

物理学报
Acta Physica Sinica
ISSN 1000-3290, CN 11-1958/O4

《物理学报》网络首发论文

题目：基于机器学习和第一性原理计算的 Janus 材料的预测
作者：张桥, 谭薇, 宁勇祺, 聂国政, 蔡孟秋, 王俊年, 朱慧平, 赵宇清
网络首发日期：2024-11-08
引用格式：张桥, 谭薇, 宁勇祺, 聂国政, 蔡孟秋, 王俊年, 朱慧平, 赵宇清. 基于机器学习和第一性原理计算的 Janus 材料的预测[J/OL]. 物理学报.
<https://link.cnki.net/urlid/11.1958.o4.20241107.1347.002>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于机器学习和第一性原理计算的 Janus 材料的预测 *

张桥¹⁾ 谭薇¹⁾ 宁勇祺¹⁾ 聂国政¹⁾ 蔡孟秋²⁾ 王俊年¹⁾ 朱慧平³⁾ 赵宇清^{1)4)†}

1) (湖南科技大学物理与电子科学学院, 智能传感器与新型传感器材料湖南省重点实验室, 湘潭 411201)

2) (湖南大学物理与微电子科学学院, 长沙 410082)

3) (中国科学院微电子研究所, 硅器件中心重点实验室, 北京 100029)

4) (中国科学院半导体研究所, 超晶格国家重点实验室, 北京 100084)

寻找尺寸小、稳定性高和易操控的纳米磁结构—磁斯格明子 (magnetic skyrmion), 是发展下一代高密度、高速度和低能耗非易失性信息存储器件核心存储单元的关键. 磁性斯格明子根据其拓扑产生机制, 可以由非中心对称结构诱导的 DMI (Dzyaloshinskii–Moriya Interaction) 作用项产生. 二维 Janus 结构具有两个不同面的原子层, 可以形成垂直内建电场, 打破中心空间反演对称性. 因此寻找具有本征磁性的二维 Janus 材料是研究新型磁存储的基础. 本文基于晶体材料数据库 Materials Project 中的 1179 种六角晶系 ABC 型 Janus 材料数据, 以其元素组分信息为特征描述符, 构建了随机森林, 梯度提升决策树, 极端梯度提升和极端随机树等四种机器学习模型, 基于上述模型对晶格常数、形成能和磁矩分类进行了预测, 并采用十折交叉验证法对模型进行了评估. 梯度提升决策树在磁矩分类预测显示出最高的精度和泛化能力. 最后, 基于上述模型对尚未发现的 82018 种二维 Janus 材料进行了预测, 筛选得到 4024 种具有热稳定性的高磁矩结构, 并基于第一性原理的方法对其中随机抽样的 13 种高磁矩结构进行了计算验证. 本研究为二维 Janus 材料磁矩分类和高通量筛选训练了有效的机器学习模型, 加速了二维 Janus 结构磁性的探索.

关键词: 机器学习, 二维 Janus 材料, 磁矩, 第一性原理计算

PACS: 02.60.Cb, 81.05.Zx, 75.20.-g, 71.15.Mb

1 引言

自 2004 年以来, 超薄的二维 (2D) 材料石墨烯^[1]、过渡金属硫化物 (TMDs)^[2]、过渡金属碳化物/氮化物 (MXenes)^[3] 及二维金属卤化物钙钛矿^[4] 的先后发现^[5], 为谷电子学、自旋电子学以及铁电^[6-8] 等新奇物理特性的研究提供了广泛的平台, 并在光电、传感、非易失性铁电和铁磁存储领域引起了巨大的关注. 其中铁电和铁磁耦合而成的多铁材料的设计是实现下一代新型逻辑存储的重要途径之一, 其存储原理为: 在 2D 材料中发现的具有本征受拓扑保护的磁斯格明子, 磁斯格明子通过铁电极化的翻转来调控其类型, 从而实现逻辑态的存储. 而本征磁斯格明子通常是由中心反演对称性破缺的结构诱导的 DMI (Dzyaloshinskii-Moriya Interaction) 项产生的, 因此中心反演对称性破缺的体系是探索磁斯格明子的前提条件^[9]. 二维 Janus 结构的上下两面为不同元素组成的原子层, 上下两层原子电负度的差异, 使其在垂直方向具有大的内建电场, 打破空间反演对称性, 在物理和化学性质上呈现出不对称性. 几个到十几个原子层厚度的二维 Janus 材料具有显著的量子限域效应, 作为一种典型的量子效应, 其与材料的组成、结构、化学键和电子结构等因素紧密相关^[10-13], 二维 Janus 材料因其独特的结构在压电传感器、催化剂^[14]、生物医学领域和能源存储等领域显示了广泛的应用前景^[15,16]. 特别是具有大的磁晶各向异性、高磁矩和本征磁拓扑的磁性 Janus 体系, 是发展新一代赛道级磁存储的重要候选者^[2,6,17-20].

传统的实验试错方法在新材料探索中昂贵且耗时, 而基于第一性原理计算的高通量筛选虽降低了实验和研发的周期及成本, 但对于复杂结构的大体系则面临算力大、时间成本高等一系列瓶颈^[21,22]. 近年来, 随着机器学习 (ML) 在凝聚态物理及微电子研究中的兴起^[23,24], 研究人员能够通过模型快速从海量未知体系中筛选出满足所需物理性质的体系, 显著减少了周期及成本, 随后, 通过第一性原理计算

进行理论研究, 验证模型预测的准确性. 机器学习与第一性原理计算的结合^[25-27], 大幅提升了材料研究的速度和精确度, 推动了新材料的快速发现和开发.

当前, 国内外科研工作者已经采用机器学习和第一性原理计算方法对 Janus 材料及其异质结的物理和化学性质开展了研究. Huang 等人利用机器学习结合密度泛函理论方法研究了 Janus 材料活性点位的局部偶极矩对单原子催化剂催化活性的显著影响, 发现活性点位的局部偶极矩是调节催化活性的关键因素, 为设计高性能催化剂提供了新策略^[28]. Chaney 等人利用密度泛函理论和机器学习方法研究了 Janus 异质结 Mo/WXY (X, Y = S, Se, Te) 表面锂的吸附和扩散, 探讨了二维 Janus 过渡金属硫化物的结构不对称性对电子性质的影响, 并对比了 Janus 与常规 TMDs 结构的差异, 从而为锂的吸附和扩散机制提供了机理解释^[29]. Yan 等人通过替换不同过渡金属的二维铁磁 Janus MXenes 材料, 显示其基态铁磁源于磁性原子 d 轨道电子的直接交换, 并基于机器学习的方法筛选出铁磁性的材料, 为自旋电子器件开发提供了新策略^[30]. 尽管上述研究中, 基于机器学习的方法对 Janus 材料性质的挖掘取得了重要进展, 但对于六角晶系 Janus 结构的磁预测仍有待探索. 此外, 如何通过特征工程及超参数优化进一步提高模型的准确性也面临着挑战. 因此, 研究和基于六角晶系 Janus 体系的高效且准确的磁预测模型为二维铁磁家族的扩展提供了平台, 并将为凝聚态物理及微电子等领域带来新的突破, 对自旋电子器件的发展具有重要意义.

