的吸附强度非线性增加，这是二次关系所捕获的。此外，使用金属的电负性间接估计我们模型中的吸附率 L ，从而进一步将金属的电子结构与吸附能联系起来。这种方法还将金属的电负性与D电子数量之间的关系结合到吸附能模型中，新引入的特征 φ 的物理和化学显着性已得到进一步确定。

此外，PCCM分析发现，最近活动位点的纤维与活动位点的整体特征之间存在很高的相关性。但是，在ML培训期间，过多的功能可以降低培训效率并影响预测准确性。为了解决此问题，必须采取适当的措施来消除数据集中的冗余信息。值得注意的是，由于“维度的诅咒”，相对较大的特征空间并不一定会导致更准确的预测。在高维空间中，数据变得稀疏，导致模型过度拟合并增加了计算时间。因此，我们提出了一个严格的特征选择过程，以消除重要性低或相关性高的特征。

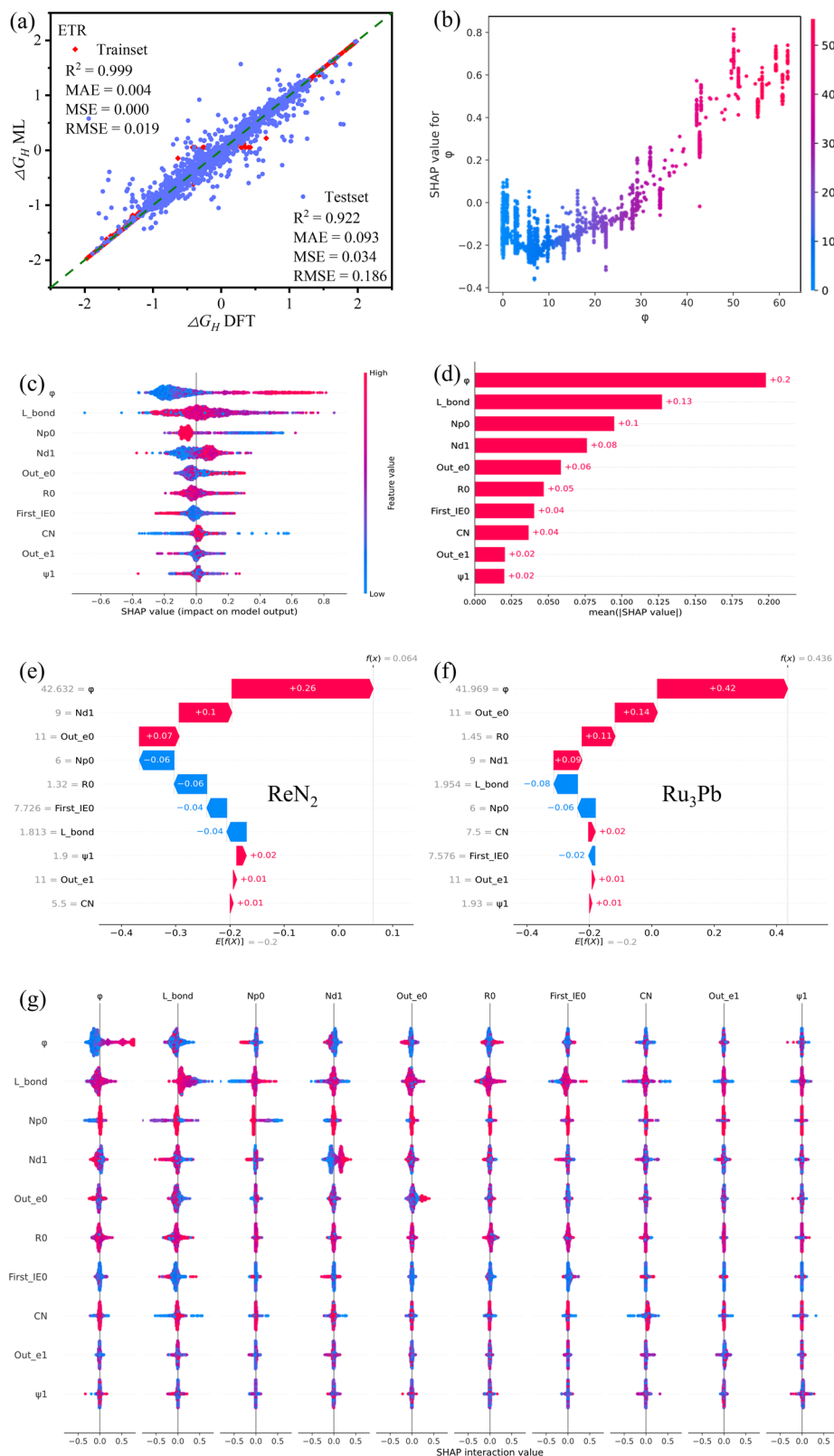
特征调整如下：1) 具有皮尔逊相关系数的特征，如果其外形值较低，则去除了大于85%的特征。2) 引入了一项新功能，即 φ ，并删除了两个相关功能 $ND0$ 和 ψ_0 。3) 消除了低外形值的特征。功能工程后，图6C中只有10个功能。随后，使用这十个特征构建了ETR模型，如图6a所示。仅使用这10个功能构建的ML模型的 R^2 值为0.922，MAE，MSE和RMSE值分别在测试集上分别为0.039、0.034和0.186。该模型使用23个功能优于ML模型，展示了我们功能工程的有效性，而新建的功能 φ 具有更高的相关性

