

基于机器学习的铝铜合金硬度与导热性能预测及联合优化 *

步彤彤^{1) #} 李霞^{1) 2) #} 娄春香¹⁾ 任平静¹⁾ 丁守军^{1) 2)} 邹勇^{1) 2) †}
熊纯明³⁾

1) (安徽工业大学, 微电子与数据科学学院, 马鞍山 243032)

2) (安徽省工业大数据分析 with 智能决策重点实验室, 马鞍山 243002)

3) (中电泰日升马鞍山科技有限公司, 马鞍山 243000)

针对铝铜合金性能预测中小样本建模困难、传统设计难以协同优化硬度与导热性能的问题, 构建物理衍生特征与多级降维筛选相结合的机器学习框架, 实现合金性能预测与成分优化. 以相同工艺制备的 24 组铝铜合金为研究对象, 结合元素理化特性构建物理衍生特征空间, 通过层次聚类、正则化回归与全子集搜索相结合的多级特征筛选方法提取核心特征, 分别建立硬度与导热率的最小绝对收缩和选择算子预测模型. 采用全局拉丁超立方体采样与局部高斯精修相结合的两阶段策略, 基于综合评分函数开展硬度与导热性能的协同优化, 在可行成分域内生成候选样本并筛选高性能种子成分, 通过高斯重采样与成分归一化定位最优成分. 最终获得四组综合性能优异的合金成分, 实现硬度与导热性能的均衡提升. 该框架可为小样本条件下高性能铝铜合金的双性能协同优化设计提供有效技术途径.

关键词: 机器学习, 铝铜合金, 特征筛选, 性能协同优化

PACS: 81.05.Bx, 82.20.Wt, 62.20.Qp, 65.40.-b

1 引言

铝合金比强度高、结构稳定性优异, 且成型加工性能良好, 同时具备低密度与高耐蚀性等特征, 在航空航天、汽车制造及各类工程领域得到广泛应用^[1-3]. 在铝合金体系中, 含镁、锌合金经时效处理可析出 η' -MgZn₂ 相, 进而获得更优异的力学性能^[4]; 铜元素能够通过形成相干析出相对铝合金起到强化作用^[5], 向铝合金中添加 0.1~0.7 wt.% 的铁元素, 可推动富铁相向 α -Fe 相转变, 有效改善铝合金的高温屈服性能^[6]. 对于高强铝合金而言, 钪元素可降低界面能并促进 Al₃Sc 相优先析出以细化晶粒, 并通过与空位强

同等贡献作者.

† 通信作者. E-mail: yongzou@ahut.edu.cn

第一作者. E-mail: 3471003223@qq.com(步彤彤); lixiaahut@ahut.edu.cn(李霞)

* 国家自然科学基金 (批准号: 52202001) 和安徽省教育厅优秀青年学者自然科学基金项目 (批准号: 2024AH030007) 资助的课题.

相互作用抑制再结晶、收窄无沉淀带，协同提升合金强韧性与耐蚀性^[7]。这些独特的元素组合使得铝合金在保持良好加工性能的同时，具备了优异的力学和热学性能。然而，传统的合金设计方法主要依赖经验和试错，这种方式不仅耗时费力，而且难以在庞大的成分空间中找到最优配比^[8-15]。

近年来，随着人工智能技术的快速发展，机器学习在材料科学领域的应用不断深入，为铝合金的研究带来了革命性的变化。机器学习能够从大量的实验数据中提取隐藏的规律，建立成分-工艺-性能之间的复杂映射关系，实现合金性能的准确预测和成分智能优化^[16-19]。Lan^[20]等基于 A380 铝合金数据集，构建并对比反向传播-人工神经网络与粒子群优化的反向传播-人工神经网络模型，结果证实后者预测精度更高，明确了关键输入参数分别对极限抗拉强度与伸长率的重要性排序。Lu^[21]等基于合金成分物理描述符，采用套索算法和基尼不纯度算法进行特征工程，构建极端梯度提升算法和支持向量机两种预测模型，实现了对铝合金导热性与极限拉伸强度的精准预测，成功开发出兼具高导热与高强度的合金材料。Bharath^[22]等建立人工神经网络与线性回归模型，实现了对 Al-Cu/B₄C 复合材料硬度的准确预测。Babaheydari^[23]等采用人工神经网络模型预测铝复合材料显微硬度，平均误差仅 5.6%，并通过敏感性分析明确合金元素体积分数、球磨速度和烧结时间为主要影响因素。Hatice^[24]等采用多种机器学习模型预测不同工艺、温度与时间下铝合金的硬度和弯曲强度，结果表明卷积神经网络模型预测精度最高。宋睿^[25]等采用梯度提升决策树模型，实现了 WC-Co 硬质合金硬度的高效预测与关键影响因子识别。相关研究表明，多种机器学习算法均能有效提升铝合金力学和热学性能预测精度，为实验设计提供可靠指导^[26,27]。

综上，采用机器学习实现铝合金性能精准预测与高效反向设计具有重要意义。但现有研究仍面临实验数据稀缺导致模型易过拟合、力学与热学性能存在固有负相关而难以协同优化、数据驱动模型缺乏物理可解释性等问题。为此，本文提出“物理衍生特征-多级降维筛选”框架，基于热力学计算与物理原理构建具有明确物理意义的特征集，并通过多级降维筛选提取核心特征，从物理层面约束模型并缓解小样本过拟合；分别建立铝铜合金硬度与导热性能的最小绝对值收缩和选择算子（Least absolute shrinkage and selection operator, Lasso）预测模型，以综合评分函数最大化为目标，采用“全局均匀探索-局部高斯精修”两阶段搜索策略实现双性能协同优化，以达到硬度与导热的均衡提升。该数据驱动框架为小样本条件下高性能铝铜系合金的设计提供方法与理论支撑。

2 实验方法

为解决多组元铝铜合金性能预测中“小样本”的核心建模难题，提出了“物理衍生特征-多级降维筛选”数据驱动合金设计路线（如图 1 所示），该路线以物理启发描述符为核心驱动，围绕特征构建、多级降维、模型评估、性能协同优化四个关键阶段展开。在特征构建阶段，通过成分加权策略提取元素级物化属性的均值与方差，将成分信息映射为合金层面的 99 维宏观描述符，实现物理背景知识的有效特征化。在多级降维阶段，针对特征间的共线性与冗余性，执行由“层次聚类初筛-Lasso 正则化精筛-全子集穷举优化”组成的递进式筛选框架，从空间中有效识别并提取具有明确物理意义的核心描述符子集。在建模与优化阶段，基于筛选特征结合留一交叉验证（Leave-One-Out Cross Validation, LOOCV）构建稳健的 Lasso 回归模型，实现对多元铝铜合金硬度与导热性能的表征，最后依托训练完成的机器学习模型，以代理模型预测的硬度与导热性能为基础构建综合评价指标函数，利用基于拉丁超立方采样^[28]（Latin Hypercube Sampling, LHS）的全局探索与基于高斯采样的局部精修相结合的两阶段搜索算法，在物理边界与总量归

