

基于高通量实验和机器学习模型的低碳熟料约束优化筛选

翟牧楠, 鄧 晓, 叶家元, 任雪红, 张文生

(中国建筑材料科学研究总院有限公司, 北京 100024)

摘 要: 水泥熟料涉及复杂的高温固/液相反应和矿相间协同水化过程, 且不同实验流程导致文献数据在结构化程度和可比性方面存在差异, 这限制了机器学习算法在新型熟料设计中的应用与发展。本工作将熟料单矿视为最小组成单元, 通过标准化高通量实验构建了基于熟料组成的数据集, 共包含 264 组样本。采用贝叶斯优化方法对 5 种非线性模型进行超参数寻优, 比较不同类别算法在所构建数据集上的预测精度, 并利用预先设计的测试集, 严格区分体系内、外两类预测情形。结果表明, 具有适度正则化的低容量人工神经网络模型表现出更高的预测精度, 且对体系外样本具有更强的鲁棒性。此外, 在 28 d 抗压强度与碳排放约束阈值条件下, 利用优选的代理模型与遗传算法耦合, 筛选得到 4 组满足约束条件的低碳熟料组成。经实验验证, 4 组低碳熟料 28 d 抗压强度结果与模型预测结果基本一致, 表明所构建的高通量实验与代理模型可用于低碳熟料的初步筛选。

关键词: 高通量制备; 熟料数据集; 预测模型; 反向设计; 低碳熟料筛选

中图分类号: TQ172.11 文献标志码: A 文章编号: 0454-5648(2026)10-3398-16

网络出版时间: 2026-07-15



水泥属于高能耗、高碳排放行业, 碳排放量约占全球的 5%~8%, 是全球主要工业碳排放源之一^[1-2]。2025 年, 中国水泥产量达到 16.9 亿 t^[3]。若按熟料系数 0.65 和吨熟料碳排放 850 kg 测算^[1, 4], 2025 年熟料生产的碳排放总量约为 9 亿 t, 由此可见, 水泥行业低碳转型的必要性和紧迫性。降低单位熟料碳排放量^[1]是水泥行业实现节能减排的重要途径之一。现有研究表明, 不同熟料体系的碳排放量存在显著差异, 如图 1 所示。硅酸盐水泥熟料(portland cement clinker, PC)的碳排放量显著高于其他熟料体系, 这主要是因 PC 中主导矿相硅酸三钙($3\text{CaO} \cdot \text{SiO}_2$, C_3S)的形成温度较高(约 1450 °C)且 CaO 含量较高(73.7%)。据此, 开发以低形成温度或低 Ca 含量矿相为主导的新型熟料体系, 是该领域的重要研究方向之一。

传统低碳熟料研发主要依赖于对熟料组成及各矿相间协同水化反应的经验性或机理认识, 包括:

1) 在已知体系中引入新矿相, 例如含富铁铁相($6\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{Fe}_2\text{O}_3$, C_6AF_2)的(硫)铁铝酸盐水泥熟料^[5], 或含硫硅酸钙($4\text{CaO} \cdot 2\text{SiO}_2 \cdot \text{CaSO}_4$, $\text{C}_5\text{S}_2\text{S}$)的贝

利特-硫铝酸钙-硫硅酸钙(BYT)熟料^[6]; 2) 优化现有体系矿相组成, 例如通过提高熟料中硅酸二钙($2\text{CaO} \cdot \text{SiO}_2$, C_2S)含量, 形成贝利特水泥熟料或高贝利特硫铝酸盐水泥熟料^[7-8]。这两类策略均旨在开发低碳水泥熟料, 拓展其组成边界。然而, 这种以实验试错与机理分析相结合的开发过程较为缓慢, 往往需要数月甚至数年才能取得实质性进展。因此, 传统研发模式已难以满足“双碳”目标下快速开发新型低碳熟料的需求。

近年来, 高通量(HT)方法与机器学习(ML)的发展为水泥基材料的研发范式提供了新思路。HT 能够在计算、制备和表征等环节快速生成大规模实验数据^[9]。这些数据既包含性能优异的样本, 也包含数量更多的“失败”样本。ML 能够充分利用全部的数据信息, 尤其是“失败”样本, 学习其中复杂的非线性关系, 建立预测与设计模型^[10]。由于能够在弱机理先验条件下从复杂数据中提取经验规律^[11], 该研发范式具备在高维组成筛选中提升研发效率的潜力, 且已在混凝土、镁质胶凝材料等领域得到应用^[12-14]。相较而言, ML 应用于熟料领域的研究主要集中于

收稿日期: 2026-04-22。 修订日期: 2026-05-10。

基金项目: “十四五”国家重点研发计划(2022YFC3803101); 国家自然科学基金双碳专项项目(52341202)。

第一作者: 翟牧楠(1994—), 男, 博士研究生。

通信作者: 叶家元(1980—), 男, 博士, 正高级工程师。

Received date: 2026-04-22. Revised date: 2026-05-10.

First author: ZHAI Munan (1994—), male, Doctoral candidate.

E-mail: zmunnann@163.com

Correspondent author: YE Jiayuan (1980—), male, Ph.D., Professor-level Senior Engineer.

E-mail: yejiayuan@cbma.com.cn

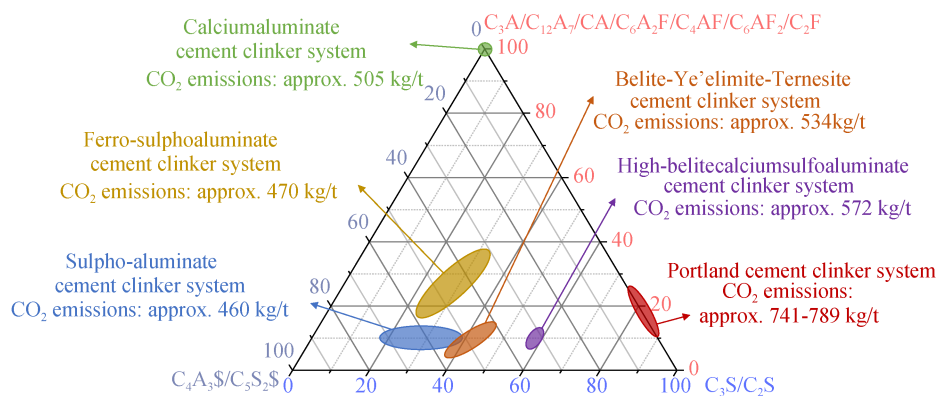


图 1 已知熟料体系及其碳排放量

Fig. 1 Known clinker systems and their corresponding carbon emissions

优化 PC 煅烧工艺参数、预测熟料 28 d 抗压强度, 以及在 PC 体系内调控矿相组成等^[15-17]。针对新型熟料矿相组成反向设计的研究仍有限, 主要原因之一在于高质量结构化熟料数据集难以获取。一方面, 熟料除后续水化过程外, 其制备还涉及煅烧过程中的高温固/液相反应, 这使得直接制备具有多元矿相组成的新型熟料极具挑战。现有数据仅覆盖熟料组成设计空间中的小部分, 大量区域仍处于尚未探索状态。另一方面, 不同制备阶段实验方法的差异又会引入额外误差, 导致数据一致性难以保证^[18]。因此, 同时提升数据可获得性和一致性, 是构建高质量熟料数据集并推动 ML 应用于熟料设计的关键。

本工作将熟料单矿作为最小组成单元, 并把多种单矿按比例干混得到的复合体系定义为熟料组成样本^[19], 以规避熟料一次煅烧制备的限制。随后, 建立标准化 HT 实验以提高数据一致性, 构建基于组成的熟料数据集, 涵盖 264 组样本的矿相组成及其抗压强度。基于该数据集, 采用支持向量回归(SVR)、高斯过程回归(GPR)、随机森林(RF)、极端梯度提升(XGBoost)和人工神经网络(ANN)5 类 ML 算法建立熟料 28 d 抗压强度预测模型, 并使用贝叶斯优化算法对模型超参数进行调优。通过比较模型预测与实验结果, 分析各模型的适用性与鲁棒性, 筛选出适宜的代理模型。最后, 优选的代理模型结合固定超参数的遗传算法(GA), 开展低碳熟料反向设计研究。本工作结果可为基于目标性能的低碳熟料反向设计提供可复用的熟料数据集, 以及经验证的研究框架。

1 基于高通量实验方法的熟料数据集

1.1 熟料组成样本实验流程

构建全面且具有良好一致性的数据集, 是开发

预测模型并实现低碳熟料反向设计的基础。为此, 将熟料单矿视为最小单元, 并将由单矿按比例干混得到的熟料组成视为一组样本。该策略将矿相间协同水化作用与高温相变反应解耦, 实现熟料在较宽组成空间内快速、可重复采样, 同时保留矿相间水化作用的可解释性。需指出, 该策略旨在构建具有较好一致性的数据集, 并学习矿相组成及其协同水化效应与抗压强度之间的映射关系。

基于上述定义, 采用高通量实验获取熟料组成数据, 如图 2 所示。样本的自动称量和干混过程由中国建筑材料科学研究总院有限公司研发的水泥基材料高通量制备与智能设计工作站完成。每组样本质量为 300 g, 最大称量误差为 $\pm 0.3\%$ 。该工作站集自动化与智能控制于一体, 共设 16 个通道, 每个通道容积为 15 L, 可满足本工作所用全部单矿的储存需求。干混完成后, 确定各样本水胶比(W/B), 完成其成型与养护, 并测得其 3 d 和 28 d 抗压强度。

1.2 原材料与试件制备

基于矿相制备及其水化反应活性的分析讨论^[20], 选取 11 种熟料单矿作为最小单元, 分别为 C_3S 、 C_2S 、铝酸钙($CaO \cdot Al_2O_3$, CA)、七铝酸十二钙($12CaO \cdot 7Al_2O_3$, $C_{12}A_7$)、铝酸三钙($3CaO \cdot Al_2O_3$, C_3A)、铁酸二钙($2CaO \cdot Fe_2O_3$, C_2F)、 C_6AF_2 、铁铝酸四钙($4CaO \cdot Al_2O_3 \cdot Fe_2O_3$, C_4AF)、富铝铁相($6CaO \cdot 2Al_2O_3 \cdot Fe_2O_3$, C_6A_2F)、硫铝酸钙($3CaO \cdot 3Al_2O_3 \cdot CaSO_4$, $C_4A_3\$$)及 $C_5S_2\$$ 。所需单矿按文献^[20]合成, 其理化特性见表 1。制备硬化试件前, 需在每一样本中额外掺入一定量二水石膏作为内掺组分, 所用二水石膏购自中国国药集团有限公司。除 $C_4A_3\$$ 外, 样本中二水石膏掺量按各矿相组成完全水化生成 AFm 所需 SO_3 的量计算。对于含 $C_4A_3\$$ 的样本, 其占比按 M 值为 0.8 掺入二水石膏, 其中 M

