

基于机器学习与第一性原理筛选锂离子电池钒基电极材料

墨云鹤 曹卫刚 赵乐泉 唐振强 郑珑 蔡宗英*

(华北理工大学冶金与能源学院 河北唐山 063210)

摘要 鉴于成本效益、资源丰富性, 钒基锂离子电池电极材料成为科研热点. 通过机器学习模型与第一性原理计算对钒基材料数据库建立了筛选-验证流程, 旨在发现优异潜在钒基电极材料. 从 Materials project 提取出了 4694 条钒基数据, 并通过 pymatgen (Python Materials Genomics) 计算了最大理论容量. 相关性研究发现密度对于钒基材料理论容量的影响比较关键, 经三种机器学习算法联合预测对比, 遗传算法确定超参数的深度神经网络算法(DNN)效果最佳, R^2 为 0.771. 并通过 DNN 算法的 SHAP 分析进一步证明. 经模型预测, 根据密度特征选取原始数据集前 0.5% 数据, 最终确定了 26 种潜在钒基电极材料. 经机器学习与第一性原理计算验证, 确定了三种理论容量均大于 650 mAh/g, 开路电压分别为 2.56、0.64、0.49 V 的钒基正负两电极候选材料. 这一流程不仅可用于对钒基电极材料的发现, 并有望在不同材料体系扩展.

关键词 机器学习; 锂离子电池; 钒基材料; 第一性原理

Screening Vanadium-Based Electrode Materials for Lithium-Ion Batteries Based on Machine Learning and First-Principles

Mo, Yunhe Cao, Weigang Zhao, Lequan Tang, Zhenqiang
Zheng, Long Cai, Zongying*

(School of Metallurgy and Energy, North China University of Science and Technology, Tangshan, Hebei 063210)

Abstract This study investigates the discovery of vanadium-based electrode materials for lithium-ion batteries using a combination of machine learning and first-principles calculations. The study extracted data for 4694 vanadium-based materials from the Materials Project database and calculated their maximum theoretical capacities using pymatgen. Correlation analysis revealed that material density has the highest correlation with theoretical capacity (Comprehensive score is 0.905), indicating that density is a key factor affecting the performance of vanadium-based materials. Based on this finding, three machine learning models (Deep Neural Network - DNN, Random Forest - RF, and Support Vector Regression - SVR) were constructed, with hyperparameters optimized using a genetic algorithm. The results showed that the DNN model performed best, achieving an R^2 value of 0.771. SHAP analysis further confirmed the significant contribution of density to the model predictions. Based on model predictions, the top 0.5% of high-density materials were selected from the original dataset, ultimately identifying 26 potential vanadium-based electrode materials. Among these, V_6O_{13} , MgV_5O_{12} , and V_2MoO_8 exhibited theoretical capacities exceeding 650 mAh/g, with open-circuit voltages of 2.56, 0.64, and 0.49 V, respectively. First-principles calculations indicated that V_6O_{13} possesses metallic properties with high electron density of states, which is beneficial for enhancing conductivity; MgV_5O_{12} and V_2MoO_8 exhibit metallic and semi-metallic characteristics, respectively. Adsorption energy calculations showed that V_6O_{13} has the lowest adsorption energy, indicating its strongest lithium-ion adsorption capability. Based on open-circuit voltage calculations, V_6O_{13} is suitable as a cathode material, while MgV_5O_{12} and V_2MoO_8 are more appropriate as anode materials. This study successfully identified high-performance vanadium-based electrode materials through the integration of machine learning and first-principles calculations, significantly shortening the experimental process. It provides new insights into the design of lithium-ion battery electrode materials and demonstrates the enormous potential of machine learning in materials science.

Keywords machine learning; lithium-ion battery; vanadium-based material; first-principles

1 引言

可再充电的锂离子电池(LIB)具有高效、无记忆效应和高能量密度而备受关注^[1]. 它被视为电动汽车、

便携式电子设备及固定储能系统主要动力来源^[2-3]. 然而, 随着锂离子电池应用场景的扩展及其性能标准的不断提升, 对电池性能的需求也相应增加^[4]. 这推动了科研工作者持续探索新型技术方案.

* E-mail: czy1106@sina.com

Received March 12, 2025; published April 24, 2025.

Supporting information for this article is available free of charge via the Internet at <http://sioc-journal.cn>.

Project supported by the National Natural Science Foundation of China (No. 52104330).

国家自然科学基金(No. 52104330)资助项目.

在锂离子电池中,电极材料扮演着核心角色,其性能直接决定了电池的容量、充放电速率以及循环稳定性^[5]. 相比于传统的 LIB 电极材料,诸如磷酸铁锂(LiFePO_4)^[6]、镍钴锰氧化物(NMC)^[7]、镍钴铝氧化物(NCA)^[8]等正极材料,又如石墨^[9]、钛酸锂基^[10]、硅基^[11]等负极材料. 钒基电极材料凭借较高的理论比容量、较宽的电化学稳定窗口、较低的环境毒性进入研究人员的视野之中^[12-13]. 如, Teo 等^[14]设计了一种锂锡钒氧化物(LiSnVO_4)负极材料,通过溶胶-凝胶法使锂离子电池获得了优异的电化学性能,在首次不可逆放电容量达到 1270 mAh/g 后,在第二次循环结束时可逆容量为 305.4 mAh/g,并在第 53 次循环结束时仍保持 211 mAh/g 的容量. De Juan Corpuz 等^[15]通过水热反应和煅烧制备了多孔 ZnV_2O_4 纳米线,作为锂离子电池负极材料,该材料在 1 A/g 电流密度下经 100 次循环后比容量达 460 mAh/g,在 5 A/g 电流密度下经 1000 次循环后比容量为 149 mAh/g,表现出优异的循环稳定性和高比容量. McNulty 等^[16]通过钠离子交换法从钒氧化物纳米管(VONTs)制备出 NaV_2O_5 晶体,并将其作为锂离子电池负极材料,从第 100 个循环到第 1000 个循环, NaV_2O_5 晶体能够保持其容量的 93%,并且循环后的比容量分别为 174 mAh/g 和 162 mAh/g. Liu 等^[17]通过第一性原理计算预测了二维 V_2N MXene 单层作为锂、钠和镁离子电池负极材料的潜力,对 Li、Na、K 和 Mg 离子的扩散能垒分别为 0.025、0.014、0.004 和 0.058 eV,理论容量分别为 925、463 和 1850 mAh/g,平均开路电压分别为 0.32、0.24 和 0.34 V. 因此,如何快速设计并筛选出具有合适电压平台、高理论比容量的新型钒基电极材料成为当前任务.

近年来,随着材料科学的进步和计算能力的增强,基于大数据技术的机器学习和第一性原理计算技术在材料研发领域逐渐兴起^[18]. 研究者能够以更高的效率筛选和优化电极材料,进而加速新材料的发现及其性能优化过程^[19]. 在新型材料探索方面, Wang 等^[20]通过机器学习方法预测固态电解质的剪切模量和体积模量,选择了轻梯度提升机(LGBM)模型作为主要预测模型,实现了高预测精度并能用于探索新材料. 在预测氧化还原电位方面, Allam 等^[21]开发了机器学习模型,以高效准确地预测各种有机材料的氧化还原电位,并分析了各种分子描述符如何影响氧化还原电位. Jinnouchi 等^[22]提出了一种结合第一性原理计算和机器学习的方法来预测半电池反应的氧化还原电位,并通过机器学习力场进行热力学积分,实现了在广泛相空间上的高效统计抽样,并通过从粗略近似方法到精确电子结构计算的逐步精细化,提高了自由能的精度. 对于工作电压预测, Joshi 等^[23]利用 Materials Project 数据库中的密度泛函理论(DFT)计算数据,通过深度神经网络、支持向量回归和核岭回归等机器学习方法来预测电极材料的工作电压. Moses 等^[24]利用 Materials Project 数据库的数据训练模

型,通过 10 折交叉验证评估模型性能,提出了一种基于深度神经网络回归的机器学习模型,用于预测金属离子电池电极材料的平均电压(V_{av})和充放电过程中的体积变化($\Delta V\%$).

