

## 综述

基于数据挖掘技术的稀土磁性材料研究进展<sup>\*</sup>刘丹<sup>1)</sup><sup>†</sup> 李渊<sup>1)</sup> 孙若瑄<sup>1)</sup> 漆星<sup>1)</sup> 沈保根<sup>2)</sup><sup>‡</sup>

1) (北京工商大学计算机与人工智能学院, 物理系, 北京 100048)

2) (中国科学院物理研究所, 磁学国家重点实验室, 北京 100190)

(2025年4月3日收到; 2025年4月15日收到修改稿)

稀土元素的原子结构特殊, 具有内层未成对4f轨道电子多、原子磁矩高、自旋轨道耦合作用强的性质, 故其电子能级极为丰富, 易形成多种价态、多种配位的化合物, 通常表现出特殊的磁学性质和丰富的磁畴结构, 成为高新技术产业发展的关键材料。这类材料中复杂的磁结构形式、多样的磁耦合类型及多种直接或间接的磁交换作用, 为开发新型功能器件提供便利的同时, 也对基础研究提出了严峻挑战。随着数据挖掘技术的快速发展, 大数据和人工智能的出现给研究人员提供了一个新的选择, 可以高效地分析大量实验和计算数据, 从而加速稀土磁性材料的研究与开发。本文围绕稀土永磁材料、稀土磁致冷材料、稀土磁致伸缩材料等, 详细阐述了数据挖掘技术在其性能预测、成分与工艺优化、微观结构分析等方面的应用进展, 深入探讨了当前面临的挑战, 并对未来发展趋势进行展望, 为推动数据挖掘技术与稀土磁性材料研究的深度融合提供理论基础。

**关键词:** 数据挖掘, 稀土磁性材料, 机器学习, 性能预测

**PACS:** 71.20.Eh, 75.50.-y, 88.30.gg

**DOI:** [10.7498/aps.74.20250431](https://doi.org/10.7498/aps.74.20250431)

**CSTR:** [32037.14.aps.74.20250431](https://cstr.cn/32037.14.aps.74.20250431)

## 1 引言

随着新一代信息技术、轨道交通、智能装备、新能源等新兴领域的加速迭代, 稀土新材料及其应用需求变得更加凸显。国务院在《国民经济和社会发展“十四五”规划和2035年远景目标纲要》以及《稀土管理条例》中强调要重点培育稀土行业的新动能, 相关研究不仅涉及基础前沿科学问题, 更是我国科技发展的重大战略需求<sup>[1,2]</sup>。稀土元素的原子结构特殊, 具有内层未成对4f轨道电子多、原子磁矩高、自旋轨道耦合作用强的性质, 故其电子能级极为丰富, 易形成多种价态、多种配位的化合物, 通常表现出特殊的磁学性质和丰富的磁畴结构, 成为高新技术产业发展的关键材料。因此, 实

现新型稀土磁性材料的科学设计和高技术开发, 全面释放稀土新材料的科技动能, 为稀土资源平衡利用和节能降碳提供新质生产力, 具有重要的科学价值和战略意义。

稀土磁性材料作为稀土领域发展最快、应用量最大的功能材料, 其化学组分和结构复杂, 大离子半径稀土元素的掺杂会引起晶格常数的变化和晶胞畸变, 加之电子、自旋、轨道、晶格等多个自由度的耦合, 材料磁性能很容易发生改变并受到外场调控<sup>[3-5]</sup>。稀土磁性材料根据功能大致可以分为稀土永磁材料、稀土磁致冷材料、稀土磁致伸缩材料、稀土软磁材料和稀土磁光材料等。其中, 稀土永磁材料(如钕铁硼、钐钴)具有高磁能积和矫顽力, 广泛应用于电动汽车、风力发电和电子设备; 稀土磁致冷材料(如钆硅锗合金)在磁制冷技术中展现出

\* 国家重点研发计划(批准号: 2023YFA1610400)和国家自然科学基金(批准号: 52088101, 52371169)资助的课题。

† 通信作者。E-mail: [liudan@btbu.edu.cn](mailto:liudan@btbu.edu.cn)

‡ 通信作者。E-mail: [shenbg@iphy.ac.cn](mailto:shenbg@iphy.ac.cn)

高效节能的潜力; 稀土磁致伸缩材料(如铽镝铁合金)则在传感器和换能器领域具有重要应用。然而, 高新科技的迅猛发展伴随着人们对稀土磁性材料性能的要求也日益提高, 其基础研究与技术开发面临成分设计复杂、工艺参数优化困难等挑战, 传统分析方法在处理海量复杂数据时也愈发力不从心。数据挖掘技术作为强大的数据分析工具, 能够从大规模、不完全、有噪声的原始数据中, 通过机器学习、数据库、统计等方法, 提取隐含且具有潜在价值的信息。其中, 机器学习是数据挖掘的重要技术支撑, 其核心是“泛化能力”, 即从训练数据推广到未知数据的能力。数据挖掘是目标导向的应用体系, 而机器学习是实现数据驱动知识发现的核心技术手段, 可以为稀土磁性材料的研究提供新的途径和方法, 加速了材料设计、性能预测和新材料发现的进程。

在磁性材料研究里, 常用的数据挖掘技术十分多元且各有优势, 其从不同角度助力研究工作<sup>[6-11]</sup>, 如表1所示。1) 对于离散型磁性能参数的预测, 比如晶粒取向、相态分类等, 可以优先选用支持向量机(SVM)或决策树(DT)算法。SVM在小样本、非线性及高维模式识别中表现出色。DT在大规模数据中分类速度快, 可解释性强。2) 对于连续型磁性能参数的预测, 比如磁能积、矫顽力等, 可以优先选用随机森林(RF)或人工神经网络(ANN)算法。人工神经网络(ANN), 如多层感知器(MLP)、径向基函数(RBF)神经网络等技术可以模仿生物神经网络结构和功能, 具有强大的非线性映射能力。在磁性材料研究中, 可以处理复杂的、高度非线性的数据关系, 通过大量数据训练学习, 精准预

测不同条件下材料的磁性能, 大大提高预测准确性, 减少实验次数。3) 对于图像型磁性能参数的预测, 比如透射电子显微镜图、磁力显微镜图等, 可以优先选用卷积神经网络(CNN)或主成分分析(PCA)算法。CNN能够通过多层卷积核自动提取晶粒边界、取向对比度等图像特征, 比如对稀土永磁材料扫描电子显微镜(SEM)图像进行自动识别, 精准分辨主相、晶界相及其他第2相, 并定量计算各相比例与尺寸分布, 以帮助科研人员深入了解微观结构与磁性能的关联。PCA可用于数据降维, 消除数据间的多重共线性, 与多元线性回归(MLR)结合建立偏最小二乘回归(PLSR), 揭示微观结构对磁性能的影响机制。4) 对于多目标优化问题, 比如成分和工艺协同优化, 可以优先选用遗传算法(GA)叠加支持向量回归(GA+SVR)或深度强化学习。GA通过模拟生物进化过程中的遗传、变异和自然选择等机制, 从众多可能的组合中快速找到具有良好综合磁性能的方案, 加速新型磁性材料的研发进程。此外, K均值聚类算法是将数据分成K个簇, 使同一簇内数据相似度高, 不同簇间差异大。比如在对磁性材料微观结构的特征参数进行聚类时, 其可以挖掘出不同制备工艺下微观结构的聚类特征, 为工艺调控提供关键微观依据。

## 2 稀土基磁性材料的高通量数据分析

### 2.1 数据挖掘技术在稀土永磁材料中的应用

稀土永磁材料, 如Nd-Fe-B, Sm-Co等作为高性能磁性材料的代表, 广泛应用于新能源汽车、

表1 常用的机器学习算法的优势与缺陷  
Table 1. Advantages and disadvantages of commonly used machine learning algorithms.

模型	优势	不足
MLR	简单易懂, 计算效率高; 可解释性强, 参数直接反映特征重要性	无法捕捉非线性关系; 对异常值敏感
SVM	擅长处理高维空间中的非线性分类; 小样本数据表现优异	核函数选择依赖经验; 计算成本高, 扩展性差
DT	无需数据标准化; 可视化清晰; 擅长处理分类问题	易过拟合, 需剪枝优化; 对连续特征处理能力有限
RF	抗过拟合能力强, 适合高维数据; 可输出特征重要性, 便于特征选择	计算复杂度较高; 对小样本数据表现不佳
K-means	无需标注数据, 自动发现数据模式; 计算效率高	需要预先指定聚类数
GA	全局搜索能力, 避免局部最优; 并行处理, 适合高维问题	计算成本高; 复杂问题的收敛速度慢; 依赖高质量的训练数据
粒子群优化算法(PSO)	适用于目标函数不连续的场景; 参数少, 易于与实验工具集成	收敛速度慢; 参数敏感性; 局部搜索能力弱
PCA	有效降维, 保留关键信息; 可视化高维数据的潜在结构	无法捕捉非线性关系; 忽略特征间的相关性
ANN	强大的非线性建模能力; 适用于复杂模式识别, 如图像、分子结构	需要大量数据和计算资源; 可解释性差, 存在“黑箱”问题
CNN	无需人工设计特征; 擅长高维数据处理; 鲁棒性强	计算资源需求大; 数据依赖性; 解释性差

风力发电、轨道交通等战略领域,是实现低碳经济转型的关键基础材料。然而,传统研发模式依赖试错法,面临资源利用率低、高温稳定性不足、成分设计周期长等瓶颈。数据挖掘技术通过整合实验数据、多尺度模拟与人工智能算法,为稀土永磁材料的研发带来显著变革,推动稀土产业向低碳、高端模式转型。

### 2.1.1 成分设计与优化

材料成分的优化设计一直是稀土永磁领域研究的核心目标之一。机器学习(ML)技术的应用能够加速材料研发进程,显著降低新材料研发成本,从而推动高性能低成本稀土永磁材料的发展。2018年,李锐<sup>[12]</sup>将数据驱动技术应用到Nd-Fe-B稀土永磁体的成分设计当中,利用机器学习方法成功破

解了含有高轻稀土元素La,Ce快淬薄带的矫顽力和最大磁能积与材料成分之间的非线性关系。如图1所示,在所使用的线性回归(LR)、回归决策树(DTR)、支持向量回归(SVM)和梯度提升回归树(GBRT)4种建模方法中,以GBRT的预测准确性最高,其矫顽力和最大磁能积的拟合优度分别为0.895和0.905,相应的预测误差分别为82.8 kA/m和9 kJ/m<sup>3</sup>。基于此算法,以稀土总含量(REs)和Ce相对含量( $p$ )为特征变量,构建 $(\text{PrNd}_{1-p}\text{Ce}_p)_{\text{REs}}\text{Fe}_{93-\text{REs}}\text{TM}_1\text{B}_6$ (TM为Zr,Nb,Ga或Ti元素)矫顽力 $H_{\text{cj}}$ 和最大磁能积 $(BH)_{\text{max}}$ 的虚拟数据集。同时结合材料价格对其“成分-性能-价格”关系进行预测分析,快速定位特定性能更高性价比的材料组成空间,从而实现经济型快淬磁体的成分设计。比如,以矫顽力875 kA/m的含Ce粘接磁体为设计对

