

基于机器学习技术的超高性能混凝土性能预测与精细化设计

余 睿¹, 陈飞翔^{1,2}, 范定强³, 徐汪洋¹, 张玲艳¹, 李 旺¹, 吉朵朵¹

(1. 硅酸盐建筑材料国家重点实验室, 中国 武汉 430070; 2. 中交第二航务工程局有限公司, 中国 武汉 430040;
3. 香港理工大学土木与环境工程系, 中国 香港 999077)

摘 要: 计算机技术的快速发展推动了建筑行业的信息化与智能化进程。对此, 研究了机器学习(ML)技术在超高性能混凝土(UHPC)材料性能预测和精细化设计中的应用。针对 ML 训练数据库常存在数据缺失和异常值等问题, 提出了一个综合的 UHPC 数据处理和分析框架, 包括数据清洗、准备和模型训练模块, 旨在提高训练精度和模型泛化能力。采用改进的链式方程多重插补法实现了缺失数据的插补, 并使用孤立森林算法筛选异常数据, 从而提升了数据库质量。基于优化后的数据库, 建立了贝叶斯超参数优化的 XGBoost 模型, 实现了对 UHPC 多项性能的精准预测($R^2>0.95$)。最后, 提出了 2 种利用 ML 技术精细化设计 UHPC 配合比思路: 其一为采用物理堆积理论进行配比初步设计, 再利用建立的 ML 模型进行多项性能预测, 进一步优化配合比; 其二是将建立的 ML 预测模型与遗传算法相结合, 以实现 UHPC 的多目标优化设计。研究成果为 UHPC 的智能化研发提供了新思路和方法, 推动了 UHPC 技术的发展与应用, 并促进了工程材料与 ML 技术的深度融合。

关键词: 超高性能混凝土; 机器学习; 性能预测; 多目标优化

中图分类号: TP181; TU528 文献标志码: A 文章编号: 0454-5648(2025)05-1165-12

网络出版时间: 2025-03-26



超高性能混凝土(UHPC)是一种性能优异的建筑材料, 具有优异的力学性能、耐久性和工作性能^[1-3]。UHPC 优异性能主要源于其致密的微结构和特殊的材料组成, 其原材料通常包括水泥、硅灰、粉煤灰等多元复合胶凝材料, 并使用钢纤维进行基体增强^[1-2]。与传统混凝土相比, UHPC 具有更高的强度、更好的韧性和更长的使用寿命, 这使得它在桥梁工程、高层建筑、海洋工程等领域具有广阔的应用前景。

然而, UHPC 的研发与应用仍面临一些挑战。首先, UHPC 的制备过程复杂, 对原材料质量和配合比设计要求高, 这增加了生产成本和技术难度^[4-6]。其次, 配合比对 UHPC 的性能影响较大, 传统的试错设计方法难以快速、高效、精准地优化 UHPC 材料组成^[1]。再者, 传统的设计方法无法预测 UHPC 的材料性能。因此, 有必要建立起科学的 UHPC 配合比设计方法并实现高效的性能预测。传统的

UHPC 性能预测主要依赖于大量的实验和经验积累。研究人员通过正交试验、响应面等方法研究各种因素对 UHPC 性能的影响, 从而建立经验模型^[4]。然而, 这种方法不仅耗时耗力, 而且难以充分考虑各因素之间复杂的相互作用。近年来, 随着计算机科学与技术的高速发展, 机器学习(ML)技术为 UHPC 的研究和应用提供了新途径。

机器学习能够快速从大量实验数据中学习和发现规律, 建立准确的预测模型。不同的机器学习算法具有其相应的优势与不足。机器学习技术在 UHPC 材料中的应用已经取得了一些进展^[7-10]。例如, 研究人员利用支持向量机等算法实现了 UHPC 的抗压强度^[11]、弹性模量^[12]和耐久性^[13]的预测。此外, 基于粒子群优化等的多目标优化方法也被应用于 UHPC 配合比的优化设计^[14-16]。尽管 ML 技术在 UHPC 材料研发与应用中展现出巨大潜力, 但仍处于初期阶段, 面临着一些挑战。首先, UHPC 数据

收稿日期: 2024-10-10。 修订日期: 2025-01-05。

基金项目: 国家自然科学基金(52178249); 广东省自然科学基金青年提升项目(2024A1515030243); 武汉市知识创新专项(2023010201010094)。

第一作者: 余 睿(1986—), 男, 博士, 研究员。

通信作者: 范定强(1996—), 男, 硕士。

Received date: 2024-10-10. Revised date: 2025-01-05.

First author: YU Rui (1986—), male, Ph.D., Professor.

E-mail: r.yu@whut.edu.cn

Correspondent author: FAN Dingqiang (1996—), male, Master.

E-mail: ding-qiang.fan@connect.polyu.hk

库的质量和规模直接影响 ML 模型的性能，但高质量的 UHPC 数据往往难以获取，存在数据错误、缺失等情况^[17]。并且，现有 ML 模型的预测精度和泛化能力仍有待提高，特别是在处理新型 UHPC 配合比或极端条件下的性能预测时^[18–19]。其次，现有的 ML 优化方法往往难以协同考虑多个性能指标和约束条件，导致设计的配合比可能不符合实际工程需求。因此，如何将 ML 模型融入材料设计理论是推进 AI 技术在 UHPC 领域应用的重要方向^[5]。

基于上述背景，本工作旨在探索 ML 技术在 UHPC 性能预测和精细化设计中的应用。首先提出了一个综合的数据处理和分析框架以应对数据缺失、异常等问题。然后，建立了 UHPC 性能预测模型并通过超参数优化进一步提升模型精度。最后，结合物理堆积理论与 ML 技术，提出了两种多目标优化设计方法，以期 UHPC 的智能化设计提供新思路。

1 实验与算法

1.1 原料

采用的原材料包括普通硅酸盐水泥 P·II 52.5(OPC)、硅灰(SF)、石灰石粉(LP)和钢渣粉(SSP)，以及普通河砂(RS)，胶凝材料化学组分见表 1。骨料采用粒径为 0~0.6 mm 的超细河砂(RS-1)、0.60~1.25 mm 的细砂(RS-2)、1.25~2.36 mm 的粗砂(RS-3) 以及 2.36~4.75 的超粗砂(RS-4)。此外，采用聚羧酸减水剂(SP)改善 UHPC 的工作性能。

1.2 数据库

1.2.1 原始数据集 根据相关文献测试方法^[20–23]，测试了 UHPC 的工作性能、密实度、抗折强度和抗压强度，从而获得了机器学习的原始数据集。数据集包含 113 组数据，为一种小型数据库。数据分布包括连续型和离散型，具体细节见表 2。

表 1 胶凝材料的化学组分

Table 1 Main chemical composition of cementitious materials										w/%
Material	Na ₂ O	MgO	Al ₂ O ₃	SiO ₂	P ₂ O ₅	SO ₃	K ₂ O	CaO	Fe ₂ O ₃	LOI
OPC	0.17	1.53	4.94	20.81	0.08	1.97	0.91	64.18	3.32	1.59
LP		0.23		0.50				56.00		43.23
SF	0.15	0.52	0.11	96.36	0.19	0.18	0.54	0.34	0.06	1.49
SSP	0.12	6.14	4.21	16.69	1.50	0.34	0.06	41.00	20.46	3.94

w is mass fraction. OPC is ordinary Portland cement. LP is limestone powder. SF is silica fume. SSP is steel slag powder. LOI is loss on ignition.

