

基于机器学习探索钙钛矿材料及其应用

胡 扬¹, 张胜利¹, 周文瀚¹, 刘高豫¹, 徐丽丽¹, 尹万健², 曾海波¹

(1. 南京理工大学材料科学与工程学院, 南京 210094; 苏州大学能源与材料创新研究院, 江苏 苏州 215006)

摘 要: 经过大半个世纪的算法模型积累, 以数据科学为基础的机器学习方法, 已经可以适配多项学科的研究需求。在理论与实验积累的数据基础上, 机器学习紧跟各个领域的研究潮流, 推动数据密集型科学研究的发展, 使其成为继“理论”、“计算”、“实验”后引领科学研究的“第四范式”。在材料科学领域, 钙钛矿材料具有构成丰富、带隙可调、发展空间广阔等优势, 但还未在其适用领域内达到环境友好等实用标准。因而基于机器学习探索钙钛矿材料及其应用, 不仅可以加速新型钙钛矿材料的发现, 而且可以探究钙钛矿材料种种优异性能与其物理化学特征之间的关联, 为发展环境友好型高性能钙钛矿器件提供指导。在此总结了机器学习结合钙钛矿材料的研究优势与研究流程, 综述了机器学习在钙钛矿材料性质与器件探索方面的研究进展, 探讨了当下面临的研究困境和挑战, 展望了未来的研究方向和发展趋势。

关键词: 钙钛矿材料; 机器学习; 材料物性; 光电器件

中图分类号: TB3; TP3 文献标志码: A 文章编号: 0454-5648(2023)02-0452-17

网络出版时间: 2022-12-27



Studies on Perovskite Material and Its Applications via Machine Learning

HU Yang¹, ZHANG Shengli¹, ZHOU Wenhan¹, LIU Gaoyu¹, XU Lili¹, YIN Wanjian², ZENG Haibo¹

(1. School of Materials Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;

2. Soochow Inst Energy & Mat Innovat SIEMIS, Soochow University, Suzhou 215006, Jiangsu, China)

Abstract: The methods of machine learning based on the data science can deal with the corresponding studies in different disciplines based on the data accumulated in theory and experiments. Machine learning promotes the development of data-intensive scientific discoveries, thus making it a "fourth paradigm" that leads to the related scientific research after "theory, calculation, and experimentation". Among different materials, perovskite material has some unique advantages of rich composition, adjustable band gap, and broad development space, but this material does not reach the practical standards such as environmental friendliness in applications. Therefore, the exploration of perovskite material and its applications based on machine learning can accelerate the discovery of novel perovskite material, and explore the relationship between the physical and chemical characteristics of perovskite material, therefore providing a guidance for the development of environmentally friendly high-performance perovskite devices. This review represented the research process of machine learning for perovskite material, summarized some research work on machine learning in perovskite material properties and device exploration, and discussed the existing difficulties and challenges. In addition, the future development direction and trend were also prospected.

Keywords: perovskite materials; machine learning; material properties; photoelectric device

钙钛矿材料是一类具有相似结构化合物的总称, 化学式概括为 ABX_3 , 最早发现存在于钙钛矿石

中的钛酸钙($CaTiO_3$)化合物^[1]。钙钛矿材料的价态多变, 早期发现的钙钛矿以 X 位为 O^{2-} 占据的氧化

收稿日期: 2022-09-21。 修订日期: 2022-09-24。

基金项目: 国家自然科学基金委员会重大研究计划-培育项目(91964103)。

第一作者: 胡 扬(1996—), 女, 博士研究生。

通信作者: 曾海波(1977—), 男, 博士, 教授;

张胜利(1984—), 男, 博士, 教授。

Received date: 2022-09-21. Revised date: 2022-09-24.

First author: HU Yang (1996—), female, Doctoral candidate.

E-mail: huDFyang@163.com

Correspondent author: ZENG Haibo (1977—), male, Ph.D., Professor;

ZHANG Shengli (1984—), male, Ph.D., Professor.

E-mail: zeng.haibo@njust.edu.cn;

zhangslvip@njust.edu.cn

物为主。近 10 年来, 卤化物钙钛矿因其在光电器件方面表现出的优异性能, 被广泛研究^[2-3]。下面以卤化物钙钛矿为主, 来介绍钙钛矿材料的基本结构。相较于 A 位, 钙钛矿材料的 B 位一般为离子半径较小的阳离子(如 Pb^{2+} 或 Sn^{2+}), 其与 X 位的 6 个阴离子(如 Cl^- 、 Br^- 或 I^-)配位构成八面体占据立方结构的顶角^[2-3]。而 A 位占据立方结构中心起到稳定钙钛矿结构的作用, 通常为有机阳离子[如 CH_3NH_3^+ (MA^+), $\text{NH}_2\text{CH}=\text{NH}_2^+$ (FA^+)]或者半径较大的无机阳离子(如 Cs^+), 分别称为有机-无机杂化钙钛矿(HOIPs)与全无机钙钛矿材料^[1, 4-5]。钙钛矿材料各个位置的离子可以被半径相近的元素或基团取代, 因此类型丰富多样, 具有巨大的挖掘空间。除了上述提到的 2 种类型外, 若 B 位交替存在 1 价和 3 价阳离子(如 $\text{Cs}_2\text{AgBiCl}_6$)则称为双钙钛矿(DPs)^[6-7], 化学式为 $\text{A}_2\text{B}(\text{I})\text{B}'(\text{III})\text{X}_6$; 若 A 位、B 位或 X 位由 2 种或 2 种以上离子混合占据时, 则称为混合钙钛矿^[8-9]。为了探索满足特定应用需要的钙钛矿化合物, 各种新型钙钛矿层出不穷。如果将元素掺杂纳入考量, 那么潜在的钙钛矿数量可能达到 10^7 量级^[4, 10]。钙钛矿材料体系不仅种类丰富, 而且具有许多优越的性质, 包括较高的光吸收系数、合适且可调的带隙宽度、窄的发射带宽、较长的载流子扩散长度、高的载流子迁移率、高的光致发光量子产率和低成本的溶液可加工性等^[11-18]。这些特性使钙钛矿材料在太阳能电池^[19-20]、发光二极管^[11, 21]、光催化材料^[22]、光电探测器^[12]和热电器件^[13]等领域成为了很好的候选材料, 尤其在太阳能电池领域, 钙钛矿材料仅用 10 年就达到了传统晶硅电池 50 年才达到的转化效率, 超过了 25%^[23]。

钙钛矿材料虽然具有颇多优点, 但是现阶段其应用仍停留在实验室阶段, 一些无法忽视的缺陷严重限制其替代传统材料大规模投入工业生产。以钙钛矿太阳能电池为例, 虽然其能量转换上限理论上远高于晶硅电池, 但是目前其相较晶硅组件, 稳定性差、电池衰减率高^[24]、无法批量生产和应用于大尺寸设备、对环境不友好(目前性能最好的钙钛矿光伏电池都含有有毒元素铅)^[18, 25]、存在“深阱态”缺陷导致光能以热量耗散转化率降低等问题, 严重限制了钙钛矿优势的发挥与其在工业生产中的应用^[26-27]。尽快针对性解决领域内问题、克服相关影响因素、最大限度发挥钙钛矿材料的本征优势、使钙钛矿材料尽早从实验室迈向产业化是目前钙钛矿材料最迫切的研究需要。

传统的材料开发方式通常基于反复试验试错, 不断进行实验合成、表征与测试, 直至材料达到目标性能。有限数量的材料和复杂的实验程序往往使得该方法对资源和设备的要求很高, 需要较长的研究周期且损耗昂贵, 同时又难以弄清楚材料内在性质和目标属性之间的关系^[28-29]。材料的第一性原理模拟计算方法^[30-33], 虽然无需进行实验合成就可以获得材料的一些关键特性, 但是计算软件需要进行大量繁琐的泛函方程计算以及近似约化。大多数计算方法仅针对特定系统, 复杂材料体系的计算量过大, 导致时间和计算资源成倍占用。尤其在钙钛矿材料研究问题上, 因为其化学搜索空间巨大、晶体组分结构复杂和多目标特性, 无论是实验方法还是第一性原理计算方法都无法达到对钙钛矿材料体系大规模快速高效率预测的需求。

机器学习作为人工智能的重要分支, 结合了计算机科学、统计学、数学和工程学等领域知识, 使计算机能够从以往的科学研究数据和统计信息中, 学习挖掘出有效的关联信息及其隐含规律, 并评估或预测研究对象^[34-36]。相较于“理论”、“计算”、“实验”三大传统研究方法, 机器学习可以不依赖具体的物理化学理论机制, 仅凭数据描绘材料体系“成分-结构-工艺-性能”相互之间的关系与趋势^[37-41]。这不仅节省了大量资源与时间, 而且相比计算与实验方法更快更准确。在良好的机器学习范例中, 训练得到的模型规律不仅在训练样本上表现良好, 而且同样适用于样本外的数据。这就意味着通过机器学习不仅可以总结材料体系内部特征关联, 还可以预测新的材料体系及其性质, 加速实用性材料的发现^[42-46]。可见机器学习对钙钛矿材料的设计、合成、物性探索与应用研究具有广泛的意义。

本文首先介绍了机器学习运用在材料研究中所要经过的“数据收集-数据处理与特征工程-模型训练-模型评价及解释-模型应用”的全部流程以及各个环节, 展开叙述了流程中的各项具体内容及在各个环节的材料研究中会遇到的问题, 概述了现有的可行处理方案。然后分别介绍了机器学习在钙钛矿材料可成形性、稳定性和带隙性质目前的最新研究进展, 以及机器学习在钙钛矿太阳能电池、发光二极管和光催化剂领域的应用情况。最后, 梳理了当下机器学习结合钙钛矿材料研究的主要方向及问题, 探讨了可能的解决思路, 展望了未来发展的趋势, 为对基于机器学习探索钙钛矿材料及其应用的研究人员提供了可借鉴的科研思路。

1 机器学习方法

1.1 机器学习在材料研究中的应用流程

机器学习在材料研究的流程一般为 3 个阶段(图 1): 1) 前期, 确立研究目标与定义、数据准备、数据处理以及特征工程。首先, 确定研究问题后, 将其梳理为机器学习的建模任务, 清楚任务的目标属性、可能影响目标属性的材料特征或性质(描述符)^[47]以及模型预期的评价标准等。根据研究要求选择不同的数据获取方式。一般来说, 所参考的数据可以从高质量论文的总结性图表、同标准实验室的数据积累以及公开可用的数据库中收集到。然后对收集到的数据进行处理, 保证其呈标准分布^[48]、无数据项缺失、无异常值等^[49–51]为后面的模型学习提供高质量的数据样本^[44, 48]。特征工程以材料体系的物理化学特征信息(例如离子的有效半径, 原子的第一电离能等)以及其复合参数因子(例如下文提到的钙钛矿结构的容忍因

子)等为基础, 对数据集中的每一条数据提供其独一无二的“数据指纹”。2) 中期, 模型及算法选择与参数设置、模型训练与测试及模型评价。为验证得到的模型具有鲁棒性, 一般将数据集分割为 2 个部分。选择模型及算法后, 一部分数据作为训练数据用于训练模型, 另一部分则作为测试数据, 验证模型对训练数据以外其他数据拟合的优劣程度。再对模型进行评价, 未达到标准可以重新选择模型或调整参数。3) 后期, 模型解释以及模型应用。对得到的机器学习模型进行分析解释, 解释模型如何运作, 模型结果如何发生。例如判断特征的重要程度, 以及某些重要特征如何影响目标属性的大小等。通过对模型的解释探索模型的实际意义, 扩展模型的应用空间。最后应用模型对新的材料数据进行判断预测, 通过材料的第一性原理计算、材料合成、材料表征以及材料测试等方法进行检验, 从而发现已知材料的优异性质或者领域适用的新型材料。