一化约束下，筛选出综合性能优异的 Top20 候选合金成分。

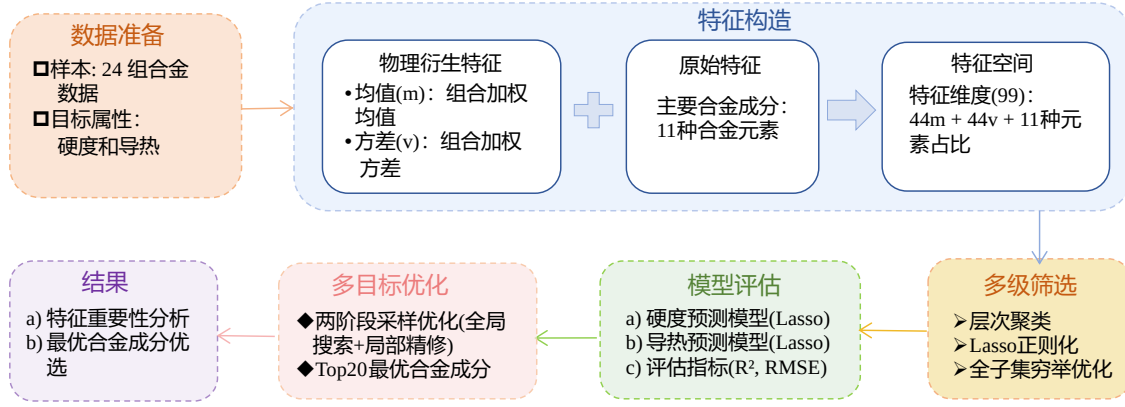


图 1 基于数据驱动的高性能铝铜合金性能预测技术路线

Fig. 1. Data-driven approach for predicting the properties of high-performance aluminum-copper alloys.

2.1 数据收集

高质量且具备一致性的数据是构建高精度机器学习模型的前提. 从文献 [29] 中收集了 24 组铝铜合金样本, 包括它的成分、硬度、导热性能以及合金中涉及的 11 种元素的 44 个物理化学特性 (如电负性、压缩模量、原子半径等). 为了消除加工工艺参数波动对性能预测产生的干扰, 确保机器学习模型能够专注于学习材料特征与性能之间的映射关系, 所有选用的合金样本均严格基于同一压铸工艺制备, 为后续基于成分加权策略构建均值、方差类物理衍生特征, 以及通过层次聚类、Lasso 正则化、穷举搜索实现多级特征降维提供了可靠的数据支撑.

2.2 特征创造

为建立合金成分与宏观性能之间的量化联系, 基于元素的物理化学属性构建了合金层面的描述符. 选取 44 项与力学及热学性能密切相关的元素物理化学属性作为底层数据 (见附录表 A1), 并统一编号为 $Ak (k = 1, 2, \dots, 44)$. 记 f_{ij} 为第 $j (j = 1, 2, 3, \dots, 11)$ 种组成元素的第 $i (i = 1, 2, 3, \dots, 44)$ 个物理化学特征值, α_j 表示第 j 种元素在合金中的质量分数 (wt.%), 数值直接从公开文献的已测合金成分数据中获得, 满足质量守恒 $\sum_{j=1}^n \alpha_j = 100\%$. 依据方程 (1) 和方程 (2) 计算了 44 个物理化学特征的成分加权均值 f_{mi} 与成分加权方差 f_{vi} :

$$f_{mi} = \sum_{j=1}^n \alpha_j f_{ij}, \quad (1)$$

$$f_{vi} = \sum_{j=1}^n \alpha_j (f_{ij} - f_{mi})^2 \quad (2)$$

其中, f_{mi} 作为各元素特征在合金中的加权平均水平, 反映了系统的宏观平均物理背景; 而 f_{vi} 则表示元素特征值偏离均值的加权平方和, 用于量化系统内部的物理化学波动程度. 依据上述计算结果, 将 44 项属性演化出的衍生特征统一命名为 Akm (均值) 与 Akv (方差) ($k = 1, 2, \dots, 44$).

通过上述的变换, 原始的 44 个基础描述符被扩展为 88 个具有明确物理意义的衍生特征. 随后, 将这些衍生特征与 11 种原始成分占比特征 α_j 进行级联, 最终构建了总维度为 99 的原始特征空间. 该方法在完整保留合金成分信息的同时, 通过均值与方差的数学表达, 深度耦合了反映固溶强化机理与能量输运特性的物理先验知识. 这不仅增强了特征空间的物理可解释性, 也为后续识别影响硬度与导热性能的关键驱动因子奠定了数据基础.

2.3 特征筛选

面对 99 维特征空间与 24 组样本之间的维度失衡问题, 直接建模极易导致严重过拟合. 为此, 设计由层次聚类初筛与 Lasso 正则化精筛、全子集穷举搜索构成的多级筛选方案, 旨在消除特征间的严重共线性并锁定最具物理意义的合金因子. 在该流程中, 首先利用能够有效捕捉非线性关系的 Spearman 秩相关系数衡量特征关联度, 然后根据相关系数计算特征间的距离矩阵并执行层次聚类. 从每个簇中精选出一个与目标性能相关性最强的特征作为该簇的代表, 剔除其余冗余特征. 定义 Spearman 秩相关系数 ρ 与特征间距离 D 分别如式 (3) 和 (4):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

$$D(X, Y) = 1 - \rho(X, Y) \quad (4)$$

其中, d_i 为两变量秩次之差, n 为样本容量.

为了在层次聚类预筛选基础上进一步识别关键特征, 引入具有嵌入式特征选择功能的 Lasso 回归算法. 该算法在普通最小二乘法的基础上引入了正则化惩罚项, 能够对模型系数进行压缩, 其目标函数定义为:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (5)$$

其中, 第一部分为衡量模型精度的残差平方和, 第二部分为受调节参数 λ 控制的惩罚项, λ 为调节正则化强度的惩罚因子. 在训练过程中, 对目标性能预测贡献较小的特征系数将被强制收缩至零, 从而实现特征的自动稀疏化与冗余剔除. 为确保筛选出的特征具有最优的预测鲁棒性, 利用 LOOCV 对 Lasso 惩罚参数 λ 进行寻优. 该过程以均方误差 (Mean Squared Error, MSE) 为核心优化目标, 通过遍历候选参数空间, 选取使 LOOCV 预测误差最小的最优惩罚参数, 实现特征空间的稀疏化与预测性能的平衡.

Lasso 回归的 L_1 的正则化虽能实现嵌入式特征选择, 但其系数压缩过程对共线性特征具有随机选择特性, 难以保证筛选出的特征子集为全局最优的最小冗余组合. 为解决这一问题, 在 Lasso 精筛的基础上引入全子集穷举搜索策略, 该策略对 Lasso 精筛后的 N 个特征进行组合, 遍历从 $k = 1$ 到 $k = N$ 维的所有特征子集 (共 $2^N - 1$ 种组合). 针对每个候选组合, 采用 LOOCV 评估其泛化性能, 以根均方误差 (Root Mean Squared Error, RMSE) 为目标函数, 最终选取 RMSE 最小的特征组合作为最优输入集, 从而在遵循“奥卡姆剃刀”原则的前提下, 实现模型预测精度与复杂度的最优平衡.

2.4 模型建立

原始数据集采用随机抽样方法, 将 24 组合金样本按 4:1 的比例划分为训练集与测试集. 其中, 19 组训练集样本用于特征筛选、模型训练及超参数寻优; 5 组独立测试集样本仅用于评估模型对未知数据的预测性能, 确保模型评估结果具备客观性与可靠性. 结合小样本数据集的建模特性, Lasso 算法凭借其 L_1 正则化特性, 能够有效压缩冗余特征权重、规避模型过拟合风险, 同时兼顾模型泛化性能与计算效率. 基于筛选所得特征子集构建 Lasso 回归预测模型, 用于完成铝铜合金硬度与导热性能的预测, 选取 RMSE 与拟合优度 (Coefficient of determination, R^2) 作为核心评价指标, 实现模型预测效果的定量评估.