凭借 Δg_H ，并且在预测各种HEC的活动中起着至关重要的作用，表S3中提供了具有十个功能的ETR模型的详细参数。图6b显示了与新引入的功能 φ 相对应的形状值，表明随着功能 φ 的增加，形状值增加。这表明功能 φ 和 Δg_H 之间存在明显的正相关。随着功能 φ 的价值增加， Δg_H 变得更大，这意味着H原子的催化剂的吸附能力较弱。图6C，d中显示了10个特征的重要性分析的形状值。这些数字还说明了能源相关的功能 φ 和 Δg_H 之间的牢固正相关。此外， φ 具有最高的绝对形状值，这表明它对 Δg_H 的影响最大。除了基于物理和化学洞察力和外形价值分析的降低性降低外，还研究了替代维度降低方法对ETR模型准确性的影响。L1常规ization通过零来缩小某些特征的权重来减轻这些挑战，从而减少了特征的数量并减轻了在高维空间中过度拟合的风险。此外，降低降低技术，例如主成分分析(PCA)和T分布的随机邻居嵌入(T-SNE)，可以有效地降低模型的复杂性并提高性能。使用L1正则化，将功能集降低到9个维度，将功能 $ND0$ ， $ND1$ ， ND ， ND ， $NP0$ ， NP ， NP ， OUT_E0 ， OUT_E1 ， OUT_E1 ， ψ_0 和 $FIRFT_IE$ 保持，ETR模型可实现测试集预测准确性 $R^2=0.786$ 。PCA将功能集降低到十个维度，从而导致测试集预测准确性为 $R^2=0.890$ 。对于T-SNE，该功能集降低到2和3维度，得出的测试集预测精度分别为 $=0.873$ 和 $R^2=0.876$ 。图S7提供了详细的模型性能指标。这些比较表明，在物理和化学见解和塑造价值分析的指导下，维度降低至10维是一种合理的方法，可以进一步提高ETR模型的准确性。

为了证明在 Δg_H catalysts上功能的预测性能，我们从测试集中选择了两个催化剂Ren和Ru pb，并使用ETR模型预测了其 Δg_H 。预测的结果如图6e，f。DFT计算(ML预测的) Δg_H 的ren和Ru PB值分别为0.063 eV (0.064 eV)和0.430 eV (0.436 eV)。在功能中， φ 为预测的 Δg_H 值做出了最大的贡献，强调了 φ 在预测 Δg_H 中的重要性。图6G划分了所有功能的互动图，其中 φ 和 $ND1$ 与 Δg_H 的正相关性很强，并且在功能本身之间没有观察到显着的相关性。

为了进一步证明ML模型的预测准确性，我们将其与两个DL图神经网络模型进行了比较。在这里，我们选择了深度学习模型CGCNN和OGCNN进行比较，因为两者在预测晶体特性方面都表现出了出色的性能，尤其是在固态材料和材料科学领域^{54,55}。例如，在有关CGCNN和OGCNN的原始文献中，CGCNN模型可实现出色的预测准确性，误差为0.039 eV和0.072 eV，用于材料形成能量和材料项目数据库中的绝对总能量。此外，带隙，费米能和泊松比的预测误差分别为0.388 eV，0.363 eV和0.030。OGCNN模型还显示出出色的预测性能，材料形成能，费米能和带隙的 R^2 值分别为0.996、0.91和0.91。此外，最近的作品也证明了它们的出色表现。例如，Jin-soo Kim及其同事利用CGCNN模型预测分解峰和无机perovskites⁵⁶的分解峰和能带的直方图，其 R^2 值为1和0.986，平均绝对错误为0.449 MEV ATOM⁻¹和0.037和0.037 eV，相应地分散。近年来，这些模型被认为是晶体结构预测领域中的基准DL模型，在准确性和一代化能力方面都表现出色。两种DL型号均经过全面训练。图7a，C分别显示了训练集和CGCNN和OGCNN的验证集的损耗函数曲线。从这些数字中可以看出，在300个时代后，两个模型的损失都融合了，表明这两个模型都有

图6 |模型性能和功能分析后，功能工程后。ET R模型在培训和测试集上的准确性是具有十个fe the Feature φ 的外形值分布。C全局解释（平均特征重要性）和本地解释（Shap值分布）和d s hap值的每个输入特征在ETR模型上具有十个功能 Δg_H 的重要性。每项功能对 Δg_H 的\$ rRu p b预测的贡献。G基于重要性的所有功能的交互图。



充分学习了数据集。图7b, D显示了CGCNN和OGCNN在培训，测试和验证集中的学习曲线。对于ETR模型而言，CGCNN和OGCNN模型的 R^2 值分别低于ETR模型的 R^2 的0.913和0.921，均低于0.922的 R^2 ，进一步确定了所提出的模型的高精度和所选功能的重要性。

此外，我们独立地评估了两个最重要的功能（ φ and L_bond）对ETR模型的准确性的影响，我们仅使用一个功能 φ 建立了一个ML模型来预测测试集。如图S8A所示，该模型的精度仅具有一项功能 φ 为 $r^2 = 0.51$ 。仅使用功能 φ 的测试集数据的 Δg_H 的预测表现出明确的线性关系，表明功能的潜在应用程序值。