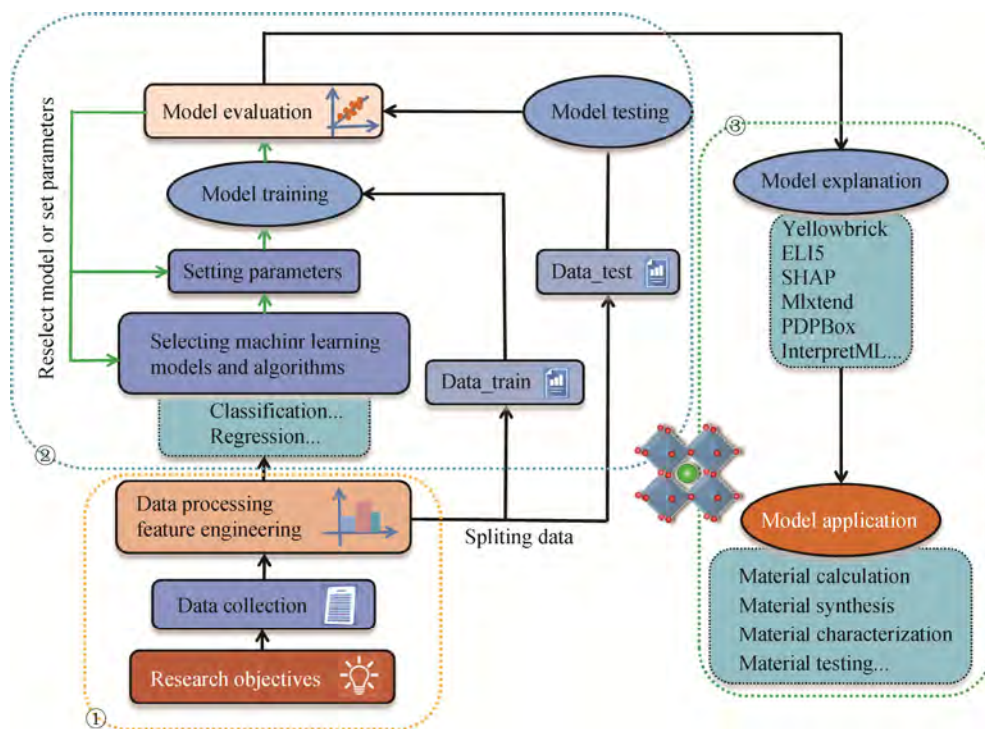


图 1 机器学习在材料研究中的应用流程示意图

Fig. 1 Schematic diagram of the application process of machine learning methods in materials research

1.2 机器学习预备: 数据采集处理及特征工程

数据质量是机器学习模型质量的基石, 它从根本上决定了模型是否可信。在机器学习的大部分工作中, 数据与特征的选择决定了模型效果的上限, 后续对模型与算法的选择等只能无限接近这个上限。模型输入的特征数据与其目标属性相关联^[52–53], 且唯一

定义了输入数据集中的每种材料。对材料特征的选择应当遵循“Occam 剃刀原理”, 剔除无用信息, 保留关键易得的信息, 达到“既简单又准确”。但是由于材料性能的优劣受多种物理化学因素的复杂影响, 所以对于未知驱动机制的材料性能, 一般会被定义很多相关的特征来提高预测的准确率。材料的物理化学特征数

据可以从公开的数据库中获取^[54], 数据库中的信息具有统一的数据类型与结构, 可为数据收集整理节约大量时间。高质量期刊论文与实验室积累的材料性质及器件性能数据通常品质较高, 但是数据样本往往较小, 尤其在一些新兴材料领域, 这与可生成的高维度材料特征矩阵产生矛盾^[55]。一般情况下需要样本数量远高描述材料的特征个数, 才能准确表达材料的目标属性^[52]。在这种高维度与小样本数据的矛盾下, 不仅不利于从数据中分割出测试数据, 且模型很容易出现过拟合(图 2a)以及效率低下等问题^[56-57]。

这一问题有 3 个解决方向^[58]: 1) 灵活运用材料领域知识, 将材料领域专家对描述符重要度的评估

进行粗粒度定量化表示, 与描述符相关评估算法结合, 判断特征与目标属性相关度^[57-60], 筛选关键的特征信息并过滤重复的依赖度高的冗余特征, 在保留重要信息的前提下缩小特征矩阵维度; 2) 通过生成对抗网络(GAN)模拟学习已知材料数据的分布与规则, 生成新的虚拟材料数据^[61], 并辅助数理模型与材料知识对虚拟数据进行分析判断, 保证生成数据的质量, 以扩大数据样本规模; 3) 通过选择主动学习和半监督学习方式, 利用未标记数据或部分标记数据优化预测模型参数解决小样本问题^[62-64], 或者利用迁移学习方式将相关模型的经验迁移过来以提升模型的性能^[65-66]。

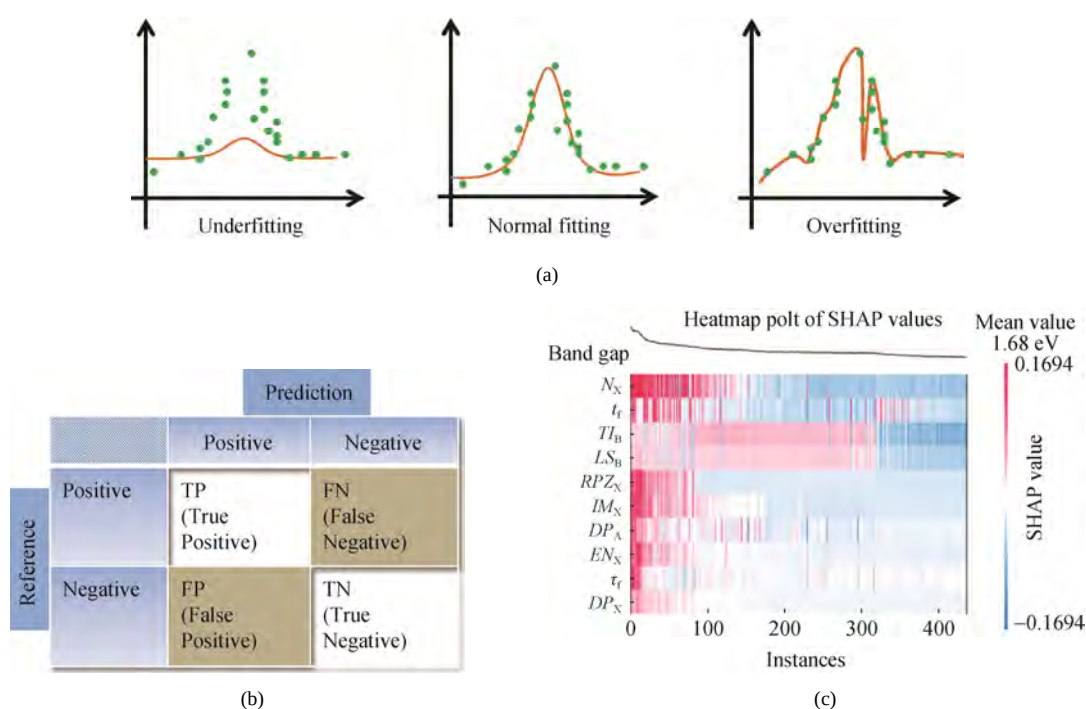


图 2 (a)机器学习模型的欠拟合、正常拟合以及过拟合情况示意图; (b)二分类问题结果的混淆矩阵; (c)SHAP 值与特征值的分布关系图举例^[67]

Fig. 2 (a) Schematic diagram of the machine learning model underfitting, normal fitting and overfitting; (b) Confusion matrix of the results of the binary classification problem; (c) Distributions of SHAP values with feature values^[67]

1.3 常见的机器学习问题与其对应的算法模型

根据学习方式的不同, 机器学习可以大致分类为监督学习^[45]与无监督学习^[68]。所谓“监督”是指对模型输入的数据是否存在标注, 也就是说对数据的预测结果是否存在“参考答案”。监督学习本质是想要学习到输入数据与“答案”之间的映射, 对应到材料研究当中就是找到材料特征参数与目标属性之间的统计规律。监督学习问题主要有 2 种: 分类与回归^[69]。分类模型的目标是预测属性的类别, 在二分类问题中, 一般将一种类别标记为正, 另一种则标记为负。例如在 2.1 节对钙钛矿成形性的判断中,

将具有钙钛矿结构的 ABO_3 化合物标记为正样本, 非钙钛矿材料的 ABO_3 化合物标记为负样本。回归模型的目标是预测一个连续值, 例如在 2.3 节中预测钙钛矿材料的带隙大小就是一个回归问题。监督学习的许多算法既可以用于分类问题也可以用于回归问题, 例如线性回归模型可以为分类问题设置阈值, 大于某个预测值时标记为正, 反之标记为负。常见的用于监督学习的算法有: K 近邻、线性回归模型、朴素 Bayes 分类器、决策树、核支持向量机和神经网络等。

无监督学习的输入数据没有被标记, 没有确定

的结果参考。例如聚类就是典型的无监督学习问题, 聚类问题需要根据样本间的相似性对其进行分类, 使类的内部数据差距最小化, 而类与类之间的差距最大化。无监督学习算法可以对数据进行预处理, 例如对数据降维与缩放^[70], 识别数据中的异常值或是异常数据点以及进行特征提取。常见的用于无监督学习的算法有: K 均值、主成分分析、非负矩阵分解和随机森林等。一些算法既可以用于监督学习也可以用于无监督学习。除了这 2 种大致分类外还有半监督学习^[63]、自监督学习^[71–72]、主动学习^[73–74]、强化学习^[75–76]和迁移学习^[66]等机器学习方式。

1.4 机器学习模型评价及算法适用性

评价机器学习分类问题模型的一般参数: 1) 精度: 用于确定正确分类的样本总数(TP+TN)占总样本数(TP+FN+FP+TN)的百分比; 2) 准确率: 表示正确判断为正的样本数(TP)占预测为正样本总数(TP+FP)的百分比, 可以反映结果的可靠性; 3) 召回率: 正确判断为正的样本数(TP)占数据正样本总数(TP+FN)的百分比, 表达了正样本被正确检出的概率(各项缩写参考图 2b: 二分类结果的混淆矩阵)。

机器学习回归模型评估一般参数: 1) 拟合优度 R 方值(R^2)。拟合优度评价了回归系数对观测值的拟合程度, 数据点分布越集中且贴合回归直线, 拟合优度越好。度量拟合优度的统计量为 R^2 。 R^2 的值在 0~1 之间, 越接近 1, 说明拟合程度越好; 越接近 0, 说明拟合程度越差。2) 均方差(MSE), 又称 L_2 损失

函数, 其值 $MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$ 度量了样本属性

实际值 y_i 与预测值 $\hat{y}_i$ 差的平方和均值(n 为样本数量)。MSE 只考虑误差的平均大小而不考虑方向, 避免了正负误差互相抵消带来的影响。但是因为使用了平方计算, MSE 会赋予异常点更大的权重, 使异常点的误差平方之后在求和中占比很大。如果该异常点没有意义且没有在上一步数据预处理中进行剔除, 那么用 MSE 评价模型就不太准确。另外, 均方根误差(RMSE)可在 MSE 的基础上开根号求得。

3) 平均绝对误差(MAE), 又称 L_1 损失函数, 其值

$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$, 与 MSE 类似可以衡量样本

整体与模型预测值偏离程度。但和 MSE 不同的是 MAE 对异常点不敏感, 鲁棒性更好。

想要提高机器学习模型的评估参数与预测性能: 1) 可以根据数据集规律与算法适用性, 挑选不同类型的模型分析对比预测精度。通过总结统计和

数据可视化结果, 识别数据范围、集中趋势、分布布局等可以帮助挑选更合适的模型算法。例如当数据集呈现高维离散特征时, 逻辑回归算法因为可以使用正则化, 避免过拟合情况而实现良好的预测结果; 随机森林模型因为对缺失值不敏感也可以有很好的预测效果。神经网络算法可以很好地处理大量数据和大量特征的问题, 但其解释性不高且内存占用密集也是运用该算法时不可忽略的问题。支持向量机算法有严格的数学理论支持, 可解释性强且鲁棒性好, 但是其训练的运算量与存储量都很高, 适合数据集样本量较小的情况。2) 可以将多个独立模型进行集成提升综合性能。例如 Lu 等^[67]提出了一种加权投票回归机器学习模型来预测 HOIP 材料的带隙。该模型综合了 4 个子模型嵌入集成模型, 对模型进行了交叉验证后得到的 R^2 值为 0.95, 测试集 R^2 值为 0.91, 均高于单个机器学习模型的预测评价。

1.5 机器学习模型的可解释性

模型的可解释性在材料领域研究中具有重要的意义。机器学习模型的可解释性可以明确材料优异性能的驱动机制, 使模型拥有更强的泛化能力。借助机器学习的可解释性结果, 可以判断对比以往领域内某些材料特征对属性影响的结果偏差, 从而调整材料的设计思路, 更明确地指导材料的设计合成与性能探索, 可以说模型的可解释性直接决定了其易用性。复杂的模型带来了更小的预测偏差以及更强的鲁棒性, 但与此同时可能使得机器学习的过程成为“黑匣子”, 而更难理解模型运作规律。

为避免在模型的准确性与解释性之间做取舍, 解决模型准确性与易用性之间的矛盾, 需要: 1) 在特征工程阶段嵌入材料领域知识, 合理筛选与转化描述符, 避免纯数据驱动的机器学习结果与领域内专家知识产生矛盾, 在模型上嵌入可借鉴的材料领域专家对材料数据潜在模式的理解, 将材料问题“化繁为简”, 建立易解释的模型框架^[77]。2) 在模型训练完成后, 灵活借助与之匹配的机器学习工具分析解释模型, 让特征描述符与目标属性的映射关系数值化可视化。

