

数据驱动范式下化学激发硅铝质胶凝材料的研究进展

王董雨¹, 张祖华¹, 施 成¹, 刘小青², 蒋正武¹

(1. 同济大学材料科学与工程学院, 先进土木工程材料教育部重点实验室, 上海 201804;
2. 中国科学院宁波材料技术与工程研究所, 海洋关键材料全国重点实验室, 浙江 宁波 315201)

摘 要:以地质聚合物为代表的化学激发硅铝质胶凝材料, 制备过程利用了化学激发活性工业固废工艺, 能耗低、碳排少, 产物力学性能与耐久性能优异, 被视为最具发展潜力的绿色胶凝材料之一。然而, 其原料组成复杂、聚合反应机制不明, 要满足多目标性能调控, 传统试错法所需实验量巨大。数据驱动范式依托人工智能与数据科学技术手段, 通过建立高维数据与材料性能之间的映射关系, 有效克服了传统实验方法高成本、长周期等局限, 正引领材料研发从实验驱动向数据驱动演进。本文综述了数据驱动范式在地质聚合物早期性能、力学性能与耐久性能预测中的应用现状, 介绍了元启发式算法驱动组成优化设计的最新研究进展, 总结、展望了数据驱动下化学激发硅铝质胶凝材料应用与发展所面临的关键挑战和未来研究重点。

关键词: 数据驱动; 化学激发硅铝质胶凝材料; 地质聚合物; 性能预测; 组成设计

中图分类号: TU528 文献标志码: A 文章编号: 0454-5648(2026)08-2702-13

网络出版时间: 2026-07-16



数据驱动范式以人工智能与数据科学技术为基础, 通过挖掘并构建高维数据与目标性能之间的统计关联, 深度解析材料成分-结构-工艺-性能间复杂的构效关系, 一定程度上突破了传统材料研发方法中存在高样本量、高成本、长周期、低效率等技术壁垒^[1-3]。这一突破性进展标志着材料研发范式正经历着从实验驱动、理论驱动和计算驱动向数据驱动的演进^[4, 5]。自 2011 年美国材料基因计划启动以来, 以机器学习为代表的驱动方法已在材料科学多个领域多种材料中得到了应用, 其中包括化学激发硅铝质胶凝材料。

化学激发硅铝质胶凝材料是以活性硅铝质原料为主要组分, 在适宜的温度条件下通过碱或酸性激发剂激发, 发生非均相化学反应, 生成的具有水硬性胶凝能力的材料体系。目前, 国内外研究最广、前景最好的一类化学激发硅铝质胶凝材料是碱激发材料, 亦被广泛称为地质聚合物^[6]。这类材料在生产过程中可大规模利用工业固废或低能耗天然

矿物原料, 一般无须高温煅烧, 显著降低了生产能耗与碳排放, 符合当前建材行业节能减排的发展需求^[7]。与传统硅酸盐水泥相比, 除节能减排效益显著外, 地质聚合物在工程应用中还表现出诸多优异性能, 例如, 特定配方下, 产物早期强度发展快、强度高、抗化学腐蚀性强、抗渗透性能好、抗冻性及耐火性能优异等^[8, 9]。

然而, 地质聚合物的研究与应用仍然面临诸多待解决的关键问题。地质聚合物原料种类繁多、来源多样, 在制备过程中的影响因素很多, 包括原材料与激发剂组分、水胶比、砂胶比、碱胶比以及养护环境与条件等, 这显然增加了性能调控的复杂性。传统配方设计多采用试错法、正交试验或单纯形质心设计等方法, 通过反复实验和经验总结研究各个因素对性能的影响, 并据此进行宏观指导。然而, 这些方法并未充分考虑各因素之间的相互影响, 且当原材料的物理化学特性或服役环境发生变化时, 仍需开展大量实验进行适配, 效率低、成本

收稿日期: 2026-02-28。 修订日期: 2026-03-19。

基金项目: 国家自然科学基金(52378257); 中央高校基本科研业务费专项资金资助(22120250415); 中国科学院宁波材料技术与工程研究所海洋关键材料重点实验室开放基金(2024K08); 上海百奥恒新材料有限公司地聚合物基因组项目。

第一作者: 王董雨(1997—), 女, 博士研究生。

通信作者: 张祖华(1983—), 男, 博士, 教授。

Received date: 2026-02-28. Revised date: 2026-03-19.

First author: WANG Dongyu(1997—), female, Doctoral candidate.

E-mail: wangdy@tongji.edu.cn

Correspondent author: ZHANG Zuhua(1983—), male, Ph.D., Professor.

E-mail: zhangzuhua@tongji.edu.cn

高,难以及时满足实际工程中大规模生产及复杂服役环境使用的需求,这也限制了地质聚合物的推广应用。当前,数字与智能技术的迅猛发展为地质聚合物的研发与应用开辟了新路径。材料基因工程、人工智能技术与地质聚合物研究的深度融合,正在推动该领域向数据驱动的研究范式转型,发展地质聚合物的智能化研发技术对于推动此类材料高效设计、制备和应用具有重要意义。

为此,本文围绕数据驱动范式下地质聚合物研究方法,系统综述了其在聚合反应过程关键参数预测、凝结硬化阶段早期性能预测、静动态抗压强度等力学性能预测,以及体积稳定性与抗碳化性能等耐久性能预测方面的研究进展,介绍了元启发式算

法地质聚合物组成正向设计和逆向优化的研究现状,探讨了当前数据驱动方法在地质聚合物应用与发展所面临的关键挑战,并展望了未来研究方向与重点,为后续材料开发与人工智能深度融合提供参考依据。

1 数据驱动方法在地质聚合物研究中的基本框架

如图 1 所示,基于数据驱动的地质聚合物性能预测的机器学习流程包含目标界定、数据集整理、数据预处理、特征工程、模型构建与模型应用 6 个基本步骤。

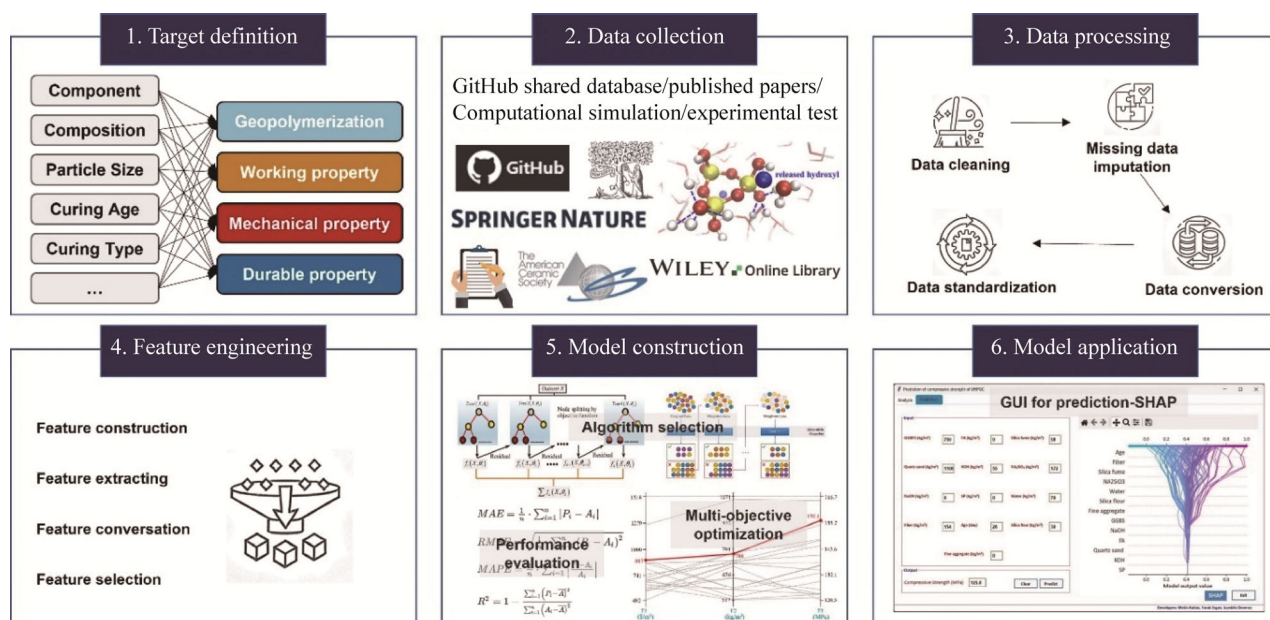


图 1 数据驱动的地质聚合物性能预测机器学习流程

Fig. 1 Machine learning workflow for data-driven performance prediction of geopolymers