表 2 原始数据集的分布

Table 2 Distribution of origin dataset					
Variables (No.)	Mean	Std	Min	Max	Missing value ratio/%
Input					
OPC (V0)	729.76	137.89	471.60	1006.08	0
SF (V1)	142.73	49.37	55.00	187.00	0
LP (V2)	148.94	108.61	0.00	405.00	0
SSP (V3)	27.36	64.23	0.00	203.00	0
RS-1 (V4)	766.18	365.24	33.82	1183.50	0
RS-2 (V5)	115.04	164.97	0.00	558.45	0
RS-3 (V6)	93.30	192.59	0.00	591.35	0
RS-4 (V7)	83.43	199.61	0.00	619.93	0
SP (V8)	37.02	7.62	27.50	49.5	0
Water (V9)	191.26	25.35	150.93	270.09	0
Output					
Workability/mm	196.40	85.58	100.00	357.00	17.70
Packing density	0.744	0.924	0.53	0.86	0
Flexure strength/MPa	16.42	3.81	8.48	25.96	35.40
Compressive strength/MPa	121.7	35.56	56.00	217.43	0

1.2.2 标准化 所使用的数据均来自于同一实验室，原材料具有一致的物理化学性质。在构建模型

时，基于材料的组成与性能关系，采用 L_1 或 L_2 正则化以防止过拟合，并且输入和输出数据进行标准

化处理^[24]。标准化方程如下:

$$X_{\text{normal}} = \frac{X - X_{\text{mean}}}{X_{\sigma}} \quad (1)$$

其中: X_{normal} 是 X 的标准化值; X_{mean} 和 X_{σ} 是训练集中每列的平均值和平均方差。为了确保结果的依赖性,测试数据也应使用上述 X_{mean} 和 X_{σ} 进行标准化。

1.2.3 数据库划分 为了评估训练模型的准确性和泛化能力,数据集被划分成训练集和测试集: 80% 用于训练, 20% 用于测试。采用 K 折交叉验证来替代简单的随机划分, 以开发更具泛化能力的模型^[9]。

1.3 人工智能算法简介

1.3.1 MICE 数据插补 采用链式方程多重插补

(MICE) 方法对数据库进行插补和完善。该方法通过生成多个插补数据集, 不仅有效填补了数据中的缺失部分, 还充分考虑了数据的不确定性, 进而增强了分析结果的稳定性与可靠性^[13], 如图 1 所示。

1.3.2 异常数据筛选 混凝土材料的测试性能通常存在波动或异常数据, 其根源在于原材料的多样性以及无法避免的实验误差^[1, 9]。因此, 采用孤立森林(IF)对插补后的数据库中的异常数据进行识别与筛选^[11]。IF 通过构建多棵孤立树(iTree)来筛选异常数据, 每棵树随机选择特征和分割点来分割样本(见图 2)^[10]。

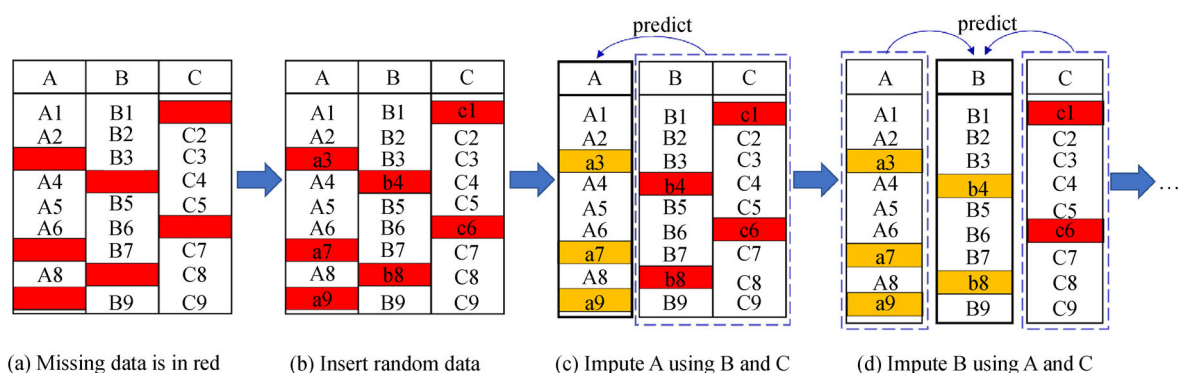


图 1 MICE 方法插补过程

Fig. 1 MICE method imputation process

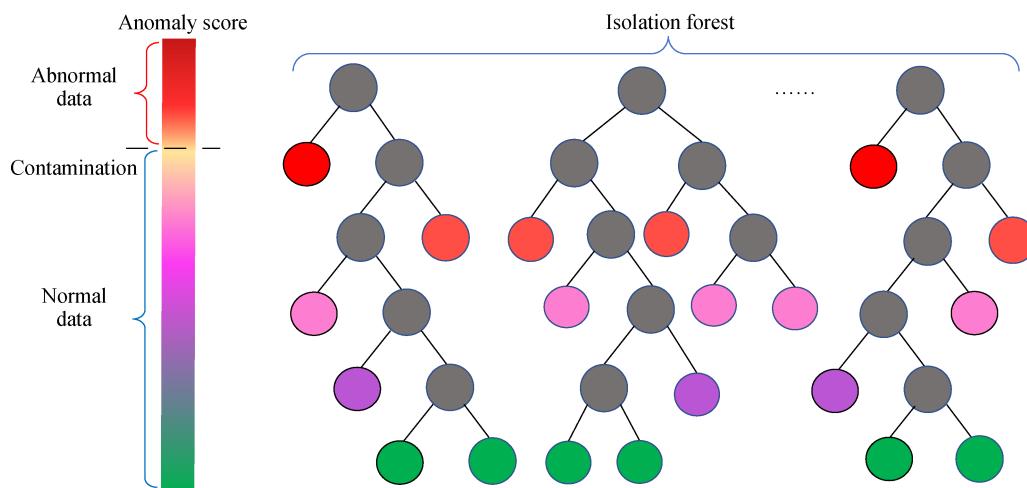


图 2 孤立森林的机理图

Fig. 2 Mechanistic diagram of isolated forests

1.3.3 XGBoost 算法 梯度提升决策树(XGBoost)是一种高效的、可扩展的机器学习算法, 特别适用于处理大规模数据集和离散数据集^[12]。XGBoost 算法内置了正则化技术, 以防止过拟合^[13-15]。

1.3.4 贝叶斯超参数优化 超参数优化是指在机器学习找到一组最优的超参数值, 以便其适应给

定数据集输入与输出之间的对应关系^[15]。

1.3.5 遗传算法 遗传算法(GA)通过模拟自然选择和遗传变异等过程, 对候选解集进行迭代优化, 使其逐步逼近最优解^[25]。遗传算法在 ML 预测模型的基础上进行多目标优化, 从而实现 UHPC 的多性能平衡^[16]。