常见的机器学习模型可解释性 Python 工具包有 Yellowbrick、ELI5、SHAP、Mlxtend、PDPBox、InterpretML 等。在此主要介绍 SHapley Additive exPlanations(SHAP)工具^[56]对模型进行可视化解释。SHAP 解释方法的目的是通过计算每个特征对预测目标属性的贡献来对模型进行理解, 根据联盟博弈理论计算 Shapley 值来度量特征对目标属性的贡献。

给定当前的一组特征值, 特征值对实际预测值与平均预测值的贡献之差就是估计的 Shapley 值。SHAP 工具不仅可以对模型做局部解释, 也就是单个样本分析, 也可进行全局解释。如图 2c 所示, 为 SHAP 工具分析绘制的热图, 图中横轴表示根据模型预测的带隙数据样本索引, 纵轴表示各项特征。色标代表 SHAP 值大小, 红色-蓝色表示每个样本和特征的正-负 SHAP 值, 表明对预测的正-负贡献。由图 2c 可以看出, 每个特征的正 SHAP 值主要位于对应为较宽带隙的左侧部分, 而负值位于与较窄带隙预测相关的右侧部分, 因此表明正-负 SHAP 值的良好分离分布以及对带隙预测的巨大贡献^[67]。另外 SHAP 工具还可以通过队列条形图、蜂群图、依赖度散点图等, 实现多样的模型结果与输入特征关系的可视化方式。

另外符号回归算法^[78]可以直接生成适应数据的解析表达式以理解材料特征参数与目标属性的映射关系。Weng 等^[79]将符号回归用于指导设计具有更好析氧反应活性的新型钙钛矿氧化物电催化剂。在 18 种同一实验标准的钙钛矿氧化物材料数据支持下, 此项工作用遗传算法实现符号回归, 跳过了以往用来表达材料析氧性能的吸附能、 d 带中心和 d/e_g 轨道电子占据等材料性质, 模型直接给出了材料特征参数与析氧性能的若干解析公式。在权衡各个公式的复杂度与 MAE 值后, 该研究选定 μ/t (八面体因子与容忍因子之比) 作为材料析氧性能的描述子, 该公式的 MAE 值仅为 0.025 3 eV, 达到了能够指导发现更好析氧性能钙钛矿氧化物材料的标准。

2 机器学习在钙钛矿材料中的应用

2.1 机器学习预测新的钙钛矿结构材料

理论上, 元素周期表中的大多数元素都可以在 ABX_3 中找到恰当的位置进行取代, 形成钙钛矿材料。然而实际上并非所有元素替换后都可以保持在固定位置上, 继承钙钛矿的对称性优势, 也就是说并非所有具有 ABX_3 化学计量的化合物都是钙钛矿材料。如何根据 ABX_3 化合物中各元素的特征判断其是否可以形成钙钛矿结构, 即钙钛矿材料的可成形性预测, 是预测钙钛矿材料首先遇到的问题。容忍因子^[80]

$$[t = \frac{r_A + r_B}{\sqrt{2}(r_B + r_X)}], \text{ 式中: } r_A, r_B, r_X \text{ 分别代表 A、B、}$$

X 位的离子半径]是广泛使用的评价钙钛矿可成形性的预测因子, 由 Goldschmidt 在 1 个世纪前提出, 一般用于确定氧化物能否形成钙钛矿结构, 随着钙

钛矿材料的发展其适用性也越来越受限。后续研究人员也提出了一些包括八面体因子^[81] ($\mu = \frac{r_B}{r_X}$), 新容忍因子 $(\mu+t)^{\eta}$ ^[82] (其中: η 为堆积比) 和新容忍因子

$$\tau \left\{ \tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_B}{r_A} \right) \right\}^{\frac{r_A}{\ln(\frac{r_A}{r_B})}} \text{ [83]}, \text{ 其中: } n_A \text{ 代表 A}$$

离子的价态} 在内的描述钙钛矿结构可成形性的预测因子, 在一些特定种类的钙钛矿材料中这些预测因子有着良好的表现。现在, 通过结合机器学习, 研究人员期望寻找到更精确控制钙钛矿结构可成形性的因素和模式, 以开发可以更快地更全面判断钙钛矿可成形性的方法。

在全无机钙钛矿材料领域, 氧化物钙钛矿因为存在浓度较大的氧空位而不会发生结构坍塌, 引起研究人员的兴趣。Talapatra 等^[84]针对氧化物单钙钛矿 ABO_3 与氧化物双钙钛矿 $AA'BB'O_6$ 的可形成性, 从文献中采集了 1 505 种实验已知的氧化物, 其中 318 种为非钙钛矿材料作为负样本, 1 187 种为钙钛矿材料作为正样本, 使用随机森林分类算法建立分类模型。训练得到的可成形性分类模型的精度、准确率和召回率分别达到 94.01%、93.44% 和 99.13%。在对模型的解释中发现, 不仅传统的容忍因子和八面体因子对可成形性贡献非常高, 还有许多 B 位元素特征, 例如 Zunger 势半径 ($Z \text{ radius}^{B+}$)、电负性 (X^{B+}) 和最低未占据分子轨道 ($LUMO^{B+}$) 对于区分形成和不能形成氧化物钙钛矿也很重要, 如图 3a 所示。利用可成形性分类模型结合热力学稳定性分类模型, 该项工作预测确定了 414 种化合物可以作为未来实验合成氧化物钙钛矿的优异候选材料。Liu 等^[87]基于 397 种 ABO_3 化合物的数据集, 通过输入 9 个关键特征以及 25 个复合描述符, 训练梯度提升决策树模型预测氧化物钙钛矿的可成形性。经过交叉验证后, 模型的分类精度达到 94.6%。将模型应用到 891 个未被开发的 ABO_3 化合物中, 发现了 331 个潜在的氧化物钙钛矿材料, 前 174 种候选材料的可成形概率超过 85.0%。

预测有机-无机杂化钙钛矿与无机钙钛矿不同, 由于有机阳离子的结构复杂半径过大, 相比无机钙钛矿的描述符, 需要额外添加 A 位有机阳离子的结构描述符。所以之前所提出的预测无机钙钛矿可成形性的影响机制在有机-无机杂化钙钛矿材料的成形性预测中会受到很大的限制。对于有机阳离子的

结构通常使用有机离子-分子半径来描述, Lu 等^[85] 分析比较了 10 种采样方法与 10 种分类模型, 通过进行不平衡学习预测有机-无机杂化钙钛矿的实验可成形性。从报道中收集到的 539 个 HOIP 样本和 24 个非 HOIP 样本, 分别作为正负样本进行模型训练, 再使用留一交叉验证计算模型精度与测试精度。达到 100% 最高精度和准确率的模型及分类的组合有 SMOTEENN-CAT 和 SMOTEENN-SVC 2 种, 如图 3b 所示, 且它们的测试精度达到 95.5%~96.1%。通过 SHAP 方法解释 SMOTEENN-CAT 模型发现了 3 个重要特征: A 位原子半径(R_A , 如图 3c 所示)、A 位离子半径和容忍因子与 HOIP 可成形性之间的关系。最后, 3 种有机-无机杂化钙钛矿材料

MASn_{0.85}Ge_{0.15}I₃、MASn_{0.74}Ge_{0.26}I₃ 和 MASn_{0.66}Ge_{0.34}I₃ 基于模型预测被合成出来, 与所预测概率的一致性超过 99.0%。Zhang 等^[86] 提出了将机器学习与 SHAP 方法相结合的可解释策略, 来预测有机-无机杂化钙钛矿的可成形性。在模型选择中, 使用 SHAP 工具判断特征个数变化后预测结果的精度。基于此判断, 该工作选择了仅用 6 项特征就达到最佳精度的极端梯度提升(XGBoost)模型。该模型在到达最佳值再增加特征数目后, 精度仍稳定在微小的波动范围内, 如图 3d 所示。SHAP 分析表明, B 位的离子半径和 B 位点的晶格常数与成形性呈正相关, 而与 A 位的离子半径、容忍因子和 B 位的第一电离与成形性呈负相关。

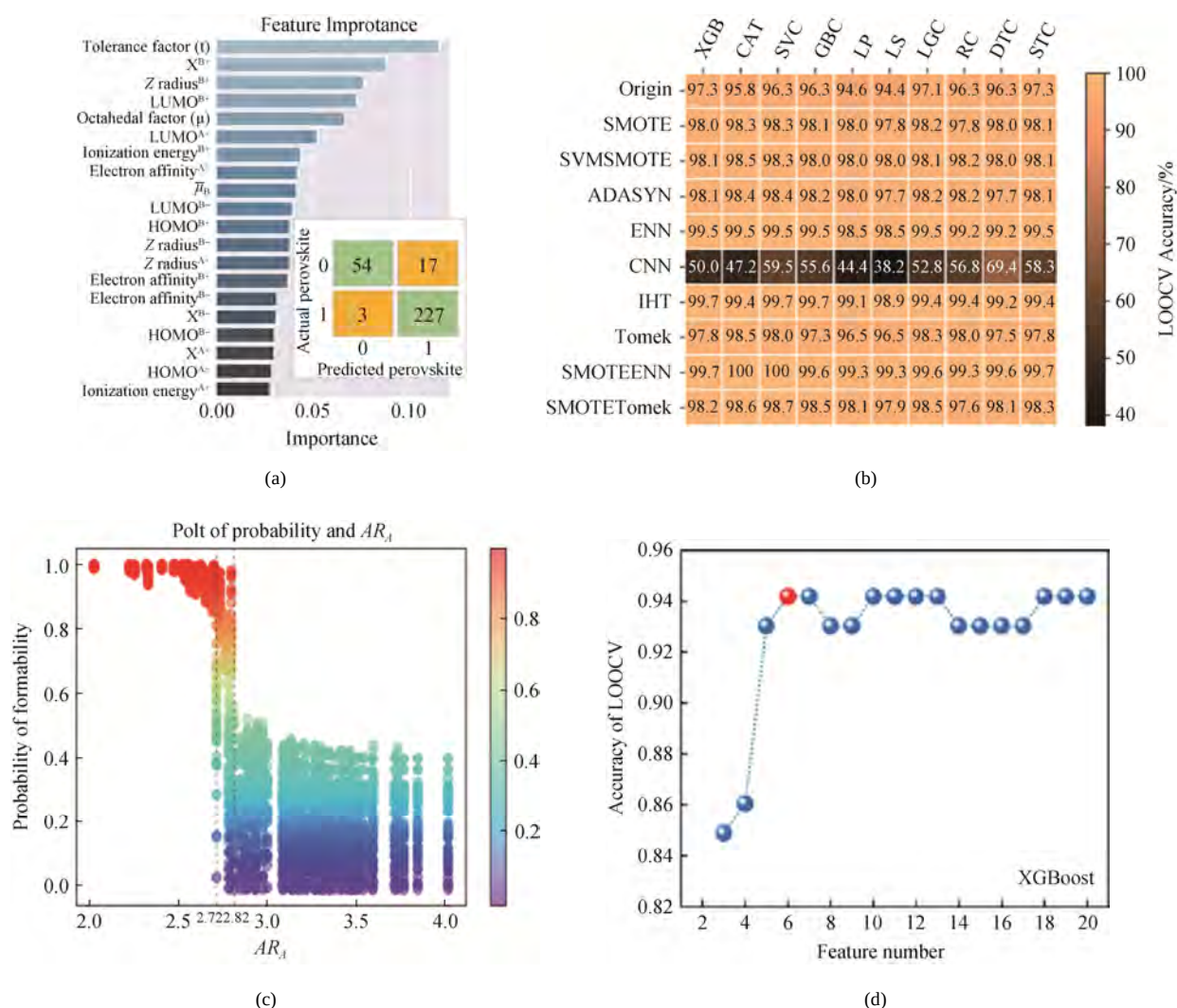


图 3 (a)钙钛矿可成形性的随机森林分类结果: 所有非零贡献值特征的重要度^[84]; (b)留一交叉验证中基于不同采样方法的各个分类模型精度^[85]; (c)可成形性关于参数 AR_A 的概率分布点图^[85]; (d)不同特征数目下, XGBoost 模型的留一交叉验证精度, 红点代表模型精度最佳位置^[86]

Fig. 3 (a) Random forest classification results for perovskite formability: Feature importance plot for all the features with non-zero values^[84]; (b) Accuracy metrics of various classification models in LOOCV based on different sampling methods^[85]; (c) Distribution of formability probability: plots of AR_A ^[85]; (d) LOOCV accuracy of the XGBoost model, the red point represent the optimal value of accuracies in model^[86]

2.2 机器学习对钙钛矿稳定性的预测

稳定性作为钙钛矿材料的基本性质, 决定了钙钛矿材料器件的使用寿命和使用环境, 从而限制其使用领域及推广范围, 是钙钛矿材料性质中决定其应用能力至关重要的一环。上一节内容讨论了 ABX_3 化合物是否能够形成钙钛矿结构, 这一节则主要探索钙钛矿材料是否具有热力学稳定性, 以及机器学习如何对钙钛矿结构与其稳定性之间的关系进行描述。材料的形成能(E_f)、分解能(ΔH_D)和超出凸包能量值^[88-89](E_{hull})都可以反应材料的热力学稳定性^[90], 也可以通过第一性原理计算获得, 但巨大的计算成本限制了第一性原理计算在化学搜索空间较大的材料中的计算能力^[91]。

