第 54 卷第 3 期
2026 年 3 月

硅 酸 盐 学 报
JOURNAL OF THE CHINESE CERAMIC SOCIETY

Vol. 54，No. 3
March，2026

http://jccsoc.cbpt.cnki.net

DOI：10.14062/j.issn.0454-5648.20250461

基于条件生成对抗网络与马尔可夫链蒙特卡洛双驱动的超高性能混凝土数据增强及智能预测方法

余 睿[1]，李 旺[1]，范定强[2,3]

(1. 武汉理工大学硅酸盐科学与先进建材全国重点实验室， 武汉 430070，中国；
2. 香港理工大学土木与环境工程系，香港 999077，中国；3. 东京大学建筑系建筑材料研究室，东京 113-8654，日本)

摘 要：随着建筑行业智能化转型，人工智能在混凝土材料研发与应用中备受关注，但超高性能混凝土(UHPC)存在数据库质量不稳定、数据量过小的问题，现有数据增强方法难以处理其高维异构数据，严重制约了 UHPC 的智能化发展。对此，本工作构建了基于条件生成对抗网络(CGAN)与马尔可夫链蒙特卡洛(MCMC)算法协同驱动的 UHPC 数据增强模型：首先对原始数据进行预处理，然后借助 CGAN 算法生成虚拟样本，并基于 MCMC 模型进行统计修正，最后采用贝叶斯优化的 XGBoost 模型验证智能预测准确性与可靠性。结果表明，增强后训练集规模可扩至 1500 组，增强数据与原始数据分布高度一致，平均 Jensen Shannon 距离<0.05，Kolmogorov-Smirnov 检验证明分布具有高度同源性。采用增强数据模型后，XGBoost 模型的各评估指标均证明了预测精度的显著提升，其中 3 d 累计水化放热量、中值孔径、抗压强度、抗折强度及工作性能的测试集的 R^2 分别提升了 62%、55%、18%、46%和 35%。最后，基于实际工程数据的模型验证结果表明，建立的模型可用于工程材料性能预测(R^2=0.92)，证明了该方法的显著可靠性和泛化能力。研究成果为解决 UHPC 数据库难题提供了新思路，对推动建材行业转型升级和促进智能建造发展具有显著价值。

关键词：人工智能；超高性能混凝土；数据增强；机器学习；智能预测
中图分类号：TP181；TU528　　**文献标志码**：A　　**文章编号**：0454–5648(2026)03–0842–15
网络出版时间：2026–02–10



在极端气候与重大工程长寿命需求的驱动下，现代结构对材料性能提出更高要求[1–3]。传统混凝土在强度与耐久性方面的不足，越来越难以满足跨海桥梁、超高层建筑、核电防护等工程应用需求[4–6]。与之相关的耐久性退化问题已引发显著经济损失，促使高性能建材的研发日益迫切[7–10]。超高性能混凝土(UHPC)因其优异的力学与耐久性能(抗压强度>120 MPa、韧性优良、耐久性强)在工程中的应用迅速增长，并在推动构件轻量化与结构可持续发展方面展现出广阔前景[11–14]。然而，UHPC 在研发与实际应用过程中仍面临诸多挑战[15]。UHPC 的制备复杂、原料敏感性强，叠加其组分–结构–性能间的强非线性耦合，严重制约了多尺度性能建模与工程推广，而传统基于线性假设的数学建模方法在表征这种多参数耦合体系时具有显著局限性。因此，开发一种新方法以实现 UHPC 性能的智能、精准预测具有重要意义。

近年来，人工智能技术在混凝土性能预测中显示出良好前景，相关智能模型在多尺度建模、非线性挖掘和主动调控方面取得积极进展[16–18]。例如，Yu 等[19]基于深度学习的跨尺度建模技术，通过卷积神经网络(CNN)解析了透水混凝土三相结构的图像(精度达 95%左右)；Iman 等[20]基于超声波的分布，使用长短期记忆网络(LSTM)预测混凝土的损伤，较传统手动提取特征的效率大幅提升。通常，基于数据驱动的智能化模型的可靠性在很大程度上依赖于高质量数据库的支撑。然而，现有 UHPC 数据库普遍存在样本规模小、实验条件异构、缺乏微观数据等

收稿日期：2025–06–11。　修订日期：2025–06–24。
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，博士研究生。