在这项研究中,构建了一个用于预测钒基电极材料最大理论容量的预测模型,并通过第一性原理计算进一步验证了预测模型. 为了构建钒基材料数据库,选取了 20 种有关材料自身特征作为描述符,最大理论容量使用 pymatgen 进行计算,从而提取出了 4694 条数据. 为了最大限度减少预测模型预测不准确性,对数据集进行了噪声处理、归一化处理和相关性处理等特征工程手段. 随后构建了三种机器学习算法用于预测钒基材料,筛选出 26 种优异候选钒基电极材料. 经第一性原理计算验证,确定出 V_6O_{13} 、 $\text{MgV}_5\text{O}_{12}$ 、 V_2MoO_8 三种性能最优的材料.

2 结果与讨论

2.1 数据集分析

本次研究所用数据特征向量来源于 Materials Project 材料数据库^[25]. 目标向量采用开源材料分析软件包 pymatgen 进行高通量计算(根据每种钒基材料在经历锂离子嵌入与脱出过程中的能量变化进行计算),能量释放值以 mAh/g 为单位量化,作为衡量材料电化学性能的直接目标变量^[26].

从 Materials Project 数据库中筛选并提取了共计 4694 条与钒基材料相关的记录作为本研究的数据集. 通过对前人在锂离子电池电极材料领域的研究成果进行总结,选定了 20 种关键特征(解释见支持信息表 S1),这些特征综合覆盖了材料的晶体结构、电子结构及热力学性能等多个维度^[27-31]. 从特征分布上可以明显看出,特征向量的分布差异巨大(支持信息图 S1). 为了避免某些特征值过大而导致量级较小的特征被模型低估或忽视的问题,对数据集中的特征向量实施了归一化处理^[32]. 归一化后的特征向量如图 1(a)所示. 值得注意的是,在归一化处理后,尽管各个特征向量分布略有不同,但是大多数服从高斯分布,这将有利于机器学习模型建立^[33]. 但仍有少部分特征不符合高斯分布,这些非高斯分布的特征向量可能会对模型的学习过程产生影响,在预测过程中产生异常值,从而干扰模型的正常工作. 因此,在后续建模过程中,需要特别注意选择合适的模型和评估指标,如深度学习、树类算法、支持向量机模型就适合处理复杂非线性数据,以确保模型的有效性和鲁棒性^[34]. 目标向量分布如图 1(b)所示. 大多数钒基材料质量比容量集中在 200~300 mAh/g 和 500~550 mAh/g 这两个区间内. 这一现象表明钒基材料相较于工业界常用的磷酸铁锂(LiFePO_4 : 大约在 170 mAh/g 左右)以及其他三元材料(指含镍、钴、锰或铝的复合材料: 约 200~250 mAh/g)具有显著的优势.

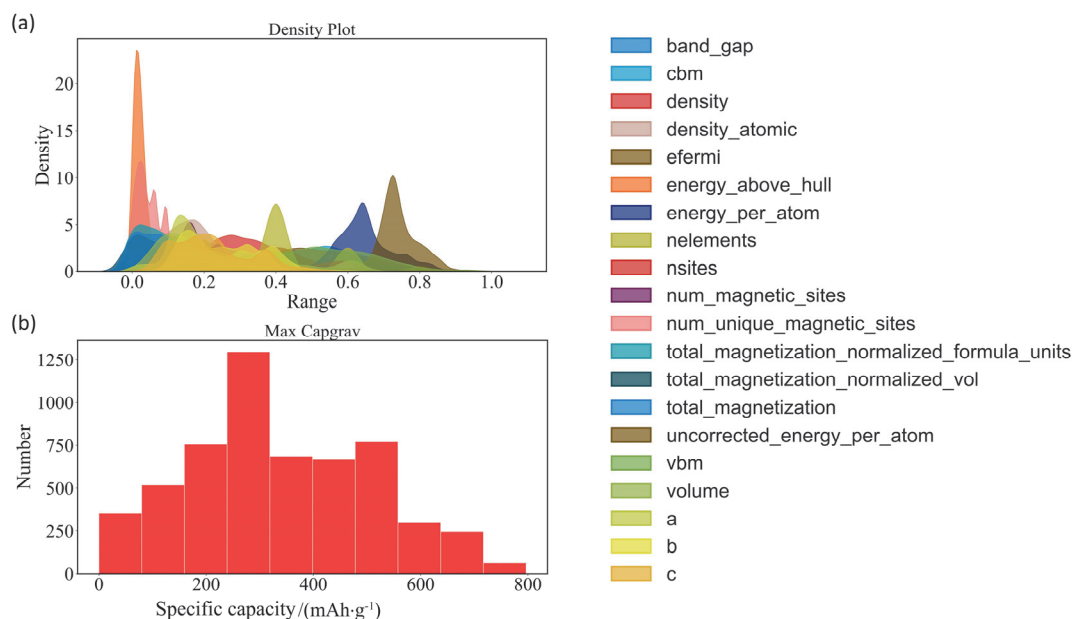


图1 特征值与目标值数据分布: (a)特征值山脊图; (b)目标值柱状图

Figure 1 Distribution of feature and target values: (a) ridge plot of feature values; (b) histogram of target values

2.2 相关性分析

面对此类多变量优化问题,本研究利用统计学中的皮尔森相关系数分析法对各特征与目标向量(即材料的比容量)之间的相关性进行了评估^[35].通过这一分析步骤,识别并排除了那些与目标向量相关性较低的特征向量,从而提高模型的解释力和预测精度.

图2为皮尔森相关系数(R 值)热力图.其中对理论容量影响最大的为密度, R 值达到0.566,具有较高线性关系.在锂离子电池的工作过程中,锂离子在正极和负极之间的嵌入与脱嵌受到材料本身晶体结构的影响,而密度对原子堆积方式、结构稳定性、缺陷形成以及相变等产生多方面影响^[36].密度对锂离子电池的比容量也有着重要影响,较高的材料密度意味着在相同的体积下可以拥有更多原子数量(图2的nsites相关性与此呼应),从而提高电池的能量密度,较高的能量密度通常伴随着较高的比容量^[37].而材料的最低导带能量、费米能级、最高价带能量与理论容量具有一定相关性, R 值分别为-0.238、-0.207、0.235.最低导带与最高价带的能量水平与材料的氧化还原行为密切相关,进而影响电池的比容量^[38].最低导带能量较低的材料更易于接受电子,促进更有效的氧化还原反应,从而在锂离子嵌入过程中更有效地储存锂离子.与之相反,最高价带能量较高的材料更易于给出电子,从而在锂离子脱嵌过程中更有效地释放锂离子.费米能级决定着材料对离子的吸附能力和稳定性,较低的费米能级意味着更强的电子-离子相互作用,有利于锂离子的稳定嵌入,从而提高材料的理论容量^[39].

为了进一步探究数据集之间的非线性关系,将各个特征的皮尔森相关系数的绝对值和互信息的值取平均

(R 值 $\in [-1,1]$, MI值 $\in [0,+\infty]$),得到一个综合特征重要性评分,如支持信息表S2所示.相关性分析揭示,在电极材料的设计中,材料的密度应该是一个关键的考虑因素.具体来说,密度与理论容量之间表现出较强的依赖关系,这由综合值为0.905所表明.此外,在随后的机器学习模型拟合过程中,应该移除与目标变量相关性较弱的变量(综合评分小于0.1,表S1中说明被移除特征),以增强模型的预测稳定性和泛化能力^[40].而且,大多数特征向量与目标向量之间的相关性都很弱,简单的线性机器学习模型不足以完全捕捉这些复杂的关系^[41].这与数据集分析相谋和.