目标界定是连接材料科学问题与机器学习方法之间的纽带,根据地质聚合物性能指标确定是回归问题(预测连续值,如抗压强度)还是分类问题(预测类别标签,如抗腐蚀等级),并利用专业知识筛选影响性能的关键参数,将复杂的材料问题转为结构化的机器学习任务^[10]。数据是机器学习的重要驱动力,其质量、数量、维度甚至分布形式直接影响模型的学习效果及最终结果的准确性。目前用于地质聚合物的研究数据主要来源于 GitHub 等平台的开放数据库、已发表的文献、计算模拟数据以及实验测试结果。然而,由于实验条件、材料组成的差异性,数据中常存在异常值、噪声、缺失值等问题,

导致数据质量较差;同时,针对特定目标构建的数据集样本规模有限,特征维度却较高。针对上述问题,数据预处理结合特征工程的协同优化为解决数据质量带来的挑战提供了有效方法。前者通过数据清洗、数据标准化及数据平衡以消除噪声,后者通过特征筛选与降维处理以减少数据冗余与离散^[11]。二者共同作用为建立面向地质聚合物目标性能的机器学习模型提供高质量的数据支撑与特征保障。近年研究进一步探索了数据填充与数据增强技术,为小样本数据的优化与扩充提供了有效途径。Abdellatief 等^[12]通过实验补充结合高斯噪声增强,生成与原始样本分布一致的新样本,实现了偏高岭

土基地质聚合物混凝土抗压强度预测数据的高效扩充。Miyan 等^[13]采用链式方程多重插补等方法填补粉煤灰基地质聚合物抗压强度数据集中的缺失值,并基于生成对抗网络进行数据增强,将数据集规模扩大 1 倍。余睿等^[14]构建了条件生成对抗网络与马尔可夫链蒙特卡洛算法协同驱动的数据增强模型,将训练集数据规模扩充至 1500 组,增强后数据分布与原始数据保持高度一致性与同源性,显著提升了训练出的模型预测精度。

如何利用优质的输入特征去关联地质聚合物目标性能、建立映射关系,则依赖于模型的有效构建。首先,应为建立模型选择合适的算法。常用算法包括:线性回归、决策树、k 邻近(kNN)、支持向量回归(SVR)等传统模型;人工神经网络(ANN)、深度神经网络(DNN)、反向传播神经网络(BPNN)、卷积神经网络(CNN)、极限学习机(ELM)等神经网络模型;以及随机森林(RF)、极限梯度提升(XGBoost)、轻量级梯度提升机(LightGBM)、类别型梯度提升(CatBoost)等集成学习算法。确定算法后,通常使用随机抽样等方法对原始数据集拆分,其中训练集用于模型训练,而测试集用于性能验证。为避免由于数据拆分带来的误差,常采用 K 折交叉验证法,通过多次迭代训练提升评估模块的有效性与适应性。训练模型时,超参数优化是提升预测性能的关键环节,常用方法包括网格搜索、随机搜索、贝叶斯优化及元启发式算法等。尤其是元启发式算法,在应对高维、非线性复杂搜索问题时展现出独特优势。典型的此类算法包括遗传算法(GA)、粒子群优化算法(PSO)、模拟退火算法(SAO)、海洋捕食者算法(MPA)、人工蜂群算法(ABC)以及混合元启发式算法等。此外,机器学习结合元启发式算法也可用于地质聚合物的逆向配合比设计,以满足针对单一甚至多目标性能优化的要求(详见第 3 节)。为了确定最优模型,在选取评价标准时可以使用决定系数(R^2)、平均绝对误差(MAE)、均方差(RMSE)、皮尔逊相关系数(R)等评价指标综合评估模型的预测性能,以及模型的适应性和泛化能力。在模型应用阶段,训练好的模型可以用来对未来新数据点进行预测,进而有效指导新型地质聚合物的设计。为辅助理解模型的学习机制与决策依据,可以通过沙普利可加性解释(SHAP)和部分依赖图(PDP)等方法对模型结果进行解释,而解释的结果需要结合领域专业知识来评估输入特征在决策过程中的可靠性^[15, 16]。

2 机器学习在地质聚合物性能预测中的应用

机器学习在地质聚合物研究中应用的报道始于 2012 年^[17],在近几年发展迅速。随着机器学习的发展和研究的深入,应用范畴已从材料的基础的工作性、力学性能及耐久性能预测,扩展至裂纹监测、断裂韧性评估等更具挑战性的结构健康监测前沿^[18, 19],显示出该领域不断深化的发展趋势。附表 1 (见补充材料)系统总结了常用机器学习算法在该领域性能预测方面的研究进展。

基于不同模型的应用特征,可将现有方法归纳为神经网络、集成学习、传统回归与统计学习、符号回归与神经模糊系统 4 类,各类方法均有其适用场景与限定条件:1) 神经网络模型适用于捕捉输入变量与目标性能之间的非线性关系^[20],在地质聚合物性能预测中应用最为广泛,涵盖反应机理、凝结时间、坍落度、干燥收缩及抗压强度、抗折强度、劈裂抗拉强度等力学性能指标。该方法要求输入变量为数值型工艺或成分参数,对分类型变量适配性较弱,若需使用应进行特征转换。神经网络通常在较大样本量下预测效果更优,而当前研究数据集多集中于小规模样本,限制了其性能发挥。此外,模型可解释性较差,难以揭示性能演化的内在机理^[21]。2) 集成学习模型通过构建多个基学习器并采用特定融合策略进行集成,可获得性能更优、泛化能力更强的综合模型。该方法适用于中到大规模样本且输入特征维度较高的场景,对数值型与分类型混合特征具有良好的适应性,无须复杂的特征转换^[22]。目前主要应用于地质聚合物抗压强度预测,少量研究拓展至坍落度与碳化性能。相较于神经网络,集成学习模型泛化能力更强、抗噪声性能更优,但模型复杂度较高、超参数众多,需进行调优,且在小样本下可能存在过拟合风险^[22, 23]。3) 传统回归与统计学习模型在地质聚合物的早期性能、力学性能与耐久性能预测中均有应用,可兼顾多类性能预测需求。线性回归、kNN 等方法模型结构简单、易于操作,适用于小样本预实验。然而,该类方法对高维特征及大样本数据适配性较差,计算效率低于集成学习,且预测精度随问题复杂度上升而下降^[24, 25]。4) 符号回归与神经模糊系统是地质聚合物性能预测中较早应用的算法,主要聚焦于抗压强度预测,输入特征以数值型为

主。该方法适用于中小规模样本, 样本量过大时无明显预测优势; 在高维输入场景下, 模糊规则数量呈指数增长, 导致模型难以扩展^[21, 26]。

2.1 早期性能预测

地质聚合物的形成是硅铝质前驱体在激发剂作用下 Si-O-Al 和 Si-O-Si 发生断裂解聚, 生成硅氧四面体 $([\text{SiO}_4]^{4-})$ 和铝氧四面体单体 $([\text{AlO}_4]^{5-})$, 最终缩聚形成三维网络结构的化学过程。由于这一反应过程较为剧烈, 反应热的研究对于了解其早期性能具有重要意义。通过等温量热法获取反应放热随时间变化的曲线, 可以反映出聚合反应发生的时间、程度和速率等关键信息。Ling 等^[27]基于粉煤灰基地质聚合物的反应热数据, 以溶解峰和聚合峰的热流率与时间为目标特征, 激发剂模数、碱浓度、碱溶液与粉煤灰的液固比和温度为输入特征, 构建 ANN 模型并系统分析了输入特征与地聚合反应过程的相关性。结果表明, ANN 模型能够准确预测地聚合反应性能与进程, 为量化地质聚合物反应程度、优化配合比设计提供了有效的技术方法。在此数据基础上, Tanyildizi^[28]进一步引入粉煤灰化学组成 (包含 CaO 、 SiO_2 、 Al_2O_3 、 Fe_2O_3 、 MgO 和 Na_2O 含量) 作为输入参数, 分别用深度长短期记忆网络 (deep LSTM)、SVR 和 kNN 进行建模并比较了不同影响因子对预测反应程度的作用, 发现 deep LSTM 模型预测效果最佳, 能够实现对聚合过程的较高准确度预测, 该模型对于聚合峰时间、聚合峰热流率、溶解峰时间和溶解峰热流率 4 个关键参数的预测精度都达到 99.55% 以上。敏感性分析表明, 液固比和养护温度是影响聚合峰时间和热流率的最主要因素, 而溶解峰参数主要受粉煤灰化学组成和养护温度的控制。Chen 等^[24]同样对上述 4 个关键参数进行了预测, 发现 RF 模型在所有模型中表现最为稳定, 预测效果最佳, 模型相关系数依次为 0.88、0.95、0.92 和 0.95 (图 2)。SHAP 重要性分析也表明, 养护温度和液固比是聚合反应最重要的影响因子, 其中液固比越小, 反应时间越短, 地聚合程度越高。此外, 激发剂的模数和碱浓度均与溶解峰指标呈负相关。显然, 准确预测与分析地质聚合反应的关键参数是实现材料组成精细化优化的重要理论依据。目前应考虑化学组成之外、探究矿物组成 (活性差异) 的影响并适当拓展至偏高岭土、矿渣等原材料及其多元复合体系, 进一步探索学习模型在不同体系地质聚合物中的普适性与可操作性。