3 结果与讨论

3.1 关键特征选取

层次聚类将特征维度压缩至 26 维, 为直观验证初筛效果, 选取 26 维特征空间中与硬度、导热性能相关系数排名前 10 的特征, 基于 Spearman 秩相关系数绘制相关性热力图 (图 2), 以展示核心特征间的关联程度. 图 2 结果显示前 10 个高相关性核心特征间仅存在少量低程度关联, 证实了该步骤有效消除了特征共线性, 在保留核心物理信息的前提下, 显著降低了数据的冗余度和模型训练的计算复杂度.

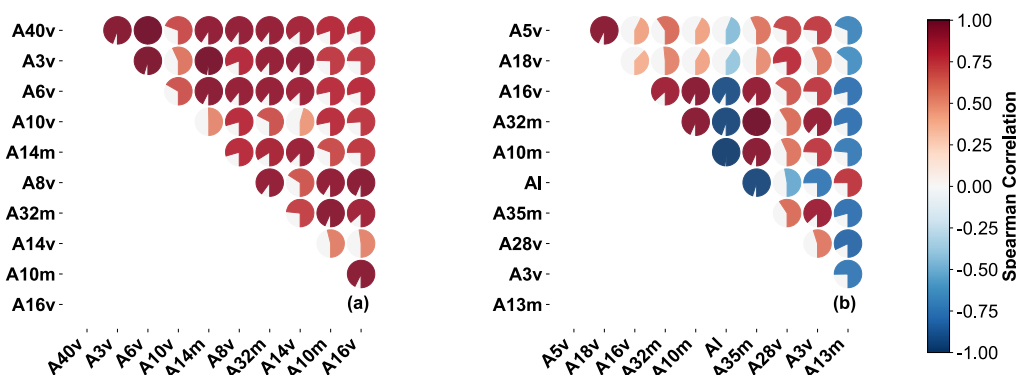


图 2 硬度与导热模型输入特征的相关性矩阵 (a) 硬度特征相关性分析热力图; (b) 导热特征相关性分析热力图

Fig. 2. Correlation matrix between hardness and thermal conductivity model inputs (a) Heatmap of hardness feature correlation analysis; (b) Heatmap of thermal conductivity feature correlation analysis.

在层次聚类预筛选的基础上, Lasso 算法利用 L_1 惩罚项对候选特征空间执行了目标导向的稀疏压缩, 分别为硬度与导热模型确定了差异化的关键物理驱动因子. 针对硬度预测, 最优惩罚参数 $\lambda=1.456$, 最终筛选出 6 个关键特征 (['A8m', 'A8v', 'A10v', 'A14v', 'A40v', 'A13m']); 针对导热性能预测, 最优惩罚参数为 0.804, 最终保留 7 个关键特征 (['A16v', 'A20m', 'A10v', 'Ni', 'Ti', 'Sr', 'A5v']). 由于层次聚类与 Lasso 的筛选评价标准不同, 部分 Lasso 关键特征未出现在层次聚类的前 10 个特征中, 这也体现了无监督预筛选与有监督精筛的互补优势.

穷举搜索的性能演化趋势如图 3 所示, 横坐标为按 RMSE 升序排列的特征组合索引, 纵坐标为对应的 LOOCV-RMSE. 在硬度性能的特征寻优中, Lasso 精筛剩余 6 个候选特征, 全空间共 63 种组合, 最

终锁定黄金组合为 [‘A14v’, ‘A10v’], 其最优 LOOCV-RMSE 为 4.0790; 在导热性能的特征寻优中, Lasso 精筛剩余 7 个候选特征, 全空间共 127 种组合, 最终锁定最优组合为 [‘A20m’, ‘A5v’, ‘Ni’, ‘Ti’], 其对应的 LOOCV-RMSE 为 3.6527. 通过全子集遍历, 该方法能够弥补 Lasso 算法在处理共线性描述符时可能存在的筛选偏差, 通过消除特征间的冗余干扰, 锁定预测性能最稳健的全局最优子集, 从而为模型输入特征的筛选提供更为精细的优化路径.

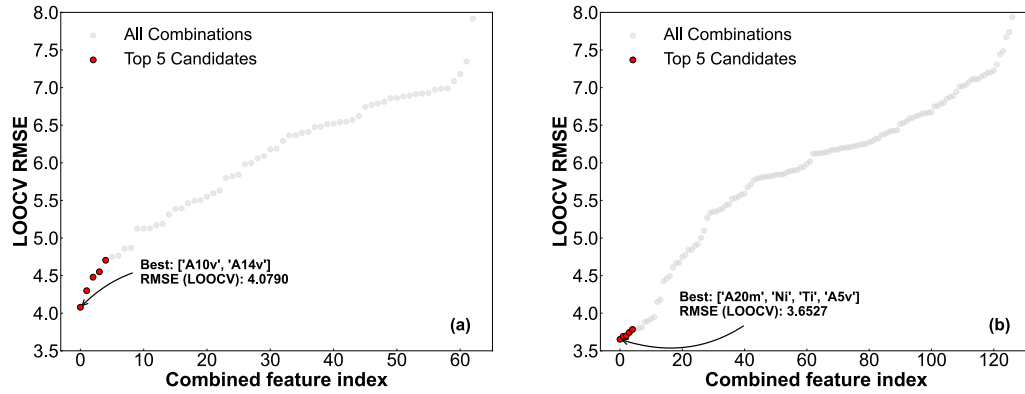


图 3 穷举搜索下的模型性能演化轨迹 (a) 硬度穷举搜索; (b) 导热穷举搜索

Fig. 3. Evolution of model performance under exhaustive search: (a) hardness exhaustive search; (b) thermal conductivity exhaustive search.

3.2 模型评估

基于物理衍生特征与多级筛选结果, 利用穷举寻优后的特征子集作为输入, 构建了铝铜合金硬度与导热性能的预测模型. 结合研究仅 24 组样本的小样本数据特征, 复杂非线性模型易引发过拟合且缺乏物理可解释性, 而 Lasso 回归凭借正则化特性, 既能有效平衡模型复杂度与泛化误差, 又能通过系数稀疏化保留特征的物理意义, 因此该实验选取 Lasso 回归作为预测模型, 以实现合金性能的高效预测. 此外, 为了定量验证本文“物理衍生特征 + 多级筛选”框架在克服小样本过拟合方面的独特优势, 引入了一个基线模型作为对照. 该基线模型在相同数据划分和建模流程下, 直接将完整的 99 维衍生特征作为模型输入, 两组模型的性能对比结果如表 1 所示.