1.3.6 评估函数 采用了系列指标对预测结果的准确性与可靠性进行评估, 具体如下^[17]:

平均绝对误差(MAE)是绝对误差的期望值, 反映出预测误差的实际情况。

$$MAE = \frac{1}{m} \sum_{i=1}^m |f_i - y_i| \quad (2)$$

均方根误差(RMSE)是平方平均误差的算术平方根, 反映结果的精度。

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (f_i - y_i)^2} \quad (3)$$

拟合优度(R^2)反映模型预测结果的准确性, 取值范围为 0~1。该评估函数结果越接近 1 表示模型拟合效果越好。

$$R^2 = 1 - \frac{\sum_{i=1}^m (f_i - y_i)^2}{\sum_{i=1}^m (\bar{y}_i - y_i)^2} \quad (4)$$

其中: y_i 是实际值; f_i 是预测值; $\bar{y}_i$ 是实际值的平均值; m 是样本数量。

2 基于 ML 的 UHPC 性能预测模型

2.1 缺失数据的插补

在实验数据集中存在缺失的数据(见表 1), 导致该数据集不能直接用于模型训练。为了保证插值数据的有效性, 一个关键指标是插值前后的流动性和弯曲强度的数据分布的一致性。在插值过程中, 将初始模型数设置为 5 个, 并在插值过程中迭代 4 次, 得到最后一组 5 组插值结果。

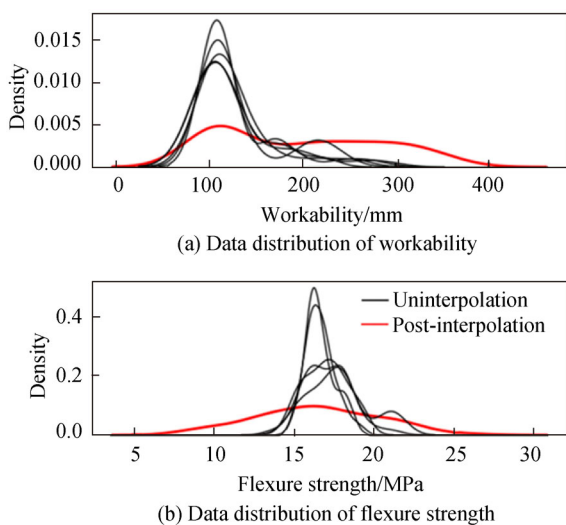


图 3 MICE 插补的数据分布与原数据分布结果对比

Fig. 3 Comparison of data distribution of MICE with origin data

采用 MICE 方法对原始数据库中的缺失数据进行

插补。从图 3 插补前后的数据变化可以看出, 插补后的数据与插补前的数据分布发生了明显的变化。虽然随着迭代次数的增加, 插补后的数据分布逐渐向真实的数据靠拢, 但是当迭代次数大于 3 时, 这一变化逐渐趋于平缓, 说明仅仅增加迭代次数并不能完全消除数据分布差异的结果。此外, 根据式 (2) 计算流动度和抗折强度插补前后的平均差异度, 结果分别为 $S=6.27\%$ 和 $S=-0.64\%$, 其中流动度的属性平均偏差度较大。这可能是因为尽管 MICE 算法尽可能地保证每一次结果的相似性, 但是模型的初始值的随机性较大, 与原数据的差异性难以消除, 因此需要调整插补过程以提高插补的拟合精度^[18]。

图 4 为原始数据的直方统计分布图。从图 4 可以看出, 数据集中大部分数据变量的分布呈现多峰和倾斜的特点, 因此简单的多重插补方法具有一定的局限性。为了提高插补后的数据与原数据之间的匹配性, 在 MICE 插补方法中添加均值匹配(PMM)内核以达到数据调优的效果^[19], 插补后的结果如图 5 所示。可以看出 2 个变量的插补前后结果具有一定的相似性。表明与简单的 MICE 插补方法相比, 内嵌 PMM 内核的 MICE 插补方法展现出更好的插补结果。这可能是由于 PMM 方法可以根据 MICE 插补结果最近的 n 个观测值自动调整数据集的拟合值, 达到自动寻优的结果^[26]。

2.2 数据筛选结果

采用了孤立森林算法对数据库中的异常数据进行筛选。在图 6 中, 不同颜色代表不同的数据分散程度, 颜色越深表明数据的分散程度越大。可以看出, UHPC 的密实度在 0.72~0.77, 抗压强度在 100~120 MPa 的数据点处于白色区域, 表明密实度和抗压强度在此范围内的数据密集度较高。此外, 标记为红色的异常数据点与其他数据的距离相差较远, 特别是抗压强度小于 70 MPa 的数据点和密度小于 0.55 的数据点被完全隔离, 红色的点表示被 IF 算法孤立出来的异常点。从图 6 可以看出, 一共有 7 组数据点被筛选出来, 表明孤立森林算法是一种可以有效筛选数据库异常值的方法, 并由此提高数据库的可靠性。

2.3 性能预测模型

采用了 XGBoost 算法建立 UHPC 材料性能预测模型。在模型建立时, 将数据集随机地划分为训练集和测试集, 其中训练集: 测试集=8:2。在模型训练的过程中, 将训练集作为模型参数拟合的训练数据, 测试集作为检验模型泛化性能的部分。