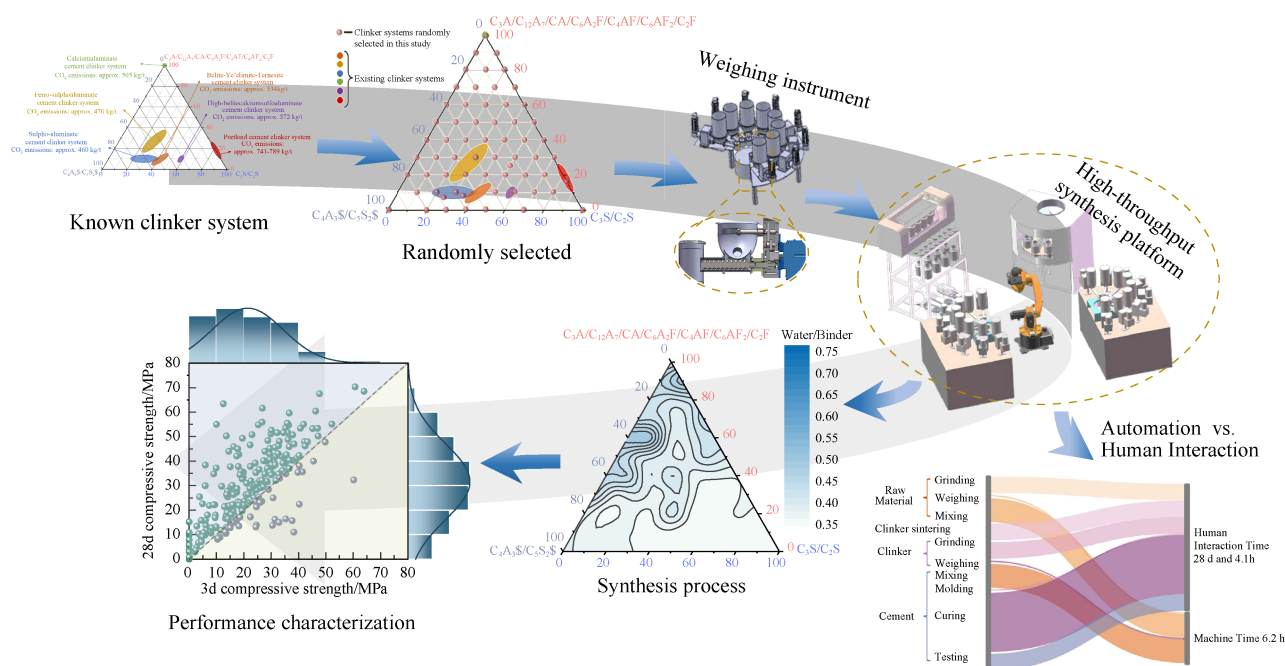


图2 熟料组成样本的高通量制备实验

Fig. 2 High-throughput preparation experiment of clinker composition samples

表1 熟料单矿的煅烧制度及其理化特性分析

Table 1 Preparation process of pure-phase clinker phases and analysis of their physical and chemical properties

Clinker phase	Calcination process		Purity of clinker phase (in mass fraction)/%	<i>f</i> -CaO content (in mass fraction)/%	Specific surface area/(m ² ·kg ⁻¹)
	Sintering temperature/°C	Hold time/h			
M3 C ₃ S	1600	10	98.80 ± 0.62	0.38 ± 0.10	408
β-C ₂ S	1450	4	98.43 ± 0.29	0.38 ± 0.02	398
CA	1300	8+(6)*	98.32 ± 0.30	0.40 ± 0.02	411
C ₁₂ A ₇	1350	4	98.06 ± 0.23	0.34 ± 0.01	398
C ₃ A-Na	1450	8+(2)*	98.41 ± 0.26	0.43 ± 0.05	412
C ₂ F	1250	4	99.64 ± 0.27	0.44 ± 0.01	410
C ₆ AF ₂	1300	4	99.58 ± 0.41	0.49 ± 0.03	382
C ₄ AF	1350	6	98.91 ± 0.42	0.37 ± 0.05	401
C ₆ A ₂ F	1350	6+(2)*	98.14 ± 0.28	0.22 ± 0.08	401
C ₄ A ₃ S	1320	3	98.37 ± 0.30	0.40 ± 0.03	417
C ₅ S ₂ S	1200	9	98.20 ± 0.23	0.47 ± 0.01	415

Note: * represents time in parentheses represents the holding time during the secondary sintering.

值为二水石膏提供的 SO₃ 质量与 C₄A₃S 水化形成 AFt 所需 SO₃ 质量之比^[21]。

由于不同单矿标准稠度需水量差异较大, 采用固定低 W/B 可能会降低部分样本的工作性, 甚至无法正常成型。因此, 为在保证工作性的同时尽量减小 W/B 变化对抗压强度的影响, 采用了 2 种 W/B 控制策略。当样本标准稠度需水量不大于 0.35 时, W/B 固定为 0.35; 当标准稠度需水量大于 0.35 时, 通过调节 W/B 使浆体达到设定表观黏度。该设定值

定义为通过数字旋转黏度计测定的硅酸盐水泥(P·I)浆体在标准稠度下的表观黏度^[22–24]。受高通量制备条件下原材料用量限制, 20 mm×20 mm×20 mm 的立方体试件用于抗压强度测试。试件成型后在恒温恒湿养护室中[(20 ± 2) °C、相对湿度 95% ± 5%]养护 24 h 后脱模, 随后置于去离子水中养护。在规定龄期, 参照 GB/T 17671—2021《水泥胶砂强度检验方法(ISO 法)》进行抗压强度测试, 取 6 个试件的平均值表示样本 28 d 抗压强度。需指出, 使用小尺寸

试件会不可避免地将尺寸效应引入 ML 模型，从而可能影响 ML 模型训练，但在相同测试条件下比较熟料 28 d 抗压强度，仍可实现对新型低碳熟料的建模与挖掘。

1.3 熟料体系及其矿相组成的选择规则

熟料矿相组成采用受约束的两阶段随机采样策略。第一阶段，从 11 种单矿中选取 3 种或 4 种定义为某一熟料体系。鉴于 C_2F 、 C_6AF_2 、 C_4AF 和 C_6A_2F 构成连续固溶体系列^[25]，限定每组样本体系中至多包含 1 种铁相。第二阶段，各单矿以 10% 步长随机赋值，总质量分数约束为 100%。为保证各矿相含量具有实际意义，规定其最小质量分数不低于 10%。该策略在有限实验条件下提高了数据集的多样性，并降低了主观选择导致的样本聚集风险。除上述生成的

三元和四元熟料体系外，该数据集还包括：熟料单矿、由单矿与二水石膏构成的二元体系、PC 和 BYT 熟料，以及为评估模型而留出的样本(其组成不受 10% 步长限制)，共计 264 个熟料组成样本，每组样本记录其组成信息、W/B 以及 3 d 和 28 d 抗压强度。各熟料组成的采样策略及样本数量汇总见表 2。训练/验证集与测试集中熟料矿相组成分布情况(不含石膏)分别见图 3a 和图 3b。由图 3 可见，训练/验证集对先前探索不足的组成空间进行了离散采样，有助于在采样稀疏区域开展筛选与规律发现。尽管测试集中熟料组成未覆盖全部组成空间，但超过 60% 样本来自未在训练阶段出现的体系，这为评估模型跨体系的预测稳定性提供了评价依据。此外，掺入二水石膏后，归一化的样本矿相组成结果如图 3c 所示。

表 2 数据集中样本的构建准则与数量
Table 2 Selection rules and numbers of specimens in the dataset

Category	Model-clinker system	Clinker phase selection rules	Number of systems and specimens selected	Total/ group
Training & validation sets	Ternary system	Select 3 phases from 11 clinker phases	9 systems/3 specimens	27
	Quaternary system	Select 4 phases from 11 clinker phases	30 systems/6 specimens	180
	Known clinker systems	PC and BYT clinker	one specimen for PC/ two specimens for BYT	3
	Unit clinker phase	Pure-phase clinker phases	11 specimens	11
	Phase-pure clinker phases + gypsum	Incorporating a certain proportion of gypsum into pure-phase clinker phases	7 specimens	7
Test set	Ternary system & Quaternary system	In-system (included in the training and validation sets)	12 specimens	12
		Out-of-system (Not included in the training and validation sets)	24 specimens	24

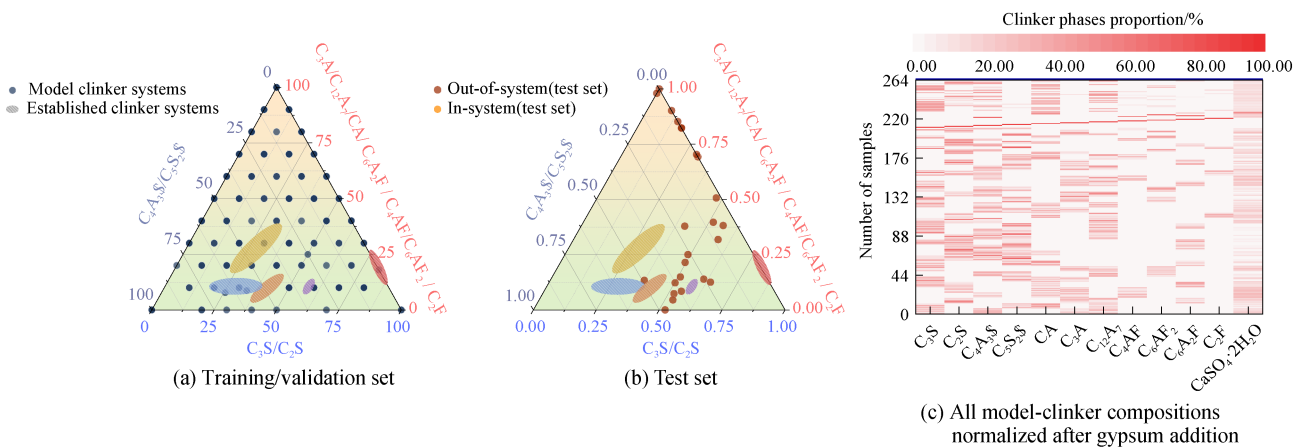


图 3 数据集中样本的矿相组成分布情况
Fig. 3 Distribution of clinker phase compositions among specimens in the dataset