Ye 等^[92]建立了仅使用 2 个描述符(Pauling 电负性和离子半径)预测密度泛函理论(DFT)所计算的石榴石($C_3A_2D_3O_{12}$)和氧化物单钙钛矿形成能的深度神经网络模型, 如图 4a 所示。模型的 MAE 分别仅为 $7\sim 10 \text{ meV}\cdot\text{atom}^{-1}$ 和 $20\sim 34 \text{ meV}\cdot\text{atom}^{-1}$ 。模型预测

精度完全在 DFT 计算的精度范围内, 但无法对深度神经网络模型的结果进行解释分析。

Li 等^[93]提出了一种机器学习结合 DFT 计算来设计稳定卤化物双钙钛矿的方法。以 354 个高通量 DFT 计算的卤化物钙钛矿分解能(ΔH_D)作为训练集, 机器学习建立起了钙钛矿稳定性与组成其离子半径之间的映射。该模型预测的稳定性趋势与独立于训练数据之外的 246 个双钙钛矿的实验数据非常吻合, 表现出比容忍因子和 $(u+t)\eta$ 描述因子更好的预测性能。在 10^4 数量级的预测样本中, 模型针对目前尚未被报道的, A 位为碱金属阳离子的卤化物双钙钛矿表现出良好的预测性能, 如图 4b 所示。使用该模型绘制的基于离子半径的钙钛矿稳定性图谱, 可以使领域内各种实验结果合理化。这项工作表明, 通过机器学习研究钙钛矿性的稳定性可以依赖来自高通量 DFT 计算的训练数据, 这远远比尝试性的材料合成实验更加经济高效。

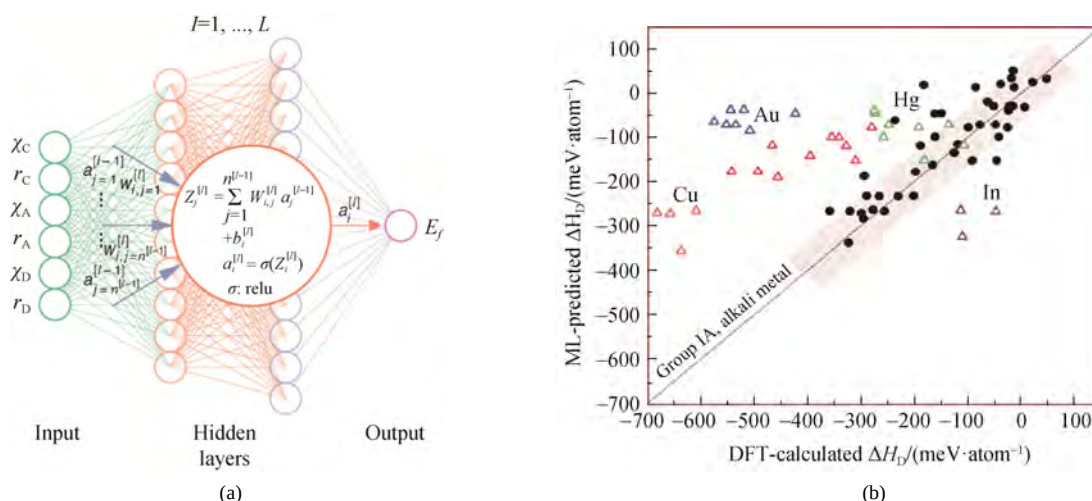


图 4 (a)神经网络示意图: 描述符输入层(每个端点上输入元素的 Pauling 电负性或离子半径)、多个隐藏层以及输出层(E_f)^[92]; (b)74 种随机选择的钙钛矿材料密度泛函计算和机器学习模型预测的分解能回归直线结果对比。黑点代表 $A_2B(I)B(III)X_6$ (A 为碱金属元素)^[93]

Fig. 4 (a) General schematic of the artificial neural network. The artificial neural network comprises an input layer of descriptors (the Pauling electronegativity and ionic radii on each site), followed by a number of hidden layers, and finally an output layer (E_f)^[92]; (b) The DFT-calculated and ML-predicted decomposition energies of 74 randomly selected perovskites. The black dots are $A_2B(I)B(III)X_6$ (A=alkali metal elements)^[93]

Schmidt 等^[94]针对 10^4 数量级全无机钙钛矿 ABX_3 样本进行了机器学习模型的大规模训练与测试, 以 E_{hull} 为参照预测材料热力学稳定性。经过 20 000 个样本训练后的模型, 在 230 000 个钙钛矿数据的测试集中, 给出 E_{hull} 的最小平均绝对误差为 $121 \text{ meV}\cdot\text{atom}^{-1}$ 。研究表明通过限制相关化学成分特征而不降低准确性, 机器学习可显著加快(至少 5 倍)高通量 DFT 计算。Li 等^[95]开发了一种机器学习

模型, 由 DFT 计算提供了超过 1 900 个样本的 E_{hull} 数据集, 来预测氧化物钙钛矿的热力学稳定性。通过分别构建的 2 个机器学习模型对 E_{hull} 进行分类和回归后, 预测出了 15 种具有热力学稳定性的氧化物钙钛矿候选材料。最后, 通过筛选提出了 4 种稳定的氧化物钙钛矿($\text{La}_{0.5}\text{Y}_{0.5}\text{Co}_{0.5}\text{Mn}_{0.5}\text{O}_3$ 、 $\text{Y}_{0.75}\text{Sr}_{0.25}\text{VO}_3$ 、 CeReO_3 和 $\text{Dy}_{0.75}\text{Nd}_{0.25}\text{RuO}_3$)。Liu 等^[87]通过机器学习回归模型筛选出 37 种热力学稳定的 ABO_3 钙钛矿

($0 \text{ meV} \cdot \text{atom}^{-1} < E_{\text{hull}} < 36 \text{ meV} \cdot \text{atom}^{-1}$)和 13 种亚稳态钙钛矿($36 \text{ meV} \cdot \text{atom}^{-1} < E_{\text{hull}} < 70 \text{ meV} \cdot \text{atom}^{-1}$), 作为进一步合成和应用的氧化物钙钛矿候选材料。

2.3 机器学习对钙钛矿带隙的预测

从材料到器件的实现进程中, 材料的本征带隙性质往往从一开始就给材料划定了可应用的范围以及方向。功能材料的带隙与太阳能电池、发光二极管、光电探测器等各种光电器件的性能直接相关^[96–99]。对于这些领域, 钙钛矿材料的带隙具有十分广泛的应用优势: 直接带隙、合适的禁带宽度以及带隙可控等^[100]。对钙钛矿材料带隙的预测可以加速实用钙钛矿的发现, 指导筛选合适的候选材料进行实验合成。

在全无机钙钛矿领域, Guo 等^[101]利用 4 种不同的机器学习模型, 预测了无铅双钙钛矿卤化物的热力学稳定性与带隙性质。其中随机森林对带隙的预测效果(R^2 为 0.941, MAE 为 0.149)最佳。经过分析发现随机森林能够很好地捕捉带隙预测中的非线性关系, 模型分析表明带隙和所提取的最高占据能级以及立方相特征有非线性关联。Gao 等^[102]从 5 796 种无机双钙钛矿材料中, 应用极端梯度提升回归算法筛选具有合适带隙的材料。根据该机器学习模型筛选, 并通过限制带隙范围、过滤有毒元素以及贵金属后, 得到了 2 种新型无铅无机双钙钛矿卤化物候选材料 K_2NaInI_6 和 $\text{Na}_2\text{MgMnI}_6$ 。 K_2NaInI_6 的直接带隙大小为 1.46 eV, $\text{Na}_2\text{MgMnI}_6$ 为 1.89 eV, 其值与性能优异的有机–无机杂化钙钛矿材料 $\text{CH}_3\text{NH}_3\text{PbI}_3$ 相近^[103], 有很高的应用价值, 而且这 2 种材料的热稳定性和良好的光学性能也被 DFT 计算所证实。

在有机–无机杂化钙钛矿领域, 结合主动搜索过程和机器学习方法, Lu 等^[67]收集了 1 201 个基于实验的 HOIP 带隙样本进行了研究。在构建了加权投票回归机器学习模型预测 HOIP 材料的带隙后, SHAP 特征分析表明, 低于 0.971 的容忍因子和 3.75~4.09 范围内的新容忍因子有助于降低带隙, 反之亦然。该研究从超过 8.20×10^{18} 种化合物组合中成功预测合成了 3 种新型 HOIP 材料, 预测和实验之间的平均带隙误差仅为 0.07 eV。Wu 等^[104]对无铅混合有机–无机双钙钛矿卤化物(MDHOIPs)材料进行了快速筛选。此项工作从 11 370 个候选材料中选出了 204 个具有理想带隙的无铅 MDHOIPs。特征工程表明, $B-B'$ 位金属的大离子半径有利于钙钛矿结构的形成、 $B-B'$ 位金属的电负性与带隙值成正比、

替换 X 位离子可以适当地调整带隙值。候选材料中的 $\text{MA}_2\text{GeSnI}_4\text{Br}_2$ 和 $\text{MA}_2\text{InBiI}_2\text{Br}_4$, 因为其具有的良好稳定性和光电特性(合适的直接带隙、轻载流子有效质量、小的激子结合能、强可见光吸收和良好的抗分解稳定性等), 被确定为有应用前景的太阳能电池候选材料。

与文献[85–86]中应用有机离子半径描述有机–无机杂化钙钛矿 A 位不同, Chen 等^[105]考虑了有机离子的各向异性, 应用具有 3 个参数的长方体替代以往一个参数的球体来描述有机阳离子的形状, 以更好地预测带隙。通过结合机器学习方法和 DFT 计算, 实现了一个从 78 400 种有机–无机杂化双钙钛矿中筛选优异光伏材料的框架, 筛选出 19 个具有适合带隙的钙钛矿材料, 其中 17 个带隙在 1.5 至 3.0 eV 范围内, 2 个具有 0.76 至 1.44 eV 的较小带隙, 并通过 HSE06 计算进行了验证。

3 机器学习在钙钛矿器件中的应用

3.1 机器学习与钙钛矿太阳能电池

太阳能直接利用效率的提高是本世纪最迫切的前沿科学目标之一, 它对解决目前严峻的生态问题和能源问题有着至关重要的意义。太阳能有多种利用方式, 其中应用广泛的光电转换方式主要依赖半导体太阳能电池, 其主要评价标准为能量转化效率(PCE)。相比其他半导体材料, 钙钛矿半导体材料具有合适且可调的带隙, 载流子有效质量小寿命长等优势^[106–108]。钙钛矿太阳能电池(PSC)的光伏性能由多个相互关联的因素决定, 如钙钛矿的组成成分、每个传输层的电子特性、界面变化和制造参数等, 这使得器件性能的优化和潜在机制的发现相当具有挑战性^[109–110]。Srivastava 等^[111]介绍了将机器学习算法应用于钙钛矿太阳能电池研发的框架, 讨论了在钙钛矿太阳能电池设计流程中的 3 个重要阶段: 成分筛选、钙钛矿层合成和分析, 以及完整的器件制造和测试, 充分说明了机器学习方法在研究钙钛矿太阳能电池方面的优势。

Odabaşı 等^[112]对 2013 至 2018 年间发表的 800 篇关于铅基钙钛矿卤化物太阳能电池的文章进行了回顾和分析, 总结了其中 1 921 个太阳能电池器件性能数据, 并将数据整理成数据库。在经过机器学习工具分析后, 这项工作利用该数据库开发了启发式的方法和模型, 预测了未来器件性能的发展趋势(图 5a), 并使用机器学习论证了 PSC 性能的影响因素。Odabaşı 等工作主要基于统计分析, 缺乏对预

测结果和所用模型背后的关系及物理特性的理解, 需要进一步对 PSC 的底层物理关系进行清晰的描述。Li 等^[113]通过机器学习算法, 预测了钙钛矿材料的带隙性质, 以及带隙及其他性质与 PSC 性能的相关性, 如图 5b 所示。虽然模型没有考虑一些其他可能影响 PSC 性能的参数, 例如晶粒尺寸和制造技术的变化等, 但模型均方根误差小且测试表现良好, 并与 Shockley-Queiser 极限的理论预测吻合得很好。以上结果均表明该模型可以为设计新的钙钛矿材料和新型 PSC 器件提供良好的指导。