本文基于材料的元素组分信息, 训练了四种针对六角晶系 ABC 型 Janus 结构的机器学习模型. 用于预测 Janus 材料的晶格常数、形成能和磁矩分类. 通过模型的高通量筛选, 发现了 4024 种未被探索的高磁矩六角晶系 Janus 材料. 最后, 基于第一性原理的计算, 进一步验证了模型筛选的可靠性, 为未来实验上制备上述高磁矩的 Janus 材料提供了理论依据. 工作流程如图1所示.

* 国家自然科学基金 (批准号: 12204166)、湖南省自然科学基金 (批准号: 2024JJ5132) 和湖南科技大学科研启动基金

(批准号: E51996) 资助的课题

† 通信作者. E-mail: yqzhao@hnu.edu.cn

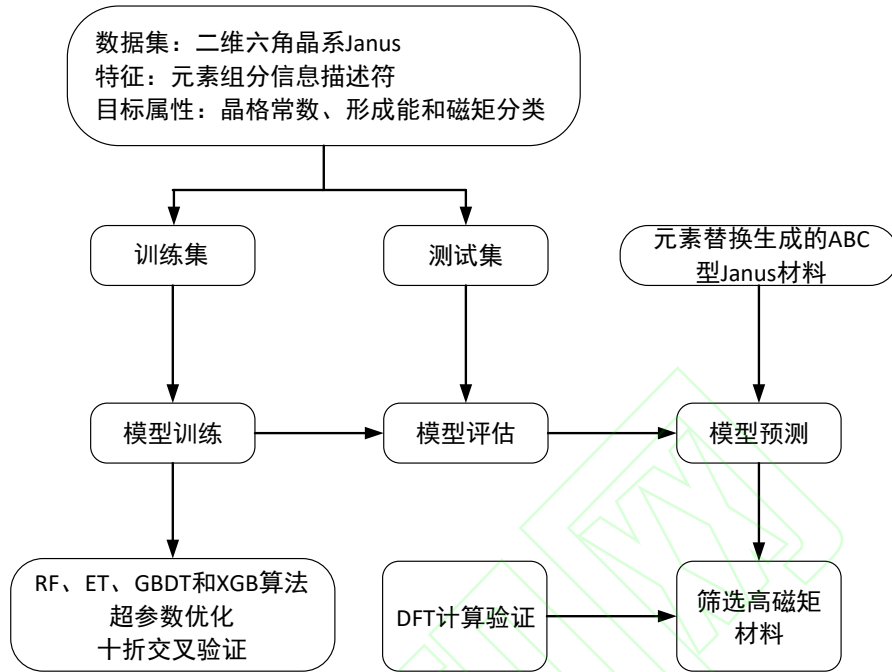


图 1 机器学习结合 DFT 发掘高磁矩 Janus 材料步骤

Fig. 1. Steps for discovering high magnetic moment Janus materials by combining machine learning with DFT.

2 数据及研究方法

2.1 数据获取、分析及预处理

本文从 Materials Project 数据库^[31]中提取 1179 种二维六角晶系 ABC 型 Janus 材料作为数据集, 图2(a) 和 (b) 分别为二维六角晶系 Janus 材

料原子结构的侧视图和俯视图. 中间层紫色原子是 B 原子, 上下两层分别为半径大小不同的绿色 A 和黄色 C 原子层. 由于 A 和 C 原子的电负度不同, 因此沿 c 轴方向会产生内置电场, 打破空间反演对称性. 该数据集包含材料的化学成分信息 (元素、化学配比)、晶体空间群、晶格常数、形成能和总磁矩等.

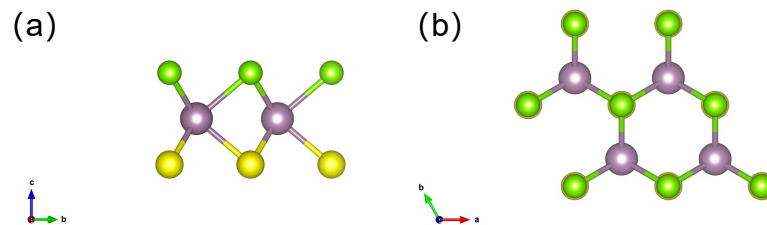


图 2 六角晶系 ABC 型 Janus 材料原子结构的 (a) 侧视图和 (b) 俯视图

Fig. 2. (a) Side view and (b) top view of atomic structures of hexagonal ABC-type Janus materials.

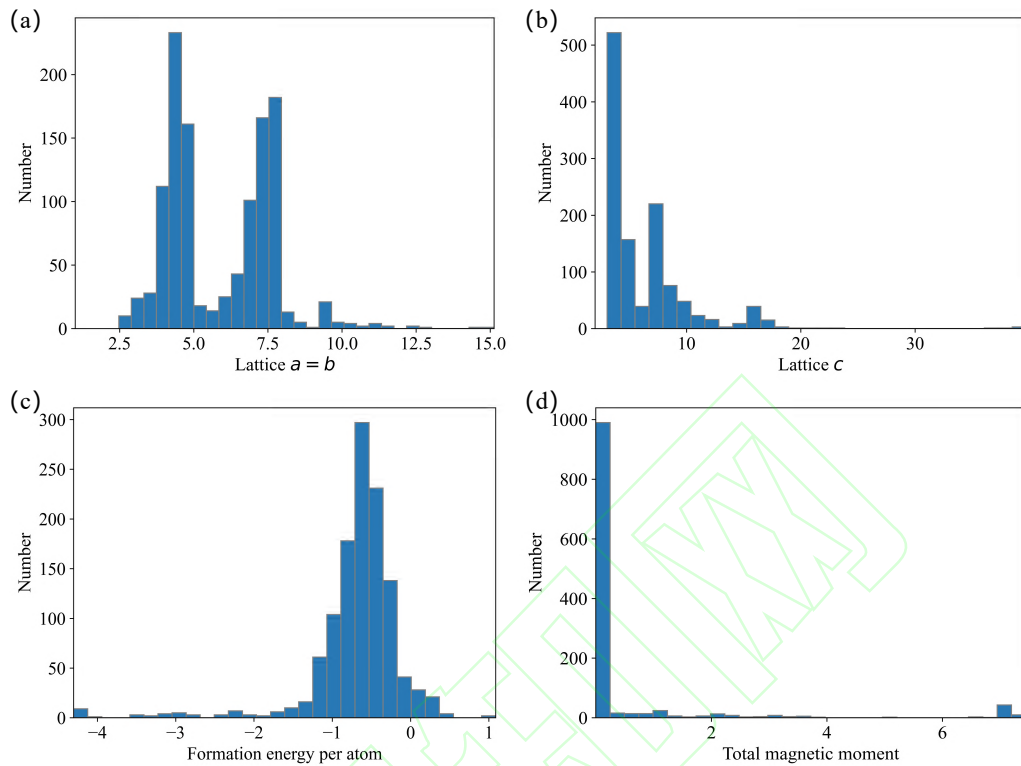


图 3 数据集中二维 Janus 材料的 (a) 晶格常数 a 和 b , (b) 晶格常数 c , (c) 形成能和 (d) 总磁矩的分布

Fig. 3. The distribution of (a) lattice constants a and b , (b) lattice constant c , (c) formation energy and (d) total magnetic moment of the dataset of 2D Janus materials .