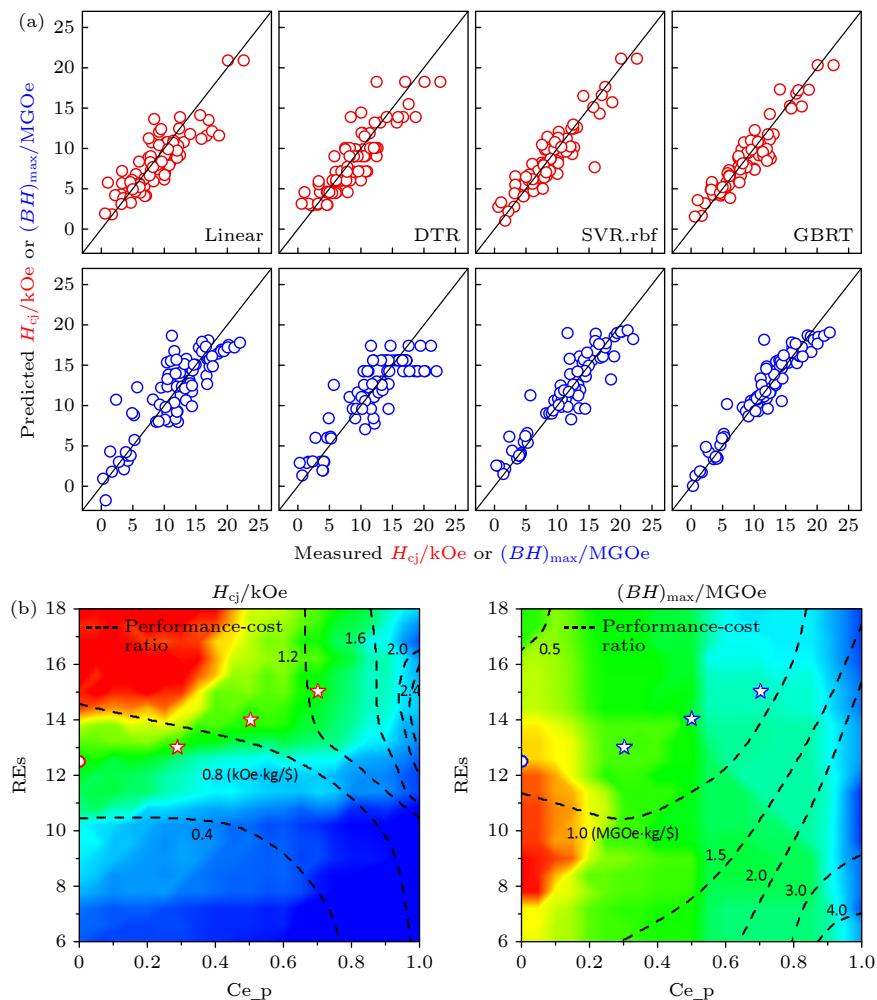


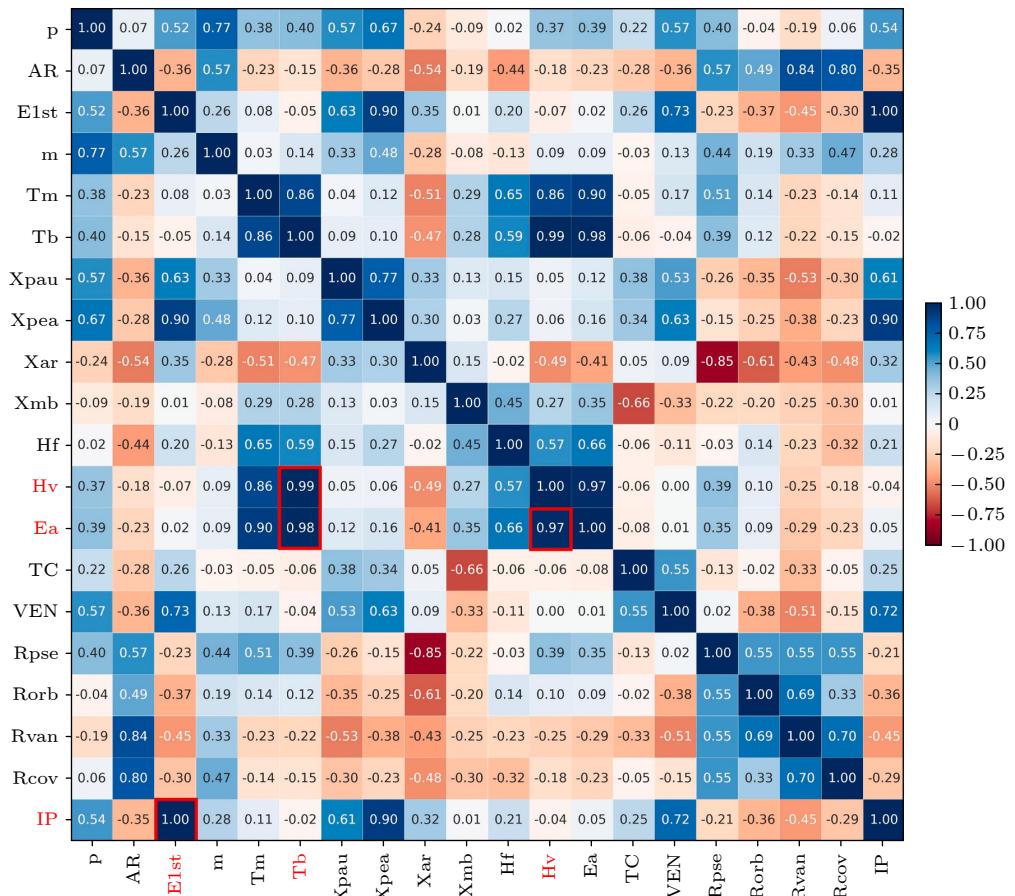
图1 (a) 不同机器学习算法的训练集预测结果对比;(b) 钕铁硼快淬带矫顽力  $H_{\text{cj}}$  和最大磁能积  $(BH)_{\text{max}}$  的预测图及其性价比等高线<sup>[12]</sup>

Fig. 1. (a) Predicted scatter plots of different machine learning algorithms; (b) predicted maps of  $H_{\text{cj}}$  and  $(BH)_{\text{max}}$  of Nd-Fe-B melt-spun magnets, and their corresponding cost performances<sup>[12]</sup>.

象, 利用 GBRT 计算出满足性价比要求的磁体成分  $(Nd_{0.7}Ce_{0.3})_{13}Fe_{80}TM_1B_6$ ,  $(Nd_{0.5}Ce_{0.5})_{14}Fe_{79}TM_1B_6$ ,  $(Nd_{0.3}Ce_{0.7})_{15}Fe_{78}TM_1B_6$  等, 为深入探究稀土磁体成分与性能的关系提供重要的理论依据和数据支撑。为进一步实现稀土资源的平衡利用, Kovacs 等<sup>[13]</sup>结合实验数据、第一性原理计算和机器学习方法, 开展低稀土含量 Nd-Fe-B 永磁体的成分设计与优化。基于偏最小二乘回归、高斯过程回归、随机森林和神经网络模型等模型, 建立材料成分与饱和磁化强度  $\mu_0 M_s$ 、磁晶各向异性  $K_1$  的关系, 预测  $(Nd/La/Ce/Pr)_2(Fe/Co/Ni)_{14}B$  磁体的本征磁性能。此外, 利用遗传算法优化成分与结构之间的关系, 发现具有核壳结构的晶粒(壳层高 Nd、芯层近零 Nd)可在降低成本的同时维持矫顽力。

2021 年, Hosokawa 等<sup>[14]</sup>整合了 800 余条关于 Sm-Fe-N 快淬磁体的数据库, 建立了以化学成分(如 Sm, Fe, Co, N 含量)、快淬工艺条件(如辊速、温度)和热处理参数为特征变量, 以剩磁( $B_r$ )、矫顽力( $H_{cj}$ )等磁性能为目标变量的机器学习模型,

并通过网格搜索来优化模型的超参数。神经网络算法模型在数据密集区表现出良好的预测精度, 但在外推场景下偏差增大。验证表明, 模型对于含 Nb, Ti 或 Zr 掺杂的 Sm-Fe-Co-N 性能预测结果与实验数据基本吻合, 仅部分极端成分设计存在显著差异。2022 年, Guo 等<sup>[15]</sup>基于机器学习技术明确了 Sm-Co 合金饱和磁化强度与掺杂元素的依赖关系。研究人员构建了一个包含材料成分、制备工艺、磁性能等多维度信息的数据集, 并利用线性回归、随机森林、支持向量机等多种机器学习算法进行模型训练。在特征处理环节, 研究人员首先引入了 20 种基本特征, 如图 2 所示, 通过 Pearson 相关系数分析法去除冗余特征, 并基于 17 种基本特征构建出 avrF 和 stdF 两种新特征。随后采用多步特征选择策略, 包括初始特征选择、子集选择以及 1000 次自助抽样等步骤, 筛选出对模型性能具有关键影响的特征, 如 std\_VEN 和 avr\_Rpse。在模型选择方面, 集成学习算法由于数据量的限制存在过拟合问题, 而线性模型在诸多模型中表现最佳。结果显示,

图 2 原子特征之间的 Pearson 相关系数热图<sup>[15]</sup>Fig. 2. Heatmap of the Pearson correlation coefficients between atomic characteristics<sup>[15]</sup>.

除 Fe, Bi 外大部分掺杂元素都会在不同程度上降低其饱和磁化强度, 优先掺杂元素为 Fe, Mn, Ni, Cu. 与此同时, 利用 SmCo<sub>6.2</sub>Fe<sub>0.5</sub>Mn<sub>0.3</sub> 和 SmCo<sub>6.2</sub>Mn<sub>0.5</sub>Cu<sub>0.3</sub> 磁体验证了该模型预测的有效性, 表明多元素掺杂有望改善合金综合磁性能.

2024 年, Wen 等 [16] 基于合金成分、剩磁  $B_r$ 、矫顽力  $H_{cj}$ 、最大磁能积  $(BH)_{max}$  等硬磁性能的实验数据构建数据库. 与传统支持向量回归 (SVR)、多元线性回归 (MLR) 和随机森林回归 (RFR) 模型相比, GBRT 在预测精度和稳定性上均表现出显著优势. 研究表明, GBRT 模型对 Nd<sub>2</sub>Fe<sub>14</sub>B 矫顽力的预测结果与实验值高度一致, 充分验证了模型的有效性和可靠性. 此外, 研究人员基于 GBRT 模型对合金成分进行了优化, 预测出了具有更高永磁性能的新合金成分, 如 Nd<sub>11</sub>Fe<sub>77</sub>B<sub>6</sub>Co<sub>3</sub>Cr<sub>3</sub>, Nd<sub>11</sub>Fe<sub>77</sub>B<sub>6</sub>Co<sub>2</sub>Cr<sub>4</sub>, Nd<sub>11</sub>Fe<sub>77</sub>B<sub>6</sub>Co<sub>3.5</sub>Cr<sub>2.5</sub> 等. 这些新合金在矫顽力、剩磁和最大磁能积等方面均展现出明显的提升, 为永磁材料的设计与开发提供了新的思路.

本研究团队通过高通量正交实验与机器学习技术相结合, 研究了 La-Co 替代锶六角铁氧体的成分设计与磁性能预测 [17]. 我们采用固相反应法制备了 145 个样品, 结合文献所获数据, 利用高斯过程回归 (GPR)、SVR 和径向基函数网络 (RBFN) 等机器学习方法建立模型, 同时通过 PSO、GA 和灰狼算法 (GWA) 优化模型的超参数. 结果表明, SVR-

GWA 组合在预测饱和磁化强度  $M_s$  和矫顽力  $H_{cj}$  时表现最优, 误差分别为 0.50 emu/g 和 0.1049 kOe, 决定系数  $R^2$  分别为 0.9919 和 0.9818. 基于该模型生成了 La-Co 成分对于磁性能影响的预测图, 如图 3 所示, 并通过实验验证了其中 5 个新成分的磁性能. 实验与预测结果高度吻合, 说明 SVR-GWA 模型能有效指导铁氧体材料的成分设计.

### 2.1.2 居里温度及磁性能预测

在永磁材料研究领域, 居里温度  $T_c$  作为表征材料磁有序-无序转变的临界参数, 是评估其磁性能热稳定性的核心指标. 近年来, 随着新能源汽车、航空航天等领域对高温环境用磁体的迫切需求, 提升  $T_c$  已成为高性能永磁材料研发的战略方向.  $T_c$  不仅直接决定了永磁体的最高工作温度, 还与材料的矫顽力温度系数、剩磁稳定性等关键磁性能密切相关. 2019 年, Nguyen 等 [18] 提出了一种基于集成学习的方法, 通过核岭回归 (KRR) 与贝叶斯优化 (BO) 联合分析稀土-过渡金属二元合金的居里温度数据. 该研究利用高斯混合模型 (GMM) 近似预测值分布, 发现某些材料的  $T_c$  预测分布呈现多峰, 表明其形成机制与其他材料存在显著差异. 即使化学组成相近, 如 CeCo<sub>5</sub> 和 LaCo<sub>5</sub>, 若其描述变量明显不同, 也会导致  $T_c$  预测的多峰性. 该研究通过差异投票机进一步量化了材料间的非

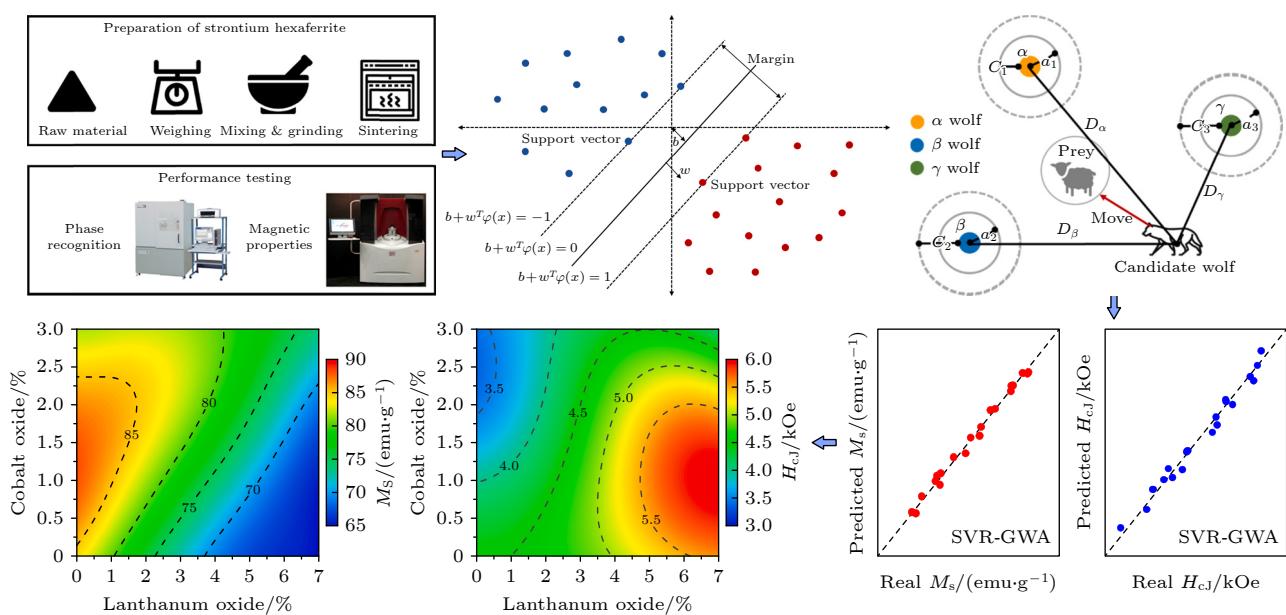


图 3 基于高通度实验和机器学习的 La-Co 取代锶六角铁氧体磁性能预测与组成设计 [17]

Fig. 3. Magnetic properties prediction and composition design of La-Co substitution Sr-hexaferrite based on high-throughput experiments and machine learning [17].