2 机器学习模型的建立

2.1 算法、模型输入与输出

采用 SVR、GPR、RF、XGBoost 和 ANN5 类机

器学习算法，构建熟料 28 d 抗压强度预测模型。上述机器学习算法在学习机制、表达能力、泛化能力以及适用数据类型和规模等方面各具特点。其理论基础已有较多文献报道^[11, 13, 26]，本文不再赘述。模

型输入为掺入二水石膏归一化后的矿相组成, 输出为 28 d 抗压强度。W/B 和 3 d 抗压强度均未作为模型的输入或输出, 仅用于数据集的描述性统计和比较分析。尽管 W/B 会影响抗压强度, 但考虑到 W/B 受工作性能及矿物组成共同约束, 并非单独设计变量, 故本工作未将其作为输入特征, 而是通过实验控制策略尽量削弱其波动对模型预测的影响。

2.2 数据集划分策略与评价指标

为评估模型跨体系泛化能力, 预先构建包括 36 组样本的独立测试集, 其中超过 60% 样本来自训练/验证集中未出现的体系(表 2)。该测试集在超参数调优和模型训练阶段保持完全隔离, 以避免信息泄露, 确保最终评估的独立性。除测试集外, 其余 228 组样本用于构建训练/验证集。鉴于数据规模有限且需尽可能保持组成空间覆盖度, 采用受约束的重复留出验证方案进行超参数选择。具体而言, 在训练集/验证集划分中, 从每个三元和四元熟料体系中随机选取 1 组样本划入验证集, 最终共有 39 组样本划入验证集, 189 组样本划入训练集。此外, 为减小验证集抽样带来的评价指标波动, 上述划分过程在 5 个确定随机种子下重复进行, 组建 5 个不同划分的数据集。该策略既保证了各体系均能参与验证评估, 又避免了验证结果被少数样本体系主导。综上, 数据集按 71.6%、14.8% 和 13.6% 的比例划分为训练集、验证集和测试集。

在完成超参数选择后, 使用合并后的训练集和

验证集重新训练模型, 最终在测试集上评价不同算法的稳定性和泛化性能。为量化预测结果与实验结果之间的偏差, 本工作在超参数调优和测试阶段均采用均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)和平均绝对百分比误差(MAPE)作为评价指标, 各评价指标的计算方法参见文献[15], 其 95%置信区间(CI)基于对应数据集的样本, 采用 bootstrap 重采样法估计, 重采样次数 $B=2000$, CI 上、下限取重采样指标分布的 2.5% 和 97.5% 分位数。

2.3 模型超参数搜索空间

超参数选择是影响模型精度和泛化能力的关键因素。对各模型部分超参数及其取值范围进行设定, 其余超参数均采用 Python 库提供的默认值, 如表 3~表 7 所示。鉴于超参数组合(θ)搜索空间较大, 采用基于树结构 Parzen 估计器(TPE)的贝叶斯优化算法对 θ 进行寻优^[27]。每种算法均开展 1000 次搜索迭代。对于具备 RF、XGBoost 和 ANN 等随机性的算法, 在每一个数据集上, 使用 5 个固定随机种子对 θ 进行搜索, 取 5 次验证集 MSE 平均值作为当前 θ 在某个数据集上的验证集 MSE。最终, 以 5 个不同数据集的验证集 MSE 均值作为当前 θ 的总体优化目标。当多个 θ 均相近时, 选择目标值和标准差均较小者作为模型 θ 。该做法可降低模型对初始化的敏感性, 并降低陷入局部最优的风险。对于 SVR 和 GPR 等确定性模型, 则不进行基于不同随机种子的重复训练。

表 3 SVR 算法的超参数寻优范围
Table 3 Hyperparameter search spaces for SVR

C	epsilon	gamma	kernel	degree	shrinking
1×10^{-3} – 1×10^3	1×10^{-3} –0.5	“scale”, “auto”	“rbf”, “linear”, “poly”, “sigmoid”	“poly”:2–5, others: 3	True, False

表 4 GPR 算法的超参数寻优范围
Table 4 Hyperparameter search spaces for GPR

alpha	n_restarts	optimizer	kernel	kernel parameter
1×10^{-8} – 1×10^4	0–5	fmin_l_bfgs_b	RBF	length_scale: 0.1–100.0
			Matern	length_scale: 0.1–100.0, nu: 0.5, 1.5, 2.5
			RationalQuadratic	length_scale: 0.1–100.0, alpha: 0.1–10.0
			DotProduct	sigma_0: 0.01–10.00

表 5 RF 算法的超参数寻优范围
Table 5 Hyperparameter search spaces for RF

n_estimators	max_depth	min_samples_split	min_samples_leaf	max_features	bootstrap
50–350	5–50	2–20	1–20	“sqrt”, “log2”, 0.3, 0.5, 0.8, 1.0	True, False

表 6 XGBoost 算法的超参数寻优范围

Table 6 Hyperparameter search spaces for XGBoost

n_ estimators	max_ depth	lr	subsample	colsample_ bytree	gamma	reg_ alpha	reg_ lambda	min_child_ weight
50–350	3–15	0.01–0.30	0.5–1.0	0.5–1.0	0–5	0–2	0–2	1–20

表 7 ANN 算法的超参数寻优范围

Table 7 Hyperparameter search spaces for ANN

hidden_dim	dropout	num_layers	lr	weight_decay	optimizer	batch_size	Epochs
4–32, step=4	0–0.3	1–5, step=1	1×10^{-4} – 1×10^{-2}	1×10^{-3} – 1×10^{-1}	"Adam", "AdamW"	16, 32	2000

2.4 基于遗传算法的熟料反向设计

为评估所选代理模型在低碳熟料反向设计中的适用性, 采用 GA 在熟料组成空间内进行约束寻优^[28]。初始种群通过稀疏随机采样生成, 对于每个个体, 随机选取 3 或 4 种单矿并赋予非零数值, 随后进行归一化处理获得熟料矿相组成, 记为 x , 用于计算 CO_2 排放量 $[E(x)]$ 并将其作为碳排放量约束条件。在此基础上, 按 1.2 节规则计算二水石膏掺量, 并再次归一化处理, 获得掺入二水石膏的熟料样本, 记为 X , 用于输入代理模型预测候选熟料 28 d 抗压强度 $[S(X)]$ 。为保证子代始终处于合理的组成范围内, 在每一代交叉和变异后, 对每个个体的矿相组成按 1.3 节规则进行评估, 随后按熟料矿相组成重新计算二水石膏掺量并归一化处理。设定的约束阈值为 $S(X) > 42.5 \text{ MPa}$, 且 $E(x) < 671$ 或 $592 \text{ kg CO}_2/\text{t}$ 。 $E(x)$ 以各矿相贡献的加权和进行近似表征, 各单矿碳排放系数参考文献^[20]。约束条件中 $E(x)$ 的取值分别对应相较于 PC 熟料(58% C_3S 、25% C_2S 、6% C_3A 和 11% C_4AF)碳排放降低 15%和 25%。种群排序采用可行性优先策略, 即满足全部约束条件的个体被判定为可行解, 优先保留, 并按 $E(x)$ 由低到高进行排序。对不可行解, 则按照约束度 V 从低到高排序, 其定义见式(1):

$$V = \max\left(0, \frac{42.5 - S(X)}{42.5}\right) + \max\left(0, \frac{E(x) - \text{CO}_2 \text{ target}}{\text{CO}_2 \text{ target}}\right) \quad (1)$$

式中: $\text{CO}_2 \text{ target}$ 取 671 或 $592 \text{ kg CO}_2/\text{t}$ 。GA 所采用的选择、交叉和变异算子及其对应系数见表 8。当迭代达到预设代数, 或可行解连续 10 代未出现改进时, 终止迭代。每次独立运行最终输出 2 个可行解(候选熟料矿相组成)。

3 结果与讨论

3.1 数据集分析

3.1.1 样本间组成比较

为量化样本在组成空间

表 8 遗传算法的超参数组合

Table 8 Hyperparameters for the GA algorithm

Category	Item	Value
Population	Population size	200
	Number of generations	400
Crossover	Probability	0.9
	Operator	BLX- α , $\alpha=0.3$
Mutation	Probability	0.30
	Noise scale	0.03
Selection	Tournament size	3
Elitism	Elite preservation	10

的局部邻域结构, 采用余弦相似度表征任意两样本 (X) 间的组成重叠程度^[29]。图 4 展示了数据集中样本对的余弦相似度分布情况, 以及体系内样本间的关系。如图 4a 所示, 数据集中以低相似度样本对为主, 其中 58.86% 样本对余弦相似度低于 0.34, 约 90% 的样本对低于 0.71。样本对相似度的平均值为 0.32, 25%、50% 和 75% 分位数分别约为 0.11、0.27 和 0.49, 这表明各样本在组成空间中较为离散, 局部邻域稀疏, 不同体系之间呈现显著异质性。相比之下, 体系内样本相似度明显高于整体样本对相似度, 其平均值约为 0.75(图 4b)。但其中, 仅约 18.2% 的体系内样本对相似度高于 0.9, 部分体系内样本对(占比约 8.9%)相似度甚至低于 0.5, 说明即使在同一体系内, 样本也非近重复样本, 组成间仍存在明显差异。这些结果间接证明了本工作采用的验证集构建策略的合理性, 在保证所有体系均纳入验证集的同时, 降低近重复样本导致评估过于乐观的风险。该设计有助于使 ML 模型学习熟料组成与强度之间的映射关系, 而非简单记忆相近配比。但是, 训练集与验证集在同一体系内仍可能包含相似样本, 导致模型预测精度被高估, 因而需要设立独立测试集, 与训练/验证集隔离, 严格地评估模型预测能力。此外, 体系内高相似度的样本占

比较低, 亦表明采用 5 次划分以获得不同的验证集来降低抽样随机性是必要的。

综上, 多样性和离散性是本工作数据集的重要特征, 这种矿相组成广覆盖的特点并非源于随机噪声, 而是为突破常规配方限制所采用的采样策略, 引入当前研究未涉及的熟料体系^[15, 30]。这一策略不