2.3 模型分析

根据热力图结果显示,密度对钒基材料比容量正相关性较高.为了量化评估机器学习模型在预测钒基材料比容量时材料密度及其它输入特征所起到的综合效果,通过在独立测试集上的决定系数(R^2)开展每个算法模型的评估,并利用最优模型测试集的shap值解释算法模型,分析各输入特征对模型预测贡献值.在相关性分析、模型评价、shap解释综合评判下筛选出26种钒基候选电极材料.

图3(a)为各个算法5折交叉验证平均打分.从图中可以看出,性能最佳的是深度神经网络算法.值得说明的是,DNN模型的 R^2 值为0.771,已经具有实用性与有效性^[42].数据驱动部分的研究重点在于筛选出对理论容量产生最大影响的关键特征,通过相关性分析和之后的SHAP分析双层验证,已经能够明确影响理论容量的关键特征.通过遗传算法进行超参数优化后,确定了采用四层隐藏层的设计最为优越,各层包含的神经元数量

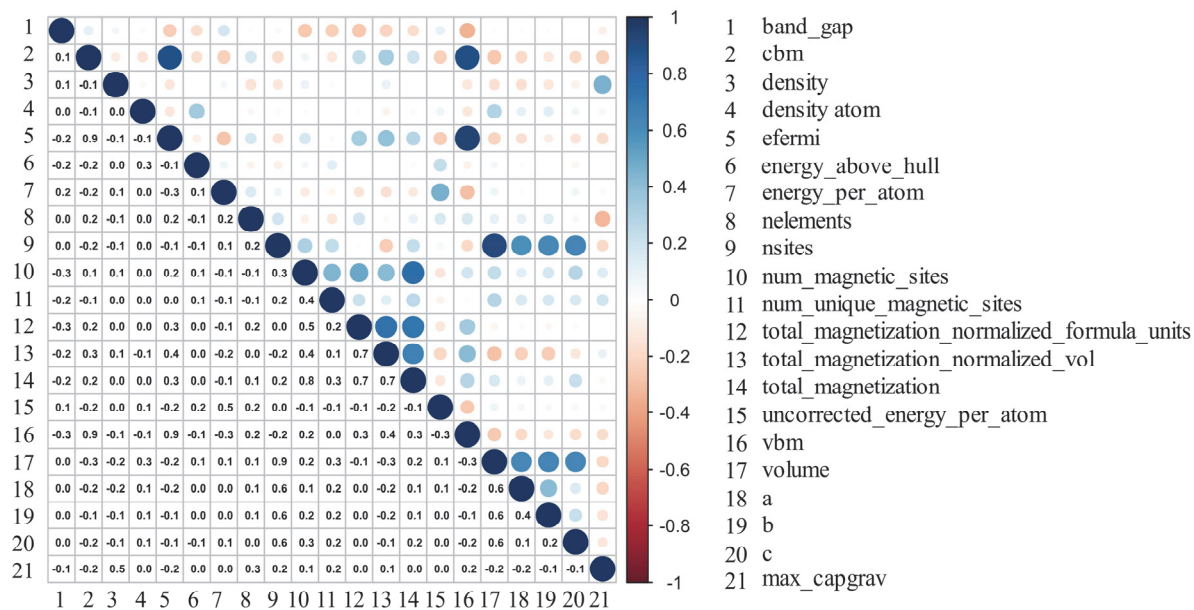


图2 特征向量与目标向量皮尔森系数(R 值)热力图

Figure 2 Heatmap of Pearson coefficients (R values) between feature vectors and target vectors

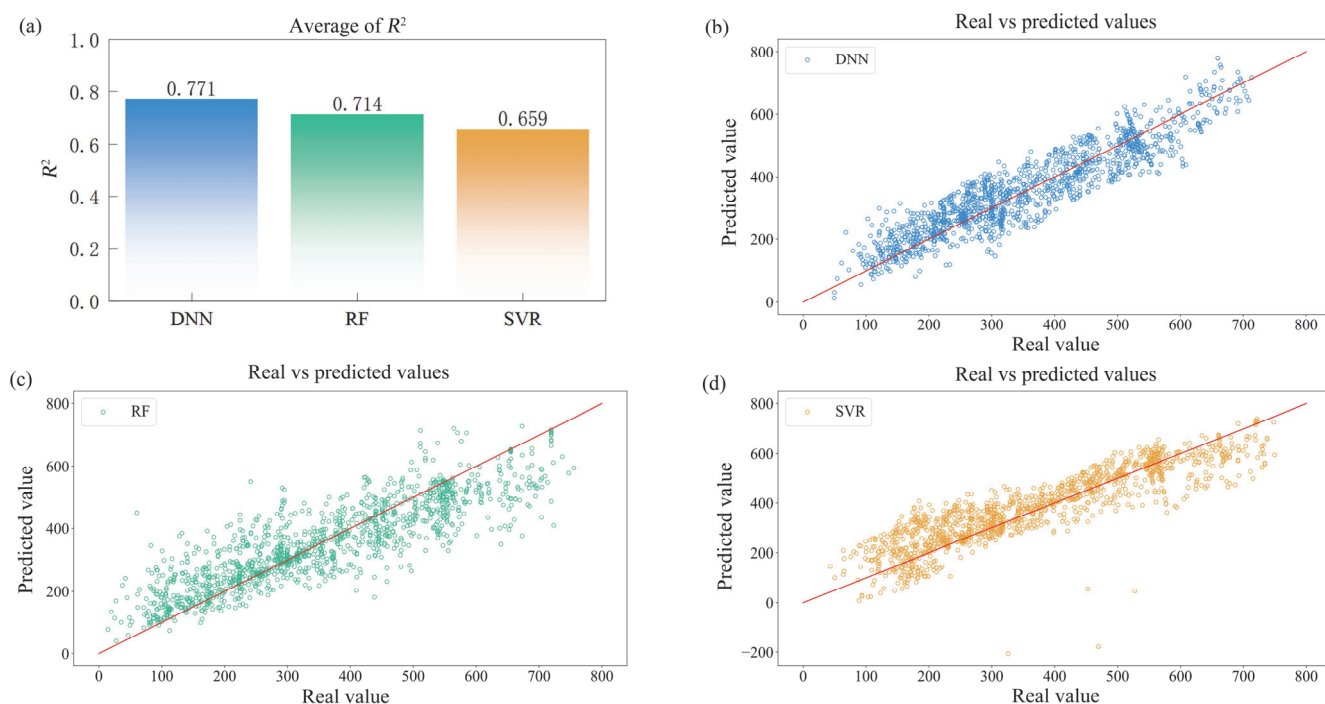


图3 对钒基材料预测比容量的机器学习算法性能评估。(a)三种算法平均 R^2 ; DNN (b)、RF (c)、SVR (d)真实值与预测值对应关系

Figure 3 Performance evaluation of machine learning algorithms for predicting specific capacity in vanadium-based materials. (a) Average R^2 scores of the three algorithms, (b) actual vs. predicted values for DNN, (c) actual vs. predicted values for RF, and (d) actual vs. predicted values for SVR

分别为 62、34、43 和 42。此外,学习率衰减最终被设定为 0.00001。其次,表现优异的是随机森林算法,决策树数量为 75、最大深度为 5、最大叶节点数设置为 30。支持向量回归算法的 C (惩罚因子)与 ϵ 均设置为 0.5、核函数选择径向基核(RBF)、 γ 参数选择 0.02,并且其 R^2 分数达到了 0.659。图 3(b)、(c)和(d)分别为 DNN、RF、SVR 算法的实际值与预测值之间的对应关系。横轴