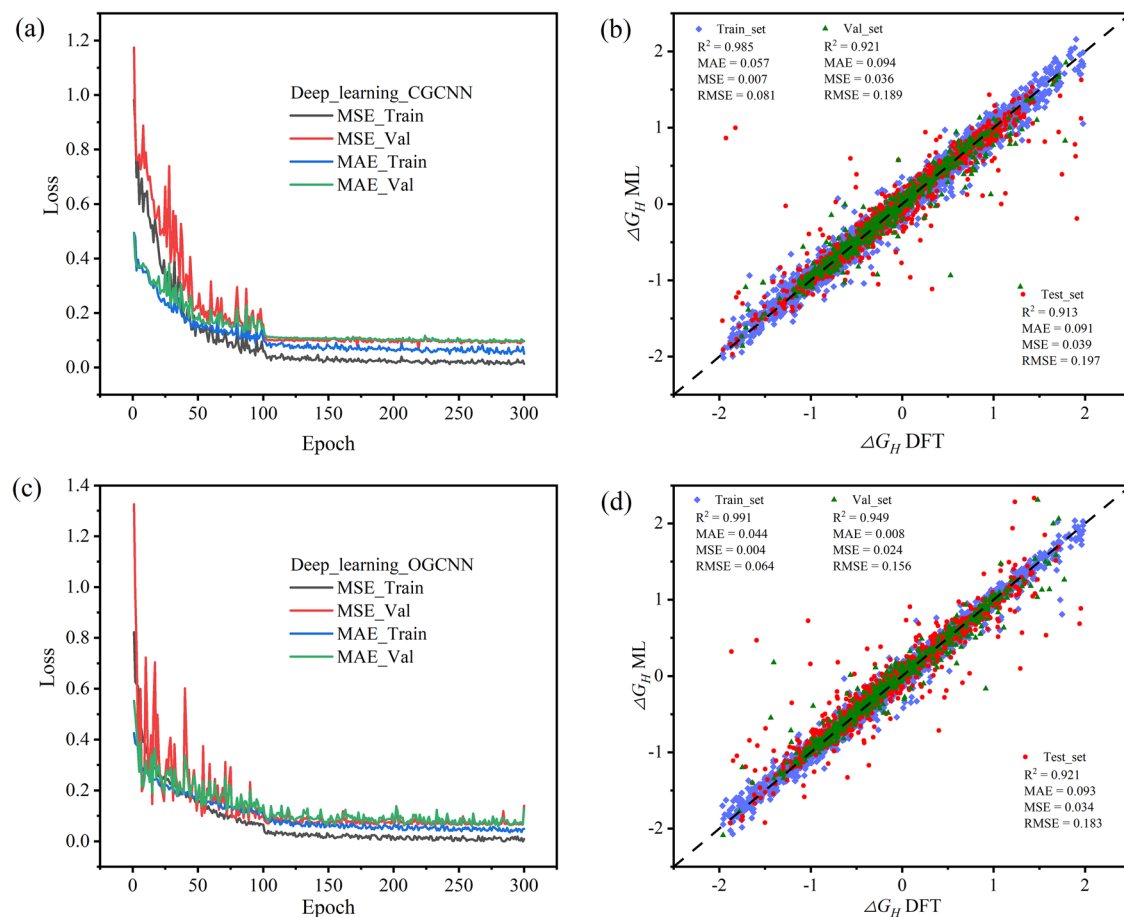


图7 | 基于图形卷积网络的两个DL模型的预测准确性。训练集和验证集的损耗函数曲线，用于CGCNN和C OGCNN，B CGCNN和D OGCNN的学习曲线在培训，测试和验证集中。

ϕ 预测 Δg_{H^*} 。为了进一步探索引起HEC性能的关键因素，我们仅使用两个功能， ϕ 和 l_{bond} 建立了ML模型，以预测测试集，如图S8B所示。预测精度达到了 R^2 值为0.741。仅使用这两个特征，准确预测了催化剂的氢演化性能，表明特征 ϕ 和 l_{bond} 对于确定催化剂的氢进化性能至关重要。

高活性催化剂筛选

该ML模型的目的是预测其潜在在催化剂的催化活性。使用我们设计的10个功能开发的ETR模型，用于预测从材料项目数据库⁵⁷中获得的 Δg_{H^*} 。随后，我们使用DFT计算验证了ML模型的预测。为了确保DFT计算设置与ML数据集的一致性，我们决定不考虑本研究中显式或隐式水的影响。猫层涵盖了各种类型，包括过渡金属金属间化合物，轻质金属间的金属间化合物，非金属组合和钙钛矿，总共进行了132个吸附位点Gibbs自由能计算。值得注意的是，我们使用的验证集与用于构建ML模型的训练集并未重叠。对此验证集的预测有助于验证我们ML模型的准确外推能力。我们采用了ML模型来预测这些HEC的 ΔG_H ，然后使用DFT进行详细的计算。为了验证ML模型的准确性，我们将DFT计算结果与ML模型的预测进行了比较。图8a说明了DFT计算与所选催化剂的ML预测之间的比较分析，而表S4列举了DFT的结果