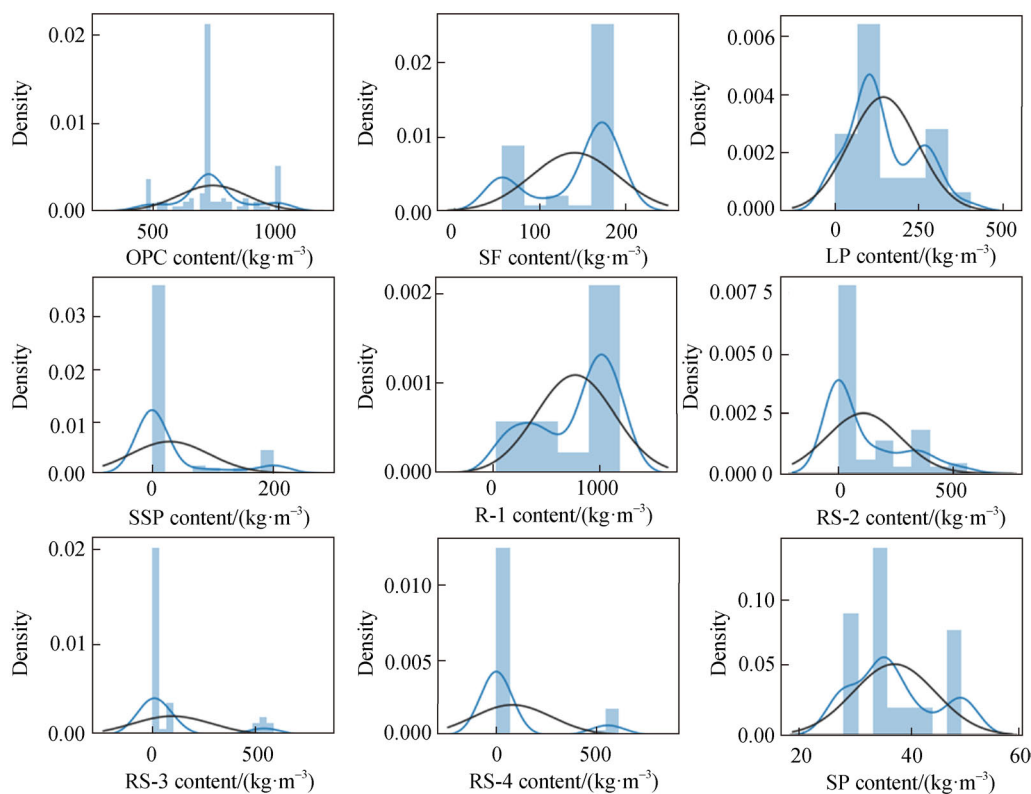


图 4 变量直方图

Fig. 4 Histogram of variables

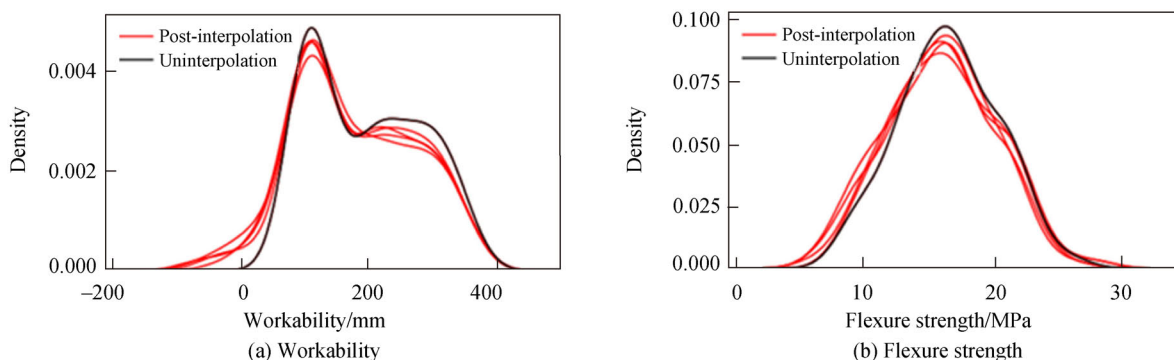


图 5 PMM 插补的数据分布与原始数据分布对比

Fig. 5 Comparison of distribution of interpolated data by PMM with origin data

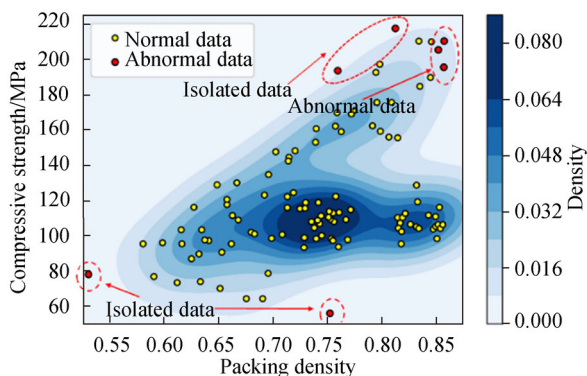


图 6 数据筛选的等高线图

Fig. 6 Contour plots filtered by data

图 7 为 XGBoost 模型的预测值与真实值的对比结果和误差值结果。图 7a 所展示的流动度训练集与训练集的结果大多分布在 10% 误差线以外。此外，可以看出建立的模型对于 UHPC 流动度和密实度预测效果不佳，但是较抗折强度的预测精度较高。从图 7c 可以看出，大部分数据点分布在 10% 误差线以内，表明模型在预测抗压强度时表现出较优性能。

表 3 进一步展示了 XGBoost 模型对不同性能的训练集与预测集的预测精度的结果。从表 3 可以发现，XGBoost 在预测抗压强度时在训练集和预测集

上均表现出较高的准确度，其中训练集和测试集的 R^2 分别为 0.873 和 0.871；在流动度的预测上表现较差，其中测试集与训练集的 R^2 均在 0.8 上下，对密度和抗折强度的预测精度介于抗压强度和流动度

之间。这可能是由于 UHPC 的抗压强度值较为集中，预测模型表现出较好的性能。虽然 XGBoost 模型对混凝土各性能具有一定的预测效果，但是其模型精度有待进一步提高。