相似性，并以实验证证了该方法的有效性，成功识别出  $\text{CeCo}_5$  和  $\text{GdFe}_5$  等异常材料，揭示了稀土元素与  $T_c$  之间的非线性关系。Sm-Co 作为目前最适用于高温服役的永磁候选材料，其居里温度的准确预测对于促进永磁体的极端环境应用至关重要。Xu 等<sup>[19]</sup> 利用数据驱动策略构建了一个包含 1050 个 Sm-Co 合金数据库，通过特征筛选发现 Co 含量、掺杂元素电导率  $\kappa$  及熔化热差异  $\Delta H_{f,\text{Sm}}$  是影响  $T_c$  的关键因素。如图 4 所示，基于遗传程序的符号回归模型 (GP-SR) 实现了对  $T_c$  的高通量预测，模型在训练后达到  $R^2 = 0.91$  的高准确性，成功揭示了 Co 含量与  $T_c$  的正相关关系。此外，借助逻辑回归模型定义了敏感系数  $Z$  来定量描述掺杂元素对  $T_c$  的影响，2% Fe 掺杂使  $\text{SmCo}_7$  的  $T_c$  从 870 °C 提升至 894.5 °C，2% Cu 掺杂使  $\text{SmCo}_{12}$  的  $T_c$  升至 1051 °C，为高温永磁材料设计提供了高效的方法论框架。

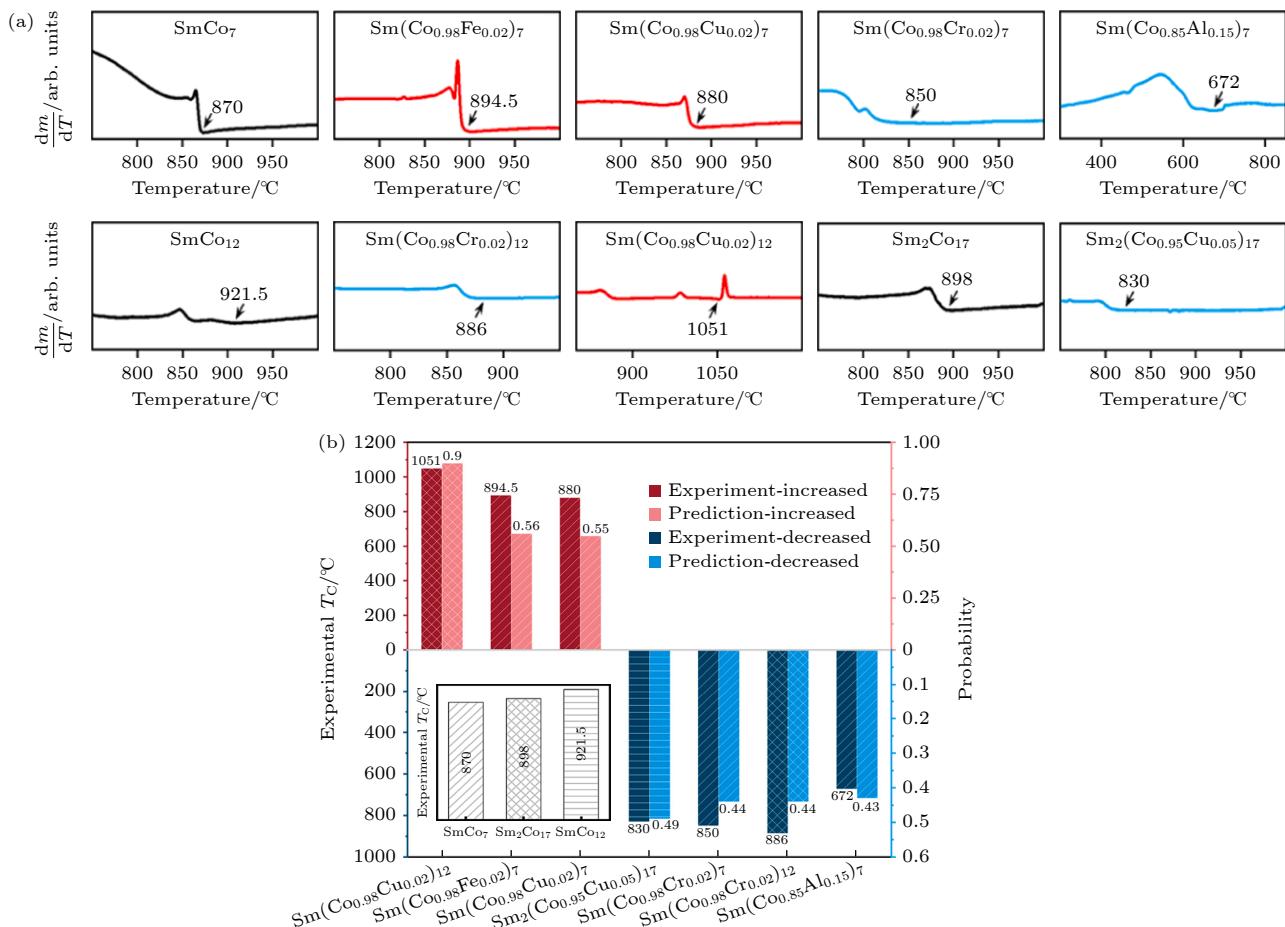
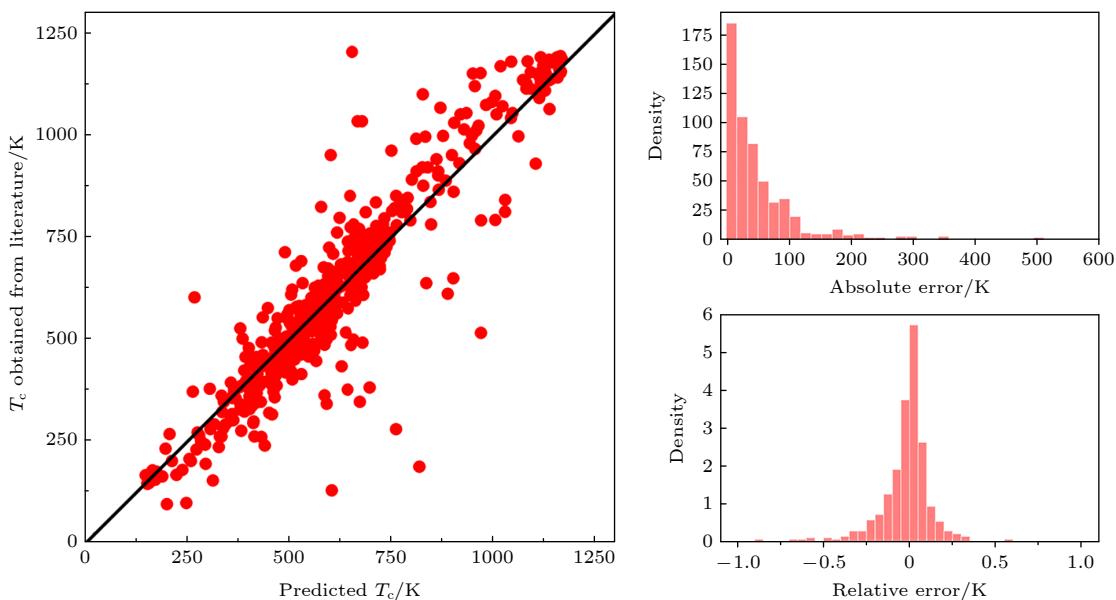


图 4 (a) 不同 Sm-Co 合金的热重分析微分曲线；(b) 实验测量和逻辑回归模型预测的  $T_c$ <sup>[19]</sup>

Fig. 4. (a) DTG curves of the Sm-Cobased alloys prepared in experiments; (b) the  $T_c$  values from experimental measurements and from the model predictions<sup>[19]</sup>.

将机器学习与第一性原理计算相结合，可以突破单一方法的局限性，为高性能材料的设计提供了高效且可解释的解决方案。第一性原理计算是基于量子力学，能够从原子尺度预测材料性质，但计算成本高，难以处理大规模材料体系。而机器学习可以通过分析大量数据，快速建立预测模型，但缺乏物理可解释性。2020 年，Halder 等<sup>[20]</sup> 利用机器学习和第一性原理计算，系统研究了低稀土含量磁体  $\text{Ce}_2\text{Fe}_{17-x}\text{Co}_x\text{CN}$  的磁性能。利用随机森林回归模型在包含 565 个化合物的数据库中，筛选出  $\mu_0 M_s > 1 \text{ T}$  且易轴各向异性  $K_u > 0$  的候选材料，并预测其居里温度  $T_c$ ，如图 5 所示。结果显示， $\text{Ce}_2\text{Fe}_{17-x}\text{Co}_x\text{CN}(x = 1-7)$  具有较高的  $M_s$  和  $T_c$  ( $> 600 \text{ K}$ )，且 Co 的替代会显著提升其磁晶各向异性。进一步基于第一性原理的 GGA+U+SOC 计算验证了预测结果，Co 优先占据 9d 和 18h 位，碳氮共掺杂显著提高了空位形成能。通过 d-p 杂化和

图 5 随机森林算法对稀土过渡族金属居里温度  $T_c$  的预测值、绝对误差分布和相对误差分布<sup>[20]</sup>Fig. 5. Model output from Random Forest algorithm for  $T_c$ , absolute error, and relative error<sup>[20]</sup>.

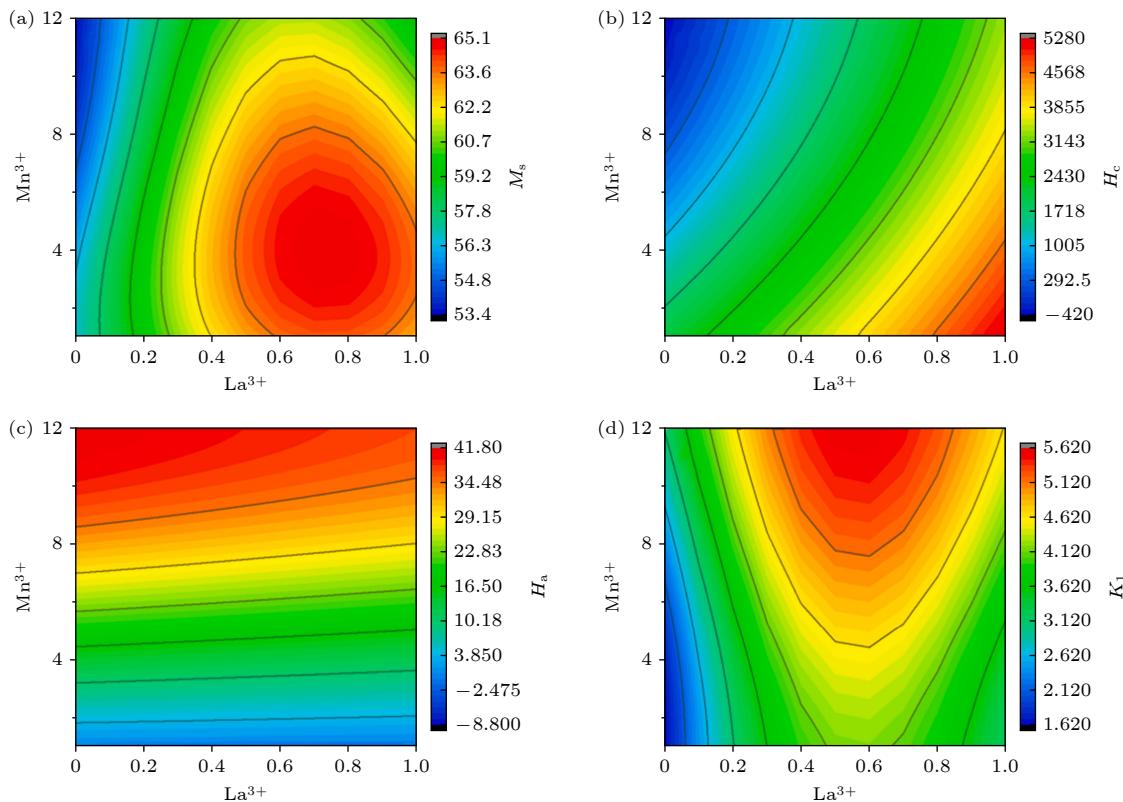
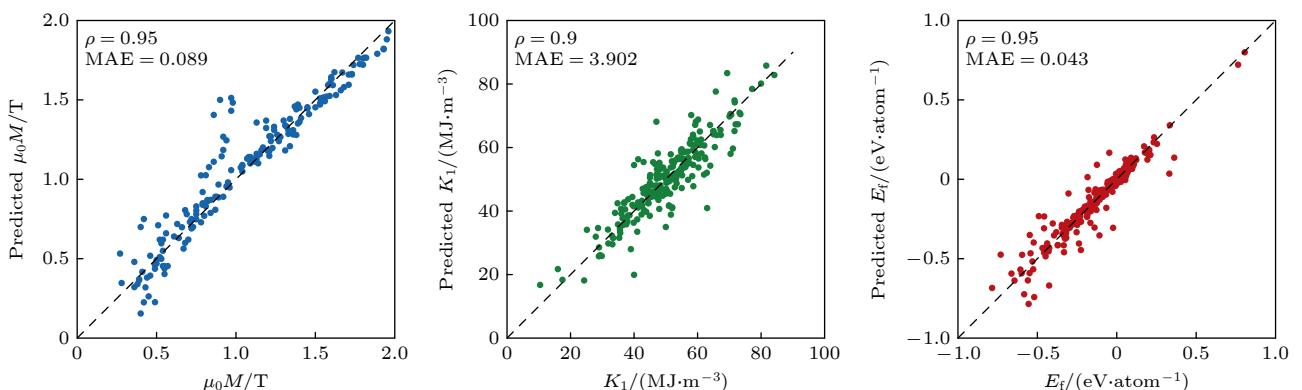
自旋-轨道耦合的增强,平面各向异性转变为轴各向异性,其中  $\text{Ce}_2\text{Fe}_{17-x}\text{Co}_x\text{CN}$ ( $x = 2, 5, 6, 7$ ) 的  $K_u$  达  $3.54 \text{ MJ/m}^3$  以上,  $\text{Ce}_2\text{Fe}_{15}\text{Co}_2\text{CN}$  等成分的最大磁能积  $(BH)_{\max}$  约  $540 \text{ kJ/m}^3$ , 有望填补传统永磁体的性能空白.