仅拓展了筛选潜在可行熟料体系的搜索空间, 还形成了具有广覆盖但局部邻域样本相对稀疏的数据结构。在该数据结构下, 严重依赖近邻假设或线性插值的模型容易产生较差的预测结果。相反, 模型需要在不同组成区域保持适当的非线性表达能力, 降低模型训练过程中的过拟合风险^[31]。

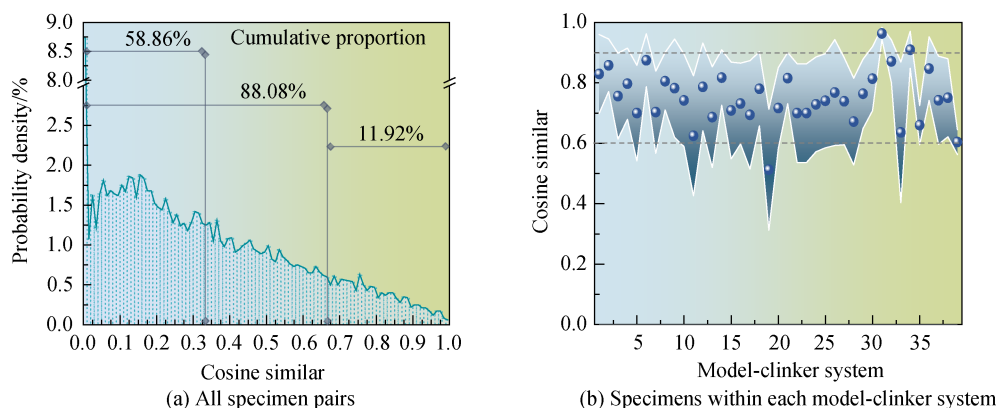


图 4 样本间余弦相似度

Fig. 4 Cosine similarity of specimen pairs

3.1.2 样本水胶比与抗压强度 样本 W/B 的变化会改变硬化浆体孔隙结构及其水化产物组成, 从而降低仅依据矿物信息预测抗压强度的准确性。图 5a 为样本的 W/B 分布情况。由图 5a 可见, W/B 呈由 C_3S/C_2S 顶端向铝酸盐矿相/铁相和 C_4A_3/C_5S_2 顶端方向逐渐增大的趋势。当样本中 C_3S/C_2S 含量超过 60% 时(即接近 C_3S/C_2S 顶端), W/B 通常保持在 0.35。W/B 变化等值线紧密且梯度较大的区域出现在铝酸盐矿相/铁相顶端附近, 表明沿该方向样本 W/B 对组成变化的敏感性较强。其中一个原因是, 铝酸盐矿相/铁相顶端共有 7 种单矿, 这些单矿在化学计量需水量、溶解-沉淀动力学及对 W/B 的敏感性方面存在差异^[32–34]; 另一个原因则是在含二水石膏体系中, 二水石膏会抑制铝酸盐矿相和铁相的水化速率, 从而改变体系由早期溶解向成核阶段转变所需用水量^[35–36]。相较之下, C_3S/C_2S - C_4A_3/C_5S_2 相边界附近的等值线存在一定弯曲, 但该区域整体梯度变化较弱, W/B 变化相对平缓。由此可见, 尽管为避免在后续反向设计阶段获得实验上难以实现的 W/B, 并同时降低其波动对模型预测精度的影响, 采用了相对统一的 W/B, 但样本间 W/B 的差异仍是影响预测精度不可忽视的重要因素之一。

图 5b 和图 5c 分别给出了训练/验证集和测试集样本的 3 d 与 28 d 抗压强度分布情况。如图 5b 所示,

3 d 和 28 d 抗压强度均分布在较宽范围内(0~70 MPa), 覆盖了常规工程应用所需的强度等级范围。大多数数据点位于对角线上方, 表明多数样本随养护龄期延长, 其抗压强度有所提高。其中, 3 d 抗压强度的峰值密度集中在 10~20 MPa, 而 28 d 抗压强度整体向高数值区间偏移。训练/验证集宽泛且相对集中的抗压强度分布, 不仅降低了模型过度偏重某一狭窄强度区间的风险, 还提高了不同强度区间预测结果的鲁棒性和泛化能力。相比之下, 测试集 3 d 和 28 d 抗压强度分别聚集于 30 MPa 和 50 MPa 附近(图 5c), 这种集中于中高抗压强度区间的情况, 既不会削弱模型的训练效果, 又可在研究重点关注的区间开展针对性评价。

此外, 部分样本存在抗压强度倒缩现象, 表现为少量数据点位于对角线下方。该现象在图 5d 中表现得更为清晰, 部分样本强度倒缩可达 30 MPa。这种强度倒缩的异常数据常出现在富铝酸盐或富硫铝酸盐熟料的相关报道中, 与其早期快速水化和后续水化产物的二次转化有关^[37–38]。此外, 在抗压强度正常发展的样本中, 还可观察到不同的增长模式。例如, 部分样本表现为早期强度低但后期增长快, 即 3 d 抗压强度约为 20 MPa, 而 28 d 可达到约 70 MPa; 另一些样本则表现出早期强度高而后期增长有限的特征, 即 3 d 抗压强度可达约 50 MPa, 而至 28 d 仅增加约 1 MPa。这些具备差异性的发展趋势为模型

提供了丰富的学习信息, 有助于促使模型进一步推断组成与性能的内在主导因素, 增强其在复杂体系

中的泛化能力^[39]。同时, 也可避免模型因过度依赖局部邻域特征而导致的局部过拟合。

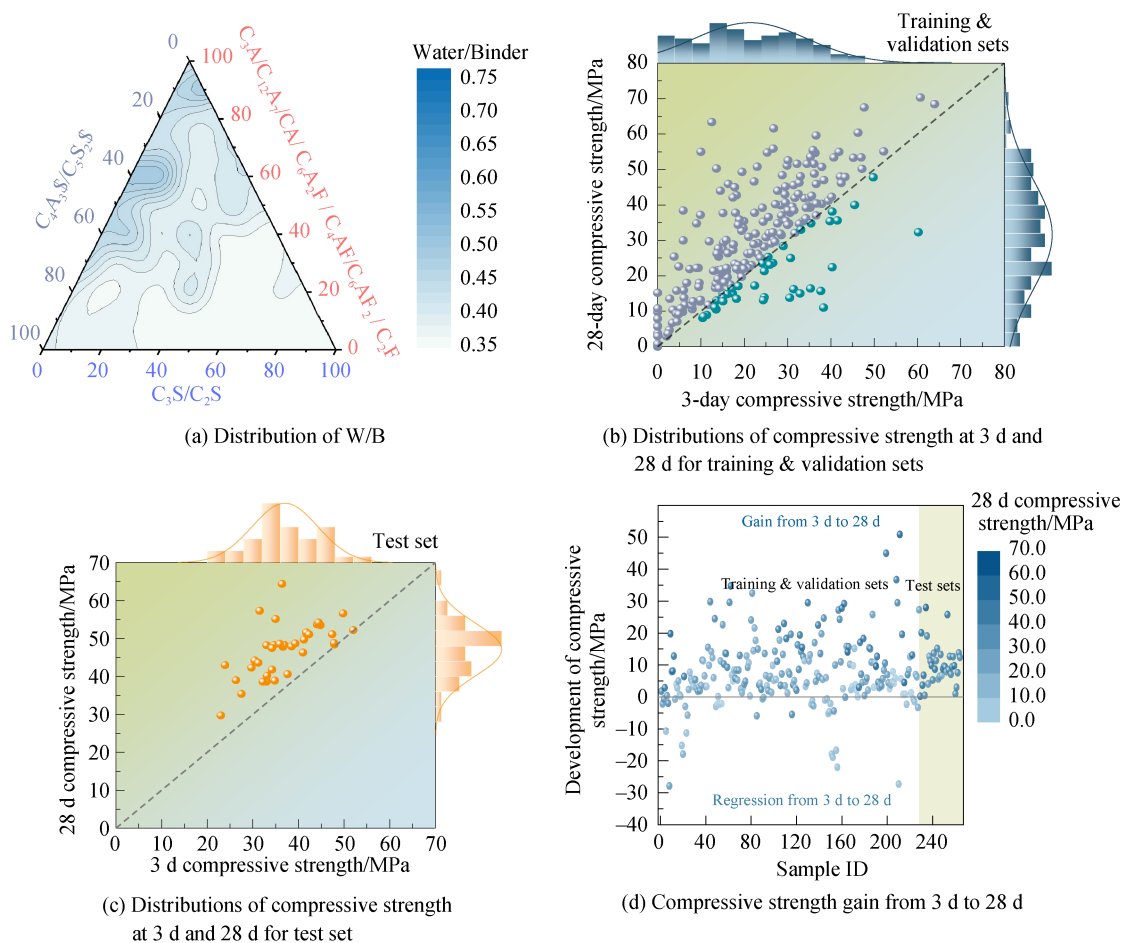


图 5 样本 W/B 及其抗压强度

Fig. 5 W/B used in the specimens and their compressive strength

3.2 验证集上的超参数优化分析

3.2.1 各模型超参数配置 表 9~表 13 列出了各模型的 θ 。对于 SVR, 最优组合为采用二次多项式核函数(degree=2)。其中, 惩罚因子为 5.41, 容忍带宽度 ϵ (epsilon)为 3.68×10^{-1} , 表明该设置可用于表征矿物组成与强度的非线性关系, 且使用 epsilon 降低模型对小偏差的敏感性。对于 GPR, 则采用 Matern 核函数(nu=1.5, length_scale 为 1.32×10^{-1}), 噪声项 α 为 3.84×10^{-1} 。较小的 length_scale 表明模型对局部变化更为敏感, 而 α 则有助于降低样本间差异带来的扰动。为降低边际似然优化过程中陷入局部最优的敏感性, 采用 fmin_l_bfgs_b 优化器, 设置 3 次重启。对于 RF 和 XGBoost 等集成算法, 模型容量主要由树数量(n_estimators)和树深度(max_depth)控制^[40]。在 RF 模型中, 所选超参数倾向于采用适中的 n_estimators(77)和 max_depth (10), 并配

合严格的节点划分约束(min_samples_split=9, min_samples_leaf=2)。这一组合有助于平衡偏差, 降低过拟合风险。相较之下, XGBoost 允许采用更深的 max_depth(11), 但需通过正则化来控制模型复杂度。具体而言, 该模型通过行采样和列采样结合正则化项(gamma=0.749, reg_alpha=0.869, reg_lambda=0.852)的方式来缓解过拟合风险。这些设置表明, 各模型均倾向于采用受约束的折中策略, 在保留足够拟合能力的同时控制模型复杂度并提升泛化性能。对于 ANN, 超参数组合亦倾向采用较小的网络结构(hidden_dim=28, num_layers=2), 以及适当的正则化设置(dropout= 1.60×10^{-1} , weight_decay= 2.21×10^{-2})来保证训练过程的稳定。上述各模型超参数结果表明, 在当前数据集规模有限且样本间异质性较高的条件下, 受控的模型容量和适度的正则化可能更有利于获得稳健预测。