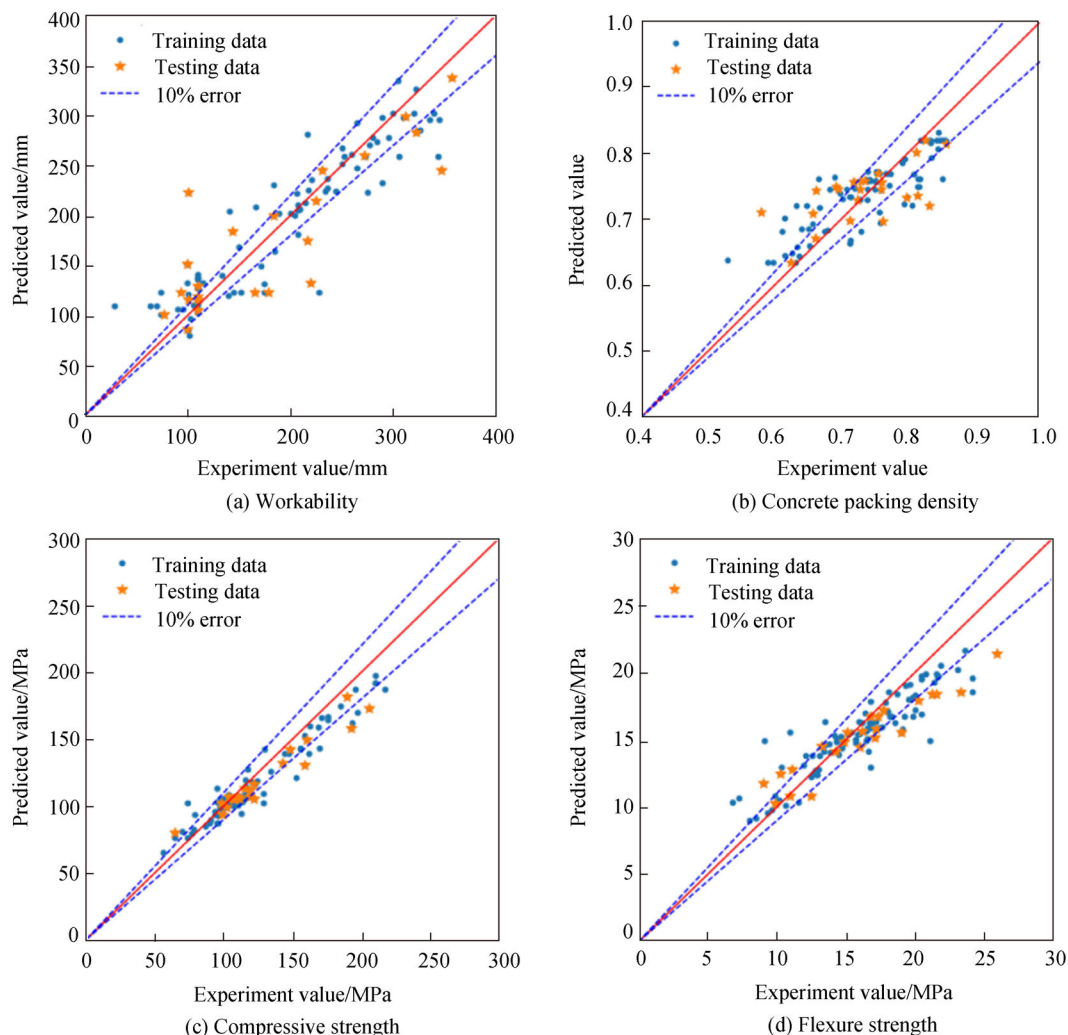


图 7 XGBoost 模型预测结果图

Fig. 7 XGBoost model prediction results strength

表 3 XGBoost 模型评价指标统计结果

Table 3 Statistical results of XGBoost model evaluation index

Property	Training data				Testing data			
	R^2	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE
Workability	0.801	1550.8	39.38	28.30	0.774	1963.7	44.31	33.26
Packing density	0.835	0.001 3	0.036 6	0.029 3	0.805	0.000 8	0.029 4	0.024 6
Compressive strength	0.873	159.62	12.634	8.815 4	0.871	112.71	10.616	7.185 3
Flexure strength	0.831	2.869 6	1.693 9	1.393 6	0.803	4.043 3	2.010 8	1.595 3

R^2 is goodness of fit. MSE is mean square error. RMSE is root mean square error. MAE is mean absolute error.

2.4 超参数优化

XGBoost 模型的性能高度依赖于超参数的选

择。超参数直接影响模型的复杂性、学习能力和泛化能力。选择合理的超参数有助于平衡模型的

偏差与方差, 进而提高模型的预测性能和泛化能力。在默认参数下, 模型的学习过程较为简单且快速, 但未必能够达到最优的预测效果。因此, 采用了贝叶斯优化算法对 XGBoost 模型进行超参数优化, 以寻找到适用于 UHPC 关键性能预测的模型超参数。表 4 为各超参数的搜索范围和最优搜寻。

超参数优化后的 XGBoost 模型对各性能训练集与测试集的预测精度评估结果如表 5 所示。可以发

现, 经超参数优化后的 XGBoost 模型的预测精度提升显著。流动度在训练集上的预测精度 R^2 由原来的 0.801 提升至 0.998, 提升了 24.6%; 对密实度、抗压强度和抗折强度的预测精度也均有提高, 分别提高了 18.8%、13.3%和 20.1%。在测试集上对流动度、密实度、抗压强度和抗折强度的预测精度则分别提升了 28.6%、19.6%、11.5%和 22.1%。从测试集结果来看, 建立的模型具有优异的预测性能且未表现出过拟合现象。

表 4 贝叶斯优化算法优化结果
Table 4 Bayesian algorithm optimization results

Property	$n_estimator$	Max_depth	Max_leaf_nodes	Learning_rate
Search scope	(1, 50, 1)	(1, 30, 1)	(0.30, 1.00)	(0.001, 0.050, 0.001)
Workability	30	37	21	0.04
Packing density	50	2	17	0.049
Compressive strength	50	5	28	0.05
Flexure strength	49	7	17	0.046

表 5 XGBoost 模型评价指标统计结果
Table 5 Statistical results of XGBoost model evaluation index

Property	Training data				Testing data			
	R^2	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE
Workability	0.998	8.691	3.169	2.098	0.995	6.336	2.517	1.779
Packing density	0.992	0.0002	0.0139	0.0030	0.963	0.0006	0.0253	0.0169
Compressive strength	0.989	13.652	3.6948	0.5701	0.971	33.890	5.7921	0.6355
Flexure strength	0.998	2.1428	1.4638	0.9391	0.980	14.989	3.8716	1.3468

2.5 混凝土材料性能预测的机器学习框架

现有的研究中通常将未经处理的数据直接用于 ML 模型的训练, 而未考虑数据缺陷或错误数据对模型潜在的负面影响^[27-28]。当机器学习技术应用到实际研究或工程中时, 数据集的处理问题是一个不可或缺的部分。特别是在混凝土材料领域, 无论是实验数据还是文献数据, 不可避免地会存在数据缺失和异常数据的情况, 这些数据往往无法提供真实信息, 并且还会影响模型的准确性。因此, 本工作提出了适用于 UHPC 数据库插补和数据库清洗的方法。此外, 为了平衡模型的预测精度和泛化能力, 基于自动化机器学习技术提出了模型自优化方法(超参数优化), 并由此总结了系统的模型框架(见图 8)。

该框架包括数据清洗模块、数据准备模块和基于自动化机器学习的模型训练模块。数据清洗模块包含数据库融合、数据库插补和数据库无监督筛选等方面的自动化流程; 数据准备模块包含有数据的

预处理、数据划分等方面的内容; 模型训练模块包含模型训练、超参数自优化、交叉验证和模型评估等方面的内容。全自动综合模型框架将多个功能模块有序集成, 解决了数据库缺陷对 ML 模型的负面影响, 从而实现混凝土材料性能的精准预测。

3 基于 ML 的 UHPC 精细化设计

建立了 UHPC 的 ML 智能预测模型后, 可以将建立的模型用以辅助 UHPC 组成设计^[29]。本工作提出了两个平行的设计思路。思路一是结合物理堆积模型与 ML 预测模型协同设计: 首先是利用修正的安德里森和安德森(MAA)模型对 UHPC 配合比进行初步设计, 然后通过基于建立的 ML 预测模型对初步配合比进行多目标性能预测, 并由此评估和验证配合比的效果, 确保其满足工程需求。思路二是将元启发式算法和建立的 ML 预测模型结合, 进行 UHPC 多目标优化设计: 采用 Pareto 最优解的概念, 采用元启发算法寻求满足多个目标的最优解。