表 1 多级特征筛选模型与全特征基线模型预测性能对比

Table 1. Comparison of the predictive performance between the multi-stage feature selection model and the full-feature baseline model

建模方案	预测目标	训练集 R^2	测试集 R^2	测试集 RMSE
实验组	硬度	0.891	0.885	4.79
实验组	导热	0.850	0.847	4.67
基线组	硬度	0.984	0.479	10.19
基线组	导热	0.963	0.773	5.70

针对硬度性能, 基于三阶段特征筛选确定的最优描述子集 [‘A14v’, ‘A10v’] 进行建模. 表 1 显示,

该模型在硬度预测任务中表现出较高的拟合精度与鲁棒性，测试集决定系数达到 0.885，RMSE 为 4.79。由图 4(a) 可知，模型预测值与实际值整体吻合良好，数据点紧密分布于对角线附近；训练集与测试集性能接近，表明模型在保持简洁性的同时，有效避免了过拟合，具有良好的泛化能力。相比之下，直接输入完整 99 维特征的基线模型则表现出严重的过拟合倾向（图 5(a)）：虽然其训练集 R^2 高达 0.984，但测试集 R^2 骤降至 0.479，且测试集 RMSE 增大至 10.19，数据点在测试集上表现出显著的离散误差。这一对比结果表明，在当前铝铜合金体系下，直接引入过多特征容易伴随冗余噪声干扰，而采用的特征筛选策略在缓解模型过拟合、保障预测稳健性方面展现出较好的适用性。

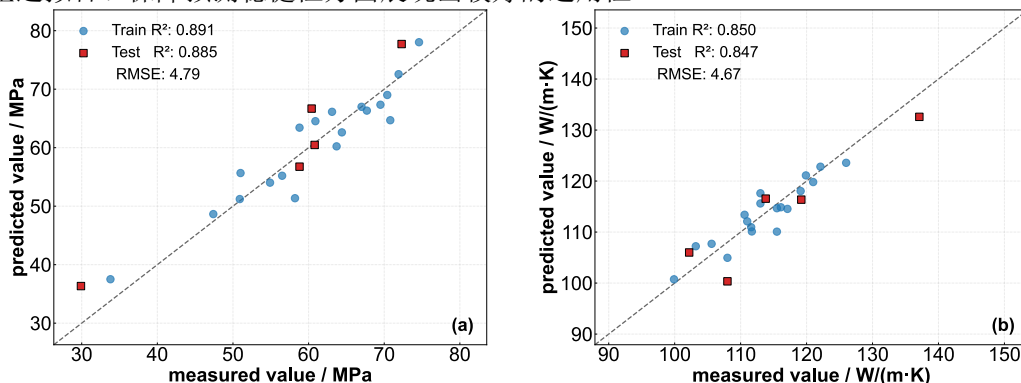


图 4 硬度 (a) 与导热 (b) 性能的 Lasso 回归预测性能评估

Fig. 4. Evaluation of the predictive performance of Lasso regression for hardness (a) and thermal conductivity (b).

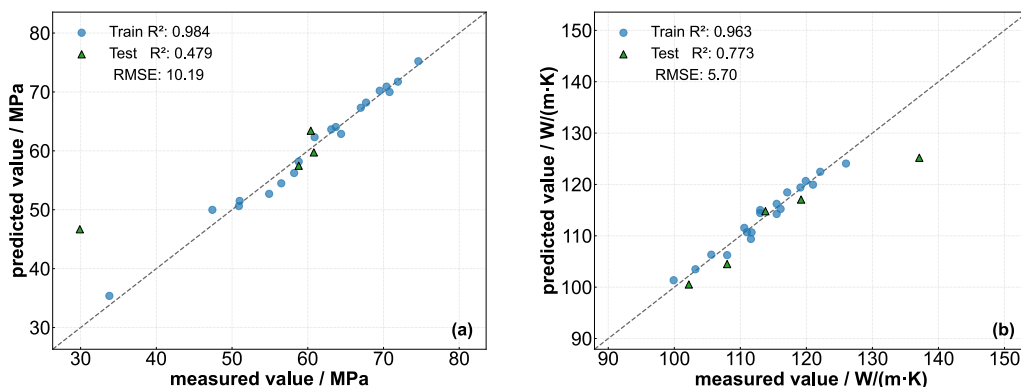


图 5 采用完整高维特征建模（基线组）的硬度 (a) 与导热 (b) 性能预测评估

Fig. 5. Prediction evaluation of hardness (a) and thermal conductivity (b) modeled with full high-dimensional features (baseline group).

对于导热性能，以最优特征子集 [‘A20m’ , ‘A5v’ , ‘Ni’ , ‘Ti’] 构建模型并展开分析。结果表明，尽管导热性能受限于电子与声子输运机制的复杂性，模型在独立测试集上的 R^2 为 0.847，RMSE 为 4.67，显示出较高的预测精度与鲁棒性。由图 4 (b) 可以看出，受样本量与数据分布影响，导热性能的测试集误差略高于硬度，但模型仍能有效捕捉其整体变化趋势。而直接采用完整特征建模的基线模型（图 5(b)），其测试集误差升至 5.70，训练集 $R^2 = 0.963$ 明显高于测试集 $R^2 = 0.773$ ，反映出模型存在过拟合现象，这表明在当前数据规模下，进行特征降维有助于模型获得更具稳健性的泛化能力。

4 合金成分探索空间构建与性能协同优化

为实现铝铜合金硬度与导热性能的联合优化,在完成高精度性能预测模型构建的基础上,设计了基于“全局空间探索 + 局部高斯精修”的两阶段成分优化策略. 该策略首先以实验样本中 11 种元素的含量极值为边界, 确立了合金设计的可行域. 这一设定既保证了探索空间的工程合理性, 又确保成分点严格落在机器学习预测模型的训练区间内, 有效规避了因“成分外推”导致的模型预测失真. 在成分空间探索阶段, 将每种元素的取值范围均匀划分为 10^4 个等宽子区间. 依据拉丁超立方采样准则, 每个元素的各子区间内仅抽取单个采样点, 保证单一维度内采样均匀分布. 最终构建出 10^4 组具备高空间覆盖率的候选合金成分样本. 相较于传统随机采样, 该方法实现了多维特征空间的均匀覆盖, 有效避免了因样本聚集导致的最优区域遗漏. 利用预先构建好的 Lasso 模型对上述候选成分进行硬度与导热性能的预测, 并引入多目标综合评分机制进行性能排序. 为消除不同物理量纲的影响, 先对两项预测的性能指标进行最小-最大归一化处理, 计算公式为:

$$y_{\text{norm}} = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (6)$$

其中 y 为模型输出的性能预测值, $y_{\max}$ 和 $y_{\min}$ 分别为对应性能指标在全局候选样本中的最大值与最小值. 在此基础上, 遵循硬度与导热性能协同优化原则, 构建幂次加权乘积形式的综合评分函数 Score (式 7), 依据综合评分由高到低排序, 从全局候选成分中筛选出综合性能占优的 Top20 种子成分.

$$\text{Score} = y_{\text{norm,hard}}^{w_1} \cdot y_{\text{norm,tc}}^{w_2} \quad (7)$$

式中, 权重系数满足约束条件 $w_1 > 0$ 、 $w_2 > 0$ 且 $w_1 + w_2 = 1$, 根据实际需要设置合适的权重系数. 该幂次加权形式对硬度或导热性能单项指标的退化都会引发总评分发生断崖式下跌. 这种“短板惩罚机制”能够强制优化算法筛选双性能均衡优异的合金成分, 锁定两项性能同步偏高的平衡区间. 从而客观表征了硬度与导热性之间的物理互抑效应.