表 9 SVR 的最优超参数设置

Table 9 Optimal hyperparameter settings for SVR

Hyperparameter	Value
C	5.41
epsilon	3.68×10^{-1}
gamma	scale
kernel	poly
degree	2
shrinking	False

表 10 GPR 的最优超参数设置

Table 10 Optimal hyperparameter settings for GPR

Hyperparameter	Value
kernel	Matern
kernel parameter	length_scale
	1.32×10^{-1}
	nu
	1.5
alpha	3.84×10^{-1}
optimizer	fmin_l_bfgs_b
n_restarts	3

表 11 RF 的最优超参数设置

Table 11 Optimal hyperparameter settings for RF

Hyperparameter	Value
n_estimators	77
max_depth	10
min_samples_split	9
min_samples_leaf	2
max_features	0.8
bootstrap	True

表 12 XGBoost 的最优超参数设置

Table 12 Optimal hyperparameter settings for XGBoost

Hyperparameter	Value
n_estimators	61
max_depth	11
lr	9.16×10^{-2}
subsample	7.76×10^{-1}
colsample_bytree	9.28×10^{-1}
gamma	7.49×10^{-1}
reg_alpha	8.69×10^{-1}
reg_lambda	8.52×10^{-1}
min_child_weight	1

表 13 ANN 的最优超参数设置

Table 13 Optimal hyperparameter settings for ANN

Hyperparameter	Value
hidden_dim	28
Dropout	1.60×10^{-1}
num_layers	2
lr	1.33×10^{-4}
weight_decay	2.21×10^{-2}
Optimizer	AdamW
batch_size	32

3.2.2 模型在验证集上的预测性能与不确定性评估 图 6 展示了 5 类算法在最优超参数下的评价指标结果。基于 5 个不同数据划分下的验证集评估结果, 采用 bootstrap 重采样构建各评价指标 95% CI, 以此量化验证集变化带来的不确定性。结果表明, MSE、RMSE、MAE 和 MAPE 呈现相一致的趋势。ANN 在各划分上的平均 MSE 最低, 优于 RF 和 XGBoost, 而 SVR 和 GPR 相对较大。95% CI 宽度差异表明, 样本间异质性是导致指标波动的重要原因。ANN 及 RF 和 XGBoost 等集成模型在不同验证集中表现出较窄的 95%CI, 而 GPR 在部分验证集上的 95%CI 较宽且误差较大。同一验证集中, 5 个随机种子重复训练所得的均值表明(图 6a 中空圆点), 算法随机性导致的波动普遍小于不同数据划分带来的波动。这一现象与图 4 所示跨体系样本相似性显著低于体系内相似性的结果一致。RMSE 和 MAE 亦呈现相同规律, 其中 ANN 总体 RMSE 和 MAE 分别为 4.67 ± 0.92 和 3.57 ± 0.72 , RF 与 XGBoost 整体相近, SVR 和 GPR 相对较高。在 MAPE 上, ANN 为最低(15.71 ± 2.55), 其次为 XGBoost。上述结果反映了验证集上的性能评估, 但模型的预测精度与泛化能力仍需通过测试集评估。

3.3 预测模型精度与稳定性

3.3.1 测试集 MSE 及偏差分布 图 7a 比较了 5 个模型在测试集及其体系内和体系外子集上的 MSE(误差棒表示测试集 95% CI)。所有模型在体系外子集上的 MSE 均高于体系内子集, 且 95% CI 更宽。这表明测试集中存在训练与验证阶段未覆盖的熟料体系, 且体系外子集中样本在体系组成及其抗压强度分布上均与训练/验证集存在差异, 进而导致模型预测精度下降、不确定性增加。在 5 个模型中, SVR 和 GPR 在测试集上表现出较高的 MSE, 说明其对新样本的预测精度和稳定性较低。RF 在测试集体系内样本上的表现较好, 但在体系外样本上的误差明显增大, 这表明其跨体系泛化能力有限。相比之下, ANN 和 XGBoost 在测试集及其两个子集上均表现出较低的 MSE, 其中 ANN 表现最佳。对于 ANN, 在测试集体系内和体系外子集上的 MSE 分别为 9.7 和 17.7。

图 7b~图 7f 给出了各模型在测试集上的误差分布。对所有模型而言, 样本误差分布峰值多位于 $-3 \sim 3$ MPa 范围内。然而, 各模型在误差集中程度、偏差方向以及大误差样本占比方面存在差异。对于

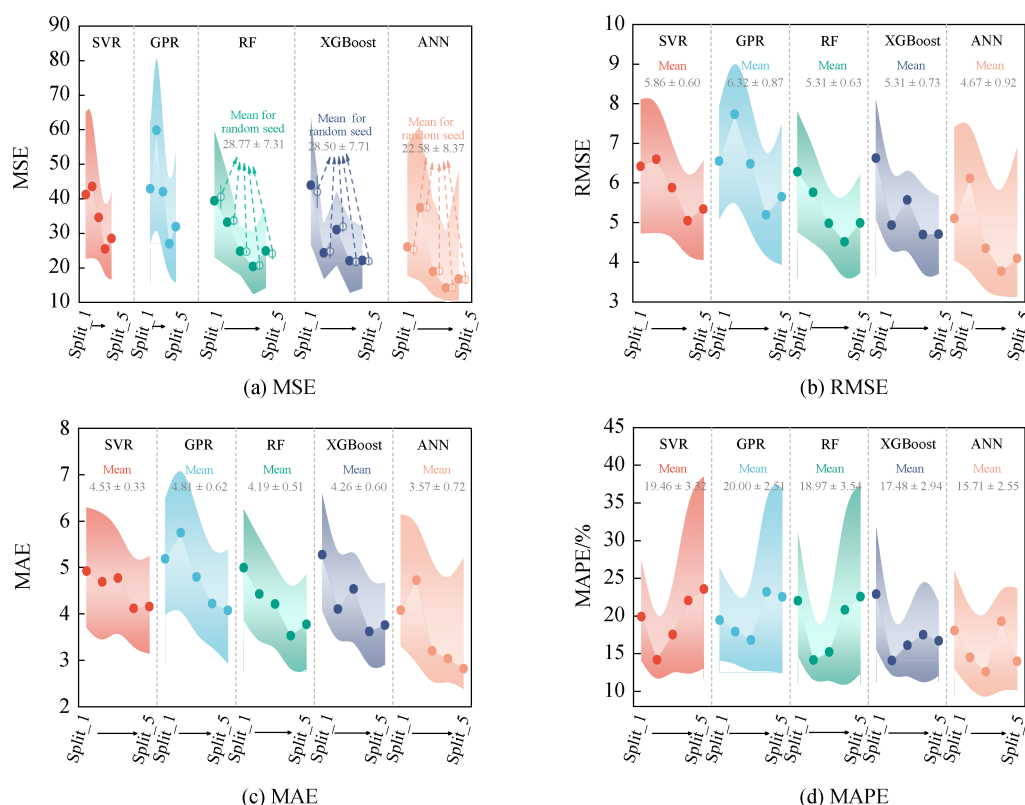


图 6 各模型在不同验证集上的评价指标(Split_1~Split_5, 按箭头所示顺序排列)。实心圆代表最优超参数组合下某个验证集上获得的结果。阴影区域表示 95% CI。空心圆代表某个验证集上 5 个随机种子的平均值。均值表示所有验证集结果的均值与标准差

Fig. 6 Performance comparison of five models across five data splits (Split_1 to Split_5, ordered as indicated by arrows). Each point represents the metric for one split under the optimal hyperparameter configuration. Shaded regions indicate the 95% CI estimated by bootstrap resampling. The hollow circles correspond to the mean across five random seeds for a given split. "Mean" denotes the overall mean $\pm$ standard deviation aggregated across the five data splits

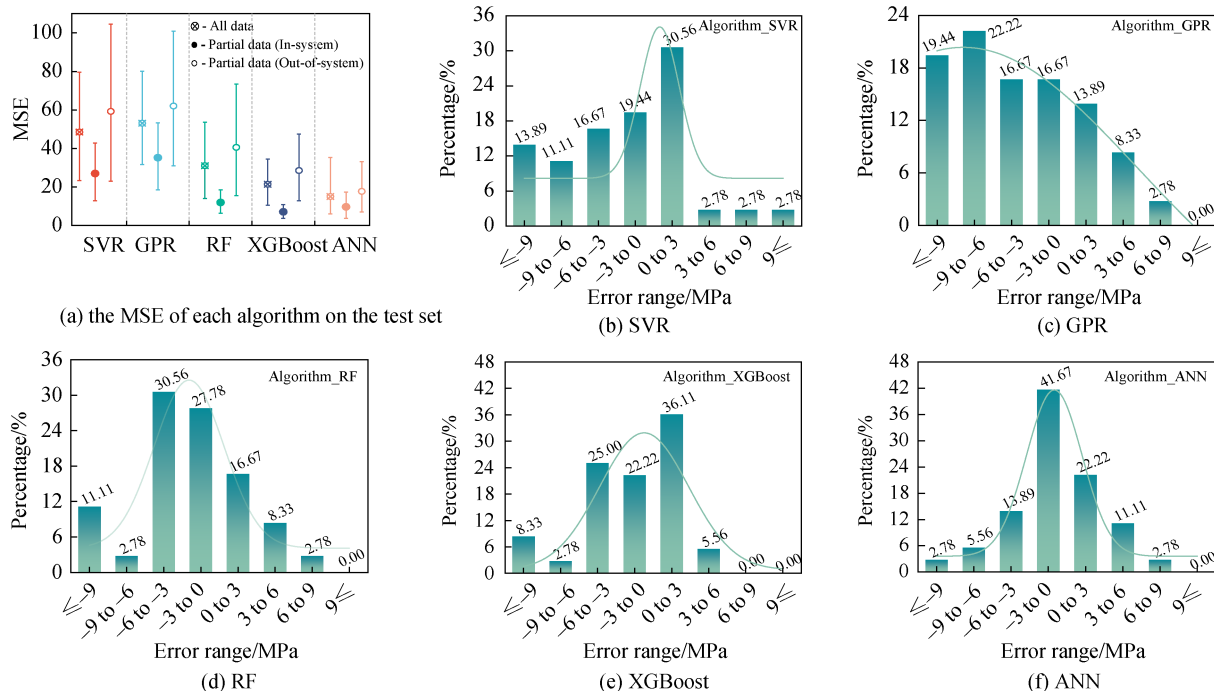


图 7 测试集上每个算法的 MSE(包括所有、体系内和体系外的 MSE)及样本残差分布

Fig. 7 MSE values of each algorithm on the test set (including the MSE for all data, the in-system data, and the out-of-system data), as well as the residual distributions