图3展示了数据集中二维 Janus 结构数量随晶格常数、形成能及总磁矩变化的分布情况. 图3(a) 显示, 数据集中 Janus 结构的晶格常数 a, b 的分布主要在 4\AA 和 8\AA 附近, 具有两个显著峰值, 表明数据集中面内晶格较为集中. 图3(b) 显示晶格常数 c 主要分布在 $3\text{--}10\text{\AA}$ 的范围内, 在 $3\text{--}7\text{\AA}$ 处具有显著峰值, 而少数二维 Janus 结构的 c 值大于 10\AA , 这可能是模型中考虑了真空层的厚度导致的^[32]. 图3(c) 显示大部分 Janus 结构的形成能低于 0 eV , 表明上述结构是热稳定的, 仅小部分是不稳定的. 图3(d) 显示大部分材料的总磁矩主要集中在 $0\text{--}0.25\mu_B$ 的狭窄区间, 只有少数具有较高的磁矩. 上述特征表明, 数据集中的大部分 Janus 结构是热稳定的, 且具有较低的磁矩. 以上数据集中目标属性的特征对于机器学习算法的选择, 超参数调试具有重要作用.

对数据集进行初步分析后, 接下来执行数据预处理步骤. 首先, 在六角晶系 ABC 型 Janus 材料中, 晶格常数 a (Lattice a) 和 b (Lattice b) 的数值始

终一致, 因此在模型训练中, 将晶格常数 a 和 b 合并为一个变量 (Lattice $a=b$), 以简化预测任务. 此外, 模型训练仅考虑晶格常数 a 和 b 小于 10\AA 的样本, 通过排除这些样本来减少噪声, 防止异常值对模型训练的干扰, 从而提高模型性能. 对于晶格常数 c (Lattice c), 也采取了相同的操作.

其次, 添加磁分类标签, 将目标属性总磁矩分为低磁矩 ($0\mu_B\text{--}0.25\mu_B$) / 中磁矩 ($0.25\mu_B\text{--}3.5\mu_B$) / 高磁矩材料 ($>3.5\mu_B$), 标签分类为 0/1/2, 每一类材料的数量分别为 899、214 及 66 个. 通过将磁矩分为三个组别, 可以更好地捕捉不同磁矩类别之间的差异, 为进一步的材料发现和研究提供更详细的信息和指导.

本研究中, 模型训练的数据集按照 9: 1 的比例, 划分为训练集和测试集. 选择 90% 的数据作为训练集能够为模型提供充足的样本量进行学习, 确保模型充分捕捉数据中的模式和特征, 从而提高其拟合能力和预测性能. 此外, 为了在分类学习

中保持训练集和测试集中各类别样本比例平衡, 采用分层抽样策略. 在模型训练过程中, 使用 SMOTEENN (SMOTE and Edited Nearest Neighbours) 采样步骤^[33]. SMOTEENN 结合了 SMOTE 的过采样和 ENN 的欠采样, 这种方法首先对数据集应用 SMOTE 技术, 对少数类样本进行过采样, 生成新的少数类样本, 然后对过采样后的数据集应用 ENN 技术, 通过删除噪声和边界样本进行欠采样. 本研究提供一种综合方法来同时处理少数类样本不足和噪声样本问题. 这种方法提高了数据的质量和代表性, 使得模型能够更准确地学习和预测, 从而在实际应用中表现出更好的性能.

2.2 特征生成及改进

本文采用 Matminer^[34] 的元素数据和化学计量组成特征器—ElementFraction, 生成了 22 组仅含材料组成信息的元素属性统计描述符 (magpie). 每组元素属性统计描述符包括: 最小值 (min)、最大值 (max)、范围 (range)、平均值 (mean)、平均偏差 (ave_dev)、众数 (mode), 共计得到了包含元素组分信息的 132 个描述符.

对每种 Janus 结构构建一个 3×3 大小的矩阵, 将 magpie 统计的 22 个元素属性分别以主对角线为元素本身的属性值, 非主对角线则是原子 i 和 j 之间元素性质之和 (如基态磁矩和与轨道电子数有关的属性) 或其比值 (其余属性) 的形式构建材料的二维元素属性矩阵特征, 得到 22 组 3×3 型元素属性矩阵描述符^[35]. 随后, 将每组元素属性矩阵按行展平并拼接, 最终获得了 198 种特征描述符.

通过将特征数量从 132 个提高到 198 个, 不仅丰富了材料的描述信息, 还引入了元素间相互作用的特征, 从而提高了模型的学习能力和预测准确性. 这样的方法能够更全面地捕捉材料的复杂性质, 提高模型在实际应用中的表现. 补充材料图 S1([online](#)), 表明了各模型在训练过程中比较重要的特征.

2.3 机器学习模型、训练及超参数优化

本研究采用了四种机器学习算法, 基于 Sklearn 库^[36] 的随机森林 (Random Forest, RF)、梯度提升决策树 (Gradient Boosting Decision Tree, GBDT)、极端随机树 (Extremely Randomized Trees, ET),

以及基于专用的 XGBoost 库^[37] 的极端梯度提升算法 XGBoost (XGB). XGB 是一种高效的梯度提升算法, 具有分块训练、自动处理缺失值、正则化防止过拟合等优化功能. 此外, 上述四种算法能有效处理高维度和复杂特征交互的数据, 具有快速计算和高精度预测的优势. 对于回归任务, 上述四种算法使用均方误差 (Mean Squared Error, MSE) 作为损失函数, 并通过拟合优度 (R^2)、平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Squared Error, RMSE) 来评估其性能. 对于分类任务, RF 和 ET 使用基尼不纯度 (Gini impurity) 作为评估准则, 而 GBDT 和 XGB 则采用对数损失 (Log-loss) 函数作为评估准则. 分类任务的性能评估指标包括准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 分数 (F1 Score). 上述评价指标的引入, 能够系统地比较和分析不同算法在分类和回归任务中的表现, 为模型选择提供了科学依据.