每个催化剂的计算和ML模型预测。验证集上ML模型预测的 R^2 值为0.878，MAE，MSE和RMSE值分别为0.167、0.040和0.200。这些结果表明，我们的ML预测与 Δg_{H^*} 的DFT计算之间具有很高的的一致性，这是我们模型的准确性。根据ML模型预测的数据，已经确定了几种高活性HEC。为了更直观地评估这些催化剂的功能，我们计算了催化剂的理论超电势。理论上的电势与 $\eta^{1/4} \Delta g_{H^*} / e$ 之间的关系。Figure 8b – d illustrate three excellent HECs, representing AB-type, AB₂-type, and perovskite-type structures: ReIr (id: mp-1219533) with vacancy H adsorption, Re₂W (id: mp-974416) with atop H adsorption, and BaNdO₃ (id: mp-54307) with atop H adsorption, with the numbers in括号对应于MP数据库中的ID。DFT- ΔG_H 的 H^* 值分别为0.009 eV，0.014 eV和-0.416 eV，而ML模型预测 ΔG_{xAS} - 0.174 \$ 0.174 \$ 0.174 eV，-0.017 eV和\$ 0.117 eV和0.1117 eV 2 25 ev25 eV。这三种催化剂的DFT计算和ML预测之间的 ΔG_H 的差异分别为0.185 eV，0.031 eV和0.299 eV。这三种催化剂的MAE为0.171 eV，非常接近图8 (a) 中0.167的MAE，反映了ML模型预测的普遍性和准确性，并进一步表明Reir和Re₂W是有希望的HEC。此外，DFT计算与ETR模型预测之间的效率比较揭示了使用ML模型预测 Δg_{H^*} 132吸附位点的效率增强了200,000倍，在图S9中设置的验证中的效率增强了132_H使用传统DFT的候选人和

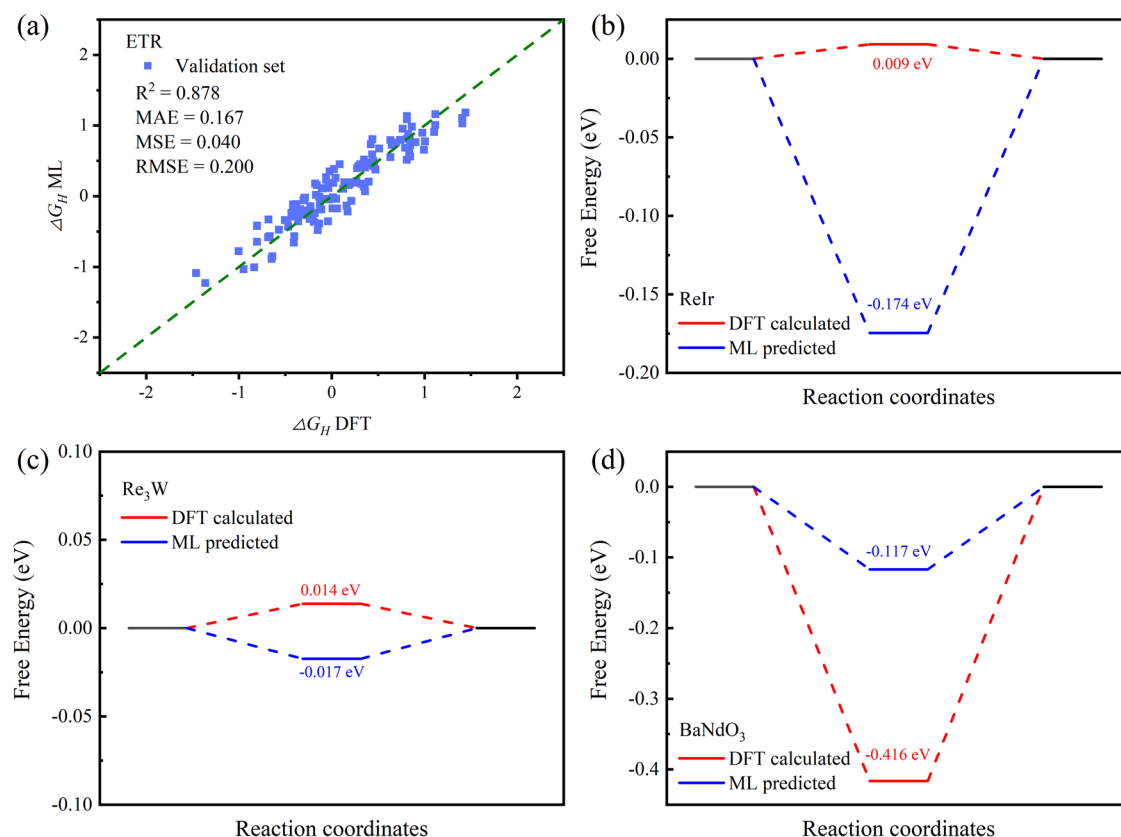


图8 | ML模型对潜在HEC的预测。ETR模型预测和DFT计算的比较图 Δg_H 132催化剂。比较DFT计算和ML预测的反应途径和 Δg_{HHH} 的值，分别为B Reir， $\text{C re}_3\text{W}$ 和DBando。