SVR(图 7b), 其误差分布较为分散, 且负误差占比较大, 表明该模型存在低估抗压强度的倾向, 尤其存在 13.89% 的样本误差低于 -9 MPa。GPR 则表现出明显向负误差方向偏移的非对称分布(图 7c), 仅有 30.56% 的样本误差位于 ± 3 MPa 范围内, 且在误差低于 -9 MPa 的所有样本中, 该模型占比最高。这反映出该模型存在系统性的低估, 且稳定性较差。与 SVR 和 GPR 相比, RF 表现出更清晰的单峰分布, 误差主要集中在较小负值附近(图 7d)。其中, 44.45% 的样本误差位于 ± 3 MPa 范围内, 11.11% 的样本误差低于 -9 MPa 或高于 9 MPa, 说明部分样本仍可能出现较大偏差。XGBoost 和 ANN 的主要误差分布均位于 $-3 \sim 3$ MPa 范围内(图 7e 和图 7f)。然而, XGBoost 仍有 8.33% 的样本误差低于 -9 MPa, 这些相对较大的偏差可能导致其整体 MSE 抬升。相比之下, ANN 在该范围内的误差占比更高, 且分布更均匀, 出现大误差的样本更少, 这表明 ANN 具有更稳定的预测能力。结合图 7 中的 MSE 结果和偏差分布可知, ANN 在体系外子集上表现出较高预测精度和稳定的误差分布, 显示出较优的鲁棒性与跨体系泛化能力。

3.3.2 样本预测准确性及模型性能评估 以样本实测 28 d 抗压强度作为基准, 设置 10% 和 15%

相对误差阈值, 分别统计体系外样本($n=24$)和体系内样本($n=12$)预测合格的样本数量(图 8)。当预测相对误差不超过相应阈值时, 即判定该样本预测结果合格。在 10% 阈值下(图 8a), ANN 在体系外和体系内子集上的合格率分别达到 75.0% 和 83.3%, 表明 ANN 具有较强的跨体系预测能力。RF 和 XGBoost 等集成算法表现同样稳健, 其体系外子集合格率分别为 70.8% 和 66.7%, 且二者在体系内子集上的合格率均达到 91.7%。相比之下, SVR 和 GPR 整体表现较差。GPR 在两个子集上的合格率均为 41.7%, 而 SVR 则表现出明显的不一致性。具体而言, SVR 在体系外子集上的合格率为 70.8%, 但在体系内子集上仅为 50.0%, 这说明其在体系内预测上的稳定性有限。当误差容限放宽至 15%(图 8b), 所有模型的合格样本数均有所增加, 但整体排序基本保持不变。对于体系外子集, ANN 的合格率进一步提升至 91.7%, 而 RF 和 XGBoost 分别达到 83.3% 和 87.5%。GPR 仍相对较低, 仅为 62.5%。这些合格率统计结果与 MSE 和残差分析结果一致, 表明在小样本且高度异质的跨体系数据集中, ANN 和集成算法能够提供更高比例的可靠预测, 尤其是在预测更具挑战性的体系外样本。

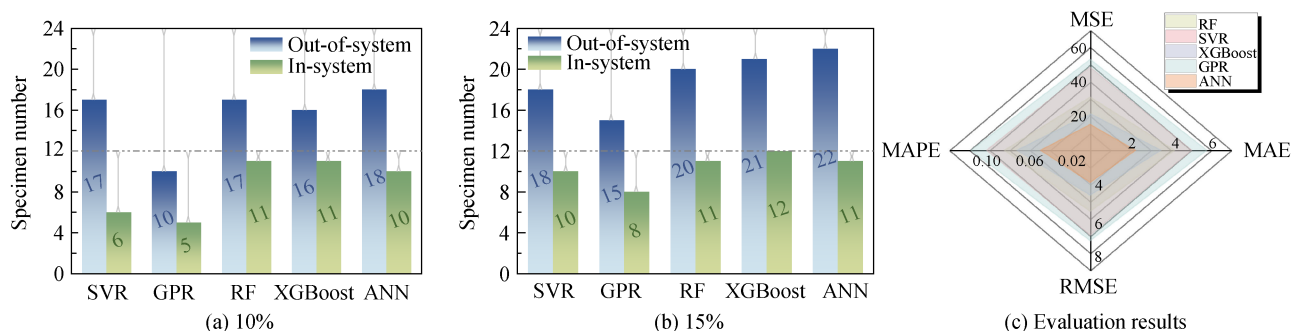


图 8 不同预测模型在测试集上的合格样本数及综合评价结果

Fig. 8 Numbers of specimens whose relative prediction errors are within the specified thresholds and overall evaluation results of different models on the test set

采用 4 个具有互补性的误差指标全面比较模型预测性能, 并以雷达图形式进行汇总展示(图 8c)。结果表明, ANN 在测试集上的 4 个指标中均保持较低误差。其中, ANN 在对大偏差更敏感的 MSE 和 RMSE 指标上表现最佳, 同时在 MAE 和 MAPE 方面也具有优势, 表明其整体性能较为均衡。其次为 XGBoost 和 RF 模型, 但其 MAE 和 MAPE 仍相对较高, 说明对部分样本的误差控制能力仍有限。相较而言, SVR 和 GPR 整体误差较高, 尤其在

MSE、RMSE 和 MAPE 指标上表现较差。这一结果表明, 在预测复杂的跨体系样本时, 二者的拟合能力和鲁棒性均较为有限。

3.4 代理模型泛化性能与适用性分析

数据集的多样性和离散性特征是影响不同模型在验证集和测试集上预测表现的重要因素, 亦是最终优选 ANN 作为代理模型的重要原因。从 3.2.2 节中的实验结果来看, RF、XGBoost 和 ANN 模型预测性能的波动主要受到数据划分差异的影响, 随机

种子变化带来的评价指标波动远小于不同验证集划分引起的波动。这说明, 模型预测性能差异与样本所处的组成区域、体系差异和局部邻域结构有关, 而非训练初始化过程中的波动。模型超参数的选择亦表明, 各模型倾向于控模型容量并采用适度正则化。若模型复杂度过高, 则可能增强其对局部体系特征的依赖, 从而降低跨体系预测稳定性^[41-42]。上述结果共同指向, 在当前数据条件下, 模型需要在保留非线性表征的同时抑制对局部样本特征的过拟合。

在测试集中, 由于存在超过 60% 的体系外样本, 这就要求模型在缺乏邻域支持的情况下进行跨体系预测。对于 SVR 和 GPR 模型, 其预测能力在较大程度上受核函数形式和局部平滑假设影响。当样本组成位于未见体系或接近体系边界时, 先验核函数可能难以充分适应不同熟料体系中组成与 28 d 抗压强度的关系, 因此在测试集中表现出较大的误差^[43-44]。RF 模型能够较处理好体系内样本的局部差异, 但其预测结果受树深度、节点划分约束和叶节点样本统计特征影响较大, 因此在体系外样本上更容易受到训练样本覆盖不足的限制^[45]。XGBoost 则可通过正则化、subsample 和 colsample_bytree 等方式控制模型复杂度, 并保持较强的非线性拟合能力, 整体表现接近 ANN, 但在部分体系外样本中仍出现较大的预测误差, 说明其对局部特征的拟合仍可能影响跨体系稳定性。

相比之下, ANN 模型在测试集上表现出最稳健的跨体系预测, 这主要源于模型表达能力与正则化约束之间的平衡。在当前数据集结构下, ANN 模型采用低容量网络结构, 并引入 dropout 和 weight_decay 进行正则化, 使其既能近似表征由矿相组成及其潜在协同水化效应共同决定的非线性组成与抗压强度间关系, 又能降低对体系特异性噪声和局部样本特征的依赖。ANN 在测试集、体系内/外子集上均保持较低误差, 其中体系内和体系外子集 MSE 分别为 9.7 和 17.7, 残差分布集中且对称, 大误差样本占比较低。

考虑到低碳熟料反向设计需要在大量候选组成中开展初步筛选, 体系外泛化能力和误差稳定性相较体系内插值预测更为关键, 本工作选取 ANN 作为后续 GA 约束优化的代理预测模型, 并在下一节中进一步验证其在低碳熟料候选组成筛选中的应用效果。

3.5 新型低碳熟料反向设计与实验验证

以 ANN 作为代理预测模型、CO₂ 排放与 28 d 抗压强度为约束阈值构建低碳熟料反向设计模型。图 9 展示了 GA 在 2 种不同碳减排约束阈值下筛选得到的候选熟料组成, 其实验验证结果采用与本工作数据集构建过程一致的实验方法进行测定。在 15% CO₂ 减排目标下, 2 个熟料组成的 CO₂ 排放量分别为 648、635 kg CO₂/t。其矿相组成表现为在降低 PC 中高 CO₂ 排放矿相(如 C₃S 和 C₃A)比例的基础上, 向低 CO₂ 排放矿相(如 CA、C₁₂A₇ 或 C₂S)方向调整。同时, 为满足 28 d 抗压强度约束, 候选熟料组成中仍保留一定比例 C₃S, 但其含量低于 PC 体系。当 CO₂ 排放约束进一步收紧至 25% 时, GA 更倾向于选择低碳排矿相, 生成以铝酸盐矿相或硫铝酸钙为主导的熟料体系。将预测与实验测试结果比较可知, 预测值与实验值较为接近, 且部分预测结果落在实验误差棒范围内。此外, 各样本预测误差均控制在实测抗压强度的 10% 以内。总体而言, 该案例展示了数据集构建、模型训练、基于 GA 的约束搜索和实验验证构成的候选熟料组成筛选方案, 实现了在碳排放与抗压强度约束阈值下对低碳熟料组成的设计, 为低碳熟料设计与探索提供支持, 后续可通过热力学计算、烧成实验和 X 射线衍射定量分析等方法进一步验证候选熟料组成的制备可行性。