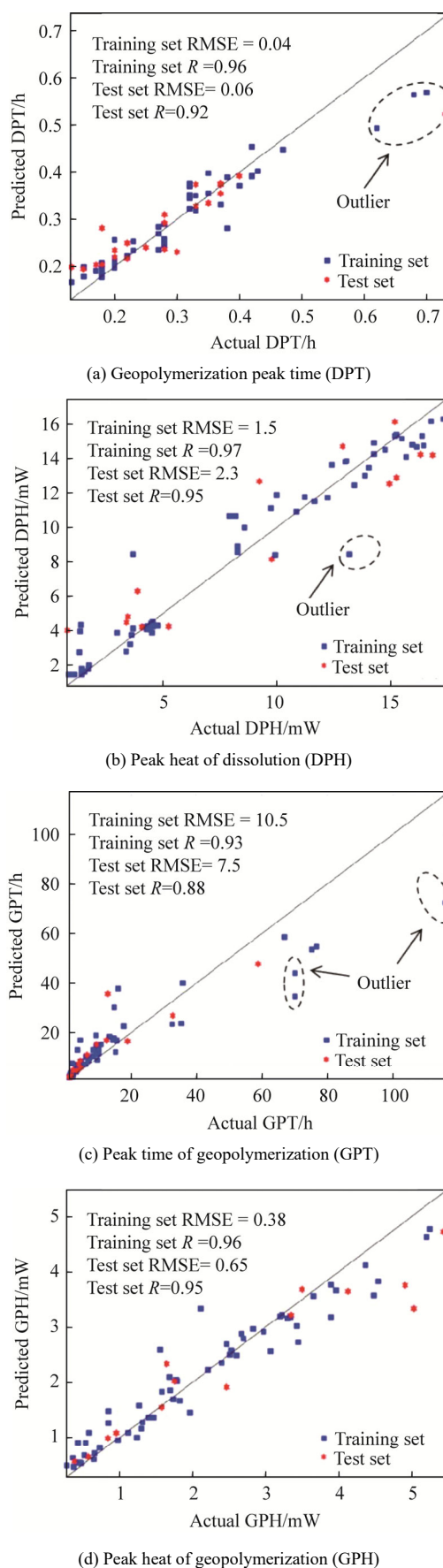


图 2 地质聚合物的聚合反应过程预测^[24]
Fig. 2 Prediction of the geopolymerization process^[24]

地聚合反应的动力学特征在宏观上表现为新拌浆体的凝结特性与流变性。Ling 等^[27]将模数、碱浓度和液固比作为 ANN 模型的输入层, 研究这三者对于凝结时间的影响, 发现 ANN 模型的预测结果与实验测试的凝结时间具有高度的一致性, R^2 达 0.99。3 种变量中, 液固比对凝结时间的影响最大, 表现为粉煤灰地质聚合物的凝结时间随着液固比的增大呈指数增长。此外, 激发剂模数也是影响凝结时间的重要因素。凝结时间随模数从 1.0 增加至 2.0 呈先减小、后增大、再减小的变化趋势。在模数的适用区间内, 凝结时间随模数增大而延长^[27]。当模数较低时, 体系 pH 值往往较高, 有利于前驱体中硅氧和铝氧四面体的溶解, 从而增大浆体的屈服应力与塑性黏度^[29, 30], 此时碱性激发剂中溶解的硅酸盐可发挥润滑与分散作用。然而, 当模数较高时, 体系 pH 值降低, 激发剂中高聚合度硅酸盐基团增多, 抑制前驱体颗粒的溶解, 导致地质聚合物凝结时间延长^[31]。

当前, 地质聚合物混凝土的工作性能研究主要集中于原材料组成设计, 而前驱体及激发剂本

身的化学物理特性, 以及搅拌方式(包括投料顺序、搅拌速度和时间), 同样是影响其工作性能的重要因素。为了考虑复杂工艺参数对工作性的影响, Gomaa 等^[32]利用机器学习方法, 整合了涵盖粉煤灰特性、激发剂组成及 8 种不同搅拌流程等数据集, 并采用 RF 模型结合网格搜索法与十折交叉验证法对模型的超参数进行严格优化, 实现了粉煤灰地质聚合物混凝土坍落度的高精度预测, 其 R 达 0.95。由此可见, 基于数据驱动方法可以更好地关联地质聚合物复杂的“组成–工艺–性能”关系, 通过组成与工艺的协同优化, 或可实现早期性能的配合比精准设计。

2.2 力学性能预测

抗压强度作为地质聚合物最基本的力学性能指标, 既是配合比设计中的首要优化目标, 也是连接材料组成、微观结构与结构性能的关键性质, 因而成为机器学习在本领域中最早实现预测、应用最广泛的输出特征。在地质聚合物抗压强度预测的早期研究中, ANN 和 ANFIS 是常用的经典模型, 如图 3 所示。Van 等^[26]以粉煤灰、NaOH、硅酸钠和