实验方法，从而有助于催化剂的加速发展和实际部署。

讨论

在这项工作中，一个包含10,855张她的催化数据的数据集是从催化中心数据库中收集的，用于培训和测试六个ML模型。功能重要性分析和功能工程技术已被用来最大程度地减少功能的数量，并引入了新的合并功能， $\varphi_{1/2}\text{nd}0^2/\psi 0 \$ 0$ ，与她的 Δg_H 相关，并且对她的活动有明确的物理解释。仅使用基于主动位点的10个功能建立了精确且有效的ML模型，而无需其他DFT计算，以预测其各种类型的催化剂。通过十倍的交叉验证，ETR模型在测试集上的 R^2 得分分别为0.922，MAE，MSE和RMSE值分别为0.039、0.034和0.186。此外，我们比较了两个深度学习模型CGCNN和OGCNN，发现我们的ML模型的预测准确性超过了这两者。为了进一步验证ML模型在预测各种催化剂方面的有效性，预测了 Δg_H 值为132个催化剂，并与DFT计算的结果进行了比较。根据ML模型的预测确定了有希望的HEC。与昂贵的DFT计算相比，ML模型在预测其catalysis方面的时间效率增长了200,000倍。这项工作中开发的ML模型可以预测各种类型的催化剂的吸附自由能，以及不同吸附位点的吸附自由能，从而有助于筛选潜在的催化剂。通过使用ML模型来预测每个吸附位点的自由能并评估这些位点的吸附强度，可以识别催化剂的最佳吸附位点并评估其催化活性。这项工作提出了实用的解决方案，可以在催化剂的广阔空间中发现高性能的催化剂，并为设计其他电催化剂的设计提供了见解。

方法密度功能理论计算

所有计算均基于密度函数理论 (DFT) 进行，通过维也纳AB INITIO仿真软件包 (VASP) 和DS-PAW^{58–60}进行。投影的波方法用于描述ION – 电子相互作用，其截止能量为500 eV⁶⁰，这对高效率进行了足够的精确测试。图S1显示了AGPD表面的总能量的变化，截止能量的函数范围从50 eV到500 eV，步骤尺寸为50 eV。结果表明，500 eV的截止能提供了足够的精度。Exchange – 相关互动由Perdew – burke – ernzerhof在广义梯度近似⁶¹的框架上确定。对于所有计算，将z方向的真空空间设置为15 Å，以避免周期性表面之间的潜在相互作用。将底部的两个原子层固定，而其余原子则完全放松，以允许结构优化。这种方法确保在催化反应期间的表面结构的准确表示，并将模拟系统保持在物理上逼真的状态。在这项研究中，我们所有的氢吸附模型均在低覆盖条件下设计，氢原子覆盖率低于25%。根据Langmuir的吸附理论和Nørskov等人的发现，^{36,62}，在低覆盖范围内，吸附原子之间的相互作用可以忽略不计，并且通常可以忽略覆盖效果。能量收敛标准为 10^{-5} eV，而在放松过程中，原子力为0.01 eV/Å。A $3 \times 3 \times 1$ 使用以伽马为中心的K点网格。使用DFT-D3方法⁶³考虑了Van der Waals (VDW) 交互的效果。

催化剂的活动趋势可以通过与氢在其表面上的吸附相关的吉布斯自由能阐明。为了使SABATIER原理，当氢吸附的Gibbs自由能接近接近零⁶⁴的热载值时，就可以实现最佳的催化活性。氢吸附的吉布斯自由能，表示

由于 Δg_H 在以下内容中，可以计算为：

$$\Delta G_H = \Delta E_H + \Delta E_{ZPE} - T\Delta S_H \quad (3)$$

其中 Δe_H 是氢吸附能，定义为：

$$\Delta E_H = \Delta E_{slab-H} - \Delta E_{slab} - \frac{1}{2}E_{H^2} \quad (4)$$

其中 Δe_{slab-H} 是氢吸附的平板结构的总能量，而 Δe_{slab} 是没有氢原子的平板结构的总能量，而 e_{H^2} 表示孤立的氢气分子的能量。 Δe_{ZPE} 和 Δs_H 是从振动频率计算中获得的零点能量（ZPE）和氢的熵的变化。为了计算H原子的振动频率，除了氢原子外，还固定了所有其他原子，并计算了氢的振动频率。获得频率计算结果后，使用VASP工具包⁶⁵确定系统的自由能校正。特定地，使用计算获得的频率来计算零点校正，然后将其与热力学公式结合使用，以估计与温度相关的自由能校正。T代表温度，并将其设置为Nørskov的工作中的298.15 K.，当 Δg_H 的绝对值为零时，她的催化活动是最好的^{64,66}。