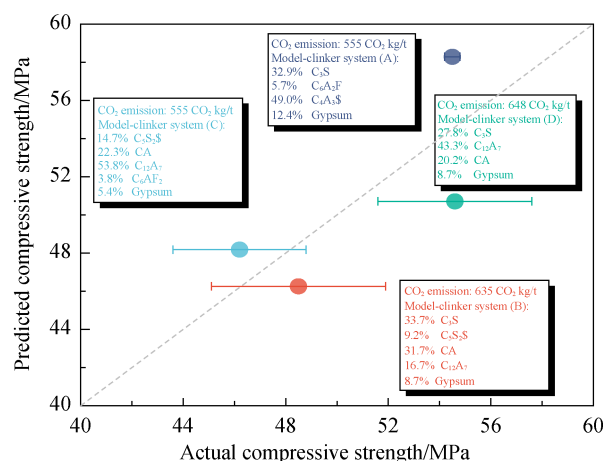


图 9 基于遗传算法和 ANN 代理模型的低碳熟料预测和实验结果间的比较, 以及相应的矿相组成(质量分数)和 CO₂ 排放量

Fig. 9 Comparison between predicted and experimental results for low-carbon clinker designed using a GA algorithm and an ANN surrogate model, together with the corresponding phase composition (in mass fraction) and CO₂ emissions

3.6 局限性

本工作基于高通量实验构建了基于组成的熟料

数据集,采用体系内/外分离的评价策略评估模型预测精度与跨体系泛化能力,并进一步结合遗传算法与代理模型构建熟料反向设计框架。尽管如此,本工作仍存在以下局限性:

1) 熟料数据集覆盖广度有限。尽管本工作样本组成差异性高,但相对于组成空间,目前数据集仍较为稀疏,对组成空间的覆盖不足 1%。此外,本工作熟料单矿也仅包含 11 种常见熟料矿相,远低于前期研究中报道的 60 余种不同矿相(包括多晶型)^[20]。这种覆盖范围较广但局部采样深度不足的方式,致使模型采用更保守的超参数寻优策略,增加了模型对先验与正则化的依赖,限制模型可达到的精度上限。目前,代理模型在 ± 4 MPa 误差范围内的预测准确性仍有待提高。

2) 熟料矿相组成存在理想化假设。本工作数据集中的熟料体系主要通过实验室合成单矿并干混获得。一方面,这种方法省略了煅烧过程,但模型中尚未纳入三元相图作为物理约束条件。因此,筛选得到的熟料是否能够实现实验室制备仍不确定。另一方面,该方法难以充分表征煅烧过程中引入的掺杂效应等,低估工业杂质(如碱金属)对不同矿相晶体结构的影响。

3) 忽略了 W/B 等关键因素。为保持反向设计的可实施性,本工作未将 W/B 作为输入特征。然而, W/B 的残余波动仍可能传递到训练模型中,从而降低抗压强度预测精度。未来研究可将 W/B 纳入特征输入,或构建将工作性(W/B)与性能联系起来的多阶段预测模型。

4 结论

1) 提出并验证了一种面向低碳熟料体系探索的高通量数据驱动的机器学习流程。其核心思想是将实验室合成的熟料单矿视为具有可解释性的最小组成单元,并按一定比例干混构建熟料样本。该策略能够在统一制备与测试条件下获取基于高通量实验的熟料数据集,有助于减少人为操作差异带来的影响。所构建的基于高通量实验的熟料数据集包含 264 组样本,明确区分测试集中体系内与体系外样本,用以严格评估模型的预测精度和跨体系预测的泛化性能。在熟料数据训练预测模型的基础上,比较了多类非线性机器学习算法的适用性,并将优选的代理模型与遗传算法相结合,构建新型低碳熟料筛选框架。

2) 在测试集上,具有适度正则化的低容量

ANN 模型表现出优异的预测精度和稳定性。在测试集上的 RMSE 为 3.9 MPa,在体系内和体系外子集上的 RMSE 分别为 3.1 MPa 和 4.2 MPa。XGBoost 可作为具有竞争力的潜在代理模型,但其在部分样本上易出现较大的预测误差。此外,样本间的内在异质性是模型偏向于受控容量和较强正则化的重要原因。

3) 将代理预测模型与遗传算法耦合,在满足 28 d 强度阈值和碳减排约束阈值下,对矿相组成空间进行筛选。筛选得到的熟料组成实验结果重现了模型预测的相对变化趋势,表明该框架可用于常规熟料组成范围之外的新型低碳熟料候选组成的筛选与探索。

4) 本工作目的在于构建自一致性的新型熟料数据集,并捕捉矿相组成与其协同水化效应的映射关系。但是,当前熟料组成并不等同于可通过煅烧获得的真实熟料组成。因此,未来研究应进一步拓展矿相种类,并开展局部致密采样。同时,还应纳入与固溶/掺杂杂质、煅烧过程及微结构演化相关的工艺和结构描述符。这些改进有望提高反向设计候选熟料体系的制备性,并增强其工程应用相关性。

(第二十八届中国科协年会学术论文)

参考文献:

- [1] BARCELO L, KLINE J, WALENTA G, et al. Cement and carbon emissions[J]. *Mater Struct*, 2014, 47(6): 1055–1065.
- [2] GUO Y Y, LUO L, LIU T T, et al. A review of low-carbon technologies and projects for the global cement industry[J]. *J Environ Sci*, 2024, 136: 682–697.
- [3] 国家统计局. 2025 年 12 月份规模以上工业增加值增长 5.2%^[EB/OL]. [2026-01-19]. https://www.stats.gov.cn/sj/zxfb/202601/t20260119_1962329.html.
- [4] International Energy Agency. Technology Roadmap: Low-carbon transition in the cement industry[R]. Paris: International Energy Agency, 2018: 18.
- [5] 廖宜顺, 叶建, 汤盛文. 海水拌和与偏高岭土对铁铝酸盐水泥水化的影响[J]. *硅酸盐学报*, 2023, 51(11): 2834–2845. LIAO Yishun, YE Jian, TANG Shengwen. *J Chin Ceram Soc*, 2023, 51(11): 2834–2845.
- [6] CAO L X, REN X H, YE J Y, et al. Optimizing cementitious material properties: An exploration of the sintering mechanism in ternesite-belite systems under metastable conditions[J]. *J Sustain Cem Based Mater*, 2025, 14(5): 1028–1039.
- [7] CUESTA A, AYUELA A, ARANDA M A G. Belite cements and their activation[J]. *Cem Concr Res*, 2021, 140: 106319.
- [8] 张文生, 安楠, 张洪滔, 等. 硼/磷复合掺杂对高硫型贝利特硫铝酸盐熟料形成与性能的影响[J]. *硅酸盐学报*, 2023, 51(11): 2741–2749. ZHANG Wensheng, AN Nan, ZHANG Hongtao, et al. *J Chin Ceram Soc*, 2023, 51(11): 2741–2749.