永磁材料作为一种结构敏感型材料,其磁性能受到烧结、热处理过程中多个因素的影响. Lambard 等<sup>[21]</sup>以机器学习和贝叶斯优化为基础建立主动学习(AlMLBO)方法,探究热变形磁体 Nd-Fe-B 的最优工艺条件.以商业粉末 MQU-F 为例,定义了热压温度、负载,热挤压温度、速度、负载限制和模具等 6 个可调参数,通过随机森林回归和高斯过程对模型进行迭代优化,并将实验数据作为数据集再进行模型的建立.在初始仅有少量样本数据条件下,经过循环迭代和优化,成功将磁体性能提升至  $\mu_0 H_c = 1.7 \text{ T}$ ,  $\mu_0 B_r = 1.4 \text{ T}$ ,  $(BH)_{\max} = 380 \text{ kJ/m}^3$ .对于 M 型六角铁氧体,Lu 等<sup>[22]</sup>构建了包含 1242 个材料的数据库,涵盖 51 种离子、工艺参数及磁性能,并对饱和磁化强度  $M_s$ 、矫顽力  $H_c$ 、磁晶各向异性常数  $K_1$  和各向异性场  $H_a$  等进行预测,如图 6 所示.支持向量回归(SVR)在数据集上表现最优,  $R^2$  达到 0.74—0.92.实验和模型预测结果一致,  $\text{Ba}_{1-x}\text{La}_x\text{Fe}_{12-\gamma}\text{Mn}\gamma\text{O}_{19}$ ( $x = 0.4$ ) 的剩磁比最高  $M_r/M_s = 0.906$ , 归因于磁体的高致密度和合适的晶粒尺寸.通过第一性原理计算揭示了  $\text{Mn}^{3+}$  在 2a, 2b, 4f<sub>1</sub>, 4f<sub>2</sub> 和 12 k 晶位的占位概率分别为 2.91%,

19.31%, 22.32%, 14.68% 和 40.78%, 同时结合实验优化了烧结密度和晶粒尺寸等工艺参数.此外, Choudhary 等<sup>[23]</sup>基于传统机器学习和先进深度学习技术开发了一种计算机辅助方法,用于定量分析烧结钕铁硼磁体的磁光克尔显微镜图像. 经过微调的深度学习模型比手动提取特征训练的传统机器学习模型具有更好的泛化能力,该方法可以在高精度与高通量之间实现比电子背散射衍射(EBSD)更优的平衡,并对晶粒取向和尺寸进行量化.此外,该方法还可以通过晶粒取向数据确定各向异性烧结磁体的取向质量/织构质量及其与剩磁的关系,有望成为快速分析和量化晶粒尺寸及取向异质性的工具.

### 2.1.3 新型永磁材料的探索

探索具有更高磁性能、更好热稳定性以及更低成本的新型材料一直是稀土永磁材料领域研究的重点与难点,尤其是低稀土、轻稀土和无稀土永磁材料的开发.2018 年,Möller 等<sup>[7]</sup>提出了一种基于机器学习的新策略,通过构建支持向量回归模型(SVR)优化低稀土硬磁相的成分设计,旨在发现具有高磁晶各向异性  $K_1$  和饱和磁化强度  $\mu_0 M_s$  的新型永磁材料,如图 7 所示.以  $\text{ThMn}_{12}$  型晶体结构为主要研究对象,利用密度泛函理论(DFT)高通量筛选生成的 234 条数据训练模型.通过十折交叉验证确定最优超参数(多项式核函数,正则化参

图 6 支持向量回归模型对  $\text{Ba}_{1-x}\text{La}_x\text{Fe}_{12-y}\text{Mn}_y\text{O}_{19}$  铁氧体 (a)  $M_s$ ; (b)  $H_c$ ; (c)  $K_1$ ; (d)  $H_a$  的预测图<sup>[22]</sup>Fig. 6. Prediction diagram of (a)  $M_s$ , (b)  $H_c$ , (c)  $K_1$  and (d)  $H_a$  of  $\text{Ba}_{1-x}\text{La}_x\text{Fe}_{12-y}\text{Mn}_y\text{O}_{19}$  ferrite by SVR<sup>[22]</sup>.图 7 支持向量回归模型 (SVR) 在  $\mu_0 M$ ,  $K_1$  和  $E_f$  上的十折交叉验证结果<sup>[7]</sup>Fig. 7. Results of the tenfold cross validation of the SVR models for  $\mu_0 M$ ,  $K_1$ , and  $E_f$ <sup>[7]</sup>.

数  $C = 100$ , 宽度  $\gamma = 0.1$ ), 模型在测试集上表现出高预测精度 ( $R^2 = 0.88$ — $0.95$ ). 研究发现,  $\text{CeFe}_6\text{Co}_8\text{Cu}_3\text{TiN}$ ,  $\text{NdFe}_{8.5}\text{Cu}_{3.5}\text{N}$  等几种潜在的候选永磁材料, 尽管其磁性能略低于  $\text{Nd}_2\text{Fe}_{14}\text{B}$  和  $\text{SmCo}_5$ , 但这些材料能够在保持较高性能的同时显著减少了稀土元素的使用.

为进一步降低稀土元素在永磁材料中的用量, Kusne 等<sup>[24]</sup>提出了一种结合实时机器学习与高通量实验的框架 (mean shift theory, MST), 如图 8

所示, 用于快速发现无稀土永磁材料. 研究团队通过同步辐射 X 射线衍射技术对 Fe-Co-Mo 三元薄膜材料库进行表征, 同时利用 MST 算法对衍射数据进行实时聚类分析, 并结合无机晶体结构数据库 (ICSD) 辅助分类.  $\text{Fe}_{78}\text{Co}_{11}\text{Mo}_{11}$  成分附近存在一种四方畸变结构 (空间群  $P4/m$ ), 其磁各向异性能达  $18.2 \text{ meV/atom}$ , 饱和磁化强度为  $850 \text{ emu/cc}$ . 结合遗传算法与密度泛函理论计算, 验证了该结构的热力学稳定性及磁性能起源. 该方法通过机器学

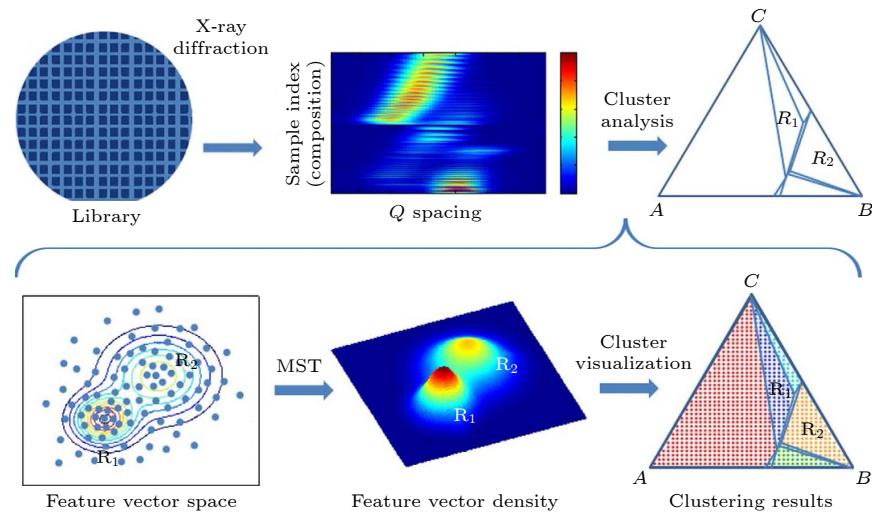


图 8 将均值漂移理论应用于成分扩展晶圆的衍射图谱, 用于快速分析相分布<sup>[24]</sup>

Fig. 8. Mean shift theory as applied to diffraction patterns taken from a composition spread wafer for rapid phase distribution analysis<sup>[24]</sup>.

显著提升了高通量实验的数据分析效率, 完成了从实验数据到新材料发现的闭环, 为无稀土永磁体开发提供了高效范式。

2016 年, Nieves 等<sup>[25]</sup>提出密度泛函理论与多尺度建模结合的高通量计算框架, 通过 3 个核心步骤实现材料设计, 重点解决稀土永磁体的资源依赖问题。首先通过 DFT 计算材料基态结构与磁性能, 如饱和磁化强度、磁晶各向异性等; 然后利用数据库存储和管理数据; 最后借助智能筛选算法快速搜索潜在候选材料。以  $\text{Fe}_{0.75}\text{Sn}_{0.25}$  为例, 成功预测其基态结构并发现多个低能亚稳相, 验证了该方法的有效性。此外, 重点关注无稀土永磁材料, 如 Fe-Co-Mo 体系的应用潜力。2022 年, Xia 等<sup>[26]</sup>进一步优化算法并加强实验验证, 利用 ML、自适应遗传算法 (AGA) 和第一性原理 (DFT) 提出了一个有效的反馈循环, 以高效筛选有永磁应用前景的化学成分和晶体结构。相比传统基于描述符的机器学习, 该方法利用晶体图卷积神经网络 (CGCNN) 直接提取晶体图结构信息, 对 Fe-Co-B 三元体系进行高通量机器学习筛选, 同时结合 DFT 计算验证候选结构的热力学稳定性与磁性能, 最终通过 AGA 搜索发现了 7 种适用于永磁应用的 Fe-Co-B 三元化合物, 如图 9 所示。其中, 实验合成的高性能无稀土磁性化合物  $\text{Fe}_3\text{CoB}_2$  具有正交结构, 室温下展现出优异的磁晶各向异性  $K_1 = 1.2 \text{ MJ/m}^3$  和高饱和磁极化强度  $J_s = 1.39 \text{ T}$ 。

## 2.2 数据挖掘技术在稀土磁致冷材料中的应用

### 2.2.1 磁熵变预测

基于磁热效应 (MCE) 的磁制冷材料因其在环境友好性和高效性方面的潜在优势, 一直是稀土功能材料领域的研究热点。磁热材料通常以等温过程中的最大磁熵变  $(\Delta S_M)_{\max}$  和材料工作温度区间的居里温度  $T_c$  为关键表征参数。2018 年, Zhang 等<sup>[27]</sup>成功利用机器学习技术构建了  $\text{La}(\text{Fe}, \text{Si}/\text{Al})_{13}$  基材料的磁热效应预测模型, 并应用梯度提升回归树实现了对居里温度  $T_c$  和最大磁熵变  $(\Delta S_M)_{\max}$  的高效预测。如图 10 所示, 通过五折交叉验证和网格搜索优化模型参数,  $T_c$  预测模型的决定系数  $R^2$  可以达到 0.96, 平均绝对误差 MAE 约 9.81 K。而  $(\Delta S_M)_{\max}$  预测模型在 2 T 和 5 T 下的  $R^2$  分别为 0.87 和 0.91, MAE 约 2.5 J/(kg·K)<sup>[28]</sup>。该研究验证了机器学习在复杂成分-性能关系中的泛化能力, 为磁制冷材料的快速筛选和优化提供了有力工具。

2022 年, Ucar 等<sup>[29]</sup>基于实验数据构建了包含多类重要磁热材料的数据集, 包括  $\text{La}(\text{Fe}, \text{Si}/\text{Al})_{13}$ , Heusler 合金、锰氧化物、 $\text{Gd}_5(\text{Si}, \text{Ge})_4$  体系、稀土及金属玻璃、Laves 相化合物等, 并整合了其磁熵变  $\Delta S_M(T, H)$  数据。通过引入线性与非线性机器学习模型, 对这些材料的磁热效应进行预测, 并通过交叉验证和测试集评估了不同模型的性能。分析表明, 随机森林模型的表现最优, 决定系数  $R^2$  达到 0.82, 如图 11 所示。基于该模型, 研究人员对近 4 万

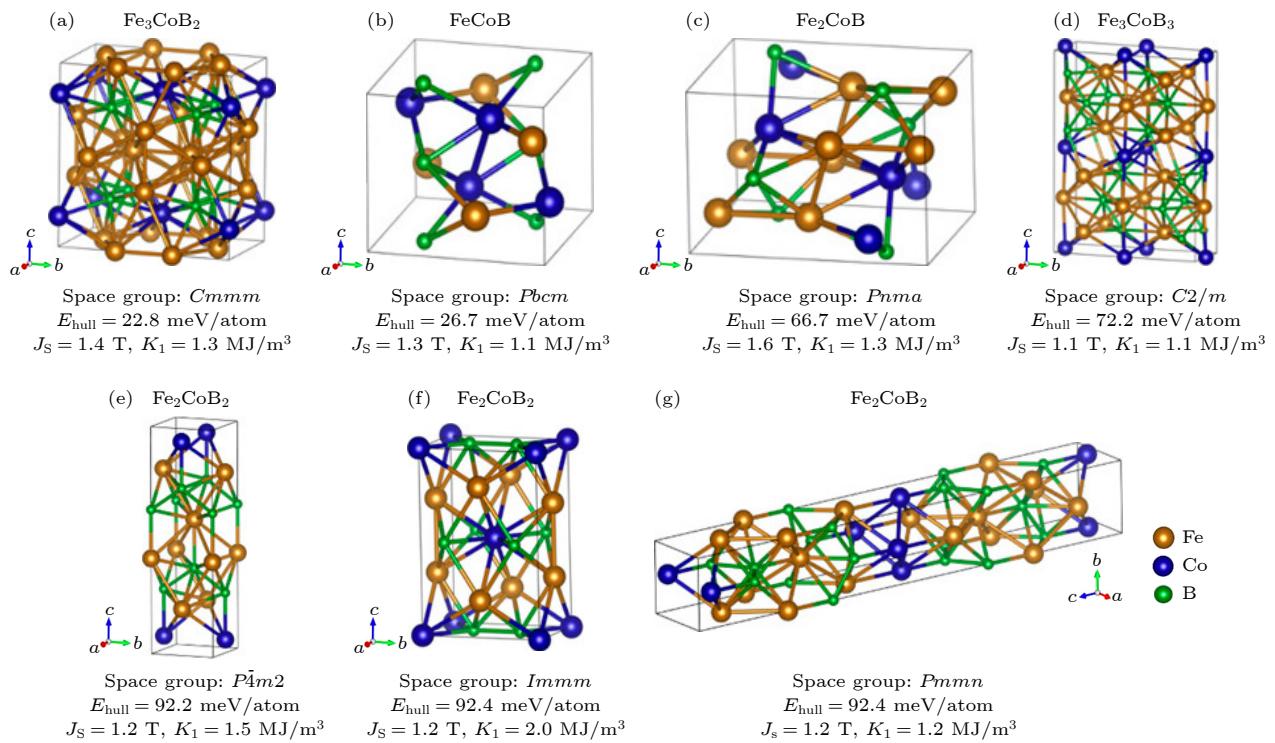
图 9 ML-AGA-DFT 反馈循环中获得的 7 种有应用潜力的永磁体 Fe-Co-B 化合物的晶体结构<sup>[26]</sup>

Fig. 9. Crystal structures of seven promising Fe-Co-B compounds for permanent magnet obtained from our ML-AGA-DFT predictions<sup>[26]</sup>.