Feng 等^[114]将机器学习与高通量计算相结合, 通过预测二维钙钛矿材料数据库中材料的带隙,

成功找到 4 种具有合适带隙的候选材料 Al_2HfO_4 、 $\text{Sr}_3\text{Si}_2\text{Se}_7$ 、 Sr_2VON_3 和 Ba_2VON_3 , 带隙预测与计算结果对比如图 5c 所示。经过稳定性测试、光吸收和理论功率转换效率评估, 最终发现了 2 种有前途的光伏器件材料, 即 Sr_2VON_3 和 Ba_2VON_3 , 其 PCE 分别高达 30.35% 和 26.03%。Cai 等^[115]运用高评价的机器学习模型(图 5d), 从 180 038 种化合物中成功筛选出 597 种具有合适带隙和高德拜温度的稳定有机-无机杂化双钙钛矿材料。基于无毒性、实验可行性条件以及 DFT 计算验证筛选后, 预测其中 3 种无铅候选材料 $(\text{CH}_3\text{NH}_3)_2\text{AgGaBr}_6$ 、 $(\text{CH}_3\text{NH}_3)_2\text{AgInBr}_6$ 和 $(\text{C}_2\text{NH}_6)_2\text{AgInBr}_6$ 拥有超宽带

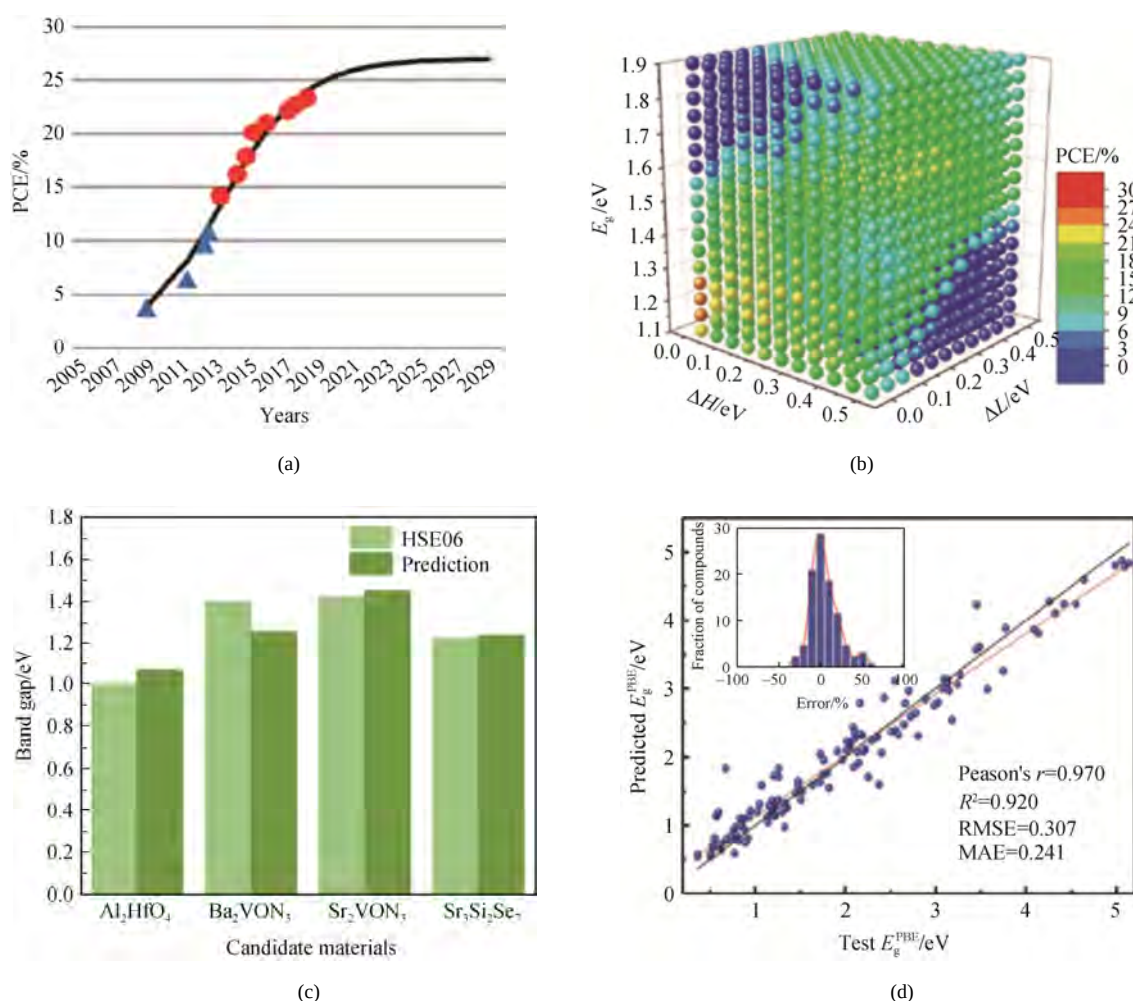


图 5 (a)正置(*n-i-p*)结构钙钛矿太阳能电池的效率演变预测^[112]; (b)能量转化效率关于带隙 E_g 以及其他两参数的 4D 散点图, 可见转化效率最高值在 1.2~1.3 eV 的带隙范围内^[113]; (c)4 种预测材料预测带隙值和 HSE06 DFT 计算值对比^[114]; (d)测试集中的实际值(DFT 计算得到的带隙) E_g^{PBE} 和预测值 Test E_g^{PBE} 的回归直线, 以及评估模型性能的几个参数。其中理想效果显示为红线, 拟合线显示为黑线^[115]

Fig. 5 (a) Evolution of efficiency for regular (*n-i-p*) structure: Logistic growth model^[112]; (b) 4D-plot of PCE with respect to E_g , ΔH , and ΔL showing that the highest PCE values are within 1.2–1.3 eV^[113]; (c) The comparison of the bandgap for the four predicted materials between ML prediction and HSE06 DFT calculations^[114]; (d) the actual value of band gap E_g^{PBE} by DFT in the test set and the predicted value E_g^{PBE} by ML. Pearson coefficient (r), R^2 , MSE, MAE are calculated to explain the training performance of the built model. The ideal line is shown as red line and the fit line is shown as dark line. The inset is the fraction of compounds according to their percent error between predicted E_g^{ML} and actual E_g^{DFT} ^[115]

吸收区域以及低激子辐射复合率,有利于抑制非辐射复合和热致降解,可以作为优异的PSC候选材料。然而,以上这些工作都是单纯进行模拟,对于模拟与实验相结合的机器学习PSC工作还很缺乏。

Cai 等^[109]提出并实现了一种基于正逆双向框架的机器学习方法(图6),以建立高性能钙钛矿光伏材料 $\text{MASn}_x\text{Pb}_{1-x}\text{I}_3$ 的关键参数与其光伏性能之间的关系。该方法探索了材料带隙与 Sn 成分之间的不对称弓形关系,并通过实验得到了准确的验证。通过建立器件的光伏参数模型,在该研究中机器学习成功捕获了 Shockley–Queiser 理论中缺陷区和低带隙平台区短路电流和开路电压与带隙

的偏差。前者是由于晶体畸变形成的“深阱态”缺陷,后者则是由于 Sn^{4+} 含量的增加增大了磁化率。由此得出了空穴提取比电子提取更困难的结论,且转换效率的预测曲线符合 Shockley–Queiser 极限。通过搜索和优化算法,此项工作最终获得了性能优良的钙钛矿太阳能电池组分,Sn:Pb 组成比接近 0.6 的优化结果,并通过实验进行了验证。该工作构建的机器学习方法框架对于涉及多个变量的复杂系统,可以更快地给出结果,也可用于其他材料的优化和高效器件的开发,有望对器件的物理现象提供更深入的理解,并探索新的功能材料和高性能器件。

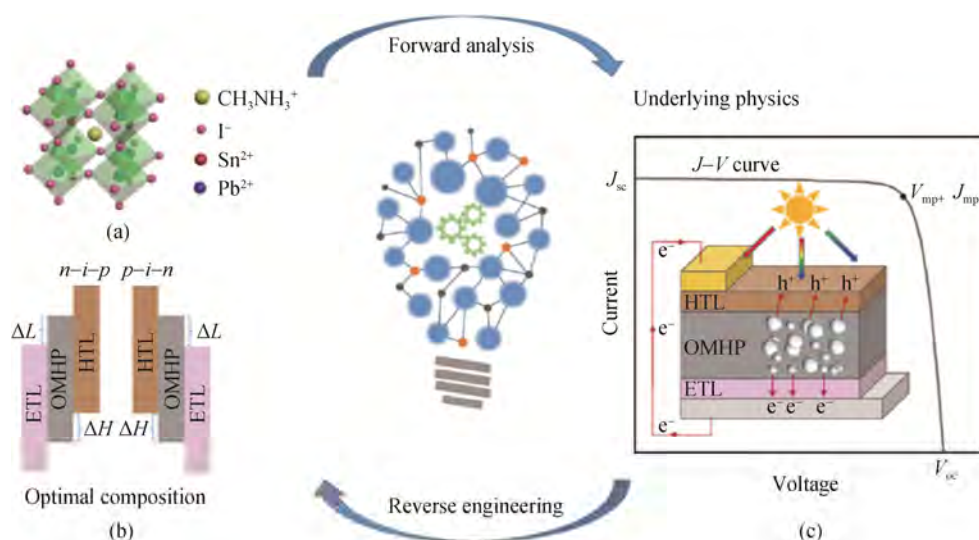


图6 正向-反向框架机器学习方法流程图。正向分析:(a)将有机金属钙钛矿卤化物材料信息输入机器学习模型预测带隙;(b)正置($n-i-p$)和倒置($p-i-n$)的器件结构,表示钙钛矿太阳能电池信息,也可以组合起来构建其他机器学习模型;(c)通过机器学习可以预测钙钛矿太阳能电池的性能,并且可以利用产生的大量数据揭示和分析潜在的物理现象。然后基于正向结果,反向推导出高性能钙钛矿太阳能电池中具有合适带隙的机金属钙钛矿卤化物材料的最佳组成^[109]

Fig. 6 Flow chart of proposed forward–reverse method. For forward analysis. (a)OMHP (Organic metal halide perovskites) material information is fed into a machine learning model to predict the band gap; (b)the device structure of regular ($n-i-p$) and inverted ($p-i-n$), which represents PSC information and is also combined to construct other ML models. (c)Through ML the performances of PSCs are predicted, and underlying physics could be revealed and analyzed with huge data produced. Then based on the forward results, the optimal composition of OMHP with suitable band gap is reversely deduced for high-performance PSC^[109]

3.2 机器学习与钙钛矿发光二极管

钙钛矿材料因其荧光量子产率高、发光波长可调、发光半峰宽窄、可通过溶液法制备等优势成为有前景的发光二极管材料^[116–117]。钙钛矿发光二极管(PeLEDs)可以用于显示屏幕与照明等领域,不仅显色纯度高、亮度高、轻薄、视角广、响应速度快还可以实现柔性显示,拥有广阔的应用前景^[118–124]。自可在室温下工作的 PeLED 器件被报道以来,PeLEDs 研究就得到了持续性的进展,如今在绿光、

红光与近红外范围的 PeLEDs 外量子效率(EQE)均超过了 20%^[123, 125–126]。但是 PeLEDs 的器件效率低,稳定性差,蓝光器件 EQE 低以及材料中含有有毒元素铅等问题限制了其进一步的工业化应用。虽然机器学习已经在钙钛矿材料及其器件的研究中,取得了一些亮眼的研究进展,但是在 PeLEDs 方面机器学习的应用才刚刚开始。

有机分子的添加剂工程对于实现高性能钙钛矿光电器件至关重要。由于添加剂的结构和相关物理

化学特征对钙钛矿结晶、形态、点缺陷和前体化学影响的复杂性,所以尚不清楚添加剂的结构和相关物理化学特征如何影响器件性能。然而,通过实验寻找合适的添加剂既昂贵又耗时,而简单的机器学习模型由于在这个相对较新的领域中可用的实验数据有限,因此难以准确预测添加剂的性能。Zhang 等^[127]运用复杂的深度学习方法,通过 132 个分子的小型数据集,以 96% 的高精度预测了 PeLEDs 中添加剂的有效性。该模型可以最大限度地利用分子的信息,并显著缓解之前机器学习模型进行分子筛选时经常出现的重复问题。通过使用预测的添加剂,实现了非常高效率的 PeLEDs 器件,其峰值外量子效率高达 22.7%,为进一步提高钙钛矿光电器件的性能开辟了一条新途径。

Kirman 等^[128]报道了一种用于发现新钙钛矿单晶的反溶剂法高通量合成和表征钙钛矿单晶的实验框架。使用机器学习来指导并不断改进机器人合成试验的顺序。该工作使用蛋白质结晶机器人进行钙钛矿单晶的高通量合成,并借助基于卷积神经网络的图像识别方法来表征结果。然后使用机器学习模型预测合成新钙钛矿单晶的最佳条件,成功地对 $(3\text{-PLA})_2\text{PbCl}_4$ 钙钛矿材料进行了合成,发现其具有 4 eV 的直接带隙和蓝光发射,证明机器学习方法在加速新型钙钛矿光电材料的预测与合成中的可行性。