- [9] WEBBER K G, CLEMENS O, BUSCAGLIA V, et al. Review of the opportunities and limitations for powder-based high-throughput solid-state processing of advanced functional ceramics[J]. *J Eur Ceram Soc*, 2024, 44(15): 116780.
- [10] RACCUGLIA P, ELBERT K C, ADLER P D F, et al. Machine-learning-assisted materials discovery using failed experiments[J]. *Nature*, 2016, 533(7601): 73–76.
- [11] CHEN W G, LIANG L, JIANG F M, et al. New opportunity: Materials genome strategy for engineered cementitious composites (ECC) design[J]. *Cem Concr Compos*, 2025, 159: 106009.
- [12] PENG Y M, UNLUER C. Interpretable machine learning-based analysis of hydration and carbonation of carbonated reactive magnesia cement mixes[J]. *J Clean Prod*, 2024, 434: 140054.
- [13] DONG Y H, TANG J H, XU X H, et al. A new method to evaluate features importance in machine-learning based prediction of concrete compressive strength[J]. *J Build Eng*, 2025, 102: 111874.
- [14] PFEIFFER O P, GONG K, SEVERSON K A, et al. Bayesian design of concrete with amortized Gaussian processes and multi-objective optimization[J]. *Cem Concr Res*, 2024, 177: 107406.
- [15] 姚歆玥, 任雪红, 叶家元, 等. 基于半监督生成对抗网络的水泥熟料游离氧化钙数据增强方法及其有效性评价[J]. *硅酸盐学报*, 2025, 53(11): 3127–3135.
YAO Xinyue, REN Xuehong, YE Jiayuan, et al. *J Chin Ceram Soc*, 2025, 53(11): 3127–3135.
- [16] CAO F L, YE J Y, SUN Z T, et al. Compressive strength prediction of Portland cement clinker using machine learning by XRD data[J]. *J Sustain Cem Based Mater*, 2024, 13(8): 1120–1131.
- [17] FAYAZ S J, MONTIEL-BOHÓRQUEZ N, BISHNOI S, et al. Industrial-scale prediction of cement clinker phases using machine learning[J]. *Commun Eng*, 2025, 4: 94.
- [18] XU P C, JI X B, LI M J, et al. Small data machine learning in materials science[J]. *npj Comput Mater*, 2023, 9: 42.
- [19] MARCHON D, JUILLAND P, GALLUCCI E, et al. Molecular and submolecular scale effects of comb-copolymers on tri-calcium silicate reactivity: Toward molecular design[J]. *J Am Ceram Soc*, 2017, 100(3): 817–841.
- [20] ZHAI M N, YE J Y, REN X H, et al. Kilogram-scale synthesis of phase-pure clinker phases with hydration reactivity benchmarks and CO₂ emissions constraints for machine-learning-assisted low-carbon clinker design[J]. *J Build Eng*, 2026, 120: 115395.
- [21] LOTHENBACH B, GEIGER C A, DACHS E, et al. Thermodynamic properties and hydration behavior of ye'elimite[J]. *Cem Concr Res*, 2022, 162: 106995.
- [22] RISSANEN J, OHENOJA K, KINNUNEN P, et al. Milling of peat-wood fly ash: Effect on water demand of mortar and rheology of cement paste[J]. *Constr Build Mater*, 2018, 180: 143–153.
- [23] YUAN Y, WANG X Y, CHEN X, et al. Mathematical models of apparent viscosity as a function of water–cement/binder ratio and superplasticizer in cement pastes[J]. *Sci Rep*, 2023, 13: 22301.
- [24] KIM Y J, CHO B Y, LEE S J, et al. Investigation of rheological properties of blended cement pastes using rotational viscometer and dynamic shear rheometer[J]. *Adv Mater Sci Eng*, 2018, 2018(1): 6303681.
- [25] WU S, YAO X L, REN C Z, et al. Effect of iron on the preparation of iron-rich calcium sulfoaluminat cement using gypsum as the sole calcium oxide source and its incorporation into mineral phases[J]. *Constr Build Mater*, 2021, 290: 123214.
- [26] HUANG Y M, ZHANG J F, TZE ANN F, et al. Intelligent mixture design of steel fibre reinforced concrete using a support vector regression and firefly algorithm based multi-objective optimization model[J]. *Constr Build Mater*, 2020, 260: 120457.
- [27] MAEDA H, WU S, MARUI R, et al. Discovery of liquid crystalline polymers with high thermal conductivity using machine learning[J]. *npj Comput Mater*, 2025, 11: 205.
- [28] CASSAR D R, SANTOS G G, ZANOTTO E D. Designing optical glasses by machine learning coupled with a genetic algorithm[J]. *Ceram Int*, 2021, 47(8): 10555–10564.
- [29] MANA S C, SASIPRABA T. Research on cosine similarity and Pearson correlation based recommendation models[J]. *J Phys Conf Ser*, 2021, 1770(1): 012014.
- [30] BOIGER R, XI B, MIRON G D, et al. Machine learning-accelerated discovery of green cement recipes[J]. *Mater Struct*, 2025, 58(5): 173.
- [31] HAAS B C, KALYANI D, SIGMAN M S. Applying statistical modeling strategies to sparse datasets in synthetic chemistry[J]. *Sci Adv*, 2025, 11: eadt3013.
- [32] ALAPATI P, KURTIS K. Carbonation in alternative cementitious materials: implications on durability and mechanical properties[C]. Sixth International Conference on Durability of Concrete Structures, Leeds, UK, 2018, ICC02: 111–119.
- [33] ZHANG K C, LU Y, RAO M J, et al. Understanding the role of brownmillerite on corrosion resistance[J]. *Constr Build Mater*, 2020, 254: 119262.
- [34] ZAJAC M, HOOCK S, STABLER C, et al. Effect of hydration kinetics on properties of compositionally similar binders[J]. *Cem Concr Res*, 2017, 101: 13–24.
- [35] YE S X, FENG P, LIU J P. Dissolution and early hydration of tetracalcium aluminoferrite (C4AF) in water and in aqueous sulfate solutions[J]. *Cem Concr Res*, 2024, 186: 107676.
- [36] SALVADOR R P, CAVALARO S H P, CINCOTTO M A, et al. Parameters controlling early age hydration of cement pastes containing accelerators for sprayed concrete[J]. *Cem Concr Res*, 2016, 89: 230–248.
- [37] JANSEN D, LOTHENBACH B. Thermodynamic prediction of complex hydration processes and verification by experiments for the Ca-Al-C (CA, calcite, water) and Ca-Al-S (ye'elimite, gypsum, water) systems[J]. *Cem Concr Res*, 2023, 174: 107337.
- [38] XIE Y D, QIAN C X. Improved ettringite stabilization by calcium carbonate and calcium nitrate additions in ternary PC-CSA-C_s systems[J]. *Cem Concr Res*, 2024, 175: 107383.
- [39] DOU B Z, ZHU Z L, MERKURJEV E, et al. Machine learning methods for small data challenges in molecular science[J]. *Chem Rev*, 2023, 123(13): 8736–8780.
- [40] LUO X, LI Y, LIN H, et al. Research on predicting compressive strength of magnesium silicate hydrate cement based on machine learning[J]. *Constr Build Mater*, 2023, 406: 133412.
- [41] MA Y Y, XU P C, LI M J, et al. The mastery of details in the workflow of materials machine learning[J]. *npj Comput Mater*, 2024, 10: 141.
- [42] YOUNG B A, HALL A, PILON L, et al. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions: New insights from statistical analysis and machine learning methods[J]. *Cem Concr Res*, 2019, 115: 379–388.

- [43] STANEV V, OSES C, KUSNE A G, et al. Machine learning modeling of superconducting critical temperature[J]. npj Comput Mater, 2018, 4: 29.
- [44] THANT Y M, WAKAMIYA T, NUKUNUDOMPANICH M, et al. Kernel regression methods for prediction of materials properties: Recent developments[J]. Chem Phys Rev, 2025, 6: 011306.
- [45] HSU A I, YTTRI E A. B-SO₂D, an open-source unsupervised algorithm for identification and fast prediction of behaviors[J]. Nat Commun, 2021, 12: 5188.

作者贡献声明:

翟牧楠: 撰写初稿, 绘制图表, 数据管理

鄧 晓: 调查研究, 设计研究方法, 项目管理, 提供资源

叶家元: 分析/综合数据并推导结论, 论文构思, 审阅与修订

任雪红: 提供资源, 获取资助, 监督指导

张文生: 审阅与修订, 监督指导, 获取资助

Constraint-Driven Optimization and Screening of Low-Carbon Clinker Based on a High-Throughput Process and Machine Learning Models

ZHAI Munan, ZHI Xiao, YE Jiayuan, REN Xuehong, ZHANG Wensheng
(China Building Materials Academy Co., Ltd., Beijing 100024, China)

Extended Abstract

Introduction Conventional development of low-carbon clinker mainly relies on empirical or mechanistic understanding of high-temperature and hydration reactions among clinker phases. This strategy, which combines experimental trial-and-error with mechanistic analysis, is generally time-consuming. Therefore, conventional development approaches are no longer sufficient to meet the demand for the rapid development of novel low-carbon clinkers. The development of high-throughput (HT) methods and machine learning (ML) has provided a new paradigm for the design of cement-based materials. However, available clinker data for ML-based clinker design are usually collected from scattered sources, and they differ substantially in preparation procedures, testing conditions, and compositional definitions, making them difficult to use directly. To address these issues, this study used clinker phases as compositional units and constructed a clinker dataset through a standardized HT preparation procedure. On this basis, the predictive performance of different ML models for the 28 d compressive strength of clinker was compared. Furthermore, the optimized surrogate model was coupled with a genetic algorithm to screen for low-carbon clinkers under constraints on 28 d compressive strength and CO₂ emissions. This approach enabled evaluation of the applicability of the proposed data-driven framework for the design of novel low-carbon clinkers.

Methods In this study, 11 clinker phases were selected as the basis for compositional design, and each clinker composition prepared by dry-mixing these phases with gypsum in specified proportions was regarded as one specimen. A clinker dataset was constructed using an HT experimental procedure. The dataset contained information on phase composition, water-to-binder ratio (W/B), 3 and 28 d compressive strengths for a total of 264 specimens. Among them, 228 specimens were used for training and validation, while 36 specimens were reserved as a test set. The test set was further divided into in-system and out-of-system specimens to evaluate model stability. For model construction, five nonlinear ML algorithms were employed. The normalized phase composition after gypsum incorporation was used as the input. The 28 d compressive strength was used as the output. Model hyperparameters were determined by Bayesian optimization, and repeated constrained holdout validation was adopted to reduce the influence of random splitting on model evaluation. Model performance was evaluated using mean squared error, root mean squared error (RMSE), mean absolute error, and mean absolute percentage error. After the surrogate model had been determined, it was coupled with a genetic algorithm to search the phase composition space for low-carbon clinkers with a predicted 28 d compressive strength higher than 42.5 MPa and CO₂ emissions lower than 671 or 592 kg CO₂/t. These CO₂-emission thresholds were reduced by 15% and 25%, respectively, relative to Portland clinker.

Results and discussion Dataset analysis showed that the specimens in the dataset exhibited pronounced compositional dispersion and system-level heterogeneity. This heterogeneity was one of the main sources of model instability. The W/B gradually increased from the C3S/C2S-rich region toward regions enriched in aluminate/ferrite phases and C4A3 $\$/C5S2 $\$. The 3 d and 28 d compressive strengths of the training/validation set covered a range of 0–70 MPa, whereas the test set was mainly concentrated in the medium-to-high strength range. Some specimens exhibited strength regression from 3 d to 28 d. These complex strength-development patterns increased the difficulty of model learning but also provided useful information for identifying nonlinear relationships between composition and compressive strength. Validation results showed that, under the conditions of limited data and high compositional heterogeneity, controlled model capacity and moderate regularization were more beneficial for improving prediction stability. The optimized hyperparameters of different models all reflected a tendency to suppress overfitting. Among the models, the artificial neural network (ANN) adopted a relatively small architecture with regularization and achieved the best performance on the validation sets. The test-set results further demonstrated that the ANN had better cross-system generalization capability. Compared with the other models, the ANN achieved an overall RMSE of 3.9 MPa on the test set. Subsequently, the ANN was selected as the$$

surrogate model and coupled with a genetic algorithm for the inverse design of low-carbon clinkers. When the CO₂-emission reduction constraint was tightened to 25%, the genetic algorithm further favored low-carbon phases and generated candidate clinkers dominated by aluminates phases or the C₄A₃S phase, with CO₂ emissions of approximately 555 kg CO₂/t. Experimental validation showed that the measured 28 d compressive strengths of the candidate clinker compositions were close to the model predictions. The prediction errors for all specimens were lower than 10% of the measured strengths.

Conclusions This study constructs a clinker dataset based on a HT preparation procedure and establishes an ML prediction and constrained-optimization framework for screening low-carbon clinkers. The results show that the dataset is characterized by broad compositional coverage, local sparsity, and pronounced system heterogeneity. Therefore, the model needs to balance nonlinear representation capability with regularization constraints. Comparisons on the validation sets and the test set show that the low-capacity ANN with moderate regularization achieves the best performance in terms of prediction accuracy. Its RMSE is 3.9 MPa on the overall test set, and the RMSE values for the in-system and out-of-system subsets are 3.1 MPa and 4.2 MPa, respectively, indicating that it can serve as the surrogate model for inverse design of low-carbon clinker. XGBoost can be regarded as a competitive potential surrogate model, but it tends to produce relatively large prediction errors for some specimens. By coupling the ANN surrogate model with a genetic algorithm, four candidate low-carbon clinkers satisfying the 28 d compressive strength and CO₂ emission constraints are screened. Experimental validation confirms good agreement between the predicted and measured compressive strengths. These results demonstrate that the proposed dataset construction and surrogate model screening strategy can be used to explore novel low-carbon clinkers.

Keywords high-throughput preparation; clinker dataset; predictive model; reverse design; low-carbon clinker screening