代表实际值,而纵轴则表示预测值。整体来看,三种算法预测值与实际值的对应关系主要集中在两个容量区间内: 200~300 mAh/g 和 500~600 mAh/g。这之前对数据集的分析结果相符,并进一步证实了模型预测的准确性。深度神经网络(DNN)算法的点分布最为接近回归线 $y=x$, 这表明其预测效果最佳。但是在 600 mAh/g 之后预测效果相对于低范围时下降,这可能是训练集在

600 mAh/g 之后的样本点稀少的原因, 在之后利用分析结果筛选材料时应注意这个现象. 随机森林(RF)算法大多数数据点都集中在拟合线附近, 但小于 200 mAh/g 以及大于 600 mAh/g 的值出现偏差, 与 DNN 算法类似, 偏差出现原因在于数据分布不均匀以及数据中非线性关系较为复杂, 也说明了钒基材料数据的高维特征之间交互性较强^[43]. 支持向量回归(SVR)算法的 R^2 值低于其他两种算法, 表明在这三个模型中, SVR 在预测任务上的表现相对较差. 其原因在于, SVR 是一种基于边界的回归方法, 试图在保证一定边界内预测误差最小的情况下, 找到一个尽可能平滑的回归超平面^[44]. 此外, 鉴于所选取的特征(如密度、电子结构相关参数等)是从材料的晶体结构、电子结构及热力学性能等多维度综合考量而确定的, 同时所采用的三种机器学习算法均为通用的数据处理和模型解释手段, 也能够为不同材料体系的研究提供相似的分析框架, 如钴基、锰基体系.

图 4(a)与(b)展示了深度神经网络算法中经相关性筛选之后各个特征的 SHAP 分析结果^[45]. 4(a)为特征重要性的分布图, 这是一种直观展示各个特征对模型预测影响程度的工具. 在先前相关性研究环节, 已经通过统计分析确定了各个特征与理论容量之间的相关性. 为了进一步验证这些发现, 采用了 SHAP 值(Shapley Additive exPlanations)分析, SHAP 值是一种解释模型预测的方法, 基于博弈论中的 Shapley 值, 能够量化每个特征对模型预测结果的贡献度^[46]. 发现材料密度的 SHAP 平均值显著高于其他特征, 这进一步证明了在数据层面上密度对于理论容量的影响程度较深. 此外, 材料中总元素种类数量(nelements)与材料中原子或分子数量(nsites)这两个特征也具有较高影响. nelements 越多, 不同元素的协同作用越复杂. 不同元素的协同作用, 不仅能够提高电导率和化学吸附能力, 还可以保持材料的结构完整性^[47]. 较多的 nsites, 可以构建出更复杂的结构网络, 从而增加无序结构的可能性, 提供更多的潜在离子迁移路径, 减少了锂离子在传输过程中的阻力^[48-49]. 图 4(b)展示了各个特征值分布的小提琴图. 从图中可以看出, 密度特征的颜色分布较深, 这表明对于 DNN 模

型的预测具有较高的影响力. 同时, 小提琴图横向形状较宽, 说明在数据集中分布较为广泛, 包含多样化的值, 数据集中的样本具有广泛的变异性^[50]. 此外, 在小提琴图中未观察到广泛数据点偏离集中分布区域的现象, 这表明异常值与极端值较少^[51].

通过密度这个重要特征进行筛选工作, 选定原始数据集前 0.5% 的数据, 确定了密度大于 3.70 的 26 种钒基材料(支持信息表 S3). 如表 1 所示, 此处展示理论容量最高的三种材料信息, 包括化学式、计算的理论比容量、密度以及在 Materials Project 数据库中的 ID 信息, 若要想进一步了解该材料基本信息, 可以根据 ID 前往官方网站(<https://next-gen.materialsproject.org>)进行检索. 可以发现, 所筛选的 26 种候选材料大部分皆为钒氧化物以及钒酸盐材料, 诸如氧化钒、钒酸镁、钒酸铈等. 钒氧化物由于其多价态特性, 在充放电过程中可以发生可逆的氧化还原反应, 从而实现较高的理论容量^[52]. 钒酸盐材料则因其层状结构和隧道结构提供了快速的离子扩散通道, 有助于提高材料的倍率性能和循环稳定性^[53]. 这些特性共同作用, 使得钒基材料成为高性能锂离子电池电极材料的理想候选者.

表 1 优异性能候选材料基本信息

Table 1 Basic information of candidate materials with superior performance

ID	Formula	Density/(g·cm ⁻³)	Specific capacity/(mAh·g ⁻¹)
mp-18896	V ₆ O ₁₃	3.927	755.468
mp-2229279	MgV ₅ O ₁₂	3.877	699.022
mp-27877	V ₂ MoO ₈	3.755	678.126

2.4 计算验证

V₆O₁₃、MgV₅O₁₂ 和 V₂MoO₈ 优化之后的晶体结构如图 S2. V₆O₁₃ 属单斜 $C2/m$ 空间群, 含 V⁵⁺ 和两种 V⁴⁺, V—O 键长 0.166~0.228 nm. MgV₅O₁₂ 属三斜 $P1$ 空间群, Mg²⁺ 与六个 O²⁻ 形成扭曲八面体, V⁴⁺ 和 V³⁺ 六配位. MoV₂O₈ 亦为单斜 $C2/m$ 空间群, V⁵⁺ 和 Mo⁶⁺ 六配位, Mo—O 键长 0.189~0.195 nm, 各化合物中 O²⁻ 配位数

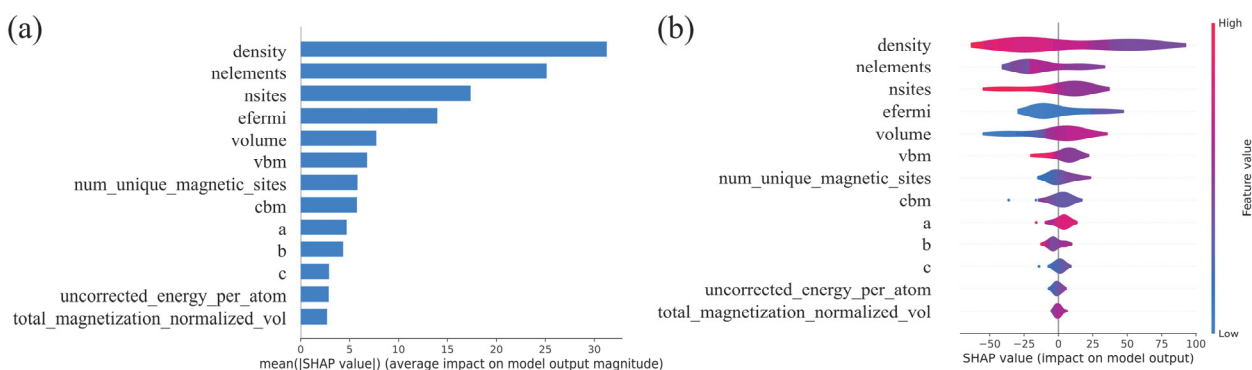


图 4 DNN 模型 SHAP 分析. (a) SHAP 特征的总体特征图; (b) SHAP 特征的重要性值(小提琴图)

Figure 4 DNN model SHAP analysis. (a) Overall feature map of SHAP features; (b) importance values of SHAP features (violin plots)