ML模型

有效和合适的ML模型对于探索结构和催化活性之间的关系至关重要，并确保模型的稳定性，可靠性和非线性处理能力。树模型的鲁棒性和可解释性在ML模型中广泛使用。ML过程使用六种基于树的算法，即RFR，GBR，XGBR，DTR，LGBMR和ETR。此外，我们还比较了两种基于非树的算法：多层感知器（MLP，神经网络模型）模型和支持向量机（SVM）模型。这些算法可以与开源Scikit-Learn库无缝集成。DFT计算的输入数据被随机分为训练集和一个测试集，比为4：1。为了证明ML模型的稳定性和准确性，将10倍的交叉验证（CV）应用于所有算法，以找到所有模型的超参数的最佳组合。网格搜索是一种详尽的搜索方法，它为每个高参数设置一组候选值，然后生成这些候选值的笛卡尔产物，以形成超参数组合的网格。随后，网格搜索列车并评估了超参数组合的每种组合，以发现结合与最佳性能，从而提高了模型的预测准确性和鲁棒性。补充信息中详细介绍了针对特定算法模型的最佳超参数。此外，采用了两种基于图的基于神经网络的DL模型CGCNN和OGCNN。两种模型均以最多300个时期的训练，批量大小为512，并将Adam Optimizer用于参数优化。该数据集分别分为培训，测试和验证集，分别占数据的80%，15%和5%。ML和DL模型的精度评估指标包括四个度量，R平方（ R^2 ），平均绝对误差（MAE），均方误差（MSE）和根平方误差（RMSE）。四个评估指标的计算公式如下：

$$R^2 = 1 - \frac{\sum_i^n (y_i - f(x_i))^2}{\sum_i^n (y_i - \bar{y})^2} \quad (5)$$

$$MAE = \frac{1}{n} \sum_i^n |f(x_i) - y_i| \quad (6)$$

$$MSE = \frac{1}{n} \sum_i^n (f(x_i) - y_i)^2 \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i^n (f(x_i) - y_i)^2} \quad (8)$$

其中 $f(x_i)$ 是该模型的预测值； y_i 是真正的价值； \bar{y} 是平均值；MAE，MSE和RMSE可以被视为预测错误， R^2 大约被视为重新填充的准确性。

为了评估我们的ML功能之间的依赖性，应用PCCM来评估两个功能之间的相关性⁶⁷，可以表示为：

$$p = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (9)$$

其中x和y是两个功能， \bar{x} 和 \bar{y} 是相应的平均值。p的范围从-1到1。当P接近1时，这两个功能之间存在线性关系。

随后，Shapley添加说明（SHAP）分析是对型号⁶⁸进行的。Shap是阐明ML模型的广泛接受方法之一。在这种方法中，每个功能都被分配一个重要的量表，其中较高的绝对形状值表示对ML模型结果的更大贡献。此外，正面或负形值表明该特征对预测产生积极或负面影响。

数据可用性

支持这项研究发现的数据可从https://github.com/wangchaobjut/multi_type_hers.git获得。

代码可用性

支持这项研究发现的数据可从https://github.com/wangchaobjut/multi_type_hers.git获得。

收到：2024年9月9日；接受：2025年3月12日；

Published online: 24 April 2025

参考

1. Chu, S. & Majumdar, A. 可持续能源未来的机会和挑战。Nature 488, 294–303 (2012)。
2. Tachibana, Y., Vayssieres, L. & Durrant, J. R. 用于太阳能水的人工光合作用。Natter. Photonics 6, 511–518 (2012)。
3. Lewis, N. S. 提高太阳能利用的研究机会。科学351, AAD1920 (2016)。
4. Zhu, W., Bian, Z. & Lu, Y. 基于移动协调员路由算法的养猪场环境控制系统。精密性畜耕作2019年-第9届欧洲精密性畜耕会议上发表的著作，2019年ECPLF, 851–857 (2019)。
5. Nørskov, J. K. & Christensen, C. H. 朝着表面上有效的氢产生。科学312, 1322–1323 (2006)。
6. Patra, B. C. 等。无金属共价有机聚合物，用于电催化氢的演化。ACS Catal. 7, 6120–6127 (2017)。
7. Liu, F. 等。更好的氢进化电催化剂的理性设计用于水分解：审查。ADV. 科学。9, 2200307 (2022)。
8. Zhang, J. 等。电催化剂设计由氢进化反应的活性中间体引导。ADV. 能源母校。13, 2302436 (2023)。
9. Strmcnik, D., Lopes, P. P., Genorio, B., Stamenkovic, V. R. & Markovic, N. M. 氢进化反应催化剂材料的设计原理。Nano Energy 29, 29–36 (2016)。