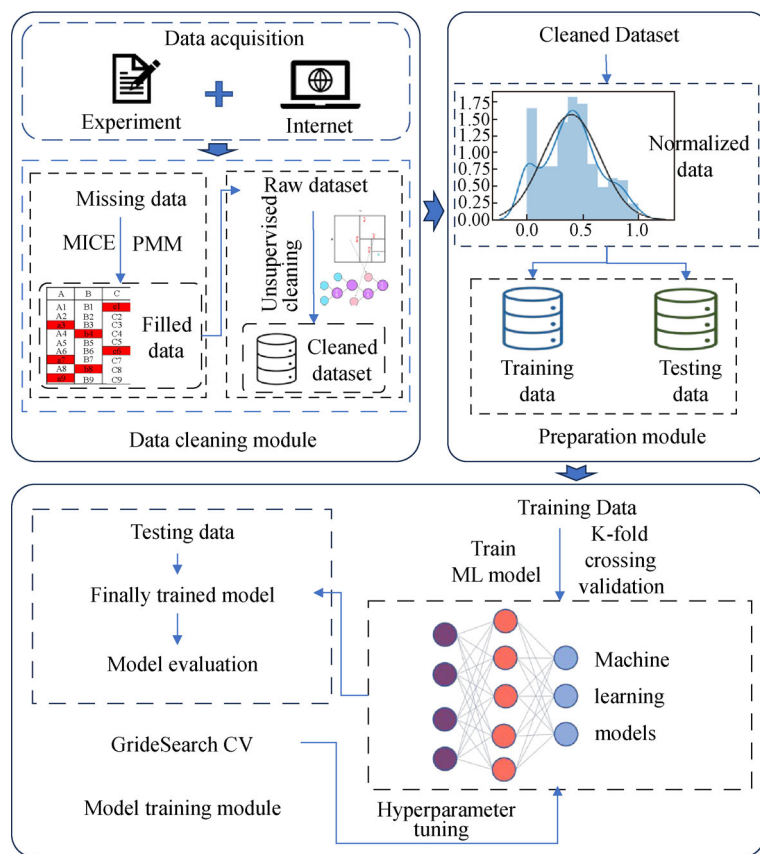


图 8 混凝土材料性能预测流程框架

Fig. 8 Framework of performance prediction process

3.1 结合物理-ML 预测模型的多目标优化设计

MAA 模型可以通过优化颗粒材料比例实现基体的最紧密堆积,从而提高其宏观密实度和力学性能^[30]。采用 MAA 颗粒堆积模型进行 UHPC 配合比的初步优化设计,公式如下^[31]:

$$P(D) = \frac{D^q - D_{\min}^q}{D_{\max}^q - D_{\min}^q} \quad (5)$$

其中: D 和 $P(D)$ 分别表示粒径,小于粒径 D 的粒子比例, $D_{\min}$ 和 $D_{\max}$ 分别表示最小和最大粒径, q 为分布模量。详细计算过程可参考作者先前文献^[30–31]。

在此,首先对四级河砂骨料进行粒径分析,并

按照一定比例进行混合,形成初步的骨料级配。然后,通过 MAA 模型计算胶凝材料的体积占比,以确定最优的 UHPC 配比组合 S1,如表 6 和表 7 所示。对照 UHPC 目标要求:强度 > 120 MPa;工作性能 > 200 mm;但是密实度偏低 < 0.7,不利于强度和耐久性的发展。因此,对 MAA 模型的参数进一步调整,并引入钢渣降低材料成本与碳排放(S2 至 S5)。最后,通过 XGBoost 模型预测配合比性能。结果表明 S5 在工作性能、强度、环境成本(低水泥用量)等方面均具有更显著的优势。因此,结合 MAA 模型和 XGBoost 预测模型,实现了 UHPC 的精细化设计。

表 6 基于 MAA 模型设计的 UHPC 配比及其预测性能

Table 6 UHPC mixtures designed by MAA model and their predicted performance

Sample No.	Mix proportion/(kg·m ⁻³)										Packing density	Compressive strength/MPa	Workability/mm
	OPC	SF	LP	SSP	RS-1	RS-2	RS-3	RS-4	SP	Water			
S1	680.72	130.70	181.53	0.00	202.69	291.92	90.18	521.94	20	180	0.625	116.3	210.5
S2	635.34	130.70	181.53	45.38	202.69	291.92	90.18	521.94	20	180	0.675	133.5	220.0
S3	589.96	130.70	181.53	90.76	202.69	291.92	90.18	521.94	20	180	0.678	135.7	200.0
S4	544.58	130.70	181.53	136.14	202.69	291.92	90.18	521.94	20	180	0.725	140.1	201.5
S5	499.20	130.70	181.53	181.53	202.69	291.92	90.18	521.94	20	180	0.734	146.5	205.0

表 7 多目标优化设计 UHPC 配比

Table 7 Multi-objective optimization design of UHPC mixtures

kg/m³

Sample No.	OPC	SF	LP	SSP	RS-1	RS-2	RS-3	RS-4	SP	Water
S6	530.33	192.25	165.45	120.87	995.34	26.17	284.14	240.02	25.82	161.23
S7	530.32	192.25	165.60	120.88	995.35	16.02	285.60	240.02	25.84	161.27
S8	533.41	192.51	165.60	120.87	995.35	16.79	285.60	240.02	25.84	161.27
S9	530.33	192.25	165.45	120.87	995.34	26.17	284.14	240.02	25.82	161.23
S10	532.95	192.42	165.38	120.81	995.39	1.35	285.37	494.25	24.41	160.98

3.2 基于元启发模型的多目标优化设计

本节探讨了结合遗传算法与 XGBoost 模型实现 UHPC 密实度和抗压强度多目标优化的方法。采用了 GA 算法结合建立的 XGBoost 模型对 UHPC 的抗压强度和孔隙率(1-密实度)进行双目标协同设计(配合比如表 7 所示)。帕累托前沿上的每个解都代表了在抗压强度和孔隙率之间实现的最优权衡, 其中 S6-S10 展示了 UHPC 抗压强度和孔隙率的在接近最优解时的逐步权衡变化。帕累托前沿分析后结果趋于相近, 意味着此时已经寻得最优解。所有最优解都收敛到一个相似值, 抗压强度为 189.39 MPa, 孔隙率为 0.15。此外, 帕累托前沿模型反馈得到 RS-2 的掺量变化波动较大, 表明二级砂(0.6~1.2 mm)对 UHPC 强度和孔隙的平衡具有明显影响。

图 9 为帕累托前沿(Pareto front), 即优化过程中得到的一组非支配解。横轴表示抗压强度(Objective 1), 纵轴表示密实度(Objective 2)。每个红点代表一个非支配解, 即没有其他解在所有目标上都优于它们, 这些解代表了优化过程中得到的最优解集。从图 9 可以发现, 随着抗压强度的增加, 孔隙率有所降低, 全部位于 0.2 以下。这表明在优化过程中, 需要在这两个目标之间实现有效平衡。帕累托前沿的演化过程中, 不同颜色的点代表不同代数的解, 颜色从深到浅依次表示从初始代到最终代的演化过程。从图 10 可以看出, 随着代数的增加, 帕累托前沿逐渐向左下角移动, 说明算法在不断优化, 找到了更优解。尽管可以通过迭代图观察算法如何进行寻优, 但是机器学习算法的“黑盒”效应依旧存在。在实际的工程应用中, UHPC 优化设计往往需要在多个目标中寻找平衡。例如, 随着孔隙率的进一步降低, 材料的工作性(如流动度)可能会受到影响。因此, 除了抗压强度和孔隙率, 未来的优化研究可以进一步引入流动度等性能指标, 以形成更加全面的多目标优化框架。此外, 后续研究有必要进一步验证设计的配合比的综合性能以满足实际工程需求。