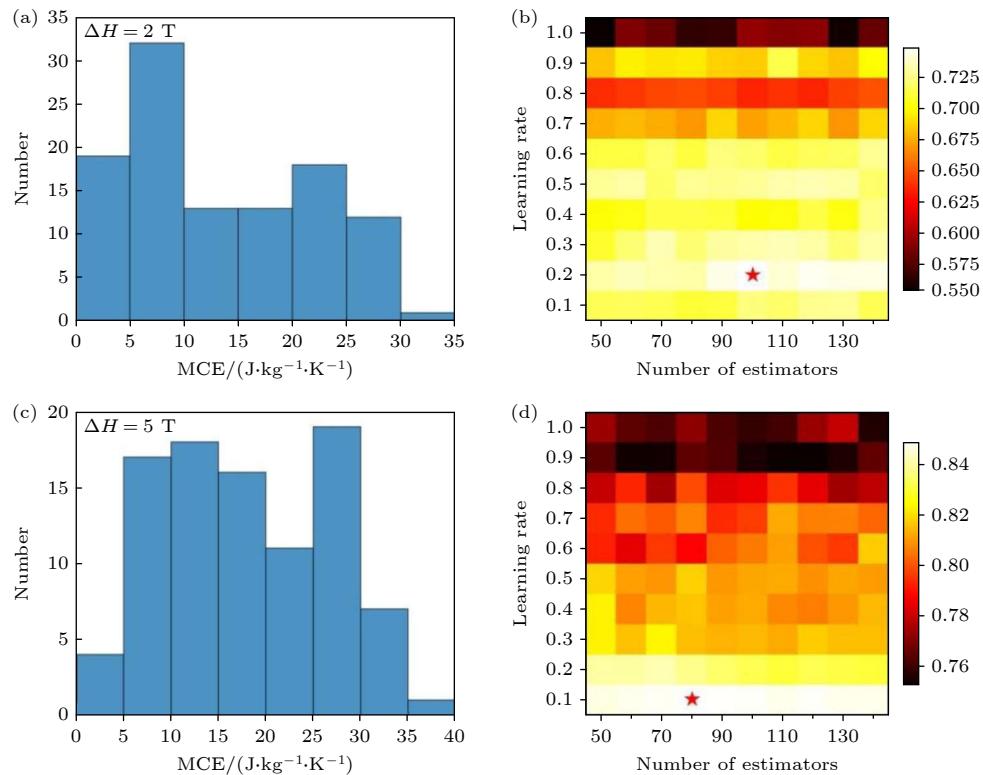
图 10 (a) 0—2 T 和 (c) 0—5 T 下最大磁熵变  $(\Delta S_M)_{\text{max}}$  的数据分布; (b) 0—2 T 和 (d) 0—5 T 预测模型的训练性能表现, 红星为最佳参数组合<sup>[28]</sup>

Fig. 10. Data distribution for  $(\Delta S_M)_{\text{max}}$  at (a) 0—2 T and (c) 0—5 T; performances of prediction models at (b) 0—2 T and (d) 0—5 T trained with different parameters combinations<sup>[28]</sup>.

种化合物的大型数据库进行筛选,发现了MnGa<sub>2</sub>Sb<sub>2</sub>, CrGa<sub>2</sub>Sb<sub>2</sub>, SbSCl<sub>0.1</sub>I<sub>0.9</sub>, Sm<sub>3</sub>Te<sub>4</sub>, LaRhSn, SbSI, Tl<sub>0.58</sub>Rb<sub>0.42</sub>Fe<sub>1.72</sub>Se<sub>2</sub>, Cs<sub>0.86</sub>Fe<sub>1.66</sub>Se<sub>2</sub>, La<sub>2.1</sub>MnGe<sub>2.2</sub>等新型室温磁热材料,这些材料有望实现高效磁制冷性能.

### 2.2.2 相对冷却功率预测

相对冷却功率(relative cooling power, RCP)是评估磁热材料性能的另一关键参数, Alqahtani<sup>[30]</sup>

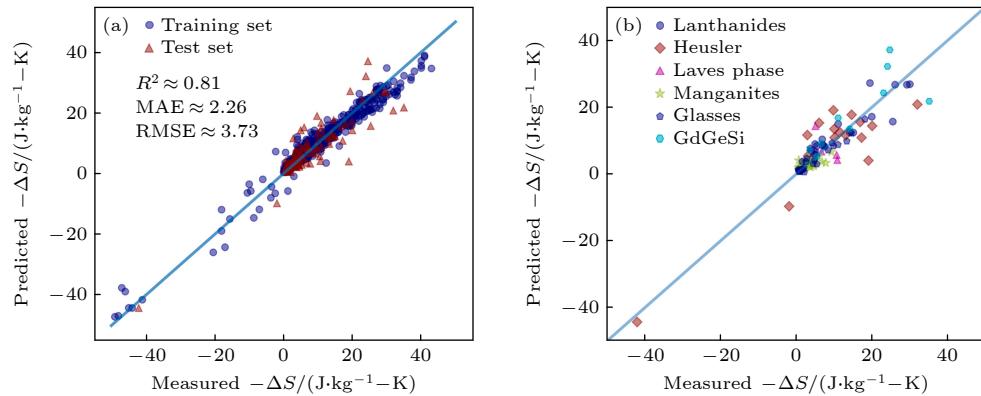


图 11 随机森林模型对材料磁熵变 $-\Delta S_M$ 的预测结果<sup>[29]</sup> (a) 数据集中所有材料 $-\Delta S_M$ 的实验值和预测值对比; (b) 测试集中按材料类型细分的实验 $-\Delta S_M$ 与预测对比

Fig. 11. RF model for  $-\Delta S_M$  of magnetocaloric materials<sup>[29]</sup>: (a) Experimental vs predicted  $-\Delta S_M$  for the materials in dataset; (b) experimental vs predicted  $-\Delta S_M$  in the test set by material type breakdown.

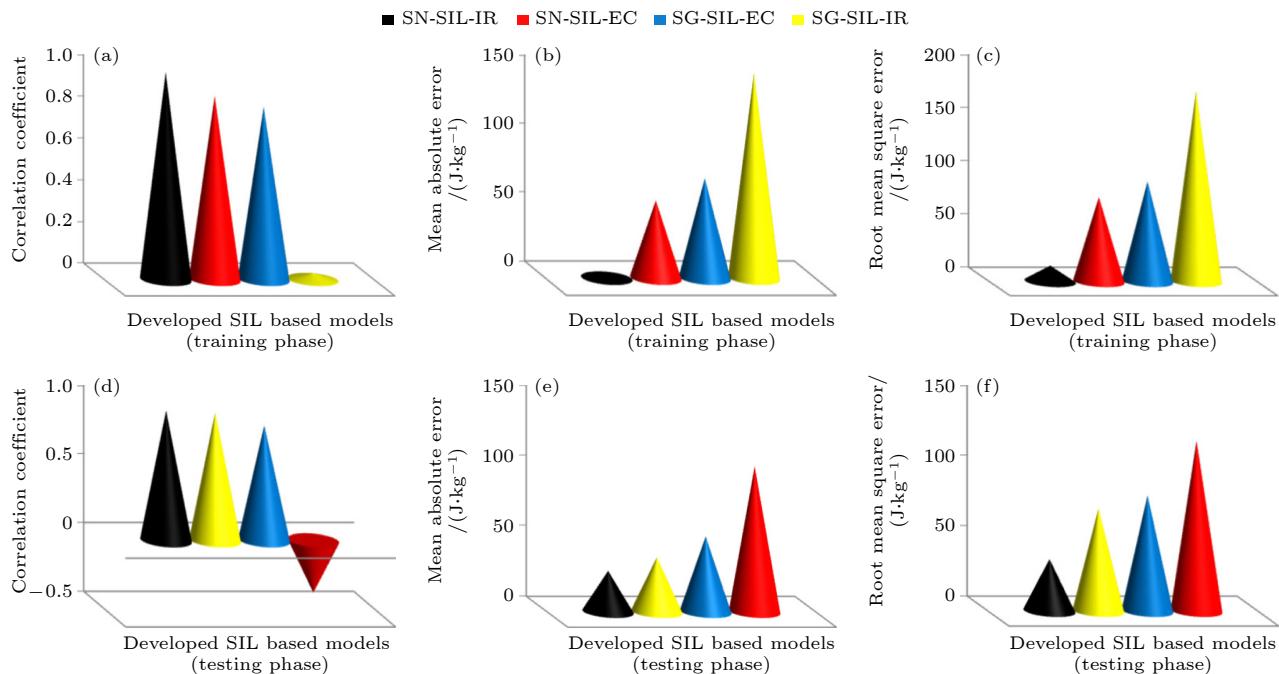


图 12 单隐层模型 SIL 的泛化能力评估<sup>[30]</sup> (a) 训练样本的相关系数; (b) 训练样本的平均绝对误差; (c) 训练样本的均方根误差; (d) 测试样本的相关系数; (e) 测试样本的平均绝对误差; (f) 测试样本的均方根误差

Fig. 12. Generalization capacity evaluation of developed SIL-based models<sup>[30]</sup>: (a) Training samples coefficient of correlation; (b) training samples mean absolute error; (c) training samples root mean square error; (d) testing samples coefficient of correlation; (e) testing samples mean absolute error; (f) testing samples root mean square error.

通过开发单隐层极端学习机(SIL-ELM)的智能模型,用于预测稀土-过渡金属基磁制冷材料,如Dy基化合物的RCP,以推动可持续磁制冷技术的发展.如图12所示,研究人员基于38种Dy基金属间化合物的实验数据,构建了两种模型:以元素离子半径(IR)为描述符的SN-SIL-IR模型(正弦激活函数)和以最大磁熵变(EC)为描述符的SN-SIL-EC模型.结果表明,SN-SIL-IR模型表现最优,测试集上的相关系数CC达0.9462,平均绝对误差

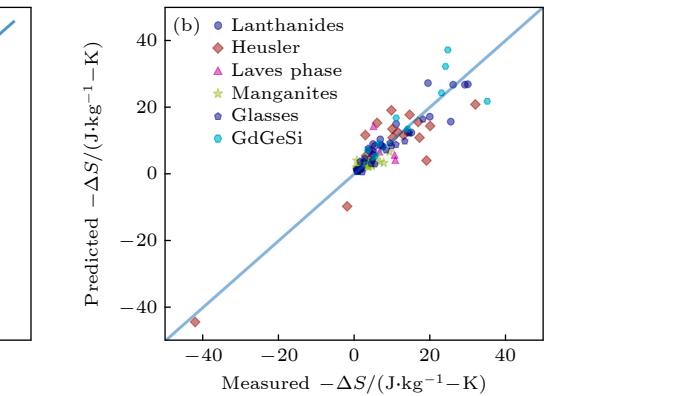


图 12 单隐层模型 SIL 的泛化能力评估<sup>[30]</sup> (a) 训练样本的相关系数; (b) 训练样本的平均绝对误差; (c) 训练样本的均方根误差; (d) 测试样本的相关系数; (e) 测试样本的平均绝对误差; (f) 测试样本的均方根误差

MAE 为 28.99 J/kg, 均优于基于 Sigmoid 激活函数的模型. 当离子半径作为描述符时, 可以更有效地捕捉材料晶体结构对于 RCP 的影响, 且模型预测与实验数据高度吻合. 此外, 模型揭示了 RCP 随外加磁场的增大呈现先增后减的趋势, 验证了  $\text{Dy}_2\text{Cr}_2\text{C}_3$  和  $\text{DyGa}$  等材料的实验结果.

2024 年, Shamsah<sup>[31]</sup> 针对稀土基三元金属间化合物  $\text{RE}_2\text{TM}_2\text{Y}$  在磁制冷应用中的相对冷却功率问题, 提出了基于极端学习机 (ELM) 和遗传优化支持向量回归 (GSVR) 的机器学习模型. 该研究以元素离子半径和磁场为描述符, 构建了正弦激活函数 (SELM)、高斯核 (GU-GSVR) 和多项式核 (PY-GSVR) 模型. 如图 13 所示, GU-GSVR 模型在测试集上表现最优, 相关系数 CC 达 0.9131, 均方根误差 RMSE 为 30.13 J/kg, 平均绝对误差 MAE 为 21.94 J/kg, 显著优于 PY-GSVR(RMSE = 41.84 J/kg) 和 SELM(RMSE = 140.05 J/kg). 基于 GU-GSVR 模型,  $\text{RE}_2\text{TM}_2\text{Y}$  的 RCP 随磁场增强呈现先增大后稳定的趋势, 验证了  $\text{Ho}_2\text{Cu}_2\text{Cd}$ ,  $\text{Dy}_2\text{Cu}_2\text{Cd}$  等高场下 RCP 可达 400 J/kg 以上. 该研究通过离子半径与磁场的协同作用分析, 有助于加速无稀土磁制冷技术的实用化进程.