及几何形态多样。

图 5 为三种材料在费米能级附近能带结构、态密度图(图 S3 为整体图)。V₆O₁₃ [图 5(a)]的能带与费米能级相交, 显示出固有的金属性质^[54]。其态密度图在费米能级附近形成明显峰值, 表明该材料具有较高的电子态密度, 有利于提高导电性, 其中 V(d)轨道对电子态密度的贡献尤为显著。MgV₅O₁₂ [图 5(b)]同样显示了金属特性, 费米能级附近的电子态密度主要由 V(s, p, d)轨道贡献, 特别是 d 轨道, 而 O(s, p)和 Mg(s, p)轨道也有一定贡献但相对较小。相比之下, V₂MoO₈ [图 5(c)]的费米能级位于价带顶和导带底之间, 表现出半金属特性^[55], O(p)轨道对其电子态密度贡献最大。总体上, 这三种材料均展现了良好的导电能力, 但 V₆O₁₃ 在费米能级附近的电子态密度最高, 显示出更优的导电性能^[56]。图 5(d)展示了三种材料的电子密度分布情况, 右侧的颜色条表示相对电子密度, 红色代表较高密度, 蓝色表示较低密度。对于 V₆O₁₃, 电子密度主要集中在氧上, 由于锂离子通常是通过与阴离子(在这里主要是氧原子)之间的库仑力进行吸附的, 因此高电子密度区域能够增强锂离子的吸附强度^[57]。此外, 电子云的密集分布可能有助于形成稳定的锂插入位点, 这对于实现高效的充放电循环至关重要^[58]。针对 MgV₅O₁₂, 除了氧原子外, 还可以观察到电子密度聚集在镁原子周围, 表明镁的存在会改变局部电子环境, 影响锂离子的吸附位置或方式^[59]。有研究表明, 如果某些电子密度高的区域靠近镁原子, 则可能成为锂离子优先选择的吸附位点^[60]。至于 V₂MoO₈, 虽然电子密度依然主要集中在氧原子上, 但钼原子周围的电

子密度不容忽视, 作为过渡金属, 未填满的 d 轨道能够参与成键并影响电子云的形状, 这可能导致特定的电子密度模式, 使得锂离子更容易找到合适的吸附位点^[61]。

图 6 为三种材料吸附能及开路电压计算图。图 6(a)为计算出的最低构型, 产生的吸附位点及吸附能如表 S4 所示。MgV₅O₁₂ 产生的吸附位点为三个, 而 V₆O₁₃ 与 V₂MoO₈ 产生的吸附位点为一个, 可能由于 MgV₅O₁₂ 结晶于三斜晶系的 P1 空间群, 其结构较为复杂, 包含多种不同位置的 Mg²⁺、V⁴⁺和 V³⁺, 而 V₆O₁₃ 与 V₂MoO₈ 结晶于单斜晶系的 C2/m 空间群, 其结构相对简单, 导致了 MgV₅O₁₂ 的表面可能存在多种不同的活性位点^[62]。其中 V₆O₁₃ 的吸附能相较于 MgV₅O₁₂ 与 V₂MoO₈ 更低, 为 -3.304 eV, 可能因为 MgV₅O₁₂ 和 V₂MoO₈ 中的电荷分布相对集中, 特别是在离子键形成的区域, 这种集中的电荷分布使得外来分子难以有效地与表面发生相互作用, 从而降低了吸附能力^[63]。

电极材料的开路电压(OCV)是锂离子电池中的重要参数^[17]。在此构建超胞模型来研究三种电极材料的开路电压, V₆O₁₃、MgV₅O₁₂ 和 V₂MoO₈ 通过构建 2×1×2、2×2×1、1×1×4 超胞模型来研究开路电压, 通过在之前计算出的三种材料吸附位点上放置不同数目的锂离子, 探索了一系列化学式为 MLi_c (M 为 V₆O₁₃、MgV₅O₁₂ 和 V₂MoO₈)的吸附构型, 其中 $c=0.25, 0.5, 0.75$ 和 1 定义为每个化学式 MLi_c 中锂离子的数目, c 越大, 表明锂离子在 MLi_c 上的覆盖率越高(见支持信息图 S4、S5、S6)。图 6(b)左侧显示了吸附能对 c 的依赖性。可以看出, 由于锂离子之间的排斥作用增强, 吸附能随着吸附锂离子

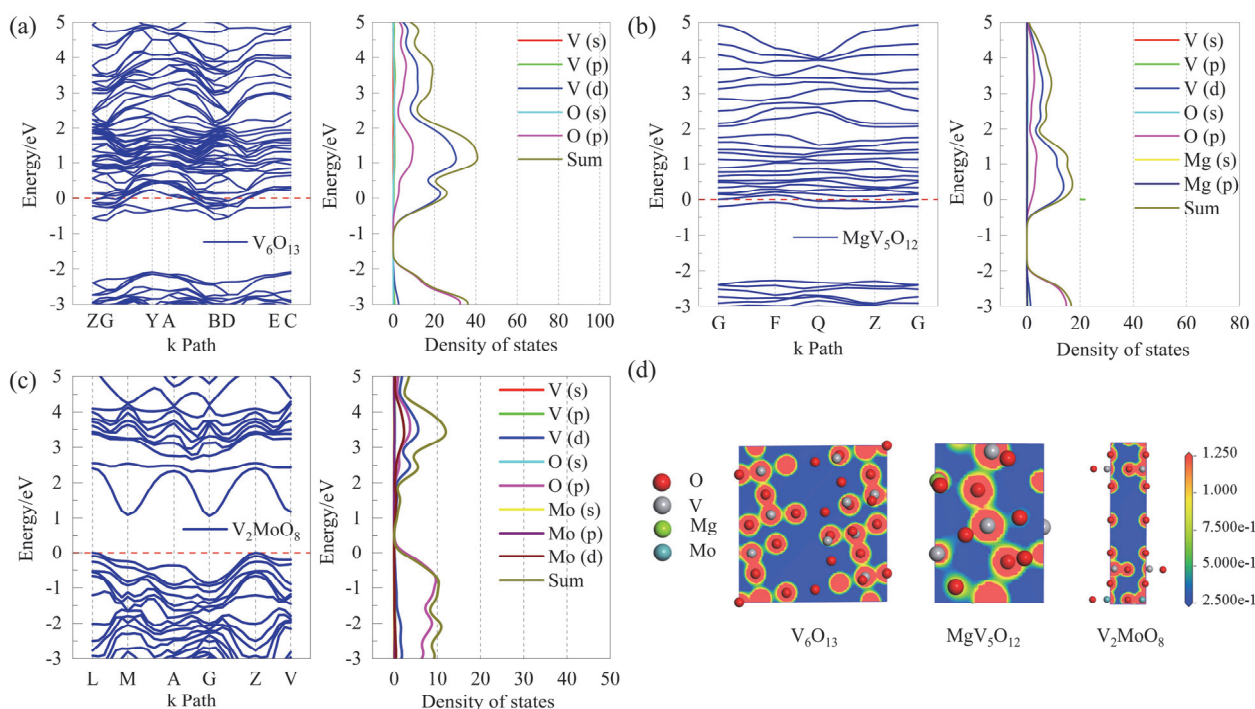


图 5 V₆O₁₃ (a)、MgV₅O₁₂ (b)和 V₂MoO₈ (c)能带结构与态密度和三种材料电子密度(d)

Figure 5 Electronic structure and density of states (DOS) for V₆O₁₃ (a), MgV₅O₁₂ (b), and V₂MoO₈ (c), along with the electron density of the three materials (d)