为进一步精炼最优成分窗口, 继续以该 20 个成分为基础, 开展自适应截断高斯采样进行局部精修. 首先对种子成分中第 i 种元素含量求取算术平均值, 作为该元素的高斯采样的中心; 同时以元素全局极值范围的 5% 作为高斯扰动的标准差 σ_i , 计算公式分别为:

$$\mu_i = \frac{1}{20} \sum_{k=1}^{20} x_{i,k} \quad (8)$$

$$\sigma_i = 0.05 \cdot (\text{Max}_i - \text{Min}_i) \quad (9)$$

式中, $[\text{Min}_i, \text{Max}_i]$ 为第 i 种元素含量极值范围. 以 μ_i 为中心、 σ_i 为标准差, 按照正态分布 $N(\mu_i, \sigma_i^2)$ 生成 10^4 组重采样样本, 并对超出元素极值范围的采样点进行边界截断处理. 为满足合金成分配比的物理守恒约束, 对截断后的有效样本进行总量归一化处理, 计算公式为:

$$x_i = \frac{x_{i,\text{init}}}{\sum_{j=1}^n x_{j,\text{init}}} \times 100\% \quad (10)$$

其中, $x_{i,\text{init}}$ 为元素 i 的初始采样含量, n 为合金体系所包含的元素种类数.

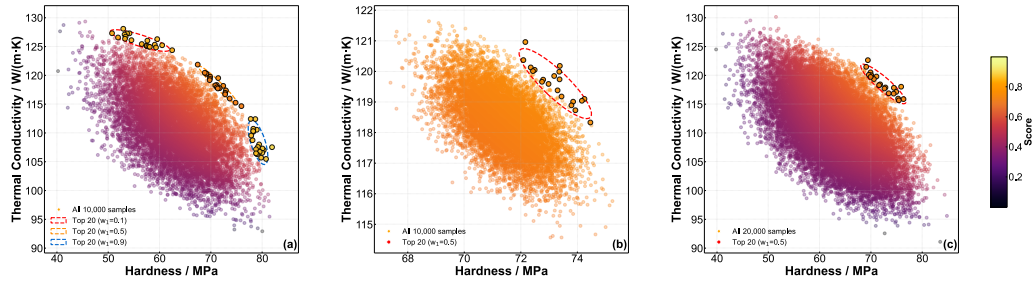


图 6 全局初筛 (a) 加局部精修 (b) 两阶段优化策略与纯全局采样 (c) 下的性能分布对比

Fig. 6. Performance distribution comparison of the two-stage optimization strategy, including global screening (a) and local refinement (b), with pure global sampling (c).

表 2 全局初筛与局部精修阶段关键性能指标对比

Table 2. Comparison of key performance indicators in global exploration and local reafinement stages.

优化阶段	最高综合评分	核心性能区间 (硬度 / 导热率)
全局初筛	0.739	67~76 MPa / 114~122 W/(m·K)
局部精修	0.771	72~74 MPa / 118~121 W/(m·K)

图 6 展示了铝铜合金硬度与导热性能协同优化的全局初筛与局部精修对比结果. 图中彩色渐变散点代表由 LHS 与局部高斯精修生成的全部候选成分, 颜色映射至综合评分数值 (黄色代表性能相对较优, 紫色代表评分相对较低), 黄色带黑边的高亮散点为各阶段综合评分 Top20 的候选组合. 对于全局初筛阶段而言, 如图 6(a) 所示, 通过 LHS 采样实现了成分空间的均匀覆盖. 随着 w_1 的增加, 初步筛选的 Top20 的种子成分在 10^4 个成分空间所构成的椭圆外边缘从左上角向右下角移动. 以 $w_1 = 0.5$ 为例进行局部精修, 如图 6(b) 所示, 局部高斯精修阶段的采样点明显向高分性能方向收敛. 为更直观地量化两阶段优化差异, 将全局初筛与局部精修阶段的关键性能指标汇总于表 2. 由表 2 可知, 全局阶段最高综合评分为 0.739, 性能分布覆盖 67~76 MPa 的硬度区间与 114~122W/(m·K) 的导热率区间. 经局部优化后, 最高综合评分进一步提升至 0.771, 高分候选点集中分布于硬度 72~74MPa 与导热率 118~121 W/(m·K) 的核心高性能区间, 与全局阶段的宽泛分布相比, 实现了对优质成分组合的聚焦式搜索.

为了进一步评估所提“全局初筛 + 局部精修”两阶段寻优策略相比于单阶段全局采样策略的适用性, 本研究构建了大规模拉丁超立方采样对照组, 其预测成分与性能分布对比结果如图 6(c) 所示. 对照组直接利用拉丁超立方采样在全局成分空间中均匀生成 2×10^4 个样本, 并提取其在综合评分下的 Top 20 候选点. 从图 6(c) 可以看出, 同样以 $w_1 = 0.5$ 筛选得到的 Top 20 候选点, 其硬度主要分布在 69~76 MPa, 导热主要分布在 116~123 W/(m·K). 这些候选点在性能空间中表现较为分散, 说明在高维成分空间中, 仅依靠单阶段全局均匀采样, 即使增加采样规模, 也难以在有限计算预算下实现对高性能区域的精确聚焦. 相比之下, 提出的两阶段优化策略 (图 6(a) 和 (b)) 首先通过全局初筛识别潜在高性能区域, 再基于优选种子点开展局部高斯精修, 使候选点进一步向高分区域聚集, 在保持候选合金成分多样性的同时, 实现了更优的综合性能.

采用“全局初筛 + 局部精修”两阶段策略和单阶段大规模拉丁超立方采样，分别从候选池中筛选出 $w_1 = 0.5$ 的 4 组综合评分最优的候选合金，其成分组合及性能列于表 3 和表 4。从综合评分来看，两阶段优化策略获得的候选合金综合评分介于 0.766~0.771 之间，最高达到 0.771，高于单阶段大规模拉丁超立方采样的最高评分 0.768。同时，两阶段优化所得候选合金的硬度主要集中于 72.16~74.25 MPa，导热率稳定分布于 119.10~120.96 W/(m·K)，性能波动较小，表现出较好的综合性能一致性；相比之下，单阶段大规模拉丁超立方采样所得候选合金虽然包含部分硬度较高或导热率较高的样本，但两项性能之间存在较明显的性能权衡关系，导致其综合评分整体低于两阶段优化结果。这一对比体现了“全局初筛 + 局部精修”两阶段策略可稳定捕获高性能合金成分组合。

表 3 基于高斯精修策略优化的四组高性能合金成分

Table 3. Four groups of high-performance alloy compositions optimized based on Gaussian refinement strategy.

Si	Mg	Fe	Cu	Ni	Cr	Mn	Ti	Zn	Sr	Al	硬度 (MPa)	导热系数 (W/(m·K))	Score
11.18	0.48	0.66	4.51	0.40	0.09	0.28	0.01	0.60	0.01	81.79	73.38	120.18	0.771
12.08	0.47	0.58	4.40	0.36	0.08	0.30	0.01	0.81	0.01	80.90	73.36	119.99	0.768
11.26	0.49	0.56	4.40	0.34	0.08	0.29	0.01	0.83	0.01	81.73	72.16	120.96	0.768
11.82	0.48	0.71	4.45	0.37	0.07	0.30	0.01	0.71	0.01	81.06	74.25	119.10	0.766

表 4 单阶段大规模拉丁超立方采样的四组高性能合金成分

Table 4. Four groups of high-performance alloy compositions from single-stage large-scale Latin hypercube sampling.