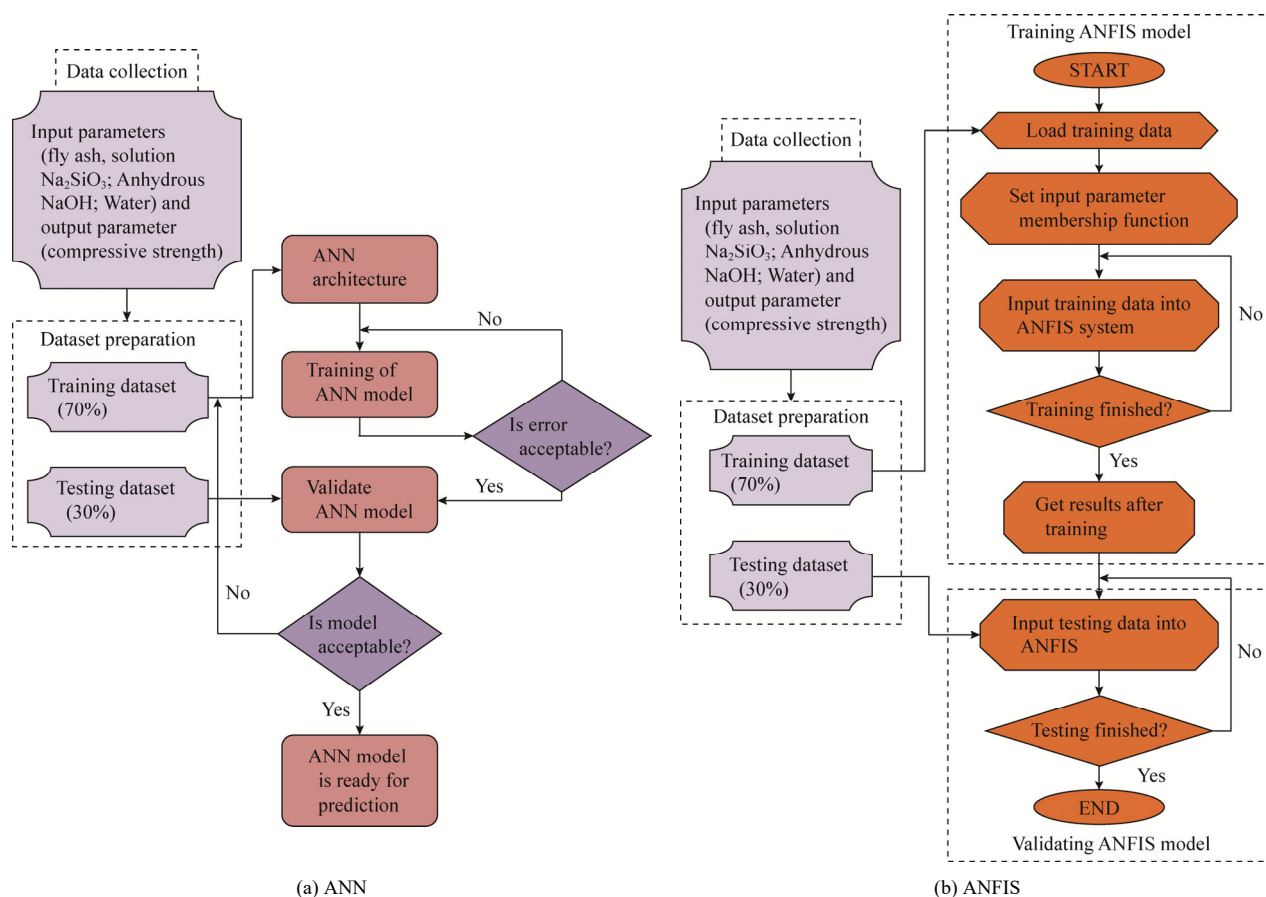


图 3 粉煤灰地质聚合物抗压强度预测方法^[26]

Fig. 3 Prediction methodology for compressive strength of fly ash based geopolymer^[26]

水 4 个变量作为输入参数,按 25、35、45 MPa 3 种目标强度制备了 210 组试件,并以实测抗压强度作为模型的输出目标,发现 ANN 和 ANFIS 均能有效预测抗压强度,且 ANFIS 模型的表现优于 ANN,其抗压强度评价指标分别为 $MAE=1.99$ MPa、 $RMSE=2.42$ MPa、 $R^2=0.85$ 。随着人工智能技术的发展,地质聚合物力学性能预测模型逐渐引入关联结构更深的神经网络,如 DNN、BPNN、CNN 等,以及 RF、XGBoost、LightBoost 等树模型。研究方法也从单一模型比较发展到多模型对比乃至模型堆叠融合。例如, Peng 等^[33]采用 BPNN、SVM 和 ELM 模型,以粉煤灰(掺量及化学组分)、激发剂(NaOH 与硅酸钠用量、碱含量、模数)、粗细骨料、液固比、水、外加剂、温度与养护龄期等 14 个参数作为输入变量,通过划分训练集、验证集和测试集预测地质聚合物的抗压强度,发现 SVM 模型具有最稳定的预测效果。尽管 BPNN 和 ELM 模型的整体 R^2 高于 SVM,但二者对验证集的预测结果显著下降,模型泛化性较弱。

集成学习算法在优化整体性能上可实现进一步突破。Abdellatif 等^[23]将集成学习算法(XGBoost、RF)在超高性能地质聚合物混凝土(UHPC)体系抗压强度预测性能与支持向量回归(SVR)进行了对比,发现 XGBoost 模型在 UHPC 体系的训练集和测试集中对抗压强度预测的 R 值高于 SVR,且 MAE 和 RMSE 更低,表明该模型在高强度预测时仍能保持良好的可靠性与泛化能力。Dai 等^[20]建立了 SVR、RF、GB 及堆叠模型用于预测地质聚合物混凝土的动态抗压强度,发现堆积模型通过融合多种机器学习算法表现最优,其测试集评价指标分别为 $R^2=0.98$ 、 $RMSE=5.42$ MPa、 $MAE=3.84$ MPa、平均绝对百分比误差(MAPE)=5.92%。尽管堆叠模型可以综合各方法的优势并发挥协同效应,但在小样本数据集上,其适用性仍存在一定局限^[34,35]。

综上所述,机器学习在地质聚合物的力学性能研究中的应用,已由最初抗压强度预测逐步拓展至抗折强度、劈裂拉伸强度、直接拉伸强度、弹性模量、动态抗压强度等多个指标。然而,面向新型地质聚合物体系及其应用拓展,原材料复杂与数据稀缺仍是制约机器学习实现精准预测的关键瓶颈。迁移学习方法通过将源域知识迁移至数据稀缺的目标域模型构建中,或可有效克服数据不足问题,提升目标域内模型预测的稳定性。Nguyen 等^[36]以 1029 组普通地质聚合物混凝土(抗压强度为

2.7~89.6 MPa)作为源域,101 组 UHPC (抗压强度为 56~165 MPa)作为目标域,通过计算数据集间差异值确保特征分布相近后,构建了用于预测 UHPC 抗压强度的迁移学习模型。结果表明,该迁移学习模型的预测精度优于传统机器学习模型,其 RMSE 降低了 10%~30%。此外,基于 SHAP 分析显示,迁移模型能够挖掘特征之间更为复杂的关系,相较于传统模型能更有效地揭示输入特征与抗压强度之间的关联。

2.3 耐久性能预测

地质聚合物在凝结硬化过程及硬化后期存在多种形式的体积变形,其收缩率可达硅酸盐水泥基材料的 3~6 倍^[37,38]。地质聚合物的收缩受聚合反应路径、凝胶结构演化及外部养护条件等诸多因素的影响,相较于传统水泥基材料更为复杂^[39]。传统经验模型难以精准描述这种非线性关系,因此亟须找到一种新的研究方法。机器学习技术的兴起为此提供了有力工具。Adesanya 等^[40]以矿渣与粉煤灰掺量、激发剂模数和养护温度为输入,建立了基于 ANN 的干燥收缩预测模型,用于训练、验证、测试的模型 R 值均高于 0.99 (图 4)。该模型能有效预测 28 d 干燥收缩值,通过与多元线性回归模型对比发现,ANN 模型的预测精度更高,与实验数据相关性更强。Shen 等^[41]则进一步采用 SVR、GPR、CART、RF、GB 和 XGB 6 种机器学习算法,构建了针对矿渣-粉煤灰地质聚合物净浆和砂浆的自收缩预测模型。这些模型综合考虑了包括胶凝材料组成、化学氧化物物质的量比、激发剂模数、碱含量、养护龄期等 10 项关键参数,并利用特征重要性分析量化了各因素对自收缩性能的影响。基于优选模型,其预开发了图形用户界面(GUI),为工程应用提供了高效便捷的预测工具。

除体积稳定性外,碳化是导致地质聚合物与钢筋混凝土性能劣化的另一关键因素。地质聚合物钢筋混凝土的碳化速率是普通硅酸盐水泥体系的 5 倍左右,抗碳化性明显偏低^[42](碳化后的地质聚合物护筋性能不在本文讨论范围),准确预测其碳化行为具有重要意义。近年来,有研究尝试基于机器学习方法构建地质聚合物碳化深度预测模型^[43]。该模型以材料配比、养护条件与环境参数作为输入,碳化深度作为输出,采用堆叠集成技术融合 RF、径向基函数、SVM 3 种算法,以提成预测性能。结果显示,堆叠模型有效整合了各基础算法的优势,在碳化性能预测中表现出更高的准确性和可靠性,其 R^2

提升至 0.92, MSE 降为 0.025 mm。此外, 该模型在衍生数据上表现良好, 具备较强的泛化能力, 能够通过新增数据持续优化整体预测性能。

机器学习在地质聚合物耐久性研究中的应用尚处于初步探索阶段。尽管现有研究在体积稳定性与碳化性能预测方面已展现出潜力, 但研究数据量

依然较少, 在抗氯离子侵蚀、耐硫酸盐侵蚀等核心工程耐久性问题上仍存在显著空白。未来研究亟须向这些关键性能拓展, 通过构建覆盖多场景的耐久性预测模型, 将机器学习从理论预测工具转化为保障结构长期安全的有效手段, 这应是领域内下一阶段发展和突破的重点方向。