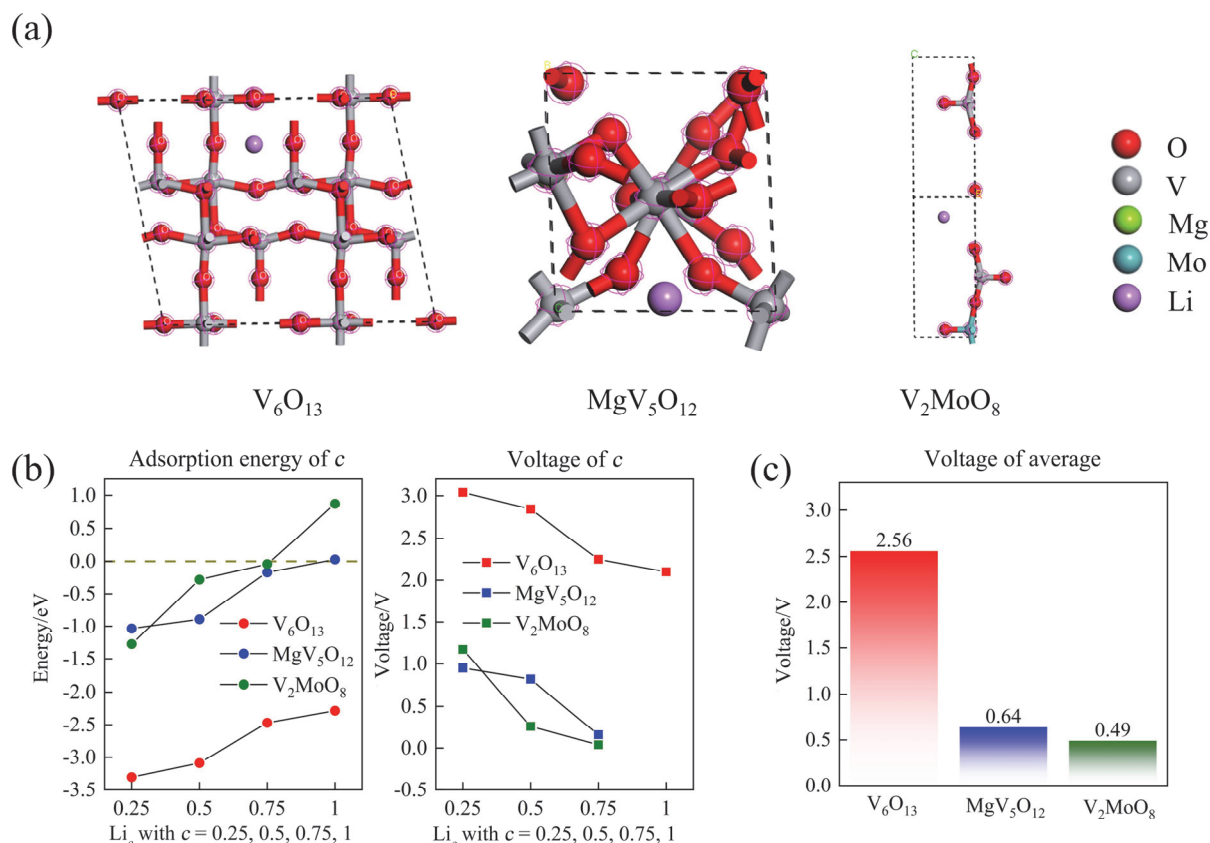


图6 三种材料吸附能及开路电压计算图。(a)最低吸附构型；(b)吸附能及嵌锂电压；(c)开路电压

Figure 6 Adsorption energy and open-circuit voltage for three materials. (a) Lowest-energy adsorption configurations; (b) adsorption energy and intercalation voltage; (c) open circuit voltage

数目的增大而减小, V_6O_{13} 、 MgV_5O_{12} 这种趋势并不明显, 吸附能在 c 从 0.25 到 1 分别增长了 1.021 和 1.063 eV, 值得注意的是, MgV_5O_{12} 在 $c=1$ 时, 吸附能增长到了 0.0296 eV (>0 eV), 表明吸附构型开始变得不稳定^[64]. 而 V_2MoO_8 增长了 2.142 eV, 尤其在 $c \geq 0.75$ 时尤为明显, 从 -0.042 eV 增长到了 0.87 eV, 这可能是因为 Mo 的引入改变了电子的分布, 尤其是 Mo 的 d 轨道对总态密度的贡献, 影响了吸附过程中电子的转移和共享, 从而影响吸附能^[61]. 图 6(b)右侧显示了开路电压随 c 变化图. 吸附锂离子最大数目对于 $V_6O_{13}Li_c$ 是 $c=1$, 而对于 $MgV_5O_{12}Li_c$ 和 $V_2MoO_8Li_c$ 是 $c=0.75$. 因此可以算出锂离子的 V_6O_{13} 、 MgV_5O_{12} 和 V_2MoO_8 的平均 OCV 分别为 2.56 V (这与实验值极其相近)^[65]、0.64 V、0.49 V, 如图 6(c)所示. 经平均 OCV 计算得出, V_6O_{13} 适合作为锂离子电池正极材料, 而 MgV_5O_{12} 和 V_2MoO_8 适合作为锂离子电池负极材料^[17].

3 结论

通过机器学习模型与第一性原理计算构建了一种筛选-验证流程. 从 4694 条钒基数据中发现了 26 种潜在的候选材料. 具体而言, 从 Materials Project 数据库中提

取了包含 20 种关键特征的 4694 条钒基材料信息, 并使用 pymatgen 高通量计算以确定每种材料的最大理论比容量. 为了筛选最相关特征向量, 通过皮尔森系数(R)分析了每种特征与比容量的相关性, 并初步筛选出有效特征并排除了冗余特征. 其中, 密度显示出与比容量的高度相关性. 通过对三种不同算法的综合比较, 深度神经网络算法(DNN)获得了最高的预测评分. 利用 SHAP 解释模型揭示了密度对 DNN 模型预测结果的贡献度最高. 基于这些发现, 按照密度特征选取原始数据集前 0.5% 数据, 最终确定了 26 种有潜力的电极候选材料. 经第一性原理计算验证, 验证了 3 种理论容量均大于 650 mAh/g, 平均电压分别为 2.56、0.64、0.49 V 的钒基正负两电极候选材料. 这项研究表明, 这一筛选-验证流程对于探索新型钒基候选材料是一种有效的替代方法, 并有可能扩展应用于其他材料的研发领域.

4 实验部分

4.1 算法模型与评估

构建了一种以深度神经网络(DNN)为核心架构的预测模型, 该模型被设计用于与支持向量回归(SVR)、随机森林(RF)算法进行比较, 以确定最优的特征预测模

型. 为了进一步优化模型的性能, 应用了遗传算法来调整三种算法超参数(包括但不限于学习率和各隐藏层中神经元的数量). 遗传算法的种群大小与进化代数皆设置为 50, 交叉与异变概率分别为 0.5 与 0.2(优化流程参见支持信息 S2.4). 为了确保模型评估结果的准确性及增强模型的泛化能力, 采取了 5 折交叉验证以减少模型对于特定数据集的依赖性^[66]. 为了量化模型性能, 选取了决定系数(R^2)作为评价标准, 该系数能够直观地反映模型解释变量间关系的能力, 其计算方法如公式(1)所示^[67]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

其中 y_i 为目标变量真实值, $\hat{y}$ 为目标变量预测值, $\bar{y}$ 为目标变量预测值的平均值.

4.2 第一性原理计算方法

构建了 V_6O_{13} 、 MgV_5O_{12} 、 V_2MoO_8 超胞模型, 交换关联函数采用广义梯度近似(GGA)中的 Perdew-Burke-Ernzerhof (PBE)泛函处理, 离子核与价电子之间的相互作用通过超软赝势方法处理. V_6O_{13} 、 MgV_5O_{12} 、 V_2MoO_8 的截断能设置皆为 571.40 eV; 积分波函数积分点数量(k-point)分别为 $2 \times 7 \times 3$, $5 \times 5 \times 3$, $1 \times 7 \times 6$. 几何优化过程中采用了 Broyden-Fletcher-Goldfarb-Shanno (BFGS)方法. 结构优化过程中, V_6O_{13} 、 MgV_5O_{12} 、 V_2MoO_8 的所有原子位置和晶格参数都进行了完全弛豫.