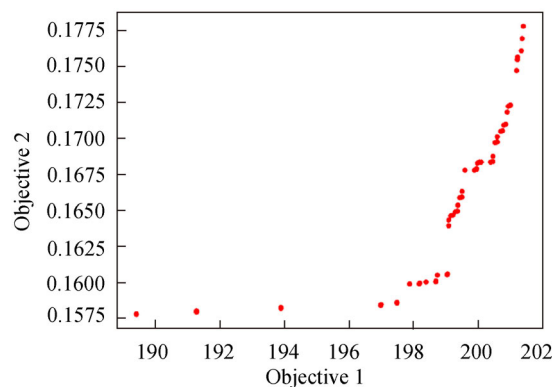


图 9 帕累托前沿

Fig. 9 Pareto front

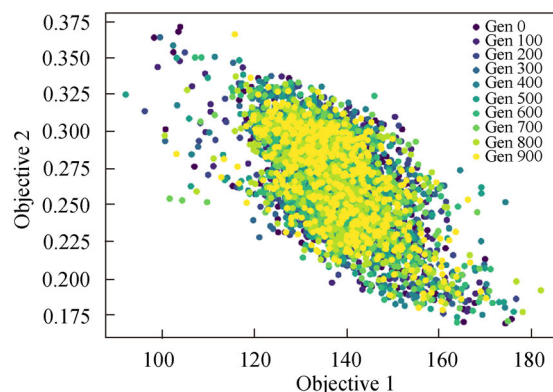


图 10 帕累托前沿的演化

Fig. 10 Pareto front evolution

综上, 可以将 XGBoost 预测模型作为遗传算法的目标函数, 进行 UHPC 孔隙率和抗压强度多目标优化。在实际应用中, 可以根据具体需求选择合适的最优解, 从而提高 UHPC 的整体性能。

4 结论

1) 提出了一个基于机器学习技术的 UHPC 性能预测与优化设计框架, 涵盖了数据插补、异常值筛选、机器学习建模、多目标优化模块, 有效解决了 UHPC 数据库问题, 提高了预测模型的准确性和可靠性, 实现了多目标配合比寻优。

2) 采用 MICE 方法对原始数据库中的缺失数据进行插补, 并使用孤立森林算法识别异常值。经过数据清洗和优化处理后的数据库, 为 XGBoost 模型的性能预测提供了可靠基础, 实现了 UHPC 多项性能的精准预测。

3) 针对 UHPC 的多目标优化问题, 提出了 2 种平行的设计方法: 方法一结合物理堆积理论和 XGBoost 模型, 通过理论设计与预测模型的协同作用, 实现了 UHPC 配合比的精细化设计; 方法二直接结合遗传算法与 XGBoost 模型, 通过帕累托前沿获得最优解, 能够在满足工程性能需求的基础上实现快速寻优。

4) 通过物理模型和 ML 模型的结合, 为 UHPC 多目标设计提供了更多理论支撑; 结合遗传算法和 ML 模型能够揭示帕累托前沿, 提高设计效率。两种设计思路各有优劣, 需要结合实际工程需求进行选择。

参考文献:

- [1] FAN D Q, ZHU J Y, FAN M X, et al. Intelligent design and manufacturing of ultra-high performance concrete (UHPC)—A review[J]. Constr Build Mater, 2023, 385: 131495.
- [2] 刘一帆, 吴泽媚, 张轩翰, 等. 超高性能混凝土流变特性及调控研究进展[J]. 硅酸盐学报, 2023, 51(11): 3025–3038.
LIU Yifan, WU Zemei, ZHANG Xuanhan, et al. J Chin Ceram Soc, 2023, 51(11): 3025–3038.
- [3] 韩方玉, 吉旭平, 刘加平, 等. 高温对不同钢纤维掺量超高性能混凝土性能及钢纤维拔出机制的影响[J]. 硅酸盐学报, 2024, 52(11): 3565–3571.
HAN Fangyu, JI Xuping, LIU Jiaping, et al. J Chin Ceram Soc, 2024, 52(11): 3565–3571.
- [4] SUN G R, DU M H, SHAN B H, et al. Ultra-high performance concrete design method based on machine learning model and steel slag powder[J]. Case Stud Constr Mater, 2022, 17: e01682.
- [5] AMINI PISHRO A, ZHANG Z R, AMINI PISHRO M, et al. UHPC-PINN-parallel micro element system for the local bond stress-slip model subjected to monotonic loading[J]. Structures, 2022, 46: 570–597.
- [6] DU J, MENG W N, KHAYAT K H, et al. New development of ultra-high-performance concrete (UHPC)[J]. Compos Part B Eng, 2021, 224: 109220.
- [7] 陈庆, 马瑞, 蒋正武, 等. 基于 GA-BP 神经网络的 UHPC 抗压强度预测与配合比设计[J]. 建筑材料学报, 2020, 23(1): 176–183.
CHEN Qing, MA Rui, JIANG Zhengwu, et al. J Build Mater, 2020, 23(1): 176–183.
- [8] LI Y Q, SHEN J L, LI Y, et al. The data-driven research on the autogenous shrinkage of ultra-high performance concrete (UHPC) based on machine learning[J]. J Build Eng, 2024, 82: 108373.
- [9] ZOU D L, WU L L, HAO Y F, et al. Composition-strength relationship study of ultrahigh performance fiber reinforced concrete (UHPFRC) using an interpretable data-driven approach[J]. Constr Build Mater, 2023, 392: 131973.
- [10] XU L L, FAN D Q, LIU K N, et al. A machine learning framework for intelligent development of Ultra-High performance concrete (UHPC): From dataset cleaning to performance predicting[J]. Expert Syst Appl, 2024, 242: 122790.
- [11] ALSINI R, ALMAKRAB A, IBRAHIM A, et al. Improving the outlier detection method in concrete mix design by combining the isolation forest and local outlier factor[J]. Constr Build Mater, 2021, 270: 121396.
- [12] SUN Z, WANG X, HUANG H, et al. Predicting compressive strength of fiber-reinforced coral aggregate concrete: Interpretable optimized XGBoost model and experimental validation[J]. Structures, 2024, 64: 106516.
- [13] WAKJIRA T G, KUTTY A A, ALAM M S. A novel framework for developing environmentally sustainable and cost-effective ultra-high-performance concrete (UHPC) using advanced machine learning and multi-objective optimization techniques[J]. Constr Build Mater, 2024, 416: 135114.
- [14] AMINI PISHRO A, ZHANG S Q, HU Q X, et al. Advancing ultimate bond stress-slip model of UHPC structures through a novel hybrid machine learning approach[J]. Structures, 2024, 62: 106162.
- [15] 张文生, 曹传荔, 郑晓, 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.
ZHANG Wensheng, CAO Fuli, ZHI Xiao, et al. J Chin Ceram Soc, 2024, 52(11): 3617–3630.
- [16] 刘晓, 王思迈, 卢磊, 等. 机器学习预测混凝土材料耐久性的研究进展[J]. 硅酸盐学报, 2023, 51(8): 2062–2073.
LIU Xiao, WANG Simai, LU Lei, et al. J Chin Ceram Soc, 2023, 51(8): 2062–2073.
- [17] ZHANG K J, LIN W Q, ZHANG Q T, et al. Evaluation of anisotropy and statistical parameters of compressive strength for 3D printed concrete[J]. Constr Build Mater, 2024, 440: 137417.
- [18] ALI A, ABU-ELKHEIR M, ATWAN A, et al. Missing values imputation using fuzzy K-top matching value[J]. J King Saud Univ Comput Inf Sci, 2023, 35(1): 426–437.
- [19] CAI M Y, VAN BUUREN S, VINK G. Graphical and numerical diagnostic tools to assess multiple imputation models by posterior predictive checking[J]. Heliyon, 2023, 9(6): e17077.
- [20] 中国混凝土与水泥制品协会. 行标-超高性能混凝土基本性能与试验方法[S]. T/CBMF37—2018, 2019.
- [21] YIN T Y, JIN L, LIU K N, et al. Fractal evolution model of proton spatial distribution in low water/binder cement-based composites (LW/B-CC) during early-age hydration process[J]. Constr Build Mater, 2024, 427: 136287.
- [22] 中国混凝土与水泥制品协会. 行标-超高性能混凝土预混料[S]. T/CBMF 96-2020/ T/CCPA 20—2020, 2021.
- [23] 代柱端, 陈健, 周晓阳. 基于机器学习回归模型的 UHPC 抗压强度预测研究[J]. 混凝土与水泥制品, 2023(12): 23–27.
DAI Zhuduan, CHEN Jian, ZHOU Xiaoyang. China Concr Cem Prod, 2023(12): 23–27.
- [24] MAHJOUBI S, BARHEMAT R, MENG W N, et al. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC)[J]. Resour Conserv Recycl, 2023, 189: 106741.
- [25] FAN D, YU R, FU S, et al. Precise design and characteristics prediction of Ultra-High Performance Concrete (UHPC) based on artificial intelligence techniques[J]. Cem Concr Compos, 2021, 122: 104171.
- [26] SAMAD M D, ABRAR S, DIAWARA N. Missing value estimation using clustering and deep learning within multiple imputation