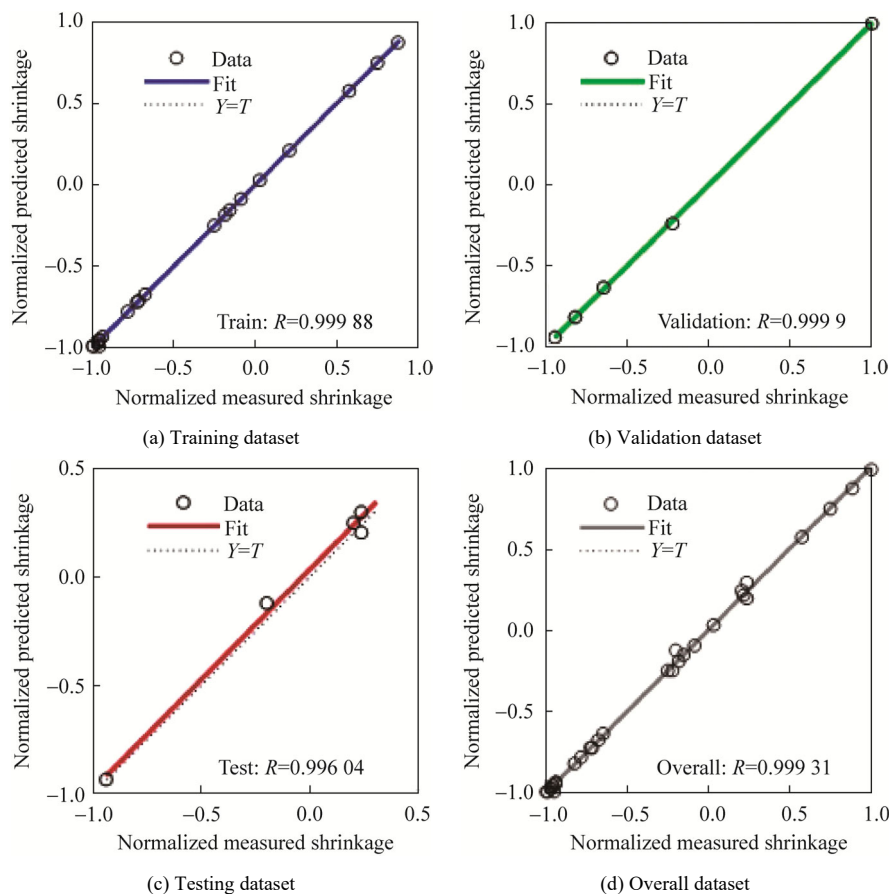


图 4 基于 ANN 模型的矿渣-粉煤灰地质聚合物砂浆干燥收缩回归图^[40]

Fig. 4 Regression plot of drying shrinkage for slag-fly ash based geopolymer mortar by ANN model^[40]

3 元启发式算法驱动的地质聚合物组成优化

近年来, 将元启发式算法与机器学习模型相融合, 通过超参数调优, 成为提升模型预测精度与适应性的重要手段。Diksha 等^[44]使用 GA 算法对 SVR 模型进行超参数寻优, 深度挖掘了地质聚合物配合比与性能间的非线性关系, 实现了模型准确性、适应性与泛化性的平衡^[44]。为进一步提升优化效果, Dhawan 等^[45]创新式融合了具有局部精细化优势的 SAO 与全局搜索能力强的海洋捕食者算法(MPA), 构建了 MPA-SAO 混合算法用于机器学习模型的超

参数调优, 发现优化后的模型对地质聚合物抗压强度的预测精度显著提高。优化后模型的 R^2 达到 0.98, 较基准模型平均提升 6.52%; RMSE 降至 0.02 MPa, 平均降幅达 66.67%; MAE 为 0.01 MPa, 降幅达 66.67%。此外, Parhi 等^[46]采用斑鬣狗优化算法对 AdaBoost、RF 和 XGBoost 3 种基模型进行超参数优化, 并引入 ELM 作为元学习器, 将优化后的基模型堆叠集成混合模型, 显著提升了模型的准确性、适应性与泛化能力。得益于高精度的预测模型, 将模型与可解释性方法或响应面法相结合, 可进一步识别并量化关键影响因素, 有望正向指导地质聚合物配合比设计。基于此模型, 结合准蒙特卡洛敏感性分析, 发现养护温度、水玻璃与 NaOH

浓度比、NaOH 模数及减水剂掺量是关键性影响因素。进一步借助响应面法,建立了以粉煤灰和矿渣作为前驱体的抗压强度响应面模型。结果表明,当养护温度控制在 25~90 °C、水玻璃与 NaOH 浓度比为 8~16 mol/L、NaOH 模数为 1.5~4.0、减水剂掺量为 2%~10%时,地质聚合物的抗压强度可达 39~43 MPa^[46]。

机器学习与元启发式算法的融合不仅限于地质聚合物性能预测及正向组成优化指导,还可拓展至逆向配合比设计(见补充材料中附表 2)。相较于

正向性能预测,逆向设计流程需另外包含目标函数确定、约束条件构建、元启发式算法选择等关键步骤,并通过收敛性分析与试验验证对生成的配合比进行评估。如图 5 所示, Ji 等^[22]基于 XGBoost 算法建立了地质聚合物混凝土坍落度与抗压强度预测模型,随后将训练好的模型作为目标函数,引入 PSO 算法进行配合比自动寻优。针对 20、40 和 60 MPa 3 个目标强度等级,成功生成了同时满足强度要求与坍落度(>180 mm)的优化配合比,实现了高效、精准的配合比自动设计^[22]。

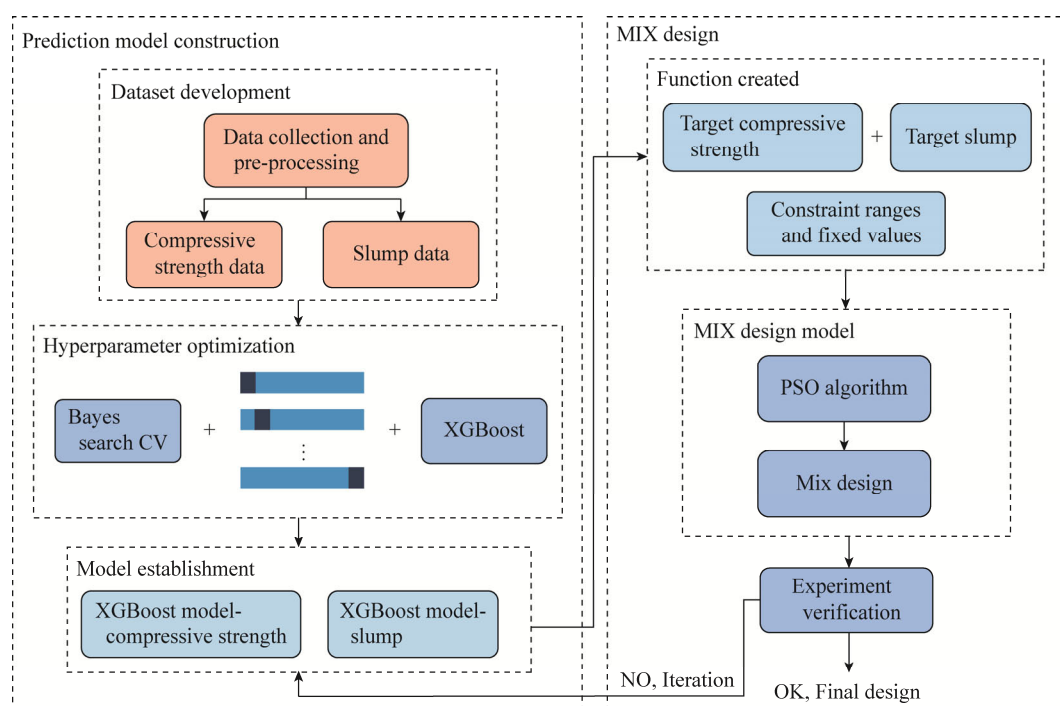


图 5 基于 PSO 算法的工作性与强度协同优化的地质聚合物配合比设计框架图^[22]

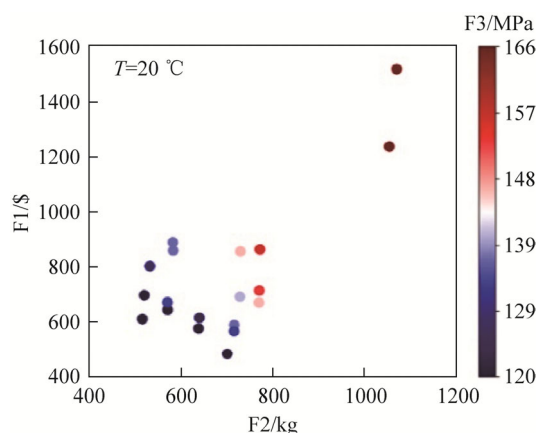
Fig. 5 A framework for geopolymer mix proportion design with collaborative optimization of workability and strength based on PSO algorithm^[22]