3.3 机器学习与钙钛矿光催化剂

光催化剂能够在光照下发挥其氧化还原能力,对污染物等进行降解净化还可以促进物质的合成和转化等,是直接有效利用太阳能的主要方式之一^[129]。光催化二氧化碳还原、光催化固氮、光催化有机污染物降解以及光解水制氢为 4 种典型的光催化应用^[130-132]。光催化氧化反应利用光生电子以及空穴的迁移和俘获与反应物反应,一般以半导体为材料。钙钛矿材料因为其独特的结构、成分多样性、优异的物理化学性质、适应性强以及稳定的催化活性成为提高光催化剂效率,开发新型光催化剂材料的热门候选^[133-135]。机器学习不仅可以加速催化剂的发现,而且作为一种强大的工具,可以帮助更深入了解材料的特征与其催化活性、选择性和稳定性之间的关系,有助于建立设计催化剂的框架以及了解其效率提高的原因。但是因为催化的复杂性以及催化过程对时间的动态依赖,与钙钛矿发光二极管相似,机器学习在光催化剂材料上的研究目前还处于起步阶段。

光催化剂的比表面积在光催化反应中起重要作用。

一般来说,光催化剂的比表面积越大,反应位点越多,光催化性能越好。Shi 等^[136]使用机器学习算法来探索钙钛矿的比表面积与其化学成分、技术参数及实验条件之间的关系。通过机器学习预测比表面积与钙钛矿样品的实验比表面积的回归直线对比,如图 7a 所示。在使用模型进行虚拟筛选后,提出了 5 种具有较大比表面积和光催化潜力的钙钛矿材料。此外,作者还将预测模型开发成在线应用程序,使研究人员利用该模型预测大型比表面积钙钛矿更加方便。

Tao 等^[137]通过机器学习模型分别建立了针对光催化剂材料带隙与产氢率的回归模型,以加速发现用于光解水制氢的高效钙钛矿光催化剂。光催化剂产氢率预测模型使用材料带隙、比表面积、微晶尺寸、制备方法和煅烧时间等特征输入。经不同模型评估对比后,反向传播人工神经网络模型具有最佳评价,其交叉验证与测试集的 R^2 值分别为 0.989 7 和 0.974 0。在结构稳定性、导带位置、价带位置和产氢率等标准筛选下,如图 9b 所示,此工作从 30 000 种无机钙钛矿材料中筛选出了 14 种潜在的光催化钙钛矿候选材料。这 14 种材料的平均产氢率比训练数据集中的最高值高出 6.4%。表明了机器学习对光催化材料预测的准确性以及广阔的光催化应用前景。

Tao 等^[138]也开发了一种多目标逐步设计策略,以加速具有增强光催化活性的潜在钙钛矿的设计开发。该策略包括模型构建、逐步筛选和预测候选钙钛矿的产氢率。基于该策略,预测得到了 35 种具有高比表面积(大于 $60\text{ m}^2\cdot\text{g}^{-1}$)、合适带隙(1.4~2.6 eV)和微小晶尺寸(小于 36 nm)的候选钙钛矿。候选材料的产氢率预测结果显示其均具有高产氢率(大于 $6\,000\text{ }\mu\text{mol}\cdot\text{h}^{-1}\cdot\text{g}^{-1}$)。根据模型统计分析结果表明,当 A 位的主要元素为 Bi、La 和 Pr, B 位的主要元素为 Fe、Ti 和 Mn 时,更容易获得合适的带隙值。使用溶胶凝胶法、聚合络合物法、燃烧法和较低的煅烧温度制备的钙钛矿具有较高的产氢率和较小的微晶尺寸。目标特性与产氢率之间的关系表明,具有 E_g 在 1.9~2.4 eV 范围内、微晶尺寸在 20~32 nm 范围内和高比表面积(大于 $50\text{ m}^2\cdot\text{g}^{-1}$)的候选钙钛矿具有更高的产氢率。此外,对于以上 2 项机器学习成果,该课题组开发了在线网络服务器来共享预测模型,这些共享模型可能会加速 ABO_3 型钙钛矿在光催化领域的实验研究与开发。

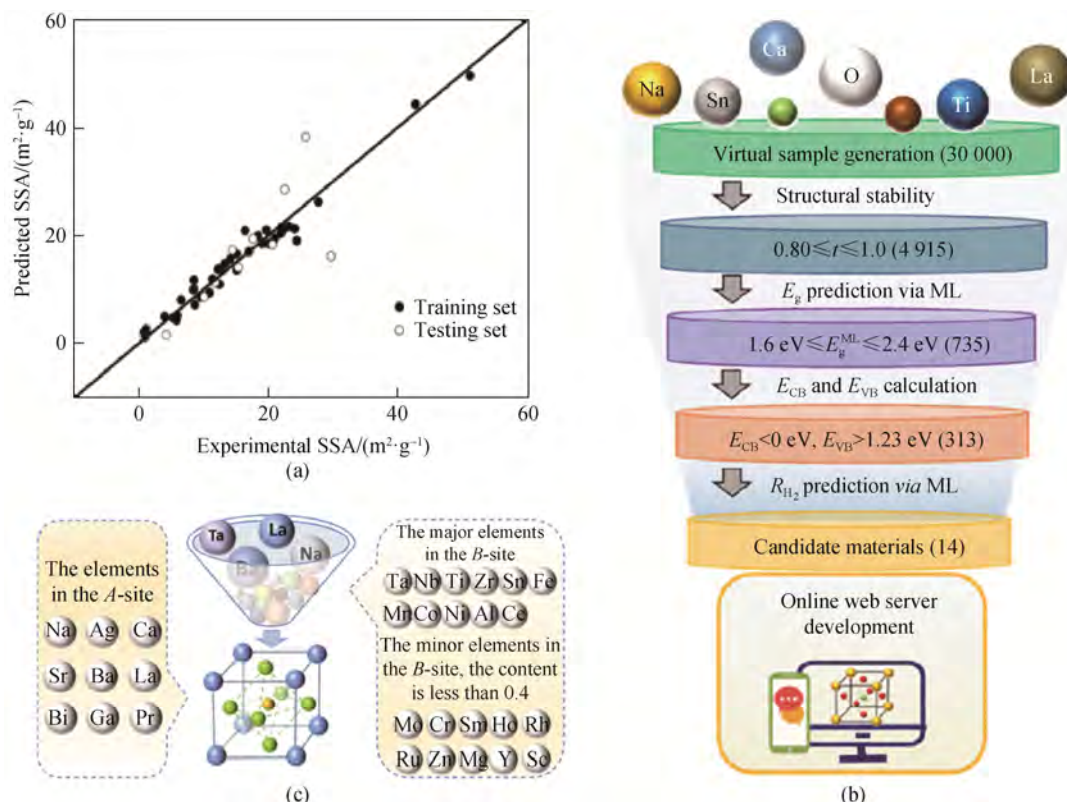


图 7 (a)通过机器学习预测的比表面积(SSA)与钙钛矿样品的实验 SSA^[136]; (b)结合机器学习筛选高效光解水制氢的钙钛矿光催化剂流程图^[137]; (c)用于虚拟筛选的 ABO₃ 型钙钛矿材料化合物以及各个位置的待筛选元素^[138]

Fig. 7 (a) Predicted specific surface area (SSA) vs experimental SSA of perovskite samples by support vector regression^[136]; (b) Flowchart of screening perovskite photocatalysts for efficient photo-splitting of water for hydrogen production combined with machine learning^[137]; (c) ABO₃-type perovskite material composition for virtual screening and elements to be screened at various positions^[138]

4 总结与展望

本文总结了机器学习在材料发现和设计中的基本过程,回顾了机器学习在钙钛矿材料性质以及器件方面的部分应用。机器学习在钙钛矿中的应用主要可分为以下 3 类:

1) 利用机器学习预测具有优异性质的钙钛矿新材料,为实验合成提供清晰、可靠、快速的指导。

2) 通过解释机器学习模型,更深层次认识各项特征和不同实验条件下,钙钛矿材料以及器件性能的影响机制。

3) 将机器学习用于对实验结果的智能识别,实现更快速的高通量实验进程,加快高性能钙钛矿材料及器件的研发应用。

但是目前机器学习在钙钛矿材料的性质方面预测还不够全面,对大部分钙钛矿器件的预测才刚刚起步,并缺乏实际学习所需的数据量支撑。在此提出一些未来的发展方向:

1) 简化特征项,不仅限于剔除冗余项,还可以

合并有数理关联的特征项。将探索合并特征项,发现对目标属性更精确的描述因子作为一项重要环节,而不是仅仅依赖更复杂的机器模型来达到预测的高精度。

2) 添加特征项,不仅依靠数据库提供的基本物理化学参数,还可以结合材料实际应用以及器件合成等条件,积极发挥材料领域知识与实验实践经验。让计算机科学与材料科学内容更加融合,增加特征在实际应用中的可操作性。

3) 对材料进行综合的全面的筛选。对材料的筛选不仅依靠 1、2 项单一的性质参数,可以使用不同预测模型,排列组合起整套钙钛矿材料及器件的筛选框架。将各个筛选流程规范化、模板化,加快预测效率与实用性。

4) 随着机器学习在钙钛矿材料应用中的不断发展,将极大加速钙钛矿实验数据量的积累。将这些数据再次运用到机器学习模型中,进一步发展优化模型,使其预测能力进一步提高并更具泛化能力。促进数据积累的更大步跨越,提供更有帮助和准确

的指导, 形成良好的正反馈循环。

5) 建立在线的模型预测网站, 共享共建优质机器学习模型。推动跨区域合作, 让不同细分领域的研究人员可以分享材料合成以及器件制造经验, 取长补短, 建立从材料预测到第一性原理计算再到器件应用的全流程机器学习在线预测框架, 加速环保型高性能钙钛矿材料的开发与应用。

机器学习作为科学研究工具开发钙钛矿材料与器件的进程才刚刚起步, 并且具有巨大的待探索空间。在机器学习方法的助力下, 钙钛矿材料定能以最大限度发挥自身优势, 为解决能源、生态等领域问题推动材料领域迅猛发展做出不可估量的贡献。

参考文献:

- [1] WANG Y, LV Z, ZHOU L, et al. Emerging perovskite materials for high density data storage and artificial synapses[J]. *J Mater Chem C*, 2018, 6(7): 1600–1617.
- [2] TRAVIS W, GLOVER E N K, BRONSTEIN H, et al. On the application of the tolerance factor to inorganic and hybrid halide perovskites: A revised system[J]. *Chem Sci*, 2016, 7(7): 4548–4556.
- [3] ZHAO X G, YANG D W, REN J C. Rational design of halide double perovskites for optoelectronic applications [J]. *Joule*, 2018, 2(19): 1662–1673.
- [4] YIN W J, WENG B, GE J, et al. Oxide perovskites, double perovskites and derivatives for electrocatalysis, photocatalysis, and photovoltaics[J]. *Energy Environ Sci*, 2019, 12(2): 442–462.
- [5] KÖRBEL S, MARQUES M A L, BOTTI S. Stability and electronic properties of new inorganic perovskites from high-throughput ab initio calculations[J]. *J Mater Chem C*, 2016, 4(15): 3157–3167.
- [6] SLAVNEY A H, HU T, LINDENBERG A M, et al. A bismuth-halide double perovskite with long carrier recombination lifetime for photovoltaic applications[J]. *J Am Chem Soc*, 2016, 138(7): 2138–2141.
- [7] FILIP M R, HILLMAN S, HAGHIGHIRAD A A, et al. Band gaps of the lead-free halide double perovskites $\text{Cs}_2\text{BiAgCl}_6$ and $\text{Cs}_2\text{BiAgBr}_6$ from theory and experiment[J]. *J Phys Chem Lett*, 2016, 7(13): 2579–2585.
- [8] SYZGANTSEVA O A, SALIBA M, GRÄTZEL M, et al. Stabilization of the perovskite phase of formamidinium lead triiodide by methylammonium, Cs, and/or Rb doping[J]. *J Phys Chem Lett*, 2017, 8(6): 1191–1196.
- [9] KUBICKI D J, PROCHOWICZ D, HOFSTETTER A, et al. Cation dynamics in mixed-cation $(\text{MA})_x(\text{FA})_{1-x}\text{PbI}_3$ hybrid perovskites from solid-state NMR[J]. *J Am Chem Soc*, 2017, 139(29): 10055–10061.
- [10] LI C, HAO H, XU B, et al. A progressive learning method for predicting the band gap of ABO_3 perovskites using an instrumental variable[J]. *J Mater Chem C*, 2020, 8(9): 3127–3136.
- [11] VELDHUIS S A, BOIX P P, YANTARA N, et al. Perovskite materials for light-emitting diodes and lasers[J]. *Adv Mater*, 2016, 28(32): 6804–6834.
- [12] WANG Y, TANG Y, JIANG J, et al. Mixed-dimensional self-assembly organic–inorganic perovskite microcrystals for stable and efficient photodetectors[J]. *J Mater Chem C*, 2020, 8(16): 5399–5408.
- [13] EKSTRÖM E, LE FEBVRIER A, BOURGEOIS F, et al. The effects of microstructure, Nb content and secondary Ruddlesden–Popper phase on thermoelectric properties in perovskite $\text{CaMn}_{1-x}\text{Nb}_x\text{O}_3$ ($x=0-0.10$) thin films[J]. *RSC Adv*, 2020, 10(13): 7918–7926.
- [14] ORÓ-SOLÉ J, CLARK L, KUMAR N, et al. Synthesis, anion order and magnetic properties of $\text{RVO}_{3-x}\text{N}_x$ perovskites ($\text{R}=\text{La}, \text{Pr}, \text{Nd}$; $0 \leq x \leq 1$)[J]. *J Mater Chem C*, 2014, 2(12): 2212–2220.
- [15] SHIOGAI J, CHIDA T, HASHIMOTO K, et al. Signature of band inversion in the perovskite thin-film alloys $\text{BaSn}_{1-x}\text{Pb}_x\text{O}_3$. [J]. *Phys Rev B*, 2020, 101(12): 125125.
- [16] SYDORCHUK V, LUTSYUK I, SHVED V, et al. $\text{PrCo}_{1-x}\text{Fe}_x\text{O}_3$ perovskite powders for possible photocatalytic applications[J]. *Res Chem Intermediat*, 2020, 46(3): 1909–1930.
- [17] JI T, NIU T, WANG J, et al. Crystallization regulation of solution-processed two-dimensional perovskite solar cells[J]. *J Mater Chem A*, 2022, 10(26): 13625–13650.
- [18] ZHU T, YANG Y, GONG X. Recent advancements and challenges for low-toxicity perovskite materials[J]. *ACS Appl Mater Inter*, 2020, 12(24): 26776–26811.
- [19] YIN W J, SHI T, YAN Y. Unusual defect physics in $\text{CH}_3\text{NH}_3\text{PbI}_3$ perovskite solar cell absorber[J]. *Appl Phys Lett*, 2014, 104(6): 063903.
- [20] ZHAO D, WANG B, LIANG C, et al. Facile deposition of high-quality $\text{Cs}_2\text{AgBiBr}_6$ films for efficient double perovskite solar cells[J]. *Sci China Mater*, 2020, 63(8): 1518–1525.
- [21] LI L, TIAN G, CHANG W, et al. A novel double-perovskite LiLaMgTeO_6 : Mn^{4+} far-red phosphor for indoor plant cultivation white LEDs: Crystal and electronic structure, and photoluminescence properties[J]. *J Alloy Compd*, 2020, 832: 154905.
- [22] ALVES G A S, CENTURION H A, SAMBRANO J R, et al. Band gap narrowing of Bi-doped NaTaO_3 for photocatalytic hydrogen evolution under simulated sunlight: A pseudocubic phase induced by doping[J]. *ACS Applied Energy Materials*, 2021, 4(1): 671–679.
- [23] GREEN M A, DUNLOP E D, HOHL-EBINGER J, et al. Solar cell efficiency tables (version 58)[J]. *Prog Photovolt: Res Appl*, 2021, 29(7): 657–667.
- [24] MA S, YUAN G, ZHANG Y, et al. Development of encapsulation strategies towards the commercialization of perovskite solar cells[J]. *Energy Environ Sci*, 2022, 15(1): 13–55.
- [25] WANG M, WANG W, MA B, et al. Lead-free perovskite materials for solar cells[J]. *Nano-Micro Lett*, 2021, 13(1): 62.
- [26] SCHMIDT-MENDE L, DYAKONOV V, OLTROF S, et al. Roadmap on organic–inorganic hybrid perovskite semiconductors and devices[J]. *APL Materials*, 2021, 9(10): 109202.
- [27] WANG R, MUJAHID M, DUAN Y, et al. A review of perovskites solar cell stability[J]. *Adv Funct Mater*, 2019, 29(47): 1808843.
- [28] GRASER J, KAUWE S K, SPARKS T D. Machine learning and energy minimization approaches for crystal structure predictions: A review and new horizons[J]. *Chem Mater*, 2018, 30(11): 3601–3612.
- [29] RAJAN K. Materials informatics[J]. *Mater Today*, 2005, 8(10): 38–45.
- [30] HUSSAIN A, YANG L H, ZOU Y B, et al. Monte Carlo simulation study of electron yields from compound semiconductor materials[J]. *J Appl Phys*, 2020, 128(1): 015305.
- [31] ALDER B J, WAINWRIGHT T E. Studies in molecular dynamics. I. General method[J]. *J Chem Phys*, 1959, 31(2): 459–466.
- [32] HOHENBERG P, KOHN W. Inhomogeneous electron gas[J]. *Phys*

- Rev, 1964, 136(3B): B864-B871.
- [33] PERDEW J P, BURKE K, ERNZERHOF M. Generalized gradient approximation made simple[J]. *Phys Rev Lett*, 1996, 77(18): 3865–3868.
- [34] AGRAWAL A, CHOUDHARY A. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science[J]. *APL Mater*, 2016, 4(5): 053208.
- [35] JORDAN M I, MITCHELL T M. Machine learning: Trends, perspectives, and prospects[J]. *Science*, 2015, 349(6245): 255–260.
- [36] RUPP M. Machine learning for quantum mechanics in a nutshell[J]. *Int J Quantum Chem*, 2015, 115(16): 1058–1073.
- [37] SUN W, ZHENG Y, YANG K, et al. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials[J]. *Sci Adv*, 2019, 5: eaay4275.
- [38] STANEV V, OSES C, KUSNE A G, et al. Machine learning modeling of superconducting critical temperature[J]. *NPJ Comput Mater*, 2018, 4(1): 29.
- [39] ZHOU T, SONG Z, SUNDMACHER K. Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design[J]. *Engineering*, 2019, 5(6): 1017–1026.
- [40] DAI D, XU T, WEI X, et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys[J]. *Comp Mater Sci*, 2020, 175: 109618.
- [41] GANTZER P, CRETON B, NIETO DRAGHI C. Inverse-QSPR for de novo design: A review[J]. *Molecular Inform*, 2020, 39(4): 1900087.
- [42] RACCUGLIA P, ELBERT K C, ADLER P D F, et al. Machine-learning-assisted materials discovery using failed experiments[J]. *Nature*, 2016, 533(7601): 73–76.
- [43] BALACHANDRAN P V, KOWALSKI B, SEHIRLIOGLU A, et al. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning[J]. *Nat Commun*, 2018, 9(1): 1668.
- [44] LIU Y, ZHAO T, JU W, et al. Materials discovery and design using machine learning[J]. *J Materiomics*, 2017, 3(3): 159–177.
- [45] SCHMIDT J, MARQUES M R G, BOTTI S, et al. Recent advances and applications of machine learning in solid-state materials science[J]. *NPJ Comput Mater*, 2019, 5(1): 83.
- [46] GOLDSMITH B R, ESTERHUIZEN J, LIU J-X, et al. Machine learning for heterogeneous catalyst design and discovery[J]. *AIChE J*, 2018, 64(7): 2311–2323.
- [47] RAMPRASAD R, BATRA R, PILANIA G, et al. Machine learning in materials informatics: Recent applications and prospects[J]. *NPJ Comput Mater*, 2017, 3(1): 54.
- [48] HAIXIANG G, YIJING L, SHANG J, et al. Learning from class-imbalanced data: Review of methods and applications[J]. *Expert Syst Appl*, 2017, 73: 220–239.
- [49] WOHLRAB L, FÜRNKRANZ J. A review and comparison of strategies for handling missing values in separate-and-conquer rule learning[J]. *J Intell Inf Syst*, 2011, 36(1): 73–98.
- [50] XU X, LIU H, YAO M. Recent progress of anomaly detection[J]. *Complexity*, 2019: 2686378.
- [51] WANG Y, PAN Z, PAN Y. A training data set cleaning method by classification ability ranking for the k-nearest neighbor classifier[J]. *IEEE Transact Neur Network Learn Syst*, 2020, 31(5): 1544–1556.
- [52] CHEN C, ZUO Y, YE W, et al. A critical review of machine learning of energy materials[J]. *Adv Energy Mater*, 2020, 10(8): 1903242.
- [53] BRAHAM E J, CHO J, FORLANO K M, et al. Machine learning-directed navigation of synthetic design space: A statistical learning approach to controlling the synthesis of perovskite halide nanoplatelets in the quantum-confined regime[J]. *Chem Mater*, 2019, 31(9): 3281–3292.
- [54] TAO Q, XU P, LI M, et al. Machine learning for perovskite materials design and discovery[J]. *npj Comput Mater*, 2021, 7(1): 23.
- [55] ZHANG Y, LING C. A strategy to apply machine learning to small datasets in materials science[J]. *npj Comput Mater*, 2018, 4(1): 25.
- [56] WEI J, CHU X, SUN X Y, et al. Machine learning in materials science[J]. *InfoMat*, 2019, 1(3): 338–358.
- [57] LIU Y, WU J M, AVDEEV M, et al. Multi-layer feature selection incorporating weighted score-based expert knowledge toward modeling materials with targeted properties[J]. *Adv Theor Simul*, 2020, 3(2): 1900215.
- [58] 刘悦, 邹欣欣, 杨正伟, 等. 材料领域知识嵌入的机器学习[J]. *硅酸盐学报*, 2022, 50(03): 863–876.
- LIU Yue, ZOU Xinxin, YANG Zhengwei, et al. *J Chin Ceram Soc*, 2022, 50(03): 863–876.
- [59] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large data sets[J]. *Science*, 2011, 334(6062): 1518–1524.
- [60] LUNDBERG S M, LEE S I, A unified approach to interpreting model predictions[C]//31st Annual Conference on Neural Information Processing Systems, Long Beach, CA, 2017.
- [61] MA B, WEI X, LIU C, et al. Data augmentation in microscopic images for material data mining[J]. *npj Comput Mater*, 2020, 6(1): 125.
- [62] MIN K, CHO E. Accelerated discovery of potential ferroelectric perovskite via active learning[J]. *J Mater Chem C*, 2020, 8(23): 7866–7872.
- [63] CHEN D, SUN D, FU J, et al. Semi-supervised learning framework for aluminum alloy metallographic image segmentation[J]. *IEEE Access*, 2021, 9: 30858–30867.
- [64] PRUKSAWAN S, LAMBARD G, SAMITSU S, et al. Prediction and optimization of epoxy adhesive strength from a small dataset through active learning[J]. *Sci Technol Adv Mater*, 2019, 20(1): 1010–1021.
- [65] SHEN S, SADOUGHI M, LI M, et al. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries[J]. *Appl Energy*, 2020, 260: 114296.
- [66] TOGO R, SAITO N, OGAWA T, et al. Estimating regions of deterioration in electron microscope images of rubber materials via a transfer learning-based anomaly detection model[J]. *IEEE Access*, 2019, 7: 162395–162404.
- [67] LU T, LI H, LI M, et al. Inverse design of hybrid organic-inorganic perovskites with suitable bandgaps via proactive searching progress[J]. *ACS Omega*, 2022, 7(25): 21583–21594.
- [68] CERIOTTI M. Unsupervised machine learning in atomistic simulations, between predictions and understanding[J]. *J Chem Phys*, 2019, 150(15): 150901.
- [69] ORUPATTUR N V, MUSHRIF S H, PRASAD V. Catalytic materials and chemistry development using a synergistic combination of machine learning and ab initio methods[J]. *Comp Mater Sci*, 2020, 174: 109474.
- [70] TOYAO T, MAENO Z, TAKAKUSAGI S, et al. Machine learning for catalysis informatics: Recent applications and prospects[J]. *ACS Catal*,