在机器学习的分类与回归任务中, 虽然 RF、GBDT、XGB 和 ET 使用的基本算法结构和超参数类型相同, 但这些超参数的调整重点因任务性质的不同而有所区别. 分类任务侧重于提高算法的泛化能力和准确预测类别, 而回归任务则强调预测值的精确度和误差的控制. 对于上述四种算法, 树的数量 ($n_estimators$)、最大深度 (max_depth) 和最大特征数 ($max_features$) 是共通的关键参数, 影响模型的学习能力和复杂性. 在分类任务中, RF 与 ET 算法的叶节点上所需的最小样本数 ($min_samples_leaf$) 及内部节点再划分所需的最小样本数 ($min_samples_split$) 是特别重要的、而对于 GBDT 和 XGB, 样本子集比例 ($subsample$) 和最小化损失的增益 (γ) 的作用亦如此, 这些参数降低模型对训练数据的噪声敏感性, 减少模型的深度, 提高对数据变化的适应能力, 防止过拟合、欠拟合及提升泛化能力. 另一方面, 对于回归任务, RF 和 ET 算法所需的最大特征数 ($max_features$) 及 XGB 与 GBDT 算法所需的每棵树所采样的特征比例 ($colsample_bytree$) 和学习率 ($learning_rate$) 是特别重要的, 这些超参数的精确调整可以帮助模型找到更好的局部最优解, 加速模型训练, 减少对训练数据的依赖, 从而降低过拟合的可能性. 因此, 虽然共通参数为模型提供了基础结构, 但是特定任务相关的参数调整是优化性能、提高精确度和防止过拟合的关键.

本研究采用贝叶斯超参数搜索技术^[38], 在超参数优化过程中, 每次迭代都基于贝叶斯统计预测各参数组合的潜在性能, 优先评估那些最有可能提升模型表现的参数组合. 这种策略通过持续更新概率模型, 并根据获得的性能数据优化搜索策略, 有效地

识别出最佳参数配置. 通过在十折交叉验证中综合评估各参数组合的表现, 在最终迭代中选出表现最优的超参数配置, 显著提升了模型的预测精度和泛化能力. 不同学习任务中最优算法模型的超参数见表1.

表 1 不同训练任务中机器学习最优模型的超参数

Table 1. The hyperparameters of the optimal machine learning models in various training tasks.

模型	超参数
GBDT(磁矩分类)	learning_rate = 0.01603011, max_depth = 5, n_estimators = 272, subsample = 0.69895067
GBDT(形成能)	learning_rate = 0.02, max_depth = 6, n_estimators = 353, subsample = 0.93030056
ET(晶格常数 a 和 b)	max_depth = 10, max_features = 0.60, n_estimators = 100, min_samples_leaf = 2, min_samples_split = 4
XGB(晶格常数 c)	learning_rate = 0.02, n_estimators = 300, max_depth = 5, subsample = 0.8, colsample_bytree = 0.49613519

2.4 第一性原理计算

本研究采用基于密度泛函理论 (Density functional theory, DFT) 的第一性原理计算方法^[39,40], 通过维也纳从头计算模拟包 (Vienna ab initio simulation package, VASP) 进行结构优化和静态计算. 电子交换关联能由 Perdau-Burke-Ernzerhof(PBE) 泛函描述. 布里渊区中采用 $5 \times 5 \times 5$ k 点网格进行优化和自洽, 平面波截断能量设置为 500 eV. 每个原子上所受力的收敛阈值设置为 $-0.1 \text{ eV}/\text{\AA}$, 能量收敛标准为 $1 \times 10^{-6} \text{ eV}$. 为了描述局域化电子之间的库仑排斥力, 对含有 d 、 f 轨道电子的强关联体系设置了有效场库仑相互作用参数 Hubbard U . 所有的计算中, 原子的初始磁矩设置为零, 通过原子尺度的结构优化和静态计算, 提取二维 Janus 结构中的原子磁矩并与

模型预测结果对比, 验证模型预测可靠性.

3 结果与讨论

3.1 晶格常数

本研究对晶格常数的模型训练分为两类: 一类是针对晶格常数 a 和 b 的训练, 另一类是针对晶格常数 c 的训练. 为了全面评估这两个部分的模型性能, 使用十折交叉验证评估模型. 此方法不仅能够评估模型的稳定性, 还能有效防止过拟合. 表2展示了四种模型分别对上述两类任务在测试集上的平均评估指标.

表 2 晶格常数预测

Table 2. Prediction of lattice constants.

模型	Lattice $a=b$			Lattice c		
	MAE	RMSE	R^2	MAE	RMSE	R^2
RF	0.5485	0.8104	0.7375	0.6491	1.0001	0.6872
GBDT	0.4477	0.7350	0.7829	0.6679	0.9924	0.6923
XGB	0.5427	0.7968	0.7462	0.5953	0.9474	0.7186
ET	0.3469	0.6808	0.8137	0.6534	1.0103	0.6817

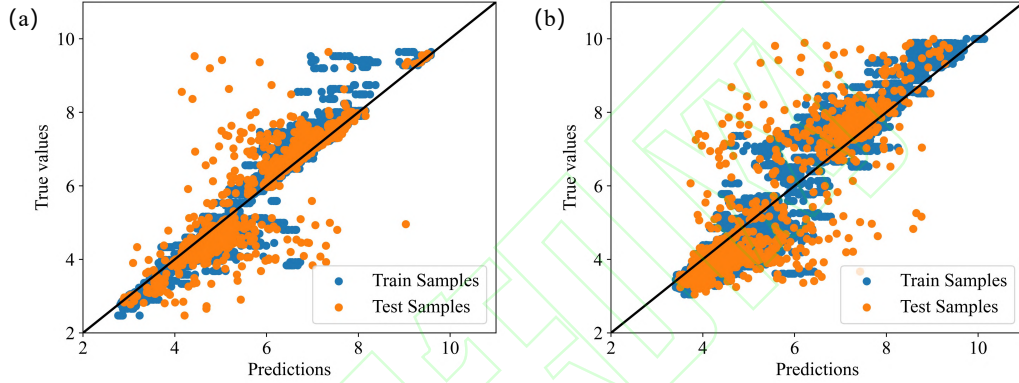


图 4 晶格常数预测：最优模型在十折交叉验证中的散点图. (a) Lattice $a=b$ 预测任务最优模型：极端随机树, (b) Lattice c 预测任务最优模型：极端梯度提升

Fig. 4. Prediction of lattice constants: scatter plots for the optimal models in ten-fold cross-validation. (a) The optimal model for the lattice $a=b$ prediction task:ET, (b) The optimal model for the lattice c prediction task:XGB.