- framework[J]. Knowl Based Syst, 2022, 249: 108968.
- [27] NGUYEN T A, TRINH S H, LY H B. Enhanced bond strength prediction in corroded reinforced concrete using optimized ML models[J]. Structures, 2024, 63: 106461.
- [28] LY H B, NGUYEN T A. Enhancing compressive strength and sustainability: A machine learning approach to recycled brick and tile concrete mix design[J]. Mater Today Commun, 2024, 40: 109642.
- [29] ZHOU M, WU Z, OUYANG X, et al. Mixture design methods for ultra-high-performance concrete-a review[J]. Cem Concr Compos, 2021, 124: 104242.
- [30] LIU K N, YIN T Y, FAN D Q, et al. Multiple effects of particle size distribution modulus (q) and maximum aggregate size (D_{max}) on the characteristics of Ultra-High Performance concrete (UHPC): Experiments and modeling[J]. Cem Concr Compos, 2022, 133: 104709.
- [31] YIN T Y, LIU K N, FAN D Q, et al. Derivation and verification of multilevel particle packing model for Ultra-High Performance Concrete (UHPC): Modelling and experiments[J]. Cem Concr Compos, 2023, 136: 104889.

Machine Learning-Based Performance Prediction and Precision Design of Ultra-High Performance Concrete

YU Rui¹, CHEN Feixiang^{1,2}, FAN Dingqiang³, XU Wangyang¹, ZHANG Lingyan¹, LI Wang¹, JI Duoduo¹

(1. State Key Laboratory of Silicate Materials for Architectures, Wuhan 430070, China; 2. CCCC Second Harbor Engineering Company LTD., Wuhan 430040, China; 3. Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong 999077, China)

Extended Abstract

Introduction Ultra-high performance concrete (UHPC) has garnered widespread attention in the field of civil engineering due to its exceptional mechanical properties, durability, and workability. These superior properties are primarily attributed to its dense microstructure and unique material composition, which typically includes cement, silica fume, quartz powder, and superplasticizers, along with reinforcing materials like steel fibers. The remarkable performance of UHPC not only meets the demands for high-performance materials in modern engineering but also contributes to reducing the life-cycle costs of structures by minimizing section sizes and extending service life.

However, the development and application of UHPC face several challenges. The complexity of the preparation process, stringent requirements for raw materials, and the high sensitivity of UHPC properties to mix design variations increase both production costs and technical difficulties. Traditional experimental design methods struggle to efficiently optimize UHPC mixtures, and issues such as shrinkage and autogenous cracking require precise design solutions. These challenges make the composition design and performance prediction of UHPC a focal point in the field. With advancements in computer technology and data science, machine learning (ML) offers new tools and approaches for UHPC research and development. Therefore, this study aims to promote the application of ML technology in the UHPC field, thereby advancing the intelligent development of advanced construction materials.

Methods The study introduced a comprehensive ML-based framework for UHPC performance prediction and precise design. The framework began with data cleaning, where the multiple imputation by chained equation (MICE) method with a predictive mean matching (PMM) kernel was employed to impute missing data in the UHPC database. Subsequently, the isolation forest algorithm was applied to identify and eliminate outlier data, thereby improving the quality of the dataset. The cleaned and optimized dataset was then used to train an XGBoost model, optimized via Bayesian hyperparameter tuning, to accurately predict various UHPC properties. Two AI-driven approaches for UHPC mix design were also proposed: one that combined the MAA model with ML predictions for multi-performance optimization, and another that integrated the ML model with genetic algorithm (GA) for multi-objective optimization.

Results and discussion The application of MICE combined with PMM resulted in a substantial improvement in the accuracy of the imputed data, as evidenced by the reduced average deviation between imputed and original data values. This enhanced imputation process allowed for a more reliable dataset, which directly contributed to the performance of the XGBoost prediction models. Outlier detection via the isolation forest algorithm effectively removed data points that exhibited significant deviation from the norm, particularly in compressive strength and density measurements. The refined dataset led to more accurate and consistent predictions across all performance metrics. After hyperparameter optimization, the XGBoost model demonstrated exceptional predictive capabilities, with notable improvements in performance metrics. The predictive accuracy of the model was further validated against experimental data, confirming its effectiveness in real-world applications. These results highlight the potential of ML techniques in advancing the field of UHPC research. The integration of MICE and PMM into the data preprocessing pipeline ensured more accurate and reliable datasets, which were crucial for developing robust predictive models. The success of the XGBoost model, particularly after Bayesian optimization, underscores the importance of hyperparameter tuning in enhancing model performance.

Furthermore, the study proposes two ML-assisted design strategies for UHPC: 1) a combined physical packing theory and ML prediction model for initial mixture design, and 2) a multi-objective optimization approach using a metaheuristic algorithm for fine-tuning UHPC compositions. These strategies provide a framework for the intelligent and efficient design of UHPC materials, aligning with the growing demand for high-performance, sustainable construction materials.

Conclusions This research demonstrates the significant potential of ML in predicting and optimizing the performance of UHPC. By addressing data challenges and enhancing prediction models, the study provides valuable insights into the application of ML in UHPC design. The proposed ML-assisted design strategies offer a practical approach to developing UHPC with tailored properties, paving the way for more sustainable and efficient construction practices.

Keywords ultra-high performance concrete; machine learning; performance predictions; multi-objective optimization