10. Lu, J., Yin, S. & Shen, P. K. 碳进化反应的碳封装的电催化剂。电化学。Energy Rev. 2, 105–127 (2019)。11. Zhong, W. 等。固定在碳纳米管上的单铂原子的超温和辅助合成, 以有效地电催化酸性氢的进化。J. 能源化学。51, 280–284 (2020)。12. Feidenhans 'l, A. A. 等。贵金属无氢进化的催化剂设计和应用。化学Rev. 124, 5617–5667 (2024)。13. Liu, J. 等。合理地设计有效的电催化剂, 以进行直接海水分裂: 挑战, 成就和承诺。安格。化学int. ed. 61, E202210753 (2022)。14. Mahmood, N. 等。碱性电解质中氢进化的电催化剂: 机制, 挑战 and 前瞻性溶液。ADV. 科学。5, 1700464 (2018)。15. Tran, K. & Ulissi, Z. W. 跨金属间学的积极学习, 以指导发现CO₂减少的电催化剂和H₂ Evolution。纳特。催化。1, 696–703 (2018)。16. Gu, Y. 等。单原子模型的杂化过渡金属碳化物作为有效的氢进化反应催化剂。A DV. 功能。母校。31, 2104285 (2021)。17. Wang, Q. 等。高渗透合金电催化剂的最新进展用于氢进化反应。ADV. 母校。接口11, 2301020 (2024)。18. Mei, J., Deng, Y., Cheng, X., Wang, X. & Wu, Q. 氧和氢进化反应的铁基电催化剂的最新进展。下巴。化学Lett. 35, 108900 (2024)。19. Brunin, G., Ricci, F., Ha, V.-A., Rignanesse, G.-M. & Hautier, G. 使用高通量计算的透明导电材料发现。NPJ计算。母校。5, 63 (2019)。20. Yang, C. 等。克服对双金属催化剂图书馆的不陈旧性。科学。ADV. 6, EAAZ6844 (2020)。21. Freeze, J. G., Kelly, H. R. 和 Batista, V. S. 通过反设计搜索催化剂: 人工智能, 登山者和炼金术士。化学Rev. 119, 6595–6612 (2019)。22. Wang, M. & Zhu, H. 基于过渡金属的氢生成电催化剂的机器学习。ACS Catal. 11, 3930–3937 (2021)。23. Wang, C. 等。有效的机器学习模型, 重点是在二进制合金中发现双功能氧电催化剂的活性位点。ACS应用。母校。接口16, 16050–16061 (2024)。24. Mai, H., Le, T. C., Chen, D., Winkler, D. A. 和 Caruso, R. A. 电催化剂和光催化剂设计与发现的机器学习。化学Rev. 122, 13478–13515 (2022)。25. Li, J. 等。机器学习辅助的低维电催化剂设计, 用于氢进化反应。纳米-麦克罗莱特。15, 227 (2023)。26. Sun, X. 等。Mbenes材料中氢进化催化剂的机器学习加速筛选。应用。冲浪。科学。526, 146522 (2020)。27. Chen, L. 等。用于电催化剂创新的通用机器学习框架: 发现用于氢进化反应合金的案例研究。ADV. 功能。母校。32, 2208418 (2022)。28. Ren, C. 等。复杂的界面对电化学还原反应的影响的通用描述。J. Am. 化学Soc. 144, 12874–12883 (2022)。29. Gao, W. 等。用吸附物和底物的内在特性确定小分子的吸附能。纳特。社区。11, 1196 (2020)。30. Lu, Z., Chen, Z. W. & Singh, C. V. 神经网络辅助高素质合金催化剂的开发: 脱耦配体和协调效应。物质3, 1318–1333 (2020)。31. Zhang, J. 等。基于结构的准确且有效的机器学习模型, 用于预测氢进化反应催化剂和合金中的电子功能工程。纳米级15, 11072–11082 (2023)。32. Umer, M. 等。机器学习有助于过渡金属单原子的高通量筛选。J. Mater. 化学A 10, 6679–6689 (2022)。33. Wei, C. 等。基于简单特征的高H₂演化活动/CO₂降低选择性的双原子催化剂的数据驱动设计。J. Mater. 化学A 11, 18168–18178 (2023)。34. Winther, K. T. 等。Catalys-Hub.org, 一种用于表面反应的开放电子结构数据库。科学。数据6, 75 (2019)。35. Torres-Beltrán, M. 等。作者校正: Saanich Inlet Water柱上的地球化学信息汇编。科学。数据6, 1 (2019)。36. Nørskov, J. K. 等。氢进化的交换电流的趋势。J. Electrochem. Soc. 152, J23 (2005)。37. Li, H. 等。氢进化反应的(100)合金表面的计算设计。J. Mater. 化学A 8, 17987–17997 (2020)。38. Ung, V., Hu, G., Wu, Z. & Jiang, D.-E. 氮掺杂石墨烯中单原子催化剂上氢进化的描述。J. Phys. 化学C. 124, 19571–19578 (2020)。39. Wexler, R. B., Martinez, J. M. P. & Rappe, A. M. 化学压力驱动的增强Ni₂p的氢化活性从通过机器学习解释的非金属表面掺杂的Ni₂p。J. Am. 化学Soc. 140, 4678–4683 (2018)。40. Lin, S., Xu, H., Wang, Y., Zeng, X. C. & Chen, Z. 直接通过机器学习从石墨烯支持的单原子电催化剂的易于获得的物理特性预测限制电位。J. Mater. 化学A 8, 5663–5670 (2020)。41. GE, L. 等。基于基于机器量子力学中的机器学习分析, 基于硅化葡萄糖剂异质结构, 预测了用于氢化核酸盐的氧气进化反应和氧气进化反应的最佳双功能电催化剂。J. Phys. 化学Lett. 11, 869–876 (2020)。42. Zheng, J. 等。在MXENE材料中对氢进化反应催化剂的高通量筛选。J. Phys. 化学C. 124, 13695–13705 (2020)。43. Parker, A. J., Opletal, G. & Barnard, A. S. 使用机器学习对铂纳米颗粒催化剂的分类。J. Appl. 物理。128, 014301 (2020)。44. Norsko, J. K. 金属表面上的化学吸附。众议员物理。53, 1253 (1990)。45. Nørskov, J. K. 催化中的电子因素。prog. 冲浪。科学。38, 103–144 (1991)。46. Oya, N., Ikezaki, T. & Yoshie, N. 具有在室温下自修复能力的结晶超分子聚合物。多聚。J. 45, 955–961 (2013)。47. Nørskov, J. K. 化学吸附和异质催化理论。Physica B + C 127, 193–202 (1984)。48. Ruban, A., Hammer, B., Stoltze, P., Skriver, H. L. & Nørskov, J. K. 过渡和贵金属的表面电子结构和反应性。J. Mol. 催化。答: 化学。115, 421–429 (1997)。49. Singh, H., Prendergast, D. & Nath, M. 通过调整阴离子电负性的调节电催化活性: 用铜硫酸盐的案例研究。J. Phys. 能源5, 045016 (2023)。50. Wang, M. 等。将2D超薄Biox/RGO (X = F, CL, B R, I) 的纳米片作为硫宿主的促进电催化活性: 对卤素元素对液化元素对锂硫粉的电化学性能的电负性效果的见解。无机化学。正面。11, 4277–4287 (2024)。51. Pan, S. 等。用于水氧化催化的无定形金属有机框架中取代基的电负性调节。int. J. 氢。能源45, 9723–9732 (2020)。