对于晶格常数 a 和 b 的预测,ET 模型表现最佳,其 MAE 为 0.3469, RMSE 为 0.6808 及 R^2 评分达到了 0.8137. 这表明 ET 模型在完成训练晶格常数 a 和 b 的任务时具有较高的准确性和稳定性. 在晶格常数 c 的预测中,XGB 模型表现最佳,其 MAE 为 0.5953, RMSE 为 0.9474 及 R^2 评分为 0.7186. 图4(a),(b) 分别展示了最优模型 ET 和 XGB 在晶格常数 a, b 和 c 的预测任务里, 在十折交叉验证过程中样本集上所有真实值与预测值拟合程度的散点图.

从上述数据可以看出, 晶格常数 c 预测的性能评估低于晶格常数 a 和 b . 这主要是由于 Materials Project 数据库中晶格常数 c 的标准不一致, 并未阐明是否包含真空层厚度. 如果考虑了真空层厚度, 其晶格会比体相的实际晶格 c 大很多, 而往往在具有大的晶格常数 c 的 Janus 体系中, 真空层时可能被包含在其中. 这就导致数据集中的晶格 c 比实际晶

格要大得多. 特别是在高晶格常数 c 的材料中, 大多增加了真空层, 导致 c 值显著偏大. 这种处理方式增加了数据分布的极端性和复杂性, 使得数据的统计特性存在显著差异和额外的变异性, 进一步加剧了数据的复杂性, 从而增加了模型的学习难度, 导致晶格常数 c 的预测效果略低于晶格常数 a 和 b .

3.2 形成能

本文分别采用了 RF, GBDT, XGB 和 ET 四种算法训练二维 Janus 材料形成能的预测模型, 并对模型性能进行了评估. 表3展示了十折交叉验证过程中各模型在测试集上的平均评估指标 MAE、RMSE 及 R^2 . 图5(a)、(b)、(c) 和 (d) 展示各模型在十折交叉验证过程中样本集上所有真实值与预测值拟合程度的散点图.

表 3 形成能预测：四种机器学习模型的评价指标

Table 3. The Prediction of formation energy: evaluation metrics of four machine learning models.

模型	MAE	RMSE	R^2
RF	0.1054	0.1697	0.8671
GBDT	0.0798	0.1411	0.9070
XGB	0.0959	0.1533	0.8930
ET	0.1120	0.1701	0.8657

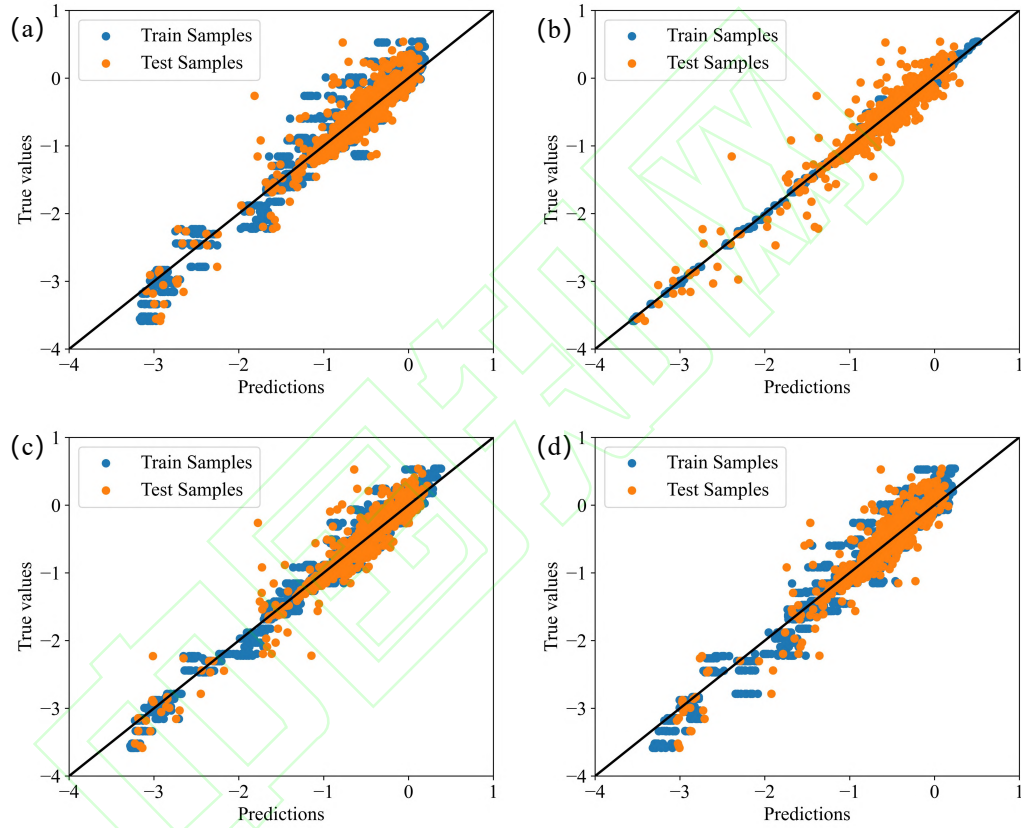


图 5 形成能预测：四种模型在十折交叉验证上的散点图. (a) 随机森林, (b) 梯度提升决策树, (c) 极端梯度提升 (d) 极端随机树

Fig. 5. Prediction of formation energy : scatter plots for four models in ten-fold cross-validation. (a) RF (b) GBDT, (c) XGB, (d) ET.

评价指标结果表明,GBDT 模型表现最为优异, 中样本集上所有真实值与预测值之间良好的拟合程度, 具有最低的 MAE 和 RMSE, 以及最高的 R^2 评分, 大多数样本点都分布在对角线上. 综上所述, 基于 GBDT 训练的形成能模型, 在各项性能评估指标上表现更优异, 能有效建立二维 Janus 材料的元素组分信息与形成能之间的映射关系, 保证了预测结果的准确性. 散点图直观地展示了 GBDT 模型在十折交叉验证过程

3.3 磁矩分类

为了评估磁矩分类模型的性能，对四种模型进行了十折交叉验证，并使用宏平均 (macro-averaging) 方式对每个类别单独计算评价指标，然后对所有类别的结果取平均，以确保每个类别的表

现都得到公平的评估。这种方法在类别不平衡的情况下，能够对模型性能提供更加全面和真实的评价。其在十折交叉验证过程中各模型在测试集上的平均评估指标如表4所示。图6(a)、(b)、(c) 和 (d) 直观展示了各模型在十折交叉验证过程中所有预测值和真实值的混淆矩阵图。

表 4 磁矩分类预测：四种机器学习模型的评价指标

Table 4. Prediction of magnetic moment classification.: evaluation metrics of four machine learning models.