- 2019, 10: 2260–2297.
- [71] JING L, TIAN Y. Self-supervised visual feature learning with deep neural networks: A survey[J]. *IEEE Transact Pattern Anal*, 2021, 43(11): 4037–4058.
- [72] KONG J G, LI Q X, LI J, et al. Self-supervised graph neural networks for accurate prediction of néel temperature[J]. *Chin Phys Lett*, 2022, 39(6): 067503.
- [73] DOAN H, AGHARWAL G, QIAN H, et al. Quantum chemistry-informed active learning to accelerate the design and discovery of sustainable energy storage materials[J]. *Chem Mater*, 2020, 32(15): 6338, 6346.
- [74] LOOKMAN T, BALACHANDRAN P V, XUE D, et al. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design[J]. *npj Comput Mater*, 2019, 5(1): 21.
- [75] RAJAK P, WANG B, NOMURA K I, et al. Autonomous reinforcement learning agent for stretchable kirigami design of 2D materials[J]. *NPJ Comput Mater*, 2021, 7(1): 102.
- [76] XUE D, BALACHANDRAN P V, HOGDEN J, et al. Accelerated search for materials with targeted properties by adaptive design[J]. *Nat Commun*, 2016, 7(1): 11241.
- [77] LIU Y, GUO B R, ZOU X X, et al. Machine learning assisted materials design and discovery for rechargeable batteries[J]. *Energy Storage Mater*, 2020, 31: 434–450.
- [78] WANG Y, WAGNER N, RONDINELLI J M. Symbolic regression in materials science[J]. *MRS Communications*, 2019, 9(3): 793–805.
- [79] WENG B, SONG Z, ZHU R, et al. Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts[J]. *Nat Commun*, 2020, 11(1): 3513.
- [80] GOLDSCHMIDT V M. Die gesetze der krystallochemie[J]. *Naturwissenschaften*, 1926, 14(21): 477–485.
- [81] LI C, SOH K C K, WU P. Formability of ABO₃ perovskites[J]. *J Alloy Compd*, 2004, 372(1): 40–48.
- [82] SUN Q, YIN W J. Thermodynamic stability trend of cubic perovskites[J]. *J Am Chem Soc*, 2017, 139(42): 14905–14908.
- [83] BARTEL C J, SUTTON C, GOLDSMITH B R, et al. New tolerance factor to predict the stability of perovskite oxides and halides[J]. *Sci Adv*, 2019, 5(2): eaav0693.
- [84] TALAPATRA A, UBERUAGA B P, STANEK C R, et al. A machine learning approach for the prediction of formability and thermodynamic stability of single and double perovskite oxides[J]. *Chem Mater*, 2021, 33(3): 845–858.
- [85] LU T, LI H, LI M, et al. Predicting experimental formability of hybrid organic-inorganic perovskites via imbalanced learning[J]. *J Phys Chem Lett*, 2022, 13(13): 3032–3038.
- [86] ZHANG S, LU T, XU P, et al. Predicting the formability of hybrid organic-inorganic perovskites via an interpretable machine learning strategy[J]. *J Phys Chem Lett*, 2021, 12(31): 7423–7430.
- [87] LIU H, CHENG J, DONG H, et al. Screening stable and metastable ABO₃ perovskites using machine learning and the materials project[J]. *Comp Mater Sci*, 2020, 177: 109614.
- [88] ARMIENTO R, KOZINSKY B, HAUTIER G, et al. High-throughput screening of perovskite alloys for piezoelectric performance and thermodynamic stability[J]. *Phys Rev B*, 2014, 89(13): 134103.
- [89] LIU M, RONG Z, MALIK R, et al. Spinel compounds as multivalent battery cathodes: A systematic evaluation based on ab initio calculations[J]. *Energy Environ Sci*, 2015, 8(3): 964–974.
- [90] ONG S P, WANG L, KANG B, et al. Li-Fe-P-O₂ phase diagram from first principles calculations[J]. *Chem Mater*, 2008, 20(5): 1798–1807.
- [91] ASKERKA M, LI Z, LEMPEN M, et al. Learning-in-templates enables accelerated discovery and synthesis of new stable double perovskites[J]. *J Am Chem Soc*, 2019, 141(8): 3682–3690.
- [92] YE W, CHEN C, WANG Z, et al. Deep neural networks for accurate predictions of crystal stability[J]. *Nat Commun*, 2018, 9(1): 3800.
- [93] LI Z, XU Q, SUN Q, et al. Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning[J]. *Adv Funct Mater*, 2019, 29(9): 1807280.
- [94] SCHMIDT J, SHI J, BORLIDO P, et al. Predicting the thermodynamic stability of solids combining density functional theory and machine learning[J]. *Chem Mater*, 2017, 29(12): 5090–5103.
- [95] LI W, JACOBS R, MORGAN D. Predicting the thermodynamic stability of perovskite oxides using machine learning models[J]. *Comp Mater Sci*, 2018, 150: 454–463.
- [96] SOE C M M, STOUMPOS C C, KEPENEKIAN M, et al. New type of 2D perovskites with alternating cations in the interlayer space, (C(NH₂)₃)(CH₃NH₃)_nPb_nI_{3n+1}: Structure, properties, and photovoltaic performance[J]. *J Am Chem Soc*, 2017, 139(45): 16297–16309.
- [97] YIN T, LIU B, YAN J, et al. Pressure-engineered structural and optical properties of two-dimensional (C₄H₉NH₃)₂PbI₄ perovskite exfoliated nm-thin flakes[J]. *J Am Chem Soc*, 2019, 141(3): 1235–1241.
- [98] PROTESCU L, YAKUNIN S, BODNARCHUK M I, et al. Nanocrystals of cesium lead halide perovskites (CsPbX₃, X = Cl, Br, and I): Novel optoelectronic materials showing bright emission with wide color gamut[J]. *Nano Lett*, 2015, 15(6): 3692–3696.
- [99] MYUNG C W, YUN J, LEE G, et al. A new perspective on the role of a-site cations in perovskite solar cells[J]. *Adv Energy Mater*, 2018, 8(14): 1702898.
- [100] YIN W J, YANG J H, KANG J, et al. Halide perovskite materials for solar cells: A theoretical review[J]. *J Mater Chem A*, 2015, 3(17): 8926–8942.
- [101] GUO Z, LIN B. Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells[J]. *Sol Energy*, 2021, 228: 689–699.
- [102] GAO Z, ZHANG H, MAO G, et al. Screening for lead-free inorganic double perovskites with suitable band gaps and high stability using combined machine learning and DFT calculation[J]. *Appl Surf Sci*, 2021, 568: 150916.
- [103] LIU Y, YANG Z, CUI D, et al. Two-inch-sized perovskite CH₃NH₃PbX₃ (X = Cl, Br, I) crystals: Growth and characterization[J]. *Adv Mater*, 2015, 27(35): 5176–5183.
- [104] WU Y, LU S, JU M G, et al. Accelerated design of promising mixed lead-free double halide organic-inorganic perovskites for photovoltaics using machine learning[J]. *Nanoscale*, 2021, 13(28): 12250–12259.
- [105] CHEN J L, XU W J, ZHNG R Q. Δ -machine learning-driven discovery of double hybrid organic-inorganic perovskites[J]. *J Mater Chem A*, 2022, 10: 1402–1413.
- [106] JU M G, DAI J, MA L, et al. Perovskite chalcogenides with optimal bandgap and desired optical absorption for photovoltaic devices[J]. *Adv Energy Mater*, 2017, 7(18): 1700216.
- [107] JU M G, DAI J, MA L, et al. Lead-free mixed tin and germanium perovskites for photovoltaic application[J]. *J Am Chem Soc*, 2017,

- 139(23): 8038–8043.
- [108] PUPPIN M, POLISHCHUK S, COLONNA N, et al. Evidence of large polarons in photoemission band mapping of the perovskite semiconductor CsPbBr₃[J]. *Phys Rev Lett*, 2020, 124(20): 206402.
- [109] CAI X, LIU F, YU A, et al. Data-driven design of high-performance MaSn_xPb_{1-x}I₃ perovskite materials by machine learning and experimental realization[J]. *Light Sci Appl*, 2022, 11(1): 234.
- [110] ZHANG L, HE M, SHAO S F. Machine learning for halide perovskite materials[J]. *Nano Energy*, 2020, 78: 105380.
- [111] SRIVASTAVA M, HOWARD J M, GONG T, et al. Machine learning roadmap for perovskite photovoltaics[J]. *J Phys Chem Lett*, 2021, 12(32): 7866–7877.
- [112] ODABAŞI Ç, YİLDİRİM R. Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools[J]. *Nano Energy*, 2019, 56: 770–791.
- [113] LI J, PRADHAN B, GAUR S, et al. Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells[J]. *Adv Energy Mater*, 2019, 9(46): 1901891.
- [114] FENG H J, MA P. Machine learning prediction of 2D perovskite photovoltaics and interaction with energetic ion implantation[J]. *Appl Phys Lett*, 2021, 119(23): 231902.
- [115] CAI X, ZHANG Y, SHI Z, et al. Discovery of lead-free perovskites for high-performance solar cells via machine learning: Ultrabroadband absorption, low radiative combination, and enhanced thermal conductivities[J]. *Adv Sci*, 2022, 9(4): e2103648.
- [116] FAKHARUDDIN A, GANGISHETTY M K, ABDI-JALEBI M, et al. Perovskite light-emitting diodes[J]. *Nat Electron*, 2022, 5(4): 203–216.
- [117] XU L, LIU G, XIANG H, et al. Charge-carrier dynamics and regulation strategies in perovskite light-emitting diodes: From materials to devices[J]. *Appl Phys Rev*, 2022, 9(2): 021308.
- [118] CAO Y, WANG N, TIAN H, et al. Perovskite light-emitting diodes based on spontaneously formed submicrometre-scale structures[J]. *Nature*, 2018, 562(7726): 249–253.
- [119] MA D, LIN K, DONG Y, et al. Distribution control enables efficient reduced-dimensional perovskite leds[J]. *Nature*, 2021, 599(7886): 594–598.
- [120] LIN K, XING J, QUAN L N, et al. Perovskite light-emitting diodes with external quantum efficiency exceeding 20 percent[J]. *Nature*, 2018, 562(7726): 245–248.
- [121] CHIBA T, HAYASHI Y, EBE H, et al. Anion-exchange red perovskite quantum dots with ammonium iodine salts for highly efficient light-emitting devices[J]. *Nat Photon*, 2018, 12(11): 681–687.
- [122] WANG N, CHENG L, GE R, et al. Perovskite light-emitting diodes based on solution-processed self-organized multiple quantum wells[J]. *Nat Photon*, 2016, 10(11): 699–704.
- [123] KARLSSON M, YI Z, REICHERT S, et al. Mixed halide perovskites for spectrally stable and high-efficiency blue light-emitting diodes[J]. *Nat Commun*, 2021, 12(1): 361.
- [124] LIU G, ZHANG S, XU L, et al. Recent advances of eco-friendly quantum dots light-emitting diodes for display[J]. *Prog Quant Electron*, 2022: 100415.
- [125] ZHAO B, BAI S, KIM V, et al. High-efficiency perovskite–polymer bulk heterostructure light-emitting diodes[J]. *Nat Photon*, 2018, 12(12): 783–789.
- [126] JIANG J, CHU Z, LI J Z, et al. Red perovskite light-emitting diodes with efficiency exceeding 25% realized by co-spacer cations [J]. *Adv Mater*, 2022, 34(36): 2204460.
- [127] ZHANG L, LI N, LIU D, et al. Deep learning for additive screening in perovskite light-emitting diodes[J]. *Angew Chem Int Ed Engl*, 2022, 61(37): e202209337.
- [128] KIRMAN J, JOHNSTON A, KUNTZ D A, et al. Machine-learning-accelerated perovskite crystallization[J]. *Matter*, 2020, 2(4): 938–947.
- [129] BRESOLIN B M, PARK Y, BAHNEMANN D. Recent progresses on metal halide perovskite-based material as potential photocatalyst[J]. *Catalysts*, 2020, 10(6): 709.
- [130] KUMAR A, KUMAR A, KRISHNAN V. Perovskite oxide based materials for energy and environment-oriented photocatalysis[J]. *ACS Catal*, 2020, 10(17): 10253–10315.
- [131] ZHAO Y, LI Y, SUN L. Recent advances in photocatalytic decomposition of water and pollutants for sustainable application[J]. *Chemosphere*, 2021, 276: 130201.
- [132] HUYNH K A, NGUYEN D L T, NGUYEN V H, et al. Halide perovskite photocatalysis: Progress and perspectives[J]. *J Chem Technol Biot*, 2020, 95(10): 2579–2596.
- [133] HUANG Y, LIU J, DENG Y, et al. The application of perovskite materials in solar water splitting[J]. *J Semicond*, 2020, 41(1): 011701.
- [134] DUBOSE J T, KAMAT P V. Efficacy of perovskite photocatalysis: Challenges to overcome[J]. *ACS Energy Lett*, 2022, 7(6): 1994–2011.
- [135] NASSAR I M, WU S, LI L, et al. Facile preparation of n-type LaFeO₃ perovskite film for efficient photoelectrochemical water splitting[J]. *Chem Select*, 2018, 3(3): 968–972.
- [136] SHI L, CHANG D, JI X, et al. Using data mining to search for perovskite materials with higher specific surface area[J]. *J Chem Inf Model*, 2018, 58(12): 2420–2427.
- [137] TAO Q, LU T, SHENG Y, et al. Machine learning aided design of perovskite oxide materials for photocatalytic water splitting[J]. *J Energy Chem*, 2021, 60: 351–359.
- [138] TAO Q, CHANG D, LU T, et al. Multiobjective stepwise design strategy-assisted design of high-performance perovskite oxide photocatalysts[J]. *J Phys Chem C*, 2021, 125(38): 21141–21150.