随着“双碳”目标与复杂工程服役环境的双向驱动,开发兼具(超)高性能、低碳环保与低成本优势的新型地质聚合物,已成为建筑材料领域的研究热点。Wu 等^[47]提出了集成自动化机器学习建模与多目标优化的智能框架。该框架首先利用 GA 算法的交叉和变异操作优化超参数,构建高精度的预测模型。继而引入帕累托(Pareto)支配准则,以强度最大化、碳排放与成本最小化为目标进行多目标优化,指导地质聚合物配合比决策。Zhang 等^[48]、Guo 等^[49]和 Dai 等^[20]也开展了基于 GA 算法与 Pareto 前沿结合的配合比优化研究。在多目标优化研究中,非支配排序遗传算法(NSGA-II)因其精英策略与快速非支配排序技术,能够在保留优质个体的同时降

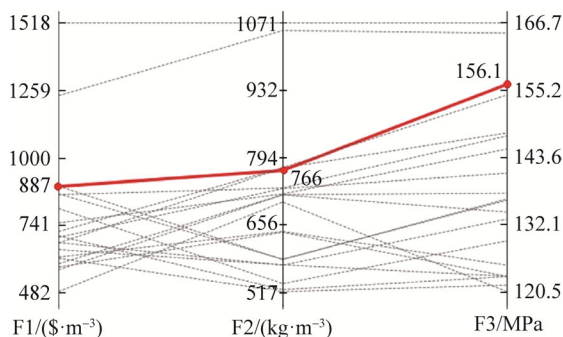
低计算复杂度,提升优化效果与适应性,因而被广泛使用^[50]。然而,多目标优化通常生成一组 Pareto 非劣解,需进一步借助优劣解距离法(TOPSIS)等进行最终决策,从中筛选出最优解,进而实现材料的高效设计(图 6)^[20, 49]。

元启发式算法的组成优化可从正向和逆向 2 个方向快速指导地质聚合物的配合比设计。随着人工智能技术的发展,基于目标性能逆向设计材料的算法日趋完善,材料逆向设计正逐步取代正向筛选,成为推动该领域从传统方法向人工智能驱动变革的核心方向。深度生成模型通过学习和模拟真实材料数据的概率分布,能够根据目标性能条件直接生成对应的材料组成,是实现可控性能逆向设计的最

前沿方向之一。其中,可逆神经网络(INN)、变分自编码器(VAE)、生成式对抗网络(GAN)、扩散模型等在材料逆向设计中受到广泛关注^[51]。以扩散模型为例,其通过具有物理启发性的逐步加噪与去噪过程学习数据分布,结合噪声预测网络与去噪算法,能够生成具有复杂约束与周期性结构的精准晶体材料^[52]。然而,该类深度生成模型在化学激发硅铝质胶凝材料体系中的研究尚属空白。若将该技术引入并发展,有望突破现有方法的设计空间局限,为实现材料体系结构的精准高效设计提供新路径。



(a) Pareto-optimal solution set from the multi-objective optimization



(b) Non-dominated solutions (in gray) and TOPSIS-recommended solution (in red)

图6 UHPGC 配合比多目标优化^[49]

Fig. 6 Multi-objective optimization of mix proportions for UHPGC^[49]

4 结论与展望

纵观数据驱动方法在以地质聚合物为代表的化学激发硅铝质胶凝材料体系的应用,可将其发展脉络归纳为以下3个方面:在原材料上,相关研究起始于粉煤灰地质聚合物体系,而后发展至矿渣-粉煤灰等二元地聚物体系,乃至矿渣-粉煤灰-硅灰等三元体系;在目标性能上,已由最初的抗压强度预测拓展至早期性能与耐久性能的预测,并进一

步结合可解释性分析与多目标优化算法,实现面向目标性能的逆向配合比设计;在材料类型上,应用范围已由传统净浆/砂浆延伸至自密实地质聚合物混凝土、工程地质聚合物混凝土、轻质地质聚合物以及 UHPGC。基于以上发展现状分析,本文的主要结论如下:

1) 数据驱动方法在地质聚合物材料研究中已初步形成了标准化流程,包括目标界定、数据集整理、数据预处理、特征工程、模型构建与模型应用6个关键步骤。

2) 数据驱动方法在地质聚合物性能预测中的应用与发展差异化显著。在早期性能方面,机器学习在实现地聚合反应过程、凝结时间、工作性能的预测效果显著;在力学性能方面,应用最为成熟,已由最初的抗压强度预测,逐步拓展至抗折强度、劈裂拉伸强度、弹性模量、动态抗压强度等多个力学指标;耐久性预测研究尚处于初步阶段,仅少量研究涉及收缩性能、碳化,尚未涉及抗氯离子侵蚀、耐硫酸盐侵蚀等问题。

3) 元启发式优化算法在地质聚合物组成优化中发挥着双重作用。一方面通过超参数寻优可显著提升模型的预测能力。另一方面基于优选预测模型结合目标函数确定、约束条件构建及算法选择可实现配合比的逆向设计。将优化算法与 Pareto 前沿、TOPSIS 决策方法相结合,已成为解决多目标(性能、成本、碳排放)协同优化的主流范式。

数据驱动方法在以地质聚合物为代表的化学激发硅铝质胶凝材料体系性能预测与组成优化中展现出巨大潜力。然而当前研究仍处于小样本验证的可行性探索阶段,如何充分发挥智能技术优势,实现材料开发与人工智能技术的深度融合与创新应用,仍需大量深入研究。纵观现有成果,人工智能技术在该材料体系的应用面临以下亟待破解的关键挑战:

1) 化学激发硅铝质胶凝材料因前驱体与激发剂组成复杂、原材料地域性差异显著,目前尚无统一规范的设计标准。而材料性能受化学组成、物理特性、物相分布等多因素影响,描述其复杂机制需引入大量特征,导致数据集呈现高维特性,易引入冗余信息,降低模型性能或效率。针对上述难题,可从数据与模型2个层面寻求突破:数据层面,亟须建立面向该材料体系的标准化设计规范与开放共享数据库系统,采集前驱体的化学组成、物理特性、物相组成及反应活性等核心指标,同时涵盖激

发剂参数与养护条件等关键变量,通过规范数据采集标准为模型构建提供高质量数据支撑;模型层面,应将聚合反应机理、热力学约束及前驱体活性指标等领域知识有机融入算法框架,将反应机理转化为模型的结构约束或损失函数项,构建“数据+知识”双驱动的混合模型,在保障预测精度的同时提升模型可解释性,实现材料知识贯穿机器学习全过程。

2) 面向新型化学激发胶凝材料体系的应用与开发,尤其是其抗氯离子侵蚀、耐硫酸盐侵蚀等核心工程耐久性问题,实验数据规模有限。破解这一困境在于:数据层面,一方面加强数据填充与数据增强技术的应用,通过生成式对抗网络、高斯噪声增强等方法扩充有效样本规模,另一方面将实验数据与分子动力模拟、热力学模拟、有限元模拟等多元数据相结合,加强数据的深度挖掘。模型层面,基于小样本实验数据,加大迁移学习等小样本学习算法的开发力度。

3) 当前逆向设计主要依赖于预测与优化相结合的两步法框架,而深度生成模型能够直接学习数据分布并生成符合目标性能的全新材料组成,是实现可控性能逆向设计的前沿方向。然而该方法在化学激发硅铝质材料体系的研究上属空白。如何将深度生成模型与化学激发胶凝材料体系的反应动力学、微结构演化相融合,揭示其从早期聚合到长期耐久的全寿命性能演化机制,从而实现静态性能预测、动态寿命评估及逆向组成设计的根本性跨越,是推动该领域智能化变革的关键问题。

补充材料

本文的电子补充材料见网络

<https://link.cnki.net/urlid/11.2310.tq.20260715.1319.001>

参考文献:

- [1] DERINGER V L, BERNSTEIN N, CSÁNYI G, et al. Origins of structural and electronic transitions in disordered silicon[J]. *Nature*, 2021, 589(7840): 59–64.
- [2] 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.
- [3] MINKYUNG B, FRANK D, IVAN A, et al. Accurate prediction of protein structures and interactions using a three-track neural network[J]. *Sci N Y N Y*, 2021, 373(6557): 871–876.
- [4] HIMANEN L, GEURTS A, FOSTER A S, et al. Data-driven materials science: Status, challenges, and perspectives[J]. *Adv Sci*, 2019, 6(21): 1900808.
- [5] SCHMIDT J, MARQUES M R G, BOTTI S, et al. Recent advances and applications of machine learning in solid-state materials science[J]. *npj Comput Mater*, 2019, 5: 83.
- [6] PROVIS J L, BERNAL S A, ZHANG Z H. The decarbonization of construction: How can alkali-activated materials contribute [J]. *Engineering*, 2024, 37: 18–21.
- [7] PROVIS J L, BERNAL S A. Geopolymers and related alkali-activated materials[J]. *Annu Rev Mater Res*, 2014, 44: 299–327.
- [8] WANG A G, ZHENG Y, ZHANG Z H, et al. The durability of alkali-activated materials in comparison with ordinary Portland cements and concretes: A review[J]. *Engineering*, 2020, 6(6): 695–706.
- [9] WANG D Y, ZHANG Z H, SHI C, et al. Comparison study of ultra-high performance geopolymer concrete (UHPGC) and ultra-high performance concrete (UHPC): Mechanical properties, durability and carbon emissions[J]. *Compos Part B Eng*, 2025, 307: 112903.
- [10] 刘悦, 邹欣欣, 杨正伟, 等. 材料领域知识嵌入的机器学习[J]. *硅酸盐学报*, 2022, 50(3): 863–876.
- [11] LIU Yue, ZOU Xinxin, YANG Zhengwei, et al. *J Chin Ceram Soc*, 2022, 50(3): 863–876.
- [12] 张文生, 曹傅荔, 郅晓, 等. 机器学习方法用于水泥基材料的研究进展[J]. *硅酸盐学报*, 2024, 52(11): 3617–3630.
- [13] ZHANG Wensheng, CAO Fuli, ZHI Xiao, et al. *J Chin Ceram Soc*, 2024, 52(11): 3617–3630.
- [14] ABDELLATIEF M, ELSAFI M, MURALI G, et al. Comparative evaluation of hybrid machine learning models for predicting the strength of metakaolin-based geopolymer concrete enhanced with Gaussian noise augmentation[J]. *J Build Eng*, 2025, 111: 113302.
- [15] MIYAN N, KRISHNAN N M A, DAS S. Integrating data imputation and augmentation with interpretable machine learning for efficient strength prediction of fly ash-based alkali-activated concretes[J]. *J Build Eng*, 2024, 98: 111248.
- [16] 余睿, 李旺, 范定强. 基于条件生成对抗网络与马尔可夫链蒙特卡罗双驱动的超高性能混凝土数据增强及智能预测方法[J]. *硅酸盐学报*, 2026, 54(3): 842–856.
- [17] YU Rui, LI Wang, FAN Dingqiang. *J Chin Ceram Soc*, 2026, 54(3): 842–856.
- [18] 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.
- [19] LYNGDOH G A, ZAKI M, KRISHNAN N M A, et al. Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning[J]. *Cem Concr Compos*, 2022, 128: 104414.
- [20] NGUYEN H A T, PHAM D H, AHN Y, et al. Machine learning and sustainable geopolymer materials: A systematic review[J]. *Mater Today Sustain*, 2025, 30: 101095.
- [21] ARAVIND N, NAGAJOTHI S, ELAVENIL S. Machine learning model for predicting the crack detection and pattern recognition of geopolymer concrete beams[J]. *Constr Build Mater*, 2021, 297: 123785.
- [22] ZIAMIAGHBI B, TOUFIGH V. Fracture toughness evaluation of ground granulated blast furnace slag concrete using experimental study and machine learning techniques[J]. *Eng Fract Mech*, 2023, 291: 109577.
- [23] DAI F, GAI W J, YANG S H, et al. Machine learning-driven multi-objective optimization and dynamic performance prediction of geopolymer concrete[J]. *J Clean Prod*, 2026, 540: 147501.
- [24] ZHENG S, LIU Z B, HUANG J K, et al. A review of machine learning

- methods for geopolymers performance prediction: Linking data-driven models with physical insights[J]. *J Build Eng*, 2025, 114: 114227.
- [22] JI H D, LYU Y H, YING W C, et al. Machine learning guided iterative mix design of geopolymer concrete[J]. *J Build Eng*, 2024, 91: 109710.
- [23] ABDELLATIEF M, HASSAN Y M, ELNABWY M T, et al. Investigation of machine learning models in predicting compressive strength for ultra-high-performance geopolymer concrete: A comparative study[J]. *Constr Build Mater*, 2024, 436: 136884.
- [24] CHEN K, CHENG Y H, YU M S, et al. Predicting the geopolymerization process of fly-ash-based geopolymer using machine learning[J]. *Buildings*, 2022, 12(11): 1792.
- [25] TOUFIGH V, JAFARI A. Developing a comprehensive prediction model for compressive strength of fly ash-based geopolymer concrete (FAGC)[J]. *Constr Build Mater*, 2021, 277: 122241.
- [26] VAN D D, LY H B, TRINH S H, et al. Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete[J]. *Materials*, 2019, 12(6): 983.
- [27] LING Y F, WANG K J, WANG X H, et al. Prediction of engineering properties of fly ash-based geopolymer using artificial neural networks[J]. *Neural Comput & Applic*, 2021, 33(1): 85–105.
- [28] TANYILDIZI H. Predicting the geopolymerization process of fly ash-based geopolymer using deep long short-term memory and machine learning[J]. *Cem Concr Compos*, 2021, 123: 104177.
- [29] ZUO Y B, NEDELJKOVIĆ M, YE G. Pore solution composition of alkali-activated slag/fly ash pastes[J]. *Cem Concr Res*, 2019, 115: 230–250.
- [30] ZHANG Z H, ZHOU Z N, ZHU H J, et al. Mechanisms of CLDH seeding on hydration kinetics of slag-based geopolymer: Towards aluminosilicate cement phase engineering[J]. *Compos Part B Eng*, 2024, 271: 111157.
- [31] LI N, SHI C J, ZHANG Z H. Understanding the roles of activators towards setting and hardening control of alkali-activated slag cement[J]. *Compos Part B Eng*, 2019, 171: 34–45.
- [32] GOMAA E, HAN T H, ELGAWADY M, et al. Machine learning to predict properties of fresh and hardened alkali-activated concrete[J]. *Cem Concr Compos*, 2021, 115: 103863.
- [33] PENG Y M, UNLUER C. Analyzing the mechanical performance of fly ash-based geopolymer concrete with different machine learning techniques[J]. *Constr Build Mater*, 2022, 316: 125785.
- [34] KHAN M I, ABBAS Y M. Intelligent data-driven compressive strength prediction and optimization of reactive powder concrete using multiple ensemble-based machine learning approach[J]. *Constr Build Mater*, 2023, 404: 133148.
- [35] 栗思琪, 茆伟杰, 王钰洁, 等. 基于机器学习和原料碱激发活性特征的矿粉基碱激发胶凝材料抗压强度预测[J]. *复合材料学报*, 2025, 42(12): 7076–7087.
- LI Siqi, MAO Weijie, WANG Yujie, et al. *Acta Mater Compos Sin*, 2025, 42(12): 7076–7087.
- [36] NGUYEN H A T, PHAM D H, LE A T, et al. Transfer learning framework for modelling the compressive strength of ultra-high performance geopolymer concrete[J]. *Constr Build Mater*, 2025, 459: 139746.
- [37] THOMAS J J, ALLEN A J, JENNINGS H M. Density and water content of nanoscale solid C-S-H formed in alkali-activated slag (AAS) paste and implications for chemical shrinkage[J]. *Cem Concr Res*, 2012, 42(2): 377–383.
- [38] THOMAS R J, LEZAMA D, PEETHAMPARAN S. On drying shrinkage in alkali-activated concrete: Improving dimensional stability by aging or heat-curing[J]. *Cem Concr Res*, 2017, 91: 13–23.
- [39] PROVIS J L. Alkali-activated materials[J]. *Cem Concr Res*, 2018, 114: 40–48.
- [40] ADESANYA E, ALADEJARE A, ADEDIRAN A, et al. Predicting shrinkage of alkali-activated blast furnace-fly ash mortars using artificial neural network (ANN)[J]. *Cem Concr Compos*, 2021, 124: 104265.
- [41] SHEN J L, LI Y, LIN H, et al. Development of autogenous shrinkage prediction model of alkali-activated slag-fly ash geopolymer based on machine learning[J]. *J Build Eng*, 2023, 71: 106538.
- [42] BERNAL S A, SAN NICOLAS R, MYERS R J, et al. MgO content of slag controls phase evolution and structural changes induced by accelerated carbonation in alkali-activated binders[J]. *Cem Concr Res*, 2014, 57: 33–43.
- [43] 谭绍恩, 马倩敏, 李平. 基于堆叠融合技术的碱激发胶凝材料碳化性能预测[J]. *材料导报*, 2024, 38(增刊2): 288–296.
- TAN Shaoen, MA Qianmin, LI Ping. *Mater Rep*, 2024, 38(Suppl 2): 288–296.
- [44] DIKSHA, REDDY N D K, DEV N, et al. Predictive modeling and multi-parameter optimization of geopolymer mixes using SVM-GA hybrid approach[J]. *Multiscale And Multidiscip Model Exp And Des*, 2026, 9: 30.
- [45] DHAWAN A, VERMA M. Optimization of geopolymer concrete mix design using machine learning for enhanced sulphate and chloride resistance[J]. *Asian J Civ Eng*, 2026, 27(1): 281–298.
- [46] PARHI S K, DWIBEDY S, PANIGRAHI S K. AI-driven critical parameter optimization of sustainable self-compacting geopolymer concrete[J]. *J Build Eng*, 2024, 86: 108923.
- [47] WU S Y, LIU X R, FU B. Explainable automl and multi-objective optimization for sustainable high-performance geopolymer concrete[J]. *Sci Rep*, 2025, 15: 33027.
- [48] ZHANG M Y, GUO P W, TAN X, et al. Cradle-to-gate assessment and optimization of sustainable geopolymer concrete[J]. *J Clean Prod*, 2026, 538: 147387.
- [49] GUO P W, MENG W N, BAO Y. Knowledge-guided data-driven design of ultra-high-performance geopolymer (UHPG)[J]. *Cem Concr Compos*, 2024, 153: 105723.
- [50] HUANG Y M, HUO Z H, MA G W, et al. Multi-objective optimization of fly ash-slag based geopolymer considering strength, cost and CO₂ emission: A new framework based on tree-based ensemble models and NSGA-II[J]. *J Build Eng*, 2023, 68: 106070.
- [51] CHENG M Y, FU C L, OKABE R, et al. Artificial intelligence-driven approaches for materials design and discovery[J]. *Nat Mater*, 2026, 25(2): 174–190.
- [52] ZENI C, PINSLER R, ZÜGNER D, et al. A generative model for inorganic materials design[J]. *Nature*, 2025, 639(8055): 624–632.