52. Calle-Vallejo, F. 等。外电子作为过渡金属及其氧化物吸附过程的描述符数量。化学科学。4, 1245–1249 (2013)。53. Xin, H., Holewinski, A. & Linic, S. 用于快速筛选基于PT的多功能电催化剂的预测结构-反应性模型, 用于氧还原反应。ACS Catal. 2, 12–16 (2012)。54. Xie, T. & Grossman, J. C. Crystal Graph卷积神经网络, 可准确且可解释的材料特性预测。物理。莱特牧师。120, 145301 (2018)。55. Karamad, M. 等。材料性质预测的轨道图卷积神经网络。物理。牧师。4, 093801 (2020)。56. Kim, J.-S., Noh, J. & IM, J. 机器学习支持的化学空间探索光伏的全无机钙钛矿。NPJ计算。母校。10, 97 (2024)。57. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. 用于预测无机材料特性的通用机器学习框架。NPJ计算。母校。2, 16028 (2016)。58. Kresse, G. & Furthmüller, J. 使用平面波基集的从头开始计算的有效迭代方案。物理。Rev. B 54, 11169–11186 (1996)。59. 物理。Rev. B 59, 1758–1775 (1999)。60. Blöchl, P. E. 投影仪增强波法。物理。Rev. B 50, 17953–17979 (1994)。61. Perdew, J. P., Burke, K. & Ernzerhof, M. 通用梯度近似变得简单。物理。莱特牧师。77, 3865–3868 (1996)。62. Langmuir, I. 固体和液体的构成和基本特性。第一部分固体。J. Am. 化学Soc. 38, 2221–2295 (1916)。63. Grimme, S. 半经验GGA型密度功能, 该密度功能性构建的远程分散校正。J. Comput. 化学27, 1787–1799 (2006)。64. Nørskov, J. K., Bligaard, T., Rossmeisl, J. & Christensen, C. H. 朝着固体催化剂的计算设计。纳特。化学1, 37–46 (2009)。65. Wang, V., Xu, N., Liu, J.-C., Tang, G. & Geng, W.-T. Vaspkit: 使用VASP代码促进高通量计算和分析的用户友好接口。计算。物理。社区。267, 108033 (2021)。66. Valdés, Á., Qu, Z. J. Phys. 化学C. 112, 9872–9879 (2008)。67. 眼科生理学。选择。39, 316–327 (2019)。68. VegaGarcía, M. & Aznarte, J. L. Shapley添加说明, 否₂预测。ECOL. 通知。56, 101039 (2020)。

致谢

这项工作得到了中国国家密钥研发计划(授予No.2021YFB3500403)和中国国家自然科学基金会(GrantNo. 52305443)的全力支持。我们非常感谢HZWTECH提供的计算设施。

作者贡献

R.Z.W.提出了工作流和科学见解。C.W.设计了代码, 执行了计算, 分析了数据并撰写了原始论文。B.W.和C.H.W.审查并编辑了论文。Z.P.C.和A.J.L.收集了验证数据集。所有作者都为讨论做出了贡献。

竞争利益

作者没有宣称没有竞争利益。

附加信息

补充信息在线版本包含<https://doi.org/10.1038/s41524-025-01607-4>的补充材料。

材料的信件和请求应与Ruzhi Wang发给Ruzhi Wang。

转载和权限信息可从<http://www.nature.com/reprints>获得

出版商的注释Springer自然在已发表的地图和机构之后的管辖权索赔方面仍然保持中立。

开放访问本文是根据创意共享归因-非商业通用-诺迪毒素4.0国际许可证的许可, 该许可允许以任何中等或格式的任何非商业用途, 共享, 分发和复制, 只要您适当地归功于原始作者和来源, 并提供了与Creative Commons许可的链接, 并指示您调整了许可的材料。您没有根据本许可证的许可来共享本文或部分内容的改编材料。本文中的图像或其他第三方材料包含在文章的Creative Commons许可中, 除非在信用额度中另有说明。如果文章的创意共享许可中未包含材料, 并且您的预期用途不得由法定法规允许或超过允许的用途, 则需要直接从版权所有那里获得许可。要查看此许可证的副本, 请访问<http://creativecommons.org/licenses/by-nc-nd/4.0/>。

© The Author(s) 2025