计算了吸附能, 锂离子在三种材料上的吸附能定义为^[68]:

$$E_{\text{ads}} = (E_{\text{total}} - E_{\text{substrate}} - cE_{\text{Li}}) / c \quad (2)$$

其中 E_{total} 是体系吸附之后的总能量, $E_{\text{substrate}}$ 是基底能量, E_{Li} 是锂单质的化学势, 并被视为在其体相中每一个客体原子的能量, c 是吸附原子数.

此外, 还计算了开路电压与平均开路电压, 这是表征可充电电池输出电压的关键参数, 开路电压公式定义为^[17]:

$$\text{OCV} = (E_{\text{M}} + cE_{\text{Li}} - E_{\text{MLi}}) / c \quad (3)$$

这里 M 表示 V_6O_{13} 、 MgV_5O_{12} 和 MoV_2O_8 三种材料, E_{M} , E_{MLi} 表示吸附锂离子前后的能量, E_{Li} 表示锂单质能量.

References

- [1] Li, Z.; Wang, Y.; Wang, J.; Wu, C.; Wang, W.; Chen, Y.; Hu, C.; Mo, K.; Gao, T.; He, Y. S.; Ren, Z.; Zhang, Y.; Liu, X.; Liu, N.; Chen, L.; Wu, K.; Shen, C.; Ma, Z.-F.; Li, L. *Nat. Commun.* **2024**, *15*, 10216.
- [2] Liu, J.; Zhu, J.; Zhang, X.; Zhang, J.; Huang, C.; Jia, G.; An, S. *Acta Chim. Sinica* **2025**, *83*, 101 (in Chinese). (刘继洪, 祝佳鹏, 张旭,

- 张纪阳, 黄超洋, 贾桂霄, 安胜利, 化学学报, **2025**, *83*, 101.)
- [3] Sun, L.; Liu, Y.; Wang, L. J.; Jin, Z. *Adv. Funct. Mater.* **2024**, *34*, 2403032.
- [4] Chen, B.; Xu, K.; Chen, Q.; Chen, L. *Acta Chim. Sinica* **2024**, *82*, 1209 (in Chinese). (陈博文, 徐克, 陈琪, 陈立桅, 化学学报, **2024**, *82*, 1209.)
- [5] Wang, Q.; O'Carroll, T.; Shi, F.; Huang, Y.; Chen, G.; Yang, X.; Nevar, A.; Dudko, N.; Tarasenko, N.; Xie, J.; Shi, L.; Wu, G.; Zhang, D. *Electrochem. Energy Rev.* **2024**, *7*, 15.
- [6] Akhmetova, K.; Sultanov, F.; Mentbayeva, A.; Umirov, N.; Bakenov, Z.; Tatykayev, B. *J. Power Sources* **2024**, *624*, 235531.
- [7] Dhanabalan, K.; Aruchamy, K.; Sriram, G.; Sadhasivam, T.; Oh, T. H. *Resour. Conserv. Recycl.* **2024**, *209*, 107825.
- [8] Xia, X. Y.; Guo, Y.; Zhao, Y. X.; Bai, X. T.; Zhu, J. J.; Yang, W. L.; Wang, Y.; Fu, L. C.; Zhou, L. P. *Ionics* **2024**, *30*, 607.
- [9] Zhao, W. B.; Zhao, C. H.; Wu, H.; Li, L. J.; Zhang, C. C. *J. Energy Storage* **2024**, *81*, 110409.
- [10] Sandhya, C. P.; John, B.; Gouri, C. *Ionics* **2014**, *20*, 601.
- [11] Zhang, X. H.; Wang, D. H.; Qiu, X. Y.; Ma, Y. J.; Kong, D. B.; Müllen, K.; Li, X. L.; Zhi, L. *J. Nat. Commun.* **2020**, *11*, 3826.
- [12] Peng, X. X.; Zhang, S. C.; Zou, Z. G.; Ling, W. Q.; Geng, J.; Zhong, S. L.; Liang, F. G. *J. Electron. Mater.* **2023**, *52*, 4413.
- [13] Wang, Y. J.; Liang, F. G.; Zou, Z. G.; Zhang, S. C.; Jia, S. K.; Nong, J. X.; Zheng, R.; Song, L. J. *J. Mater. Sci.: Mater. Electron.* **2024**, *35*, 1439.
- [14] Teo, L. P.; Buraidah, M. H.; Arof, A. K. *Ionics* **2015**, *21*, 2393.
- [15] De Juan Corpuz, L. M. Z.; Nguyen, M. T.; Corpuz, R. D.; Yonezawa, T.; Rosero Navarro, N. C.; Tadanaga, K.; Tokunaga, T.; Kheawhom, S. *ACS Appl. Nano Mater.* **2019**, *2*, 4247.
- [16] McNulty, D.; Noel Buckley, D.; O' Dwyer, C. *ACS Appl. Energy Mater.* **2019**, *2*, 822.
- [17] Liu, H. L.; Cai, Y. M.; Guo, Z. D.; Zhou, J. *ACS Omega* **2022**, *7*, 17756.
- [18] Li, J.; Song, Q.; Liu, Z.; Wang, D. *Acta Chim. Sinica* **2024**, *82*, 387 (in Chinese). (李珺卿, 宋千禧, 刘子义, 王东琪, 化学学报, **2024**, *82*, 387.)
- [19] Li, K.; Choudhary, K.; DeCost, B.; Greenwood, M.; Hattrick-Simpers, J. J. *Mater. Chem. A* **2024**, *12*, 12412.
- [20] Wang, T.; Tan, X.; Wei, Y.; Jin, H. *Mater. Today Commun.* **2021**, *29*, 102932.
- [21] Allam, O.; Kuramshin, R.; Stoichev, Z.; Cho, B. W.; Lee, S. W.; Jang, S. S. *Mater. Today Energy* **2020**, *17*, 100482.
- [22] Jinnouchi, R.; Karsai, F.; Kresse, G. *npj Comput. Mater.* **2024**, *10*, 107.
- [23] Joshi, R. P.; Eickholt, J.; Li, L.-L.; Fornari, M.; Barone, V.; Peralta, J. E. *ACS Appl. Mater. Interfaces* **2019**, *11*, 18494.
- [24] Moses, I. A.; Joshi, R. P.; Ozdemir, B.; Kumar, N.; Eickholt, J.; Barone, V. *ACS Appl. Mater. Interfaces* **2021**, *13*, 53355.
- [25] Butler, K. T.; Choudhary, K.; Csanyi, G.; Ganose, A. M.; Kalinin, S. V.; Morgan, D. *npj Comput. Mater.* **2024**, *10*, 231.
- [26] Wang, A.; Kingsbury, R.; McDermott, M.; Horton, M.; Jain, A.; Ong, S. P.; Dwaraknath, S.; Persson, K. A. *Sci. Rep.* **2021**, *11*, 15496.
- [27] Liebmann, T.; Heubner, C.; Schneider, M.; Michaelis, A. *Mater. Today Energy* **2021**, *22*, 100845.
- [28] Yang, Y. T.; Wang, W. D.; Tang, M. X.; Liu, W. H.; Liu, Q. J. *Diamond Relat. Mater.* **2025**, *151*, 111831.
- [29] Armand, M.; Axmann, P.; Bresser, D.; Copley, M.; Edström, K.; Ekberg, C.; Guyomard, D.; Lestriez, B.; Novák, P.; Petráňková, M.; Porcher, W.; Trabesinger, S.; Wohlfahrt-Mehrens, M.; Zhang, H. J. *Power Sources* **2020**, *479*, 228708.
- [30] Lu, Y.; Chen, J. *Nat. Rev. Chem.* **2020**, *4*, 127.
- [31] Erraji, A.; Masrour, R.; Xu, L. *Ionics* **2024**, *30*, 5979.
- [32] Singh, D.; Singh, B. *Pattern Recognit.* **2022**, *122*, 108307.
- [33] Raissi, M.; Perdikaris, P.; Karniadakis, G. E. *J. Comput. Phys.* **2017**, *348*, 683.
- [34] Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J. L.; Holm, E.; Ong, S. P.; Wolverton, C. *npj Comput. Mater.* **2022**, *8*, 59.
- [35] Zhang, M.; Li, W.; Zhang, L.; Jin, H.; Mu, Y.; Wang, L. *Inf. Sci.* **2023**, *639*, 118737.
- [36] Pujari, S. S.; Bobade, R. G.; Shaikh, S. F.; Al-Enizi, A. M.; Ambare, R. C.; Lokhande, B. J. *J. Mater. Sci.: Mater. Electron.* **2024**, *35*, 2162.
- [37] Manthiram A. *Nat. Commun.* **2020**, *11*, 1550.
- [38] Hikima, K.; Shimizu, K.; Kiuchi, H.; Hinuma, Y.; Suzuki, K.;