Si	Mg	Fe	Cu	Ni	Cr	Mn	Ti	Zn	Sr	Al	硬度 (MPa)	导热系数 (W/(m·K))	Score
12.12	0.50	0.70	5.12	0.24	0.06	0.07	0.02	0.55	0.01	80.62	75.89	118.03	0.768
13.48	0.50	0.63	2.41	0.33	0.15	0.09	0.01	0.51	0.01	81.88	69.40	122.64	0.755
12.21	0.36	0.69	4.51	0.52	0.10	0.52	0.02	0.80	0.02	80.25	74.47	117.84	0.749
10.81	0.45	0.70	4.83	0.50	0.02	0.24	0.02	1.37	0.02	81.03	74.93	117.13	0.744

5 特征分析

对于铝基固溶体而言，硬度的提升主要源于溶质原子引起的晶格畸变场对位错运动的钉扎作用，而筛选出的 A14v 和 A10v 这两个特征捕捉了该固溶强化行为中的弹性与化学分量。A14v 是压缩模量方差，从弹性力学角度描述了铝基体中的畸变效应，由于溶质原子与基体铝原子的压缩模量存在显著差异，这种模量失配会在晶格内部诱发非均匀的弹性应力场，显著增加了位错滑移的晶格摩擦力。与此同时，A10v 是绝对电负性方差，它量化了化学键合的异质性，电负性方差的增大意味着晶格中存在强烈的局部电荷转移，

导致金属键电子云分布发生畸变并形成化学微区,这种异质性不仅增加了位错切过基体的阻力,还往往预示着更强的早期析出倾向.A10v 和 A14v 通过强化固溶体内部的应力场波动与化学键合阻碍,共同作用于位错强化的微观过程,从而在宏观上显著提升了铝铜合金的抗变形能力.

研究中发现原子核磁频率加权均值 (A20m)、空缺焓加权方差 (A5v) 及 Ni、Ti 元素成分含量是决定铝铜合金热传输效率的核心因子. 导热性能的调控核心源于自由电子的输运效率,而电子输运过程又受电子散射效应的直接影响,筛选出的四类特征从电子迁移能力、空位散射、溶质原子散射等多维度捕捉了热传输的关键调控机制.A20m 与铝基体中自由电子的迁移能力直接相关,原子核磁频率的平均水平越高,电子在晶格中的运动阻碍越小,电子输运的流畅性越强,为热传导提供了良好的微观基础.A5v 量化了合金内部空位缺陷分布的异质性,空缺方差的增大意味着晶格中空位缺陷的分布差异显著,大量空位会形成密集的电子散射中心,显著增加电子运动的碰撞概率,大幅降低电子的热传输效率.Ni 作为溶质原子融入铝基体后,其原子半径与铝原子存在差异,会引发晶格畸变,破坏晶格势场的周期性,使电子在运动过程中发生散射,进而抑制热传导;而 Ti 元素是强效的电子散射中心,其未填满的 3d 电子层结构带来晶格势场扰动,可与基体自由电子发生强烈势散射,降低电子迁移率,同时 Ti 兼具晶粒细化作用,会使合金晶粒尺寸减小、晶界密度提升,晶界作为面缺陷会进一步形成电子传输的物理屏障,通过界面散射进一步削弱热传输效率.A20m、A5v 及 Ni、Ti 元素占比通过协同调控电子输运能力与多类型电子散射效应,共同作用于铝基固溶体的热传导微观过程,从而在宏观上决定了铝铜合金的导热性能. 铝铜合金的导热性能受控于原子级溶质势散射与微观级晶界散射的深度耦合效应,筛选出的特征组合能映射出这一复杂的多尺度能量输运过程.

6 结论

本文围绕铝铜合金硬度与导热性能的协同预测与优化展开系统研究,构建了基于“物理衍生特征-多级降维筛选”数据驱动机器学习框架,为合金性能的评估与成分定向设计提供了可行的方法,主要结论如下:

(1) 针对铝铜合金小样本数据的建模需求,建立了以 Lasso 回归为核心的机器学习预测模型. 通过层次聚类、正则化约束与穷举子集搜索相结合的特征筛选策略,有效抑制了过拟合风险,筛选出对性能影响显著的物理特征组合. 模型在测试集上表现优异,硬度与导热性能的决定系数分别达到 0.885 和 0.847,验证了该框架在合金性能预测中的可靠性与泛化能力.

(2) 通过特征重要性分析,明确了影响铝铜合金硬度与导热性能的关键调控特征. 压缩模量方差 (A14v) 与绝对电负性方差 (A10v) 通过构建微观应力场、增大位错运动阻力实现合金硬度提升; 原子核磁频率加权均值 (A20m)、空缺焓加权方差 (A5v) 及 Ni、Ti 元素共同调控电子输运与散射行为,进而改变合金导热性能,体现了特征参数与合金力学、热学性能之间的内在联系.

(3) 提出“全局拉丁超立方采样 + 局部高斯精修”的两阶段双目标优化策略,完成了铝铜合金硬度与导热性能的协同优化. 筛选得到 4 组综合评分达 0.77 的高性能合金成分,其硬度稳定在 72~74MPa、导热率处于 118~121W/(m·K) 区间,实现了两类性能的均衡提升,为铝铜合金的成分定向设计与性能协同调控提供了可行的技术方案^[30].

本研究构建的“物理衍生特征-多级降维筛选”框架在硬度与导热性能协同优化方面取得了显著成效,

但受限于样本规模与工艺耦合度. 后续研究将重点从以下三方面展开: 一是数据扩容与特征深化, 引入微观组织特征以提升模型解释性; 二是构建成分-工艺-性能一体化设计体系, 破除单一成分设计的局限, 通过理论计算和实验等手段验证候选合金体系的性能; 三是探索物理信息约束的机器学习方法, 将材料先验知识嵌入算子, 以进一步强化模型的泛化能力与应用价值.

参考文献

- [1] Lin H, Liu S J, Wei P F, Hu Z L, Ye X M 2025 *Mater. Sci. Eng. A* **945** 148967
- [2] Zhu S Q, Cui X Y, Aruga Y, Liu H W, Ringer S P 2024 *Acta Mater.* **272** 119909
- [3] Yildırım M, Özyürek D 2013 *Mater. Des.* **51** 767
- [4] Liu B, Lei Q, Xie L Q, Wang M P, Li Z 2016 *Mater. Des.* **96** 217
- [5] Zhao D D, Løvvik O M, Marthinsen K, Li Y 2018 *Acta Mater.* **145** 235
- [6] Wang E R, Hui X D, Wang S S, Zhao Y F, Chen G L 2010 *Mater. Sci. Eng. A* **527** 7878
- [7] Liu G L, Fang G L 2009 *Acta Phys. Sin.* **58** 4872 (in Chinese) [刘贵立, 方戈亮 2009 物理学报 **58** 4872]
- [8] Gao J B, Li Z C, Liu J, Zhang J L, Song B, Zhang L J 2023 *Acta Metall. Sin.* **59** 87 (in Chinese) [高建宝, 李志诚, 刘佳, 张金良, 宋波, 张利军 2023 金属学报 **59** 87]
- [9] Lin Z J, Yang C 2025 *CMC-Comput. Mater. Continua* **86** 1
- [10] Mo W Q, Wang J F, Yuan G Q, Cao D, Bai G X 2025 *Mater. Today Commun.* **44** 111933
- [11] Shahmir H, Kooshamanesh S, Langdon T G 2026 *Mater. Sci. Eng. A* **955** 149868
- [12] Nematov D, Hojamberdiev M 2025 *Comput. Condens. Matter* **45** e01139
- [13] Xue D, Balachandran P V, Hogden J, Theiler J, Xue D Q, Lookman T 2016 *Nat. Commun.* **7** 11241
- [14] Wang C S, Fu H D, Jiang L, Xue D Z, Xie J X 2019 *npj Comput. Mater.* **5** 87
- [15] Zhang S J, Wang Y, Zhong J, Gao J B, Lu Z, Zhang L J 2023 *Materials* **16** 306
- [16] Balachandran P V, Broderick S R, Rajan K 2011 *Proc. R. Soc. A* **467** 2271
- [17] Juan Y F, Niu G S, Yang Y, Dai Y B, Zhang J, Han Y F, Sun B D 2023 *J. Mater. Res. Technol.* **24** 346
- [18] Soofi Yasaman J, Rahman M A, Gu Y J, Liu J L 2022 *Comput. Mater. Sci.* **215**