作者贡献声明:

王董雨: 论文构思、调查研究、撰写初稿、综合数据并推导结论、审阅与修订

张祖华: 论文构思、审阅与修订、提供资源、获取资助、监督指导

施成: 综合数据并推导结论、软件使用、验证、绘制图表、审阅与修订

刘小青: 审阅与修订、获取资助、项目管理

蒋正武: 审阅与修订、监督指导、提供资源、项目管理

Data-Driven Research on Chemico-Activated Aluminosilicate Cementitious Materials

WANG Dongyu¹, ZHANG Zuhua¹, SHI Cheng¹, LIU Xiaoqing², JIANG Zhengwu¹

(1. Key Laboratory of Advanced Civil Engineering Materials of Ministry of Education, School of Materials Science and Engineering, Tongji University, Shanghai 201804, China; 2. State Key Laboratory of Advanced Marine Materials, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, Zhejiang, China)

Extended Abstract

In the context of global low-carbon and carbon-neutrality targets, chemico-activated aluminosilicate cementitious materials that are represented by geopolymers become one of the most promising green alternatives to ordinary Portland cement. Geopolymers are synthesized *via* the chemico-activation of reactive aluminosilicate solid wastes with an alkaline activator. Their manufacturing process is featured by low energy consumption and reduced carbon emissions through the valorization of industrial solid waste or low-energy natural mineral precursors. Furthermore, the reaction products are endowed with high early-age mechanical strength, excellent durability and fire resistance. However, the complexity of raw material composition, unclear polymerization mechanisms and intricate interdependencies among multiple variables result in low efficiency, poor reproducibility and accuracy of conventional trial-and-error, seriously restricting the multi-objective performance optimization and large-scale application of geopolymers.

Based on the materials initiative and rapid development of artificial intelligence, the data-driven paradigm initiates a fundamental shift in materials science research. Establishing the relationships between high-dimensional data (*e.g.*, materials compositions, structure, physical and chemical properties) and the material performance through learning massive multi-source data can effectively address the limitations of the conventional methods, *i.e.*, high cost, long time and numerous resources. It therefore provides intelligent and accelerated routes to revolutionize the design and investigation of chemically activated aluminosilicate cementitious materials, leading a paradigm shift from experimental-driven to data-driven approach. This review represents research progress on data-driven approaches in geopolymer, with emphasis on three aspects, *i.e.*, 1) machine learning methodologies for performance prediction; 2) application of machine learning in targeted performance prediction; and 3) metaheuristic algorithms-based composition optimization.

This review proposes a complete data-driven workflow for geopolymer research, including target definition, dataset collection, data processing, feature engineering, model construction and application. Based on the application scenarios, the existing models can be divided into four categories, *i.e.*, neural networks, ensemble learning, traditional regression and statistical learning, and symbolic regression combined with neuro-fuzzy systems. Each category offers distinct advantages and disadvantages. Neural networks capture nonlinear relationships, but require large dataset and lack interpretability. Ensemble learning handles high-dimensional mixed data with robustness, but risks overfitting with small samples. Traditional regression combined with neuro-fuzzy systems works for small to medium dataset, but faces an exponential rule growth under high-dimensional inputs. In terms of targeted performance prediction, machine learning is effectively applied to early-age properties, mechanical properties, and durability. For early-age behaviors, machine learning models can accurately predict geopolymerization heat flow, peak time, setting time and slump. For mechanical properties, compressive strength prediction is the most mature predicted performance, and the prediction is extended to flexural strength, splitting tensile strength, elastic modulus, *etc.*. For durability, the initial research mainly focuses on drying shrinkage, autogenous shrinkage, carbonation, while some studies on chloride penetration, sulfate attack and long-term durability are relatively scarce.

Metaheuristic algorithms play a significant role on geopolymer composition design. Firstly, metaheuristic algorithms serve to optimize hyperparameters and improve prediction accuracy. Subsequently, the well-trained predictive models can be leveraged to facilitate both forward guidance and inverse mix design. Based on the trained model, the importance of material composition, preparation parameters, and curing conditions to prediction accuracy can be quantitatively assessed. Furthermore, metaheuristic algorithms can also apply to reverse geopolymer mixture design targeting single or multiple objectives. For instance, a Pareto frontier is constructed *via* leveraging NSGA-II to balance high strengths, low carbon and low cost, and the optimal solution can be further selected *via* TOPSIS. This approach pioneers an effective pathway for the inverse design of geopolymers and establishes a versatile decision-support tool that is poised to become a new standard for the accelerated development of high-performance and eco-efficient cementitious materials.

Summary and prospects Despite progress on applying data-driven methods to chemically activated aluminosilicate cementitious materials, several critical challenges remain. Firstly, the inherent complexity and variability of raw materials, combined with the absence of standardized design specifications, lead to high-dimensional and redundant datasets that compromise the model performance. To address this issue, it is necessary to establish standardized open databases covering precursor composition, reactivity, activator parameters, and curing conditions, as well as the knowledge-embedded hybrid models that incorporate geopolymerization mechanisms and thermodynamic constraints. Secondly, data scarcity is a pervasive challenge in this field, with most datasets characterized by small sample sizes, severely limiting model generalizability. This issue can be alleviated through multi-source data fusion, data augmentation techniques, and transfer learning for small-sample scenarios. Lastly, the existing inverse design relies

primarily on a two-step prediction-then-optimization framework, whereas deep generative models for direct material design remain largely underexplored. Integrating these advanced models with reaction kinetics and microstructure evolution can be essential to realize end-to-end inverse design.

The evolution of artificial intelligence technology offers a significant potential for advancing chemico-activated aluminosilicate cementitious materials technology. This promises to serve as a robust reference for the deep integration of artificial intelligence into future material development. Future efforts should focus on developing standardized workflows, improving model interpretability and generalizability, as well as leveraging multi-scale simulations and deep generative models to accelerate material innovation. Data-driven methods will enable more accurate prediction and design of chemico-activated aluminosilicate materials due to these advancements, thereby contributing to the sustainable transformation of the construction industry.

Keywords data-driven; chemico-activated aluminosilicate cementitious materials; geopolymer; performance prediction; mixture design