在磁制冷材料的研究与应用中, 腐蚀问题一直是制约其性能发挥和实际应用的关键因素. 针对这一问题, Zhao 等<sup>[32]</sup> 在 2023 年提出了利用数据驱动方法加速  $\text{La}(\text{Fe}, \text{Si})_{13}$  高效腐蚀抑制剂的设计. 研究通过构建包含 170 组材料的小型数据库, 利用 LR 模型预测缓蚀剂成分 ( $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  和  $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ ) 对酸碱度 pH 和腐蚀电流密度  $J$  的影响. 如图 14 所示, 模型优化后确定最佳配方 3.0593 g/L 的  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  和 1.5245 g/L 的

$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ . 实验验证显示该缓蚀剂在材料表面形成含  $\text{LaPO}_4$ ,  $\text{FePO}_4$  和  $\text{MoO}_2$  的复合保护膜, 使材料在 60 天后仍保持光滑表面, 腐蚀电流密度仅为  $2.55 \mu\text{A}/\text{cm}^2$ , 显著优于传统试错法 (需尝试 4500 种组合). 磁性能测试显示, 经处理的  $\text{La}(\text{Fe}, \text{Si})_{13}$  材料磁熵变 ( $\Delta S_m$ ) 在 3 T 磁场下保持  $9.97 \text{ J}/(\text{kg}\cdot\text{K})$ , 与初始状态一致, 证实了缓蚀剂的实际应用潜力, 为磁制冷技术的实用化提供了关键解决方案.

### 2.3 数据挖掘技术在其他稀土功能材料中的应用

稀土磁致伸缩材料是一类在外加磁场下产生显著尺寸变化的功能材料, 其核心在于通过稀土元素 (如 Tb, Dy, Sm) 的磁晶各向异性与电子结构特性, 实现高磁致伸缩系数 ( $\lambda_s > 1500 \times 10^{-6}$ ) 和低驱动磁场响应, 在精密机械、能源转换及智能器件领域具有不可替代性. 在探索高性能磁致伸缩材料的合金设计方面, Gong 等<sup>[33]</sup> 通过结合机器学习迭代决策树 (GBRT) 与高通量微磁模拟 (HTP-MMS) 来优化其磁弹性性能. 研究通过构建包含 123 种化合物的数据库, 预测了  $\text{Tb}_x\text{Dy}_{1-x}\text{Fe}_\gamma$  合金的饱和磁致伸缩系数 ( $\lambda_s$ ). 研究表明, 高  $\lambda_s$  合金主要分布在 Tb 含量 0.26—0.60 和 Fe 含量 1.90—2.00 的成分范围内. 进一步通过 HTP-MMS 模拟不同应力下磁畴结构与应力响应的微观机制, 确定最佳压磁 (PM) 性能对应的成分范围为 Tb 含量 0.26—0.32 和 Fe 含量 1.92—1.97, 预测灵敏度达 10.22—13.61 mT/MPa, 与实验数据高度吻合. 模型解释表明, 体积磁化率 ( $\mu_{ave}$ ) 和体积模量 ( $Mb_{ave}$ ) 是影响  $\lambda_s$  的关键参数, 当  $\mu_{ave} > 0.021$  且  $Mb_{ave} > 126.36$  时, 磁致伸缩效应显著增强.

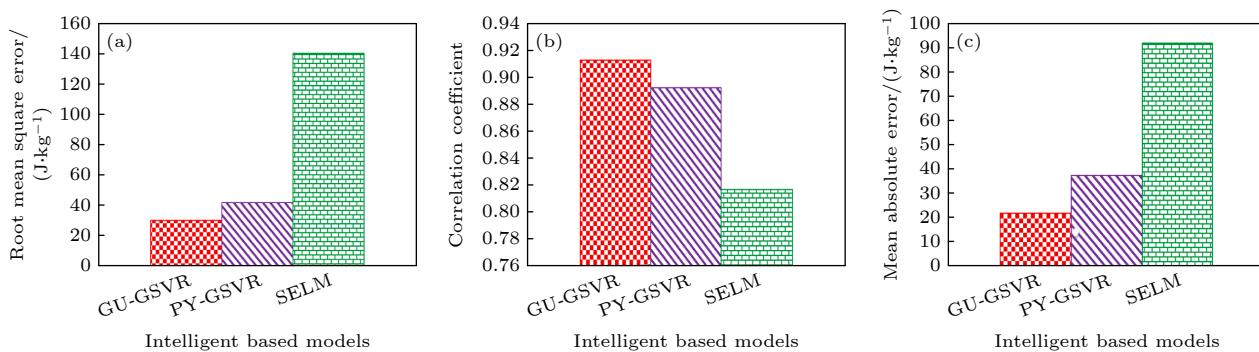


图 13 机器学习模型在预测稀土基三元金属间化合物时的性能比较<sup>[31]</sup> (a) 均方根误差; (b) 相关系数; (c) 平均绝对误差

Fig. 13. Performance of the intelligent models and their comparison for testing ternary intermetallic magnetocaloric compounds<sup>[31]</sup>: (a) Root mean square error; (b) correlation coefficient; (c) mean absolute error.

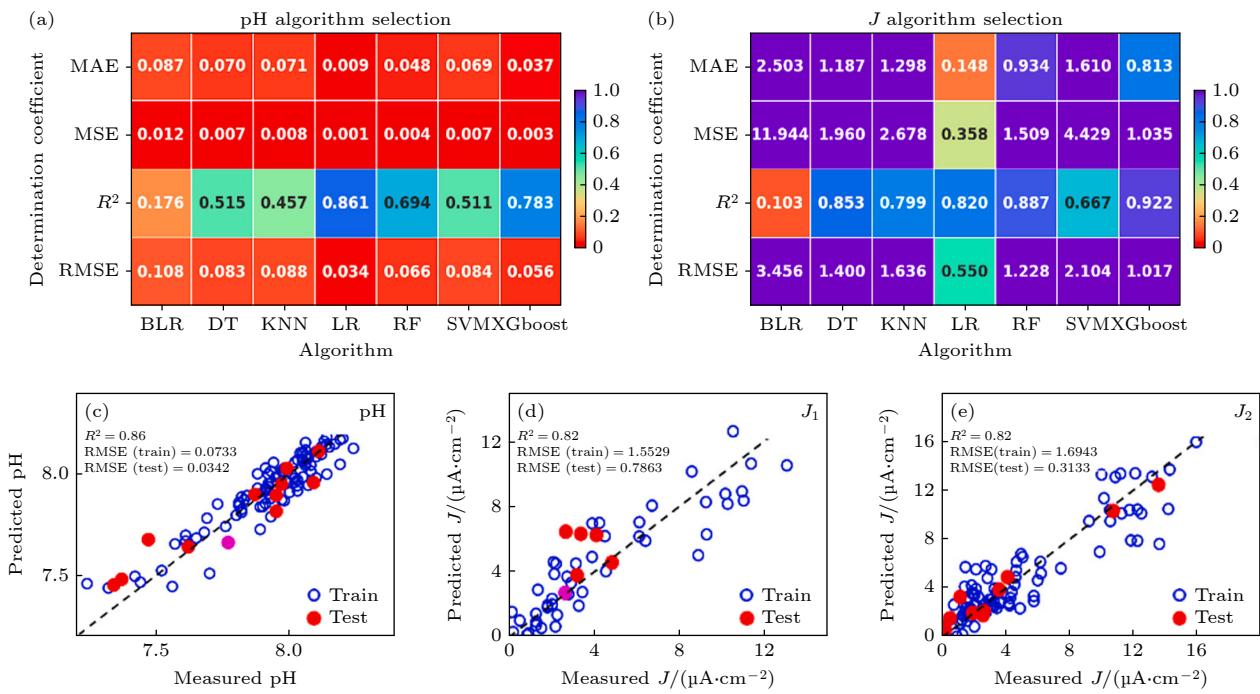


图 14 ML 建模过程及缓蚀剂最佳配比的求解<sup>[32]</sup> (a) 酸碱度 pH 算法选择; (b) 腐蚀电流密度  $J$  算法选择; (c) 基于线性回归算法 pH 值的预测值与测量值对比; (d) 基于线性回归算法的  $J_1$  值; (e)  $J_2$  值的预测值与测量值对比

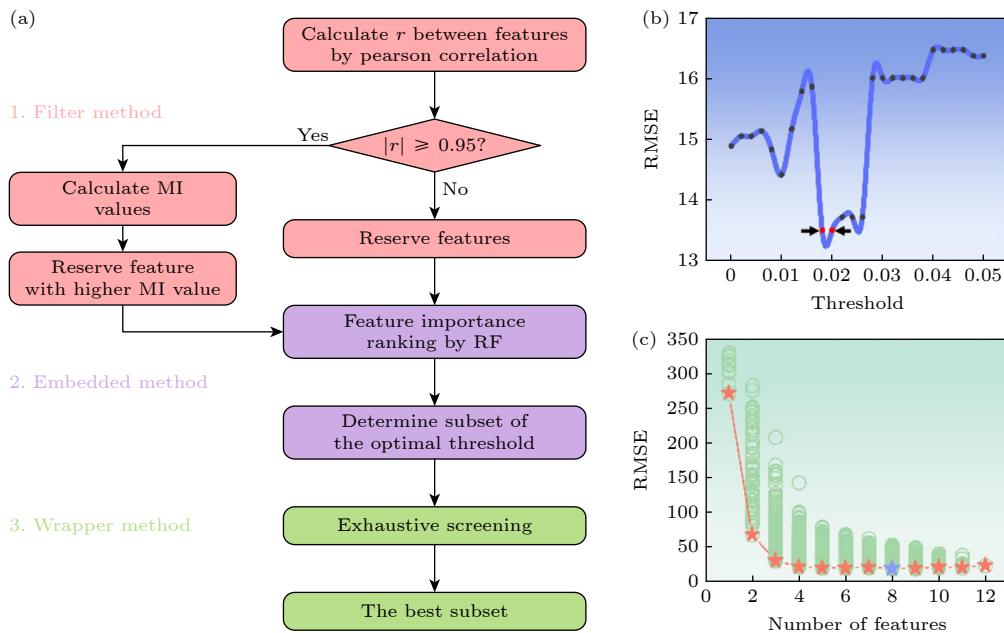
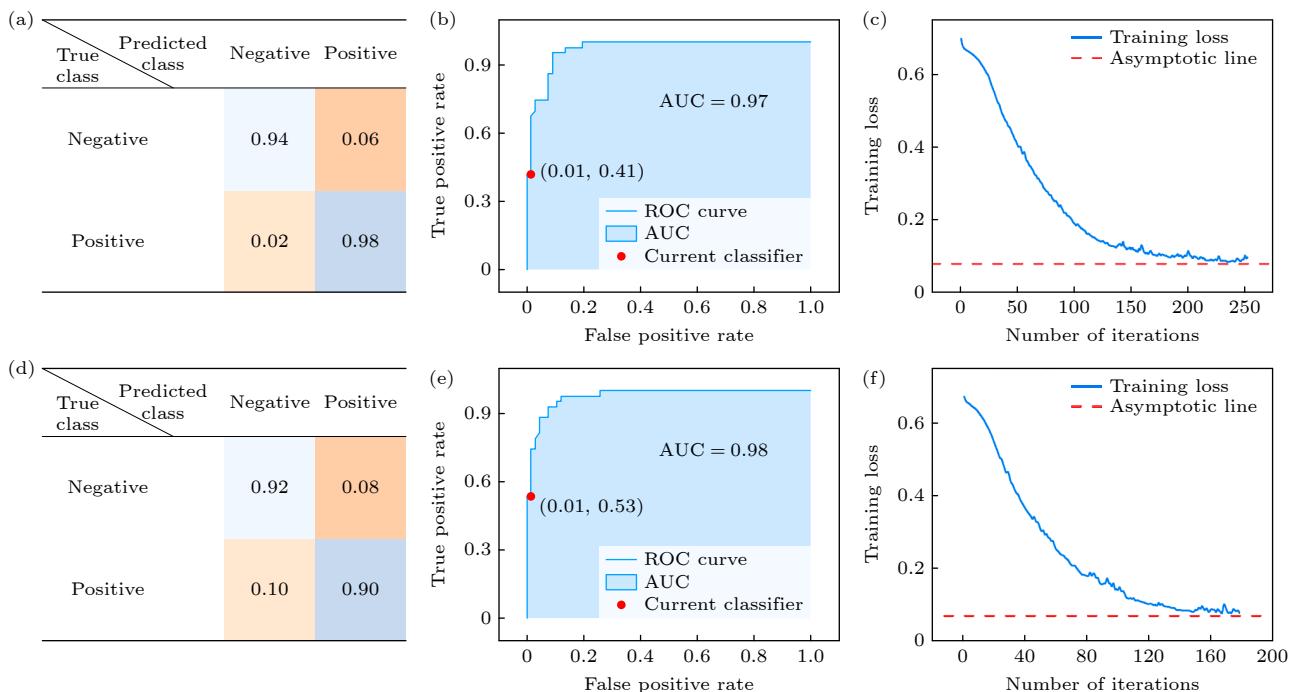
Fig. 14. ML modeling process and solving the optimal composition of corrosion inhibitor<sup>[32]</sup>: (a) The pH model algorithm selection; (b)  $J$  model algorithm selection; (c) measured-predicted pH value plot for LR algorithm; (d) measured-predicted  $J_1$  value plot for LR algorithm; (e) measured-predicted  $J_2$  value plot for LR algorithm.