- Hirayama, M.; Matsubara, E.; Kanno, R. *Commun. Chem.* **2022**, *5*, 52.
- [39] Wang, Z.; Liu, L.; Zhu, M.; Sun, Y.; Zhao, Q.; Ding, Y.; Lu, J.; Wang, C.; Li, Q.; He, A.; Ye, F. *Acta Chim. Sinica* **2024**, *82*, 589 (in Chinese). (王筑城, 刘磊, 朱梦媛, 孙悦, 赵晴, 丁玉寅, 陆继鑫, 王存国, 李奇, 贺爱华, 叶付臣, 化学学报, **2024**, *82*, 589.)
- [40] Afshar, M.; Usefi, H. *Expert Syst. Appl.* **2022**, *200*, 116928.
- [41] Aghaabbasi, M.; Chalermpong, S. *Travel Behav. Soc.* **2023**, *33*, 100640.
- [42] Kim, M.; Kang, S.; Park, H. G.; Park, K.; Min, K. *Chem. Eng. J.* **2023**, *452*, 139254.
- [43] Fife, D. A.; D'Onofrio, J. *Behav. Res. Methods* **2023**, *55*, 2447.
- [44] Sabzevar, M.; Hasheminejad, S. M. H. *Chaos, Solitons Fractals* **2021**, *144*, 110738.
- [45] Wang, H.; Liang, Q.; Hancock, J. T.; Khoshgoftaar, T. M. *J. Big Data* **2024**, *11*, 44.
- [46] Ponce-Bobadilla, A. V.; Schmitt, V.; Maier, C. S.; Mensing, S.; Stodtmann, S. *Clin. Transl. Sci.* **2024**, *17*, e70056.
- [47] Xie, Y.; Cao, J.; Wang, X.; Li, W.; Deng, L.; Ma, S.; Zhang, H.; Guan, C.; Huang, W. *Nano Lett.* **2021**, *21*, 8579.
- [48] Griffin, S. L.; Meekel, E. G.; Bulled, J. M.; Canossa, S.; Wahrhaftig-Lewis, A.; Schmidt, E. M.; Champness, N. R. *Nat. Commun.* **2025**, *16*, 3209.
- [49] Chen, Z.; Du, T.; Krishnan, N. M. A.; Yue, Y.; Smedskjaer, M. M. *Nat. Commun.* **2025**, *16*, 1057.
- [50] Tahmouresi, M. S.; Niksokhan, M. H.; Ehsani, A. H. *Sci. Rep.* **2024**, *14*, 25454.
- [51] Hullman, J.; Resnick, P.; Adar, E. *PLoS One* **2015**, *10*, e0142444.
- [52] Gamal, H.; Elshahawy, A. M.; Medany, S. S.; Hefnawy, M. A.; Shalaby, M. S. *J. Energy Storage* **2024**, *76*, 109788.
- [53] Xu, H.; Wang, Q. Y.; Jiang, M.; Li, S. S. *Anal. Chim. Acta* **2024**, *1295*, 342270.
- [54] Van Efferen, C.; Berges, J.; Hall, J.; Van Loon, E.; Kraus, S.; Schobert, A.; Wekking, T.; Huttmann, F.; Plaar, E.; Rothenbach, N.; Ollefs, K.; Arruda, L. M.; Brookes, N.; Schönhoff, G.; Kummer, K.; Wende, H.; Wehling, T.; Michely, T. *Nat. Commun.* **2021**, *12*, 6837.
- [55] Kong, S.; Ricci, F.; Guevarra, D.; Neaton, J. B.; Gomes, C. P.; Gregoire, J. M. *Nat. Electron.* **2024**, *7*, 638.
- [56] Kong, S.; Ricci, F.; Guevarra, D.; Neaton, J. B.; Gomes, C. P.; Gregoire, J. M. *Nat. Commun.* **2022**, *13*, 949.
- [57] Du, D.; Liu, P.; Tian, G.; Xu, H.; Wang, X.; Liu, S.; Fan, F.; Wang, S.; Wang, C.; Zeng, C.; Shu, C. *J. Colloid Interface Sci.* **2025**, *678*, 570.
- [58] Li, S.; Chen, B.; Shi, Z.; Tong, Q.; Weng, J. *J. Energy Storage* **2024**, *102*, 114210.
- [59] Aspinall, J.; Sada, K.; Guo, H.; Kotakadi, S.; Narayanan, S.; Chart, Y.; Jagger, B.; Milan, E.; Brassart, L.; Armstrong, D.; Pasta, M. *Nat. Commun.* **2024**, *15*, 4511.
- [60] Zhong, J.; Lin, S.; Yu, J. *Desalination* **2021**, *505*, 114983.
- [61] Zhang, W.; He, X.; He, C. *J. Colloid Interface Sci.* **2025**, *678*, 540.
- [62] Qian, F.; Peng, L.; Cao, D.; Jiang, W.; Hu, C.; Huang, J.; Zhang, X.; Luo, J.; Chen, S.; Wu, X.; Song, L.; Chen, Q. *Joule* **2024**, *8*, 2342.
- [63] Xu, D. D.; Ma, R. R.; Fu, A. P.; Guan, Z.; Zhong, N.; Peng, H.; Xiang, P. H.; Duan, C. G. *Nat. Commun.* **2021**, *12*, 655.
- [64] Lee, C. L.; Hsi, H. C.; Chang, C. M. *J. Nanopart. Res.* **2023**, *25*, 59.
- [65] Zhong, S.; Zou, Z.; Lv, S.; Zhang, S.; Geng, J.; Meng, J.; Liu, X.; Liang, F.; Rao, J. *J. Alloy Compd.* **2022**, *903*, 163845.
- [66] Adin, A.; Krainski, E. T.; Lenzi, A.; Liu, Z.; Martínez-Minaya, J.; Rue, H. *Spat. Stat.* **2024**, *62*, 100843.
- [67] Hayes, T. *Behav. Res. Methods* **2021**, *53*, 2127.
- [68] Li, H.; Pan, C.; Peng, X.; Zhang, B.; Song, S.; Xu, Z.; Qiu, X.; Liu, Y.; Wang, J.; Guo, Y. *Nat. Commun.* **2025**, *16*, 2439.

(Cheng, F.)