模型	Accuracy	Precision	Recall	F1 score
RF	0.8770	0.8459	0.7636	0.7862
GBDT	0.8948	0.8498	0.8182	0.8263
XGB	0.8762	0.8398	0.7697	0.7883
ET	0.8795	0.8392	0.7778	0.7965

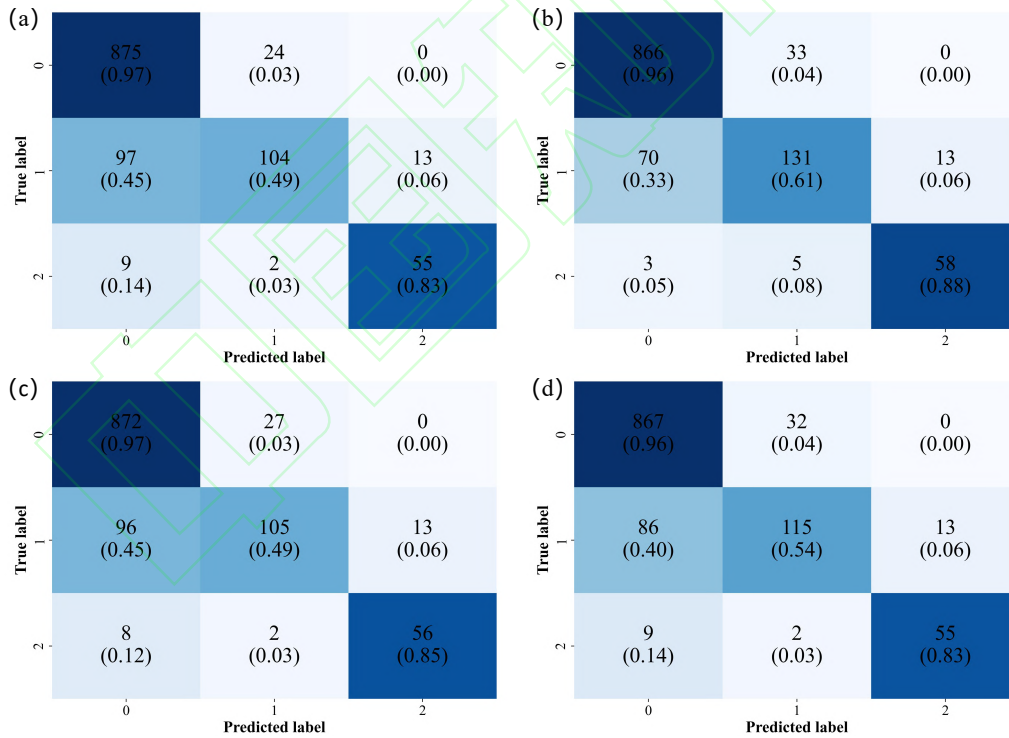


图 6 磁矩分类预测：四种模型在十折交叉验证上的混淆矩阵。(a) 随机森林, (b) 梯度提升决策树, (c) 极端梯度提升 (d) 极端随机树

Fig. 6. Prediction of magnetic moment classification: confusion matrices for four models in ten-fold cross-validation.

(a) RF, (b) GBDT, (c) XGB, (d) ET.

上述评价指标结果显示,GBDT 模型在准确率、精确率、召回率和 F1 分数上均展现出优越性,表明其能更有效地识别和分类复杂数据集中的类别. 相比之下,RF 和 ET 虽然在处理高维数据时稳健,但在某些类别的识别上存在误判,这可能影响其在实际应用中未知样本的预测效果.XGB 虽然处理速度快,适合大规模数据,但其对异常数据的敏感性可能在未探测的数据集上造成性能波动. 因此,GBDT 由于其较强的泛化能力和稳定的性能表现,成为寻找高磁矩 Janus 材料最具潜力模型.

3.4 高通量材料生成及第一性原理计算

以数据集中二维六角晶系的 ABC 型 Janus 结构为基础,在元素周期表中对 A、B 和 C 包含的的所

有元素随机替换,生成了 82018 种尚未发现的 ABC 型二维 Janus 材料,上述结构不属于原数据集.

将上述各部分评价性能最好的模型用于预测 82018 种二维六角晶系 Janus 材料的晶格常数,形成能和磁矩. 模型共预测出 4204 种包含晶格信息并具有热稳定和高磁矩的结构. 为了进一步验证其可靠性,本文从高通量筛选结果中随机挑选了 13 种涵盖了不同元素的二维 Janus 结构并结合第一性原理的研究,计算其晶格常数,形成能和磁矩. 图7分别为 ErFeTb、FeNO、HoRuSr、DyOsSr、EuSbSr、HoIrSr、LiUZn、PuSZn、GdKU、LuNbTi、GdHfSe、NaTbZn 及 HoNpSr 共 13 种二维 Janus 原子结构的侧视图. 优化后的晶格常数、形成能和磁矩如表5所示. 补充材料表 S1([online](#)) 给出了 13 种未优化结构的形成能和磁矩.

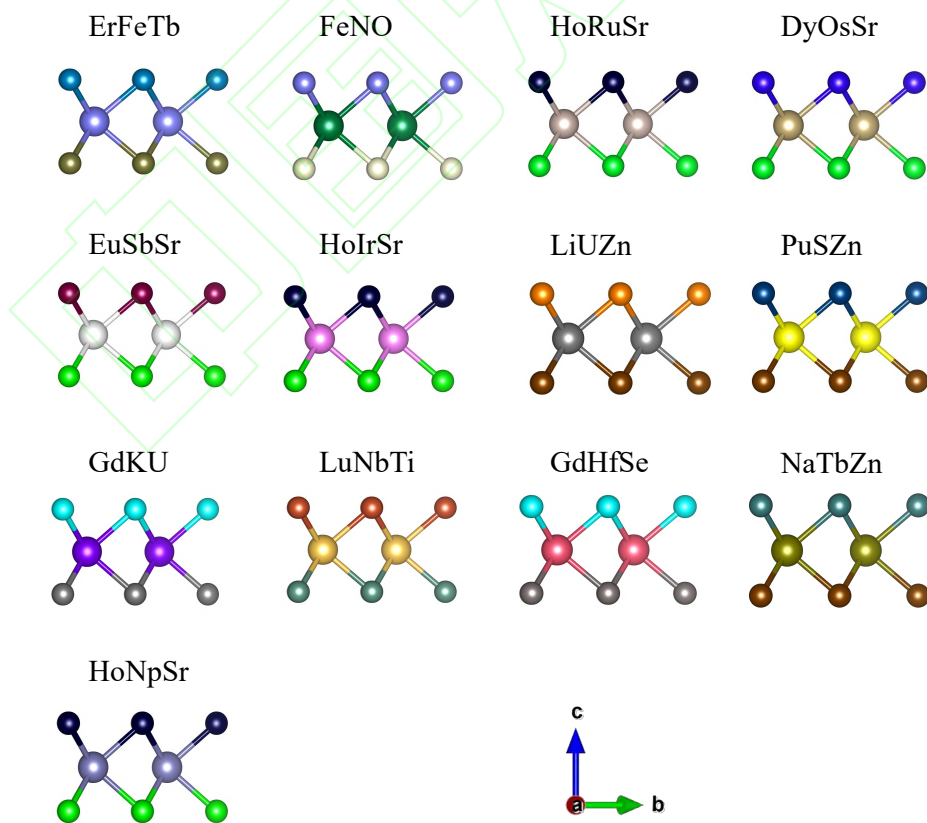


图 7 13 种二维六角晶系 Janus 原子结构的侧视图

Fig. 7. Side view of atomic structures of 13 two-dimensional hexagonal Janus materials.