为进一步优化 RFe<sub>2</sub> 型合金的磁致伸缩性能, Hu 等<sup>[34]</sup>通过引入物理信息可解释的机器学习方法, 对四元合金 Tb<sub>x</sub>Dy<sub>1-x</sub>Fe<sub>y</sub>V<sub>2-y</sub> 的磁致伸缩性能进行精准预测与优化, 如图 15 所示。研究人员首先构建了一个包含合金组成、物理参数与磁致伸缩性能之间关系的数据库, 结合特征工程和模型筛选, 发现极端梯度提升树 (XGBoost) 回归模型在预测 Laves 相 RFe<sub>2</sub> 型合金磁致伸缩性能时表现最优 ( $R^2 = 0.98$ )。研究表明, V 的掺杂抑制了有害相形成, 优化了 Laves 相结构, 提升了磁性能。高  $\lambda_s$  合金主要分布在 Tb 含量 0.23—0.38 和 V 含量 0.01—0.08 的成分范围内, 并通过实验证证了 Tb<sub>0.28</sub>Dy<sub>0.72</sub>Fe<sub>1.96</sub>V<sub>0.04</sub> 可以在 12 kOe 磁场下达到  $1619.69 \times 10^{-6}$  的巨磁致伸缩效应。SHAP 结果显示, 当体积磁化率  $\mu_{\text{ave}} > 0.021$  且体积模量  $Mb_{\text{ave}} > 126.36$  时磁致伸缩显著增强, 为高性能磁致伸缩材料及压磁传感器的设计提供了数据驱动框架。

磁存储材料的核心功能依赖于磁结构中磁畴的有序排列与动态调控。磁畴是磁性材料内部由交换作用和各向异性主导形成的自发磁化区域, 其结构决定了材料的磁性能。高密度存储要求磁畴微小且稳定以减少相邻位干扰, 而机器学习技术正被用

于解析磁畴图案与材料参数之间的关联, 助力优化磁存储材料的性能与可靠性<sup>[35,36]</sup>。2023 年, Foggiaiatto 等<sup>[35]</sup>提出一种基于机器学习的逆问题方法, 通过 CNN 和 KRR 从磁畴图案中定量提取各向异性系数  $\alpha$ 。基于时间相关的 Ginzburg-Landau 方程 (TDGL) 生成不同  $\alpha$  值和扫描速度  $v$  的模拟磁畴图像。结果表明, CNN 在小图像尺寸 (如  $32 \times 32$  像素) 和跨域迁移 (不同  $v$  值训练数据) 任务中表现更优, 平均绝对误差 MAE 低于 0.077。而 KRR 揭示了磁畴形态与  $\alpha$ ,  $v$  之间的关系, 扫描速度  $v$  越大, 图案对  $\alpha$  的敏感性越高, 模型预测精度提升。此外, CNN 在处理复杂模式时表现出更强的鲁棒性, 而 KRR 的特征选择, 如区域周长与面积比为磁畴形成机制提供了直观解释。

本研究团队对于具有不同磁结构类型的稀土磁性材料进行数据清洗和整理, 采用 LR/DTR/SVM/GBRT 等十几种算法进行对比预测<sup>[37]</sup>。我们根据相成分和晶体结构等对数据集进行分类, 在小样本空间内构建多特征变量之间的协同关系, 所有模型的预测精度均在 0.73 以上。其中, 神经网络在预测材料是否具有斯格明子磁畴 (Skyrmion) 方面表现最好, 如图 16 所示, 准确率和置信度分别为

图 15 (a) 三步特征选择流程图; (b) 随机森林特征选择; (c) 穷举筛选<sup>[34]</sup>Fig. 15. (a) Flowchart of three-step feature selection; (b) feature selection by random forest; (c) exhaustive screening<sup>[34]</sup>.图 16 神经网络在预测材料是否具有 (a)–(c) 非线性磁结构和 (d)–(f) 斯格明子结构的表现<sup>[38]</sup>Fig. 16. Performance of Neural Networks in predicting whether a material has (a)–(c) nonlinear magnetic structure and (d)–(f) Skooping substructure<sup>[38]</sup>.

95% 和 97%，具有从小数据样本构建可靠机器学习模型的潜力。与非决策树模型相比，优化后的决策树算法如梯度回归 (gradient boosting) 和 XG Boost 在二分类中具有更大的优势。值得注意的是，稀土元素对于材料拓扑结构的影响较大，其含量与非线性或 Skyrmion 结构的产生呈负相关，这意味着

着稀土的少量添加有利于非平庸拓扑结构的形成<sup>[38]</sup>。利用该模型预测稀土氧化物、稀土铁氧体等多个体系中斯格明子存在的可能，并利用 MFM 观测到稀土氧化物薄膜  $\text{La}_{2/3}\text{Ba}_{1/3}\text{MnO}_3$  中的斯格明子畴，为实现稀土磁存储材料的科学设计提供可量化的依据。

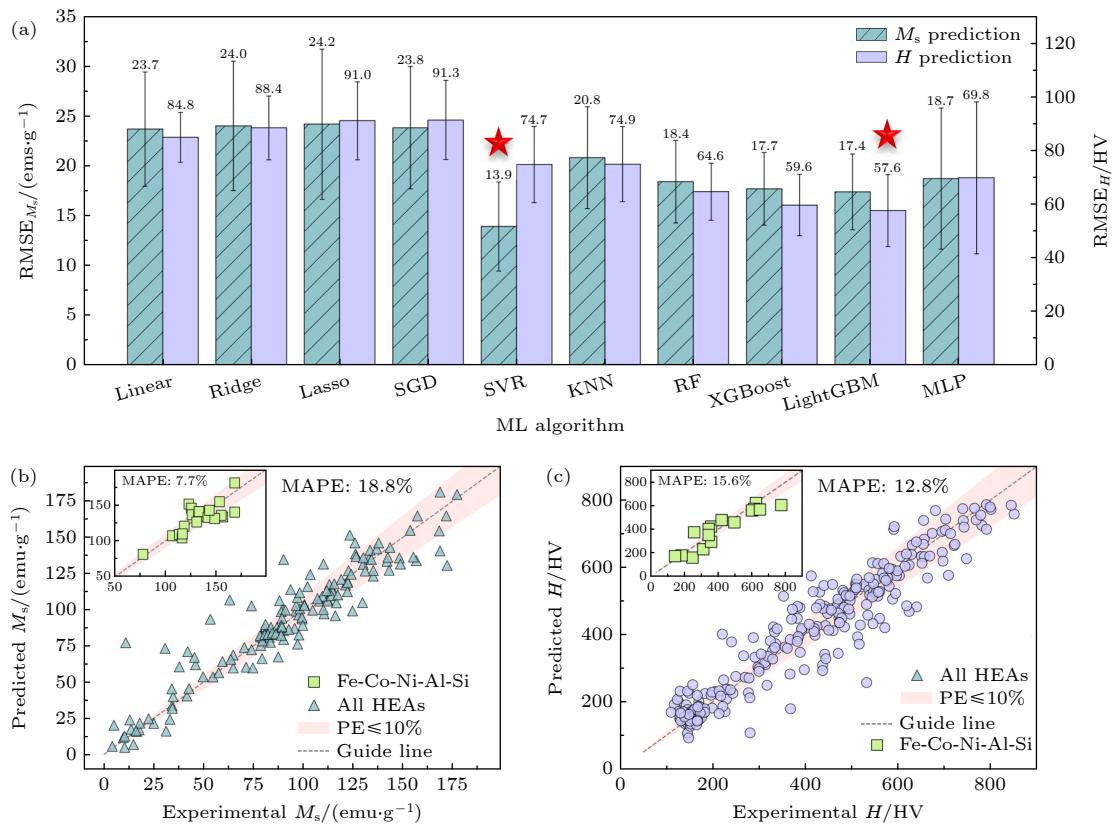
图 17 (a) 10 种机器学习算法对磁性合金  $M_s$  和硬度的均方根误差 RMSE; (b), (c)  $M_s$  和硬度  $H$  的预测值与实验值对比<sup>[39]</sup>

Fig. 17. (a) RMSE score of models based on ten ML algorithms; (b), (c) predicted values of  $M_s$  and hardness compared with the corresponding experimental values<sup>[39]</sup>.

软磁合金凭借高饱和磁通密度、高磁导率和低矫顽力, 广泛应用于磁存储、磁传感器、磁屏蔽等领域。与传统软磁合金相比, 铁基非晶/纳米晶合金和磁性高熵合金, 尤其是含稀土的软磁合金, 展现出更优的综合性能, 但化学组成复杂度更高。2024年, Li等<sup>[39]</sup>通过构建包含实验和理论数据的数据库, 结合特征工程(如原子半径、混合熵等参数)和多种机器学习算法(如RF, XG Boost), 实现了对磁性能(如饱和磁化强度)、相形成(非晶、固溶体等)及机械性能的高效预测, 如图17所示。研究发现, 机器学习模型在相预测任务中准确率超过85%, 并通过主动学习与贝叶斯优化策略, 显著减少了实验试错次数, 成功优化了材料的综合性能。多目标优化方法, 如NSGA-II进一步实现了磁性能与机械性能的协同优化。然而, 当前研究仍面临高质量数据集不足、模型可解释性差及全局优化效率等挑战。未来方向包括开发基于深度生成模型的反向设计、结合高通量实验与理论计算, 以及探索多物理场耦合下的复杂性能优化。

### 3 面临的挑战

首先, 在数据获取与数据质量方面, 稀土磁性材料研究涉及成分、工艺、微观结构和性能等多方面数据, 而且数据的来源也非常广泛, 涵盖实验测量、模拟计算等, 这就导致源数据的类型多样, 质量也参差不齐。比如, 存在数据缺失、噪声干扰、数据不一致等问题, 严重影响数据挖掘结果的准确性和可靠性。因此, 建立有效的数据清洗和预处理方法, 合理地将不同测量条件(如温度、磁场)下的物理量标准化, 从而提升数据质量, 是当前亟待解决的问题。此外, 数值型、图像型、文本型等不同数据源之间的数据格式、尺度和语义存在差异, 导致多源数据融合难度较大。开发通用的数据融合框架和技术, 实现相成分、晶体结构、微观形貌、磁性能等多源数据的高效整合与协同分析, 是推动数据挖掘技术在稀土磁性材料研究中应用的关键。

其次, 在特征工程与模型设计方面, 由于稀土材料的磁性由4f电子、自旋轨道耦合、晶体场效应

等多重物理机制共同决定,如何将晶胞参数、能带结构、内禀性质、制备工艺等信息有效编码为机器学习模型可识别的特征成为重要挑战。而且稀土功能材料的磁性能建模需要跨越原子尺度(自旋排列)、微观尺度(磁晶各向异性)及宏观尺度(磁滞回线)等多尺度耦合,同时还需要构建多目标优化框架以平衡磁能积、居里温度、资源成本等性能指标。然而,传统机器学习方法在小样本场景下易陷入过拟合陷阱,迫切需要开发基于元学习、贝叶斯优化或少样本学习的新型算法框架。

与此同时,在物理机制与模型解释方面,我们发现现有模型缺乏对磁交换作用、磁晶各向异性等理论的嵌入,导致预测结果偏离物理规律。而部分数据挖掘模型,如深度神经网络在稀土磁性材料研究中虽然具有良好的预测性能,但其模型结构复杂,内部决策过程难以理解,缺乏可解释性。这就导致研究人员难以从模型结果中获取直观的物理意义和材料设计指导,限制了模型在实际应用中的推广。尤其是稀土磁性材料研究专业性强,需要深入理解材料的物理化学原理和工艺过程。然而,目前部分从事数据挖掘工作的人员对材料专业知识掌握不足,导致数据挖掘模型的构建和应用与实际材料问题脱节。加强跨学科人才培养,促进材料专业人员与数据挖掘专家的合作,提高数据挖掘模型的可解释性,是推动数据挖掘技术在稀土磁性材料研究中深入应用的关键。

#### 4 总结与展望

数据挖掘技术在稀土磁性材料研究中的应用已取得显著进展,尤其在性能预测、成分与工艺优化、微观结构分析等方面展现出巨大优势,推动稀土磁性材料研究从“试错驱动”向“数据智能驱动”转型。然而,目前仍面临数据质量、多源数据融合、模型可解释性等方面的诸多挑战。随着计算机能力的提升和算法的不断迭代,机器学习将在稀土磁性材料研究中发挥更大作用,而实现稀土磁性材料的智能化设计与制造是未来发展的重要方向。未来突破路径可从三方面展开:1)全链条数据质量管控体系的构建。为了提高稀土磁性材料的数据质量,可以构建多维度数据清洗平台,集成相关物理约束与机器学习算法,比如,在深度学习模型中嵌入Landau-Lifshitz-Gilbert方程作为网络层约束,确保预测结果符合物理规律,实现实验数据、模拟数

据、文献数据的自动化清洗与填补;建立标准化的材料数据本体框架,实现测试条件、成分参数等元数据的统一标注;最后引入元学习模型从历史数据中学习校准规则,提升批次间数据的可比性。2)物理引导的多模态数据融合架构。在处理多源数据时,可以建立由物理思想引导的跨模态学习方法,通过对比学习将实验数据与计算数据的特征空间对齐;同时利用图神经网络挖掘显微结构特征与材料成分、制备工艺、性能指标间的关联规则;还可以采用贝叶斯神经网络等算法量化各数据源的置信度,并设置专家干预机制处理显著冲突。3)可解释智能模型的跨尺度构建。针对现有模型的局限性,研究人员应开发更具解释性的智能模型。比如在模型层和物理层分别建立可视化模型,实现从算法输出到物理机制的跨尺度解释;利用因果推理与机器学习相结合,排除混杂因子干扰,构建“因果可解释的预测模型”,提升工业应用中的决策可信度;基于数据挖掘技术、微磁学模拟、第一性原理、分子动力学模拟等分析方法,建立从原子尺度到宏观尺度的深度融合,进一步提高预测精度和分析效率,实现材料性能预测与微观机制解析的深度统一。