- [19] Ren Y C, Lin Y H, Zhu X J, Zhao S Y, Huang R J, Su T F, Wang C, Ren N Q 2026 *Polym. Degrad. Stab.* **247** 111978
- [20] Lan Q, Wang X, Sun J, Chang Z, Deng Q, Sun Q, Liu Z, Yuan L, Wang J, Wu Y, Liu B, Peng L 2022 *Mater. Today Commun.* **31** 103301
- [21] Lu Z, Kapoor I, Li Y, Liu Y, Zeng X, Wang L 2024 *J. Mater. Inf.* **4** 19
- [22] Bharath L, Kumaraswamy J, Manjunath T V, Kulkarni S K N 2024 *Multiscale Multidiscip. Model. Exp. Des.* **7** 5387
- [23] Mirahmadi Babaheydari R, Mirabootalebi S O 2020 *environ. friendly mater* **4** 31
- [24] Varol Özkavak H, İnce M, Bıçaklı E E 2023 *Arab. J. Sci. Eng.* **48** 2841
- [25] Song R, Liu X M, Wang H B, Lv H, Song X Y 2024 *Acta Phys. Sin.* **73** 254 (in Chinese) [宋睿, 刘雪梅, 王海滨, 吕皓, 宋晓艳 2024 物理学报 **73** 254]
- [26] Gao J 2020 *M.S. Dissertation*. Master's thesis, Beijing University of Technology. (in Chinese) [高建琛 2020 (北京: 北京工业大学)]
- [27] Huang X, Zheng L, Xu H, Fu H 2024 *Mater. Des.* **239** 112797
- [28] McKay M D, Beckman R J, Conover W J 1979 *Technometrics* **21** 239
- [29] Makhoulouf M M, Apelian D, Wang L 1998 Microstructures and properties of aluminum die casting alloys. *Tech. rep.*, USDOE Idaho Operations Office, Idaho Falls, ID, United States
- [30] Zou Y, Bu T T, Li X, Ding S J, Ren P J, Lou C X 2026. Chinese Patent 202611083880.2. (in Chinese) [邹勇, 步彤彤, 李霞, 丁守军, 任平静, 娄春香 2026 中国发明专利 202611083880.2]

附录

表 A1 合金元素的物理化学特性参数表

Table A1 Physicochemical property parameters of alloying elements.

特性	Si	Mg	Fe	Cu	Ni	Cr	Mn	Ti	Zn	Sr	Al
A1 原子序数	14	12	26	29	28	24	25	22	30	38	13
A2 $n_W^{1/3}$ (Miedema)	1.5	1.17	1.77	1.47	1.75	1.73	1.61	1.52	1.32	0.84	1.39

续下页...

表 A1 合金元素的物理化学特性参数表

Table A1 Physicochemical property parameters of alloying elements.

特性	Si	Mg	Fe	Cu	Ni	Cr	Mn	Ti	Zn	Sr	Al
A3 电负性	1.8	1.2	1.64	1.9	1.9	1.6	1.5	1.32	1.6	1	1.5
A4 功函数	4.8	3.7	4.5	4.6	5.1	4.5	4.1	4.3	4.2	2.4	4.3
A5 空缺焓	60	50	140	100	130	120	90	150	50	63	63
A6 原子半径/nm	1.319	1.602	1.274	1.278	1.246	1.36	1.304	1.462	1.394	2.151	1.432
A7 原子化焓	455.6	147.7	416.3	338.3	429.7	396.6	280.7	469.9	130.7	164.4	326.4
A8 原子共价半径	117	136	116	117	115	118	117	132	125	192	125
A9 电子亲和力	133.6	-21	15.7	118.5	156	64.3	-20	7.6	9	-146	44
A10 绝对电负性	4.77	8.3	4.06	4.48	4.4	3.72	3.72	3.45	4.45	2	3.23
A11 第一电离能	786.5	737.7	759.3	745.4	736.7	652.8	717.4	658	906.4	549.5	577.6
A12 导热系数	148	156	80.2	401	90.7	93.7	7.82	21.9	116	49	237
A13 距核心电子距离	1.18	1.09	0.73	1.12	1.13	0.84	0.79	0.91	1.25	1.39	1.25
A14 压缩模量	0.988	0.354	1.683	1.37	1.86	1.901	0.596	1.051	0.598	0.116	0.722
A15 与价电子距离	1.91	2.3	1.16	2.55	2.9	1.33	1.26	1.81	2.21	2.92	2.01
A16 第二电离能	1577	1451	1561	1958	1753	1592	1509	1310	1733	1064	1817
A17 融化焓/kJ·mol ⁻¹	50.6	8.9	13.8	13	17.2	21	13.4	14.55	7.28	8.2	10.5
A18 表面焓 (Miedema)	74	62	129	95	121	121	85	18.8	61	60	76
A19 线性热膨胀系数	2.6	25	11.9	16.6	13.5	4.5	21.9	8.7	30.3	22.5	23.1
A20 原子核磁频率	19.865	6.12	3.231	26.505	8.936	5.652	24.664	5.638	6.254	4.33	26.057
A21 门捷列夫化学家序列	90	13	66	16	72	60	63	56	18	11	75
A22 门捷列夫 d-t 左起	81	12	60	69	66	54	57	48	72	10	76
A23 门捷列夫 d-t 右起	25	95	46	37	40	52	49	58	34	93	30

续下页...

表 A1 合金元素的物理化学特性参数表

Table A1 Physicochemical property parameters of alloying elements.