表 5 13 种结构优化后的六角晶系 ABC 型 Janus 材料的晶格常数, 形成能和磁矩

Table 5. Optimized lattice constants, formation energies and magnetic moments of 13 two-dimensional hexagonal ABC-type Janus materials.

Formula	Lattice constants		Formation energy (eV)	$ \mu $ (μ_B)		
	$a = b(\text{\AA})$	$c(\text{\AA})$		A	B	C
ErFeTb	3.35	18.25	-2.02	2.51	3.03	6.24
FeNO	2.92	15.00	-11.87	1.17	0.08	0.47
HoRuSr	4.90	18.79	-6.66	3.79	0.02	0.05
DyOsSr	4.18	18.87	-6.89	4.89	0.00	0.13
EuSbSr	5.43	18.69	-5.53	6.85	0.01	0.05
HoIrSr	4.58	18.79	-7.24	3.72	0.00	0.05
LiUZn	2.89	18.13	-0.44	0.00	1.65	0.01
PuSZn	4.52	18.13	-6.75	5.61	0.10	0.01
GdKU	7.46	18.13	-2.39	7.33	0.00	2.96
LuNbTi	3.02	18.13	-1.76	0.02	0.28	1.67
GdHfSe	5.03	18.93	-8.46	7.33	0.34	0.02
NaTbZn	4.65	18.69	-1.87	0.02	6.00	0.00
HoNpSr	3.69	18.46	-1.80	3.81	4.38	0.08

根据上述结果, 优化后的 13 种二维六角晶系 Janus 结构的形成能均为负值, 表明上述 Janus 体系是热力学稳定的. 此外, 基于第一性原理的研究表明, 所有 13 种 Janus 结构均具有磁性, 理论计算与机器学习模型的预测值一致. 其中 ErFeTb、HoRuSr、DyOsSr、EuSbSr、HoIrSr、PuSZn、GdKU、GdHfSe、NaTbZn、HoNpSr 的单原子磁矩超过 $3 \mu_B$, 如表 5 所示. 上述具有高磁矩的原子主要属于镧系金属 (如 Pu、Np) 和稀土元素 (如 Tb、Gd、Dy). 上述元素因为未填满的 f 轨道电子的净磁矩贡献诱导了磁性. 这表明, 含有镧系金属和稀土元素的化合物是设计高磁矩六角晶系 ABC 型 Janus 材料的有力候选. (补充材料 S1 为基于机器学习模型预测的 13 种未优化的 Janus 结构, 结合第一性原理计算的形成能和磁矩.)

机器学习模型 (RF、GBDT、XGB、ET) 采用了贝叶斯超参数优化和十折交叉验证方法进行训练, 并得到了在各自任务中表现最佳的模型 (晶格常数 $a=b$ —ET、晶格常数 c —XGB、形成能—GBDT 及磁矩分类—GBDT). 通过对数据集中六角晶系结构进行元素替换, 生成了 82018 种尚未发现的六角晶系 ABC 型 Janus 材料. 应用上述训练的机器学习模型, 筛选出了 4024 种结构稳定且具有高磁矩的六角晶系 Janus 材料. 鉴于算力的限制, 随机选择了 13 种结构结合第一性原理进行计算验证, 得到了 10 种优化后热力学稳定且单原子磁矩超过 $3 \mu_B$ 的 Janus 材料, 研究表明, 含有稀土和镧系金属元素的化合物是设计高磁矩 Janus 材料的有力候选, 进一步验证了机器学习模型预测二维六角晶系 Janus 材料磁性的可靠性.

4 结论

本研究基于材料的元素组分信息, 构建了包含 22 组元素属性的矩阵特征, 每组特征通过行展平, 生成了 198 种特征描述符. 针对六角晶系 ABC 型 Janus 材料的晶格常数、形成能和磁矩分类, 对四种

参考文献

- [1] Novoselov K S, Geim A K, Morozov S V, Jiang D, Zhang Y, Dubonos S V, Grigorieva I V, Firsov A A 2004 *Science* **306** 666