同时,稀土磁性材料研究将产生海量数据,构建大数据存储和云计算平台,实现数据的高效管理、共享和分布式计算,有助于充分挖掘数据价值。“十四五”原材料工业发展规划特别提出要加快AI,5G等技术在稀土行业的融合应用,推动智能工厂建设与产业集群发展。因此,本课题组与南京航空航天大学、浙江理工大学等单位合作,自主搭建了以稀土元素为主要特征变量的磁性材料数据平台,即磁海数据库(magdatahub.cn),致力于逐步构建一个全面、权威且易于使用的稀土磁性材料数据库,为研发人员提供便捷的数据查询与机器学习分析服务,加速新材料研发进程,提升我国在该领域的国际竞争力。此外,包头稀土高新区整合行业数据资源,建立了“稀土+AI”的创新平台。相信未来随着区域性产业联盟与国家级大数据中心的不断建设,数据共享机制将得到进一步强化和完善,进而推动稀土磁性功能材料研究向着规模化和标准化的方向持续迈进。

#### 参考文献

- [1] 中华人民共和国国务院 2024-06-29 稀土管理条例  
[https://www.gov.cn/zhengce/zhengceku/202406/content\\_6960153.htm](https://www.gov.cn/zhengce/zhengceku/202406/content_6960153.htm)
- [2] 中华人民共和国国务院 2023-12-26 中华人民共和国国民经济和

- 社会发展第四个五年规划和 2035 年远景目标纲要 <https://www.gov.cn/zhuanti/shisiwuguihua/sdb.htm>
- [3] Tai D Q, Li B, Xue H Y, Zheng T, Wu J G 2024 *Acta Mater.* **262** 119411
- [4] Swamynadhan M J, O'Hara A, Ghosh S, Pantelides S T 2024 *Adv. Funct. Mater.* **34** 2400195
- [5] Ding S L, Kang M G, Legrand W, Gambardella P 2024 *Phys. Rev. Lett.* **132** 236702
- [6] Deng X W, Wu L Y, Zhao R, Wang J O, Zhao L N 2024 *Acta Phys. Sin.* **73** 210701 (in Chinese) [邓祥文, 伍力源, 赵锐, 王嘉鹏, 赵丽娜 2024 物理学报 **73** 210701]
- [7] Möller J J, Körner W, Krugel G, Urban D F, Elsässer C 2018 *Acta Mater.* **153** 53
- [8] Butler K T, Davies D W, Cartwright H, Isayev O, Walsh A 2018 *Nature* **559** 547
- [9] Zhang Q, Tan W, Ning Y Q, Nie G Z, Cai M Q, Wang J N, Zhu H P, Zhao Y Q 2024 *Acta Phys. Sin.* **73** 230201 (in Chinese) [张桥, 谭薇, 宁勇祺, 聂国政, 蔡孟秋, 王俊年, 朱慧平, 赵宇清 2024 物理学报 **73** 230201]
- [10] Wang Z, Sun Z H, Yin H, Liu X H, Wang J L, Zhao H T, Pang C H, Wu T, Li S Z, Yin Z Y, Yu X F 2022 *Adv. Mater.* **34** 2104113
- [11] Peng X K, Ji J W, Li L, Ren W, Xiang J F, Liu K K, Cheng H N, Zhang Z, Qu Q Z, Li T, Liu L, Lü D S 2019 *Acta Phys. Sin.* **68** 130701 (in Chinese) [彭向凯, 吉经纬, 李琳, 任伟, 项静峰, 刘亢亢, 程鹤楠, 张镇, 屈求智, 李唐, 刘亮, 吕德胜 2019 物理学报 **68** 130701]
- [12] Li R 2018 *Ph. D. Dissertation* (Beijing: University of Chinese Academy of Sciences) (in Chinese) [李锐 2018 博士学位论文 (北京: 中国科学院大学)]
- [13] Kovacs A, Fischbacher J, Oezelt H, Kornell A, Ali Q, Gusenbauer M, Yano M, Sakuma N, Kinoshita A, Shoji T, Kato A, Hong Y, Grenier S, Devillers T, Dempsey N M, Fukushima T, Akai H, Kawashima N, Miyake T, Schrefl T 2023 *Front. Mater.* **9** 1094055
- [14] Hosokawa H, Calvert E L, Shimojima K 2021 *J. Magn. Magn. Mater.* **526** 167651
- [15] Guo K, Lu H, Zhao Z, Tang F W, Wang H B, Song X Y 2022 *Comp. Mater. Sci.* **205** 111232
- [16] Wen J T, Hu H G, An J S, Han T, Hu J F 2024 *J. Supercond. Nov. Magn.* **37** 1443
- [17] Liu R S, Wang L C, Xu Z Y, Qin C Y, Li Z Y, Yu X, Liu D, Gong H Y, Zhao T Y, Sun J R, Hu F X, Shen B G 2022 *Mater. Today Commun.* **32** 103996
- [18] Nguyen D N, Pham T L, Nguyen V C, Kino H, Miyake T, Dam H C 2019 *J. Phys. Mater.* **2** 034009
- [19] Xu G J, Cheng F, Lu H, Hou C, Song X Y 2024 *Acta Mater.* **274** 120026
- [20] Halder A, Rom S, Ghosh A, Dasgupta T S 2020 *Phys. Rev. Appl.* **14** 034024
- [21] Lambard G, Sasakib T T, Sodeyama K, Ohkubo T, Hono K 2022 *Scripta Mater.* **209** 114341
- [22] Lu S F, Liu Y L, Yin Q S, Chen J F, Wu J, Li J, Zhan P J, Chen Z C 2024 *J. Eur. Ceram. Soc.* **44** 5677
- [23] Choudhary A K, Grubesa T, Jansche A, Bernthaler T, Goll D, Schneider G 2024 *Acta Mater.* **264** 119563
- [24] Kusne A G, Gao T, Mehta A, Ke L Q, Nguyen M C, Ho K M, Antropov V, Wang C Z, Kramer M J, Long C, Takeuchi I 2014 *Sci. Rep.* **4** 63
- [25] Nieves P, Arapan S, Hadjipanayis G C, Niarchos D, Barandiaran J M, Cuesta-López S 2016 *Phys. Status Solidi C* **13** 942
- [26] Xia W Y, Sakurai M, Balasubramanian B, Liao T, Wang R H, Zhang C, Sun H J, Ho K M, Chelikowsky J R, Sellmyer D J, Wang C Z 2022 *Proc. Natl. Acad. Sci.* **119** 2204485119
- [27] Zhang B, Zheng X Q, Zhao T Y, Hu F X, Sun J R, Shen B G 2018 *Chin. Phys. B* **27** 067503
- [28] Zhang B 2018 *Ph. D. Dissertation* (Beijing: University of Chinese Academy of Sciences) (in Chinese) [张博 2018 博士学位论文 (北京: 中国科学院大学)]
- [29] Ucar H, Paudyal D, Choudhary K 2022 *Comp. Mater. Sci.* **209** 111414
- [30] Alqahtani A 2024 *Sustainability* **16** 1542
- [31] Shamsah S M I 2024 *Int. J. Refrig.* **168** 122
- [32] Zhao Q, Yan K L, Cui Z, Wen B Y, Xue F, Li J T, Guo J N, Xu A, Qiao K M, Ye R C, Long Y, Zhang D W, Luo H, Taskaev S, Zhang H 2023 *Corros. Sci.* **216** 111115
- [33] Gong J H, Zhang Z M, Zhang C L, Hu P Q, Zhou C, Wang D H, Yang S 2024 *Rare Met.* **43** 2251
- [34] Hu P Q, Zhou C, Zhang R S, Ding S D, Guo Y J, Wang B, Xue D Z, Ma Y Z, Dai Z Y, Zhang Y, Tian F H, Yang S 2025 *Mater. Design* **252** 113799
- [35] Foggiatto A L, Mizutori Y, Yamazaki T, Sato S, Masuzawa K, Nagaoka R, Taniwaki M, Fujieda S, Suzuki S, Ishiyama K, Fukuda T, Igarashi Y, Mitsumata C, Kotsugi M 2023 *IEEE T. Magn.* **59** 2501604
- [36] Li W, Long L C, Liu J Y, Yang Y 2022 *Acta Phys. Sin.* **71** 060202 (in Chinese) [黎威, 龙连春, 刘静毅, 杨洋 2022 物理学报 **71** 060202]
- [37] Liu D, Liu Z X, Zhang J E, Yin Y N, Xi J F, Wang L C, Xiong J F, Zhang M, Zhao T Y, Jin J Y, Sun J R, Hu F X, Shen J, Shen B G 2023 *Research* **6** 0082
- [38] Liu D, Song J H, Liu Z X, Zhang J E, Chen W Q, Yin Y N, Xi J F, Zheng X Q, Hao J Z, Zhao T Y, Hu F X, Sun J R, Shen B G 2025 *Mater. Design* **251** 113710
- [39] Li X, Shek C H, Liaw P K, Shan G C 2024 *Prog. Mater. Sci.* **146** 101332

REVIEW

# Research progress of rare earth magnetic materials based on machine learning\*

LIU Dan<sup>1)†</sup> LI Yuan<sup>1)</sup> SUN Ruoxuan<sup>1)</sup> QI Xing<sup>1)</sup> SHEN Baogen<sup>2)‡</sup>

1) (*Department of Physics, School of Computer and Artificial Intelligence,*

*Beijing Technology and Business University, Beijing 100048, China*)

2) (*National Laboratory for Magnetism, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*)

(Received 3 April 2025; revised manuscript received 15 April 2025)

## Abstract

Rare-earth elements possess unique atomic structures characterized by multiple unpaired 4f orbital electrons in inner shells, high atomic magnetic moments, and strong spin-orbit coupling. These attributes endow them with rich electronic energy levels, enabling them to form compounds with different valence states and coordination environments. Consequently, rare-earth materials typically exhibit excellent magnetic properties and complex magnetic domain structures, making them critical for the development of high-tech industries. The intricate magnetic configurations, different types of magnetic coupling, and direct/indirect magnetic exchange interactions in these materials not only facilitate the development of novel functional devices but also pose significant challenges to fundamental research. With the rapid advancement of data mining techniques, the emergence of big data and artificial intelligence provides researchers with a new method to efficiently analyze vast experimental and computational datasets, thereby accelerating the exploration and development of rare-earth magnetic materials. This work focuses on rare-earth permanent magnetic materials, rare-earth magnetocaloric materials, and rare-earth magnetostrictive materials, detailing the application progress of data mining techniques in property prediction, composition and process optimization, and microstructural analysis. This work also delves into the current challenges and future trends, aiming to provide a theoretical foundation for deepening the integration of data mining technologies with rare-earth magnetic material research.

**Keywords:** data mining, rare earth magnetic material, machine learning, performance prediction

**PACS:** 71.20.Eh, 75.50.-y, 88.30.gg

**DOI:** [10.7498/aps.74.20250431](https://doi.org/10.7498/aps.74.20250431)

**CSTR:** [32037.14.aps.74.20250431](https://cstr.aphys.ac.cn/32037.14.aps.74.20250431)

\* Project supported by the National Key Research and Development Program of China (Grant No. 2023YFA1610400) and the National Natural Science Foundation of China (Grant Nos. 52088101, 52371169).

† Corresponding author. E-mail: [liandan@btbu.edu.cn](mailto:liandan@btbu.edu.cn)

‡ Corresponding author. E-mail: [shenbg@iphy.ac.cn](mailto:shenbg@iphy.ac.cn)



## 基于数据挖掘技术的稀土磁性材料研究进展

刘丹 李渊 孙若瑄 漆星 沈保根

### Research progress of rare earth magnetic materials based on machine learning

LIU Dan LI Yuan SUN Ruoxuan QI Xing SHEN Baogen

引用信息 Citation: [Acta Physica Sinica](#), 74, 137102 (2025) DOI: 10.7498/aps.74.20250431

CSTR: 32037.14.aps.74.20250431

在线阅读 View online: <https://doi.org/10.7498/aps.74.20250431>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于机器学习的无机磁性材料磁性基态分类与磁矩预测

Classification of magnetic ground states and prediction of magnetic moments of inorganic magnetic materials based on machine learning

物理学报. 2022, 71(6): 060202 <https://doi.org/10.7498/aps.71.20211625>

#### 基于机器学习和第一性原理计算的Janus材料预测

Prediction of magnetic Janus materials based on machine learning and first-principles calculations

物理学报. 2024, 73(23): 230201 <https://doi.org/10.7498/aps.73.20241278>

#### 基于机器学习的菱形穿孔石墨烯负泊松比效应预测与优化

Prediction and optimization of negative Poisson's ratio in rhombic perforated graphene based on machine learning

物理学报. 2025, 74(9): 096201 <https://doi.org/10.7498/aps.74.20241624>

#### 蛋白质计算中的机器学习

Machine learning for *in silico* protein research

物理学报. 2024, 73(6): 069301 <https://doi.org/10.7498/aps.73.20231618>

#### 机器学习辅助的WC-Co硬质合金硬度预测

Hardness prediction of WC-Co cemented carbide based on machine learning model

物理学报. 2024, 73(12): 126201 <https://doi.org/10.7498/aps.73.20240284>

#### 基于波动与扩散物理系统的机器学习

Machine learning based on wave and diffusion physical systems

物理学报. 2021, 70(14): 144204 <https://doi.org/10.7498/aps.70.20210879>