特性	Si	Mg	Fe	Cu	Ni	Cr	Mn	Ti	Zn	Sr	Al
A24 门捷列夫 H d-t 左起	80	11	59	68	65	53	56	47	71	9	75
A25 门捷列夫 H d-t 右起	26	96	47	38	41	53	50	59	35	94	31
A26 门捷列夫 H t-d 左起	78	8	57	66	63	51	54	45	69	10	73
A27 门捷列夫 H t-d 右起	24	93	45	36	39	51	48	57	33	95	29
A28 门捷列夫·佩蒂福尔	85	73	61	72	67	57	60	51	76	15	80
A29 门捷列夫·常规	84	18	63	72	69	57	60	51	75	16	79
A30 门捷列夫 t-d 左起	79	9	58	67	64	52	55	46	70	11	74
A31 门捷列夫 t-d 右起	23	92	44	35	38	50	47	56	32	94	28
A32 刚性模量	39.7	17.3	60	48.4	76	115.3	79.5	45.6	41.9	6.03	26.2
A33 摩尔热容	20	24.89	25.1	24.426	26.07	23.35	26.32	25.129	25.4	26.4	24.2
A34 价电子数	4	2	8	11	10	6	7	4	12	2	3
A35 杨氏模量	113	44.7	152.3	129.8	199.5	279	191	120.2	104.5	15.7	70.6
A36 质量/ 10^{-3} kg	28.085	24.305	55.875	63.546	58.69	51.996	54.938	47.867	65.39	87.62	26.982
A37 密度/ $\text{g}\cdot\text{cm}^{-3}$	2.33	1.738	7.874	8.96	8.908	7.19	7.47	4.507	7.14	2.63	2.7
A38 热导率/ $\text{J}/(\text{mol}\cdot\text{K})$	148	156	80.2	401	90.7	93.7	7.82	21.9	116	49	237
A39 熔点/ $^{\circ}\text{C}$	1414	97.72	1538	1084.6	1455	1907	1246	1668	419.53	777	660.32
A40 体积/ 10^{-2} nm ³	2.002	2.324	1.178	1.181	1.093	1.2	1.22	1.764	1.521	5.632	1.66
A41 族序数	4	2	8	11	10	6	7	4	12	2	3
A42 凝聚能	4.63	1.51	4.28	3.49	4.44	4.1	2.92	4.85	1.35	1.72	3.39
A43 电导率/ $\text{S}\cdot\text{m}^{-1}$	1.0e3	2.3e7	1.0e7	5.9e7	1.4e7	7.9e6	6.2e5	2.5e6	1.7e7	7.7e6	3.8e7
A44 电阻率/ $\text{m}\Omega$	1.0e-3	4.4e-8	9.7e-8	1.7e-8	7.0e-8	1.3e-7	1.6e-6	4.0e-7	5.9e-8	1.3e-7	2.6e-8

注: 特征 A1-A42 的数据提取自俄罗斯科学院巴依科夫冶金与材料研究所开发的 ELEMENTS 化学元素性质数据库 (IMET RAS, <https://im.ras.ru/>).

Prediction and Joint Optimization of Hardness and Thermal Conductivity of Aluminum-Copper Alloys Based on Machine Learning*

BU Tongtong^{1) #} LI Xia^{1) 2) #} LOU Chunxiang¹⁾ REN Pingjing¹⁾ DING Shoujun^{1) 2)} ZOU Yong^{1) 2) †} XIONG Chunming³⁾

1) (*School of Microelectronics and Data Science, Anhui University of Technology, Maanshan 243032, China*)

2) (*Anhui Provincial Key Laboratory of Industrial Big Data Analysis and Intelligent Decision, Maanshan 243002, China*)

3) (*Zhongdian Tairisheng Ma'anshan Technology Co. Ltd., Maanshan 243000, China*)

Abstract

To address the challenges of small-sample modeling and the difficulty of achieving a balance between hardness and thermal conductivity in aluminum-copper alloys through traditional design, this study proposes a machine learning framework based on physical features and multi-stage screening for performance prediction and composition optimization. The proposed framework consists of 5 sequential stages: data preparation, physical feature construction, multi-stage feature screening, predictive modeling and evaluation, and multi-objective composition optimization. A total of 24 aluminum-copper alloy samples prepared under the same die-casting process were collected, with hardness and thermal conductivity as the target properties, from which a 99-dimensional physically derived feature space was constructed using composition-weighted mean and variance descriptors of 44 physicochemical properties of 11 constituent elements. To reduce feature collinearity and redundancy and improve modeling reliability under

* Project supported by the National Natural Science Foundation of China (Grant No. 52202001) and the Natural Science Research Project of Anhui Province Education Department for Excellent Young Scholars (Grant No. 2024AH030007).

These authors contributed equally.

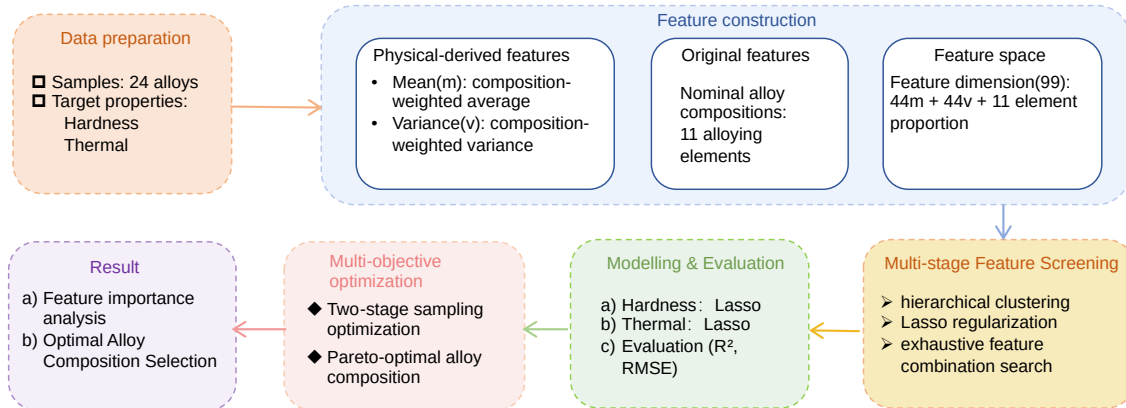
† Corresponding author. E-mail: yongzou@ahut.edu.cn

The First Author. E-mail: 3471003223@qq.com(BU Tongtong); lixiaahut@ahut.edu.cn(LI Xia)

small-sample conditions, a three-stage feature screening strategy combining hierarchical clustering, Lasso regularization, and exhaustive best subset search was employed to identify the core feature subsets for predicting hardness and thermal conductivity. Based on the screened optimal features, Lasso linear regression models with L_1 regularization were established for hardness and thermal conductivity, respectively. Model hyperparameters were optimized using leave-one-out cross-validation (LOOCV), while predictive performance was evaluated on an independent test set. Model performance was quantitatively assessed using the coefficient of determination (R^2) and root-mean-square error (RMSE), thereby characterizing prediction accuracy and generalization capability.

The results show that the hardness model achieves a test R^2 of 0.885 with an RMSE of 4.79, and the thermal conductivity model yields a test R^2 of 0.847 with an RMSE of 4.67, indicating good predictive performance on the independent test set. Furthermore, a composition optimization strategy coupling global Latin hypercube sampling (LHS) and local Gaussian refinement was proposed to achieve a balanced improvement in hardness and thermal conductivity. Within the feasible composition domain constrained by experimental boundaries, 10^4 candidate compositions were uniformly generated via LHS for global exploration. Based on the equal-weight comprehensive scoring function, the top 20 high-performance seed compositions were selected from these candidates. Using these seed compositions as local sampling centers, another 10^4 new samples were generated by Gaussian resampling, and composition normalization constraints were imposed to ensure the validity of the generated compositions. Finally, 4 groups of optimized alloy compositions were screened out, all achieving a comprehensive score of 0.77. The optimized alloys exhibit a hardness of 72~74 MPa and a thermal conductivity of 118~121 W/(m·K), indicating a balanced enhancement of the 2 target properties. Overall, the proposed method combines physically derived descriptors, multi-stage feature selection, predictive modeling, and two-stage composition optimization within a unified data-driven procedure. This combination helps retain physical interpretability while improving prediction performance under small-sample conditions and provides a data-driven approach for the collaborative optimization of hardness and thermal conductivity in Al-Cu alloys.

Data-driven approach for predicting the properties of high-performance aluminum-copper alloys.



Keywords: Machine learning, Aluminum-copper alloys, Feature selection, Synergistic optimization of properties