- [2] Zhang Z W, Lang Y F, Zhu H P, Li B, Zhao Y Q, Wei B, Zhou W X 2024 *Phys. Rev. Appl.* **21** 064012
- [3] Liu B, Feng X X, Long M Q, Cai M Q, Yang J L 2022 *Phys. Rev. Appl.* **18** 054036
- [4] Xiong X J, Zhong F, Zhang Z W, Chen F, Luo J L, Zhao Y Q, Zhu H P, Jiang S L 2024 *Acta Phys. Sin.* **73** 137101
- [5] Zhao Y Q, Liu Z S, Nie G Z, Zhu Z H, Chai Y F, Wang J N, Cai M Q, Jiang S L 2021 *Appl. Phys. Lett.* **118** 173104
- [6] Lang Y F, Zou D F, Xu Y, Jiang S L, Zhao Y Q, Ang Y S 2024 *Appl. Phys. Lett.* **124** 052903
- [7] Liao C S, Ding Y F, Zhao Y Q, Cai M Q 2021 *Appl. Phys. Lett.* **119** 182903
- [8] Tan W, Zhang Z W, Zhou X Y, Yu Z L, Zhao Y Q, Jiang S L, Ang Y S 2024 *Phys. Rev. Mater.* **8** 094414
- [9] Liang J H, Wang W W, Du H F, Hallal A, Garcia K, Chshiev M, Fert A, Yang H X 2020 *Phys. Rev. B* **101** 184401
- [10] Zhang S Q, Xu R Z, Luo N N, Zou X L 2021 *Nanoscale* **13** 1398
- [11] Dai C Y, He P, Luo L X, Zhan P X, Guan B, Zheng J 2023 *Sci. China Mater.* **66** 859
- [12] Wang P, Zong Y X, Wen H Y, Xia J B, Wei Z M 2021 *Acta Phys. Sin.* **70** 026801
- [13] Ren K, Wang K, Zhang G 2022 *ACS Appl. Electron. Mater.* **4** 4507
- [14] Peng Z L, Huang J X, Guo Z G 2021 *Nanoscale* **13** 18839
- [15] Zhang L, Yang Z J F, Gong T, Pan R K, Wang H D, Guo Z N, Zhang H, Fu X 2020 *J. Mater. Chem. A* **8** 8813
- [16] Vafaezadeh M, Thiel W R 2022 *Angew. Chem. Int. Edit.* **61** e202206403
- [17] Mukherjee T, Kar S, Ray S 2022 *J. Mater. Res.* **37** 3418
- [18] Li C Q, An Y K 2022 *Phys. Rev. B* **106** 115417
- [19] Zhang L, Zhao Y, Liu Y Q, Gao G Y 2023 *Nanoscale* **15** 18910
- [20] Xu L J, Wan W H, Peng Y R, Ge Y F, Liu Y 2024 *Ann. Phys.* **536** 2300388
- [21] Gao Z Y, Mao G Y, Chen S Y, Bai Y, Gao P, Wu C C, Gates I D, Yang W J, Ding X L, Yao J X 2022 *Phys. Chem. Chem. Phys.* **24** 3460
- [22] Liu H, Sun J T, Liu M, Meng S 2018 *J. Phys. Chem. Lett.* **9** 6709
- [23] Nelson J, Sanvito S 2019 *Phys. Rev. Mater.* **3** 104405
- [24] Belot J F, Taufour V, Sanvito S, Hart G L 2023 *Appl. Phys. Lett.* **123** 042405
- [25] Miyazato I, Tanaka Y, Takahashi K 2018 *J. Phys.: Condens. Matter* **30** 06LT01
- [26] Lu S H, Zhou Q H, Guo Y L, Zhang Y H, Wu Y L, Wang J L 2020 *Adv. Mater.* **32** 2002658
- [27] Ma X Y, Lyu H Y, Hao K R, Zhao Y M, Qian X F, Yan Q B, Su G 2021 *Sci. Bull.* **66** 233
- [28] Huang T, Yang Z X, Li L, Wan H, Leng C, Huang G F, Hu W Y, Huang W Q 2024 *J. Phys. chem. Lett.* **15** 2428
- [29] Chaney G, Ibrahim A, Ersan F, Çakır D, Ataca C 2021 *ACS Appl. Mater. Interfaces* **13** 36388
- [30] Yan X H, Zheng J M, Zhao X, Zhao P J, Guo P, Jiang Z Y 2024 *Phys. Status Solidi Rapid Res. Lett.* **18** 2300468
- [31] Jain A, Ong S P, Hautier G, Chen W, Richards W D, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson K A 2013 *APL Mater.* **1** 011002
- [32] Chen P Y, Lam C H, Edmondson B, Posadas A B, Demkov A A, Ekerdt J G 2019 *J. Vac. Sci. Technol. A* **37** 050902

- [33] Khushi M, Shaukat K, Alam T M, Hameed I A, Uddin S, Luo S, Yang X, Reyes M C 2021 *IEEE Access* **9** 109960
- [34] Ward L, Dunn A, Faghaninia A, Zimmermann N E, Bajaj S, Wang Q, Montoya J, Chen J, Bystrom K, Dylla M, Chard K, Asta M, Persson K A, Snyder G J, Foster I, Jain A 2018 *Comp. Mater. Sci.* **152** 60
- [35] Chen J, Song Y Y, Li S Z, Que Z X, Zhang W B 2023 *Sci. China Technol. Sci.* **1** 011002
- [36] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E 2011 *J. Mach. Learn. Res.* **12** 2825
- [37] Ester M, Kriegel H P, Xu X 2023 *Geogr. Anal.* **55** 207
- [38] Wu J, Chen X Y, Zhang H, Xiong L D, Lei H, Deng S H 2019 *J. Electron. Sci. Technol.* **17** 26
- [39] Ma Q Y, Wan W H, Ge Y F, Li Y M, Liu Y 2022 *J. Magn. Magn. Mater.* **605** 172314
- [40] Yin W J, Tan H J, Ding P J, Wen B, Li X B, Teobaldi G, Liu L M 2021 *Mater. Adv.* **2** 7543

Prediction of Magnetic Janus Materials Based on Machine Learning and First-Principles Calculations*

Zhang Qiao¹⁾ Tan Wei¹⁾ Ning Yong-Qi¹⁾ Nie Guo-Zheng¹⁾ Cai Meng-Qiu²⁾ Wang Jun-Nian¹⁾ Zhu Hui-Ping³⁾ Zhao Yu-Qing^{1)4)†}

- 1) (*School of Physics and Electronics, Hunan University of Science and Technology, Hunan Provincial Key Laboratory of Intelligent Sensors and New Sensor Materials, Xiangtan 411201, China*)
- 2) (*School of Physics and Microelectronics, Hunan University, Changsha 410082, China.*)
- 3) (*Institute of Microelectronics, Chinese Academy of Sciences, Key Laboratory of Silicon Device Center, Beijing 100029, China*)
- 4) (*Institute of Semiconductors, Chinese Academy of Sciences, State Key Laboratory of Superlattices, Beijing 100084, China*)

Abstract

Discovering the compact, stable and easily controllable nanoscale non-trivial topological magnetic structures—magnetic skyrmions, is the key to develop next-generation high-density, high-speed, and low-energy non-volatile information storage devices. Based on the topological generation mechanism, magnetic skyrmions could be generated through the Dzyaloshinskii–Moriya Interaction (DMI) induced by space-reversal symmetry broken. Two dimensional (2D) non-centrosymmetric Janus could generate vertical built-in electric fields to break spatial inversion symmetry. Therefore, seeking 2D Janus with intrinsic magnetism is fundamental to develop the novel chiral magnetic storage technologies. In this work, we combined detailed machine learning techniques and first-principles calculations to discover the magnetism of the unexplored 2D Janus. We first collected 1179 2D hexagonal ABC-type Janus based on the Materials Project database, and used elemental composition as feature descriptors to construct four machine learning models: Random Forest (RF), Gradient Boosting Decision Trees (GBDT), Extreme Gradient Boosting (XGB), and Extra Trees (ET). These algorithms and models were constructed to predict lattice constants, formation energies, and magnetic moment, via hyperparameter optimization and ten-fold cross-validation. GBDT exhibits the highest accuracy and best prediction performance for magnetic moment classification. Subsequently, the collected data of 82,018 yet-undiscovered 2D Janus were input into the trained models to generate 4,024 high magnetic moment 2D Janus with thermal stability. First-principles calculations were employed to validate random sample of 13 Janus with high magnetic moment. This study provides an effective machine learning framework for magnetic moment classification and high-throughput screening of 2D Janus, accelerating the exploration of magnetic properties in 2D Janus structures.

Keywords: machine learning, two-dimensional Janus materials, magnetic moment, first-principles calculations

PACS: 02.60.Cb , 81.05.Zx , 75.20.-g, 71.15.Mb

* Project supported by the National Natural Science Foundation of China (Grant No. 12204166), the Natural Science Foundation of Hunan Province (Grant No. 2024JJ5132), and the Scientific Research Start-up Fund of Hunan University of Science and Technology (Grant No. E51996).

† Corresponding author. E-mail: yqzhao@hnu.edu.cn