Received date: 2025–06–11.　**Revised date:** 2025–06–24.
First author: YU Rui (1986–), male, Ph.D., Professor.
E-mail: r.yu@whut.edu.cn
Correspondent author: FAN Dingqiang (1996–), male, Doctoral candidate.
E-mail: ding-qiang.fan@connect.polyu.hk

问题[5]，严重制约了智能模型的有效性，形成“算法超前、数据滞后”的矛盾，这已成为阻碍 UHPC 智能化转型的关键因素，迫切需要探索全新的数据库处理范式[21]。

针对 UHPC 数据库建模面临的挑战，近年来已有学者开展相关研究以提升数据质量与模型性能[22]。Li 等[23]通过理论分析指出，当训练样本规模较小(低于千级)且数据维度较高时，传统机器学习模型容易出现泛化能力下降与过拟合问题。在解决路径方面，迁移学习(TL)技术具有一定优势[24, 25]：Ford 等[25]通过迁移天然骨料混凝土的大规模数据库，有效提升了再生骨料 UHPC 抗压强度的预测精度。与此同时，数据增强策略亦取得积极进展：Miyan 等[26]采用生成对抗网络技术(GAN)有效扩展了数据规模，实现了原始数据库的倍增；Chen 等[27]结合深度增强策略与卷积神经网络架构，建立了低碳 UHPC 强度预测模型；Wang 等[28]则创新性地融合克里金插值与马尔可夫链蒙特卡洛(MCMC)方法，开发出适用于水泥基材料渗透性预测的多维数据扩充范式。

然而，由于 UHPC 材料本身存在多尺度参数突变与高维数据不均衡等特性，传统增强方法面临三重局限：一是 GAN 生成数据缺乏物理约束，可能产生违背材料规律的样本(如高孔隙率对应高强度)；二是克里金插值在高维空间下精度大幅下降；三是迁移学习在跨材料体系中存在分布失真风险，难以准确迁移普通混凝土中粗骨料粒径与强度的关系至 UHPC 体系。因此，亟需构建一种跨域分布校准的智能生成范式，以突破 UHPC 数据增强的有效性与泛化瓶颈。

基于此，本工作提出一种融合条件生成对抗网络(CGAN)[29]与马尔可夫链蒙特卡洛(MCMC)[30]方法的双驱动数据增强模型。首先，通过对原始数据库进行系统的数据预处理，消除噪声和维度灾难；其次，利用 CGAN 在材料组分约束下合成具有物理一致性的虚拟样本；然后，借助 MCMC 进行统计修正，确保增强数据库与原始数据空间保持拓扑同构；最后，基于贝叶斯优化的 XGBoost 算法对增强后的数据进行建模与预测，并通过海洋漂浮工程平台的 UHPC 工程数据验证数据增强方法的普适性与泛化能力。本工作为 UHPC 材料的智能建模与小样本问题提供新解法，助力高/超高性能混凝土材料的数字化设计与工程实践。

1　数据库与算法

1.1　数据来源与信息

本工作数据由网络开源数据集构成：网络数据集包含 145 组 UHPC 全组分配比及性能参数(如表 1)，涵盖水泥、硅灰(SF)、第三元粉料(除硅灰外的其他辅助胶凝材料，如粉煤灰、石灰石粉等，Th)、砂(SA)、水胶比(w/b)、减水剂(SP)、钢纤维(S-Fibre)等原材料参数，以及 3 d 累计水化放热量、中值孔径(不同孔径的孔以孔体积的大小进行排布，取中值孔体积对应的孔径的大小)、抗压强度、抗折强度和工作性能等参数的网络数据库。其中，关键性能指标，如抗压强度(66.68~201.47 MPa，均值 115.84 MPa 与标准差 28.03)与工作性能(112.71~314.00 mm，均值 206.25 mm 与标准差 51.54)均存在显著离散性，有必要对数据库进行增强与优化。固体颗粒材料(水泥、硅灰、第三元粉料)配比中，采取了用固体颗粒质量/水泥质量的值代替原数据，以消除量纲差异(其中部分特征是无量纲的值)。

表 1　原始数据库分布特征

Table 1　Distribution of origin dataset

Variables	N	Min	Max	Mean	Std
SF	145.00	0.00	0.83	0.17	0.14
Th	145.00	0.00	4.07	0.51	0.62
SA	145.00	0.00	4.08	1.44	0.74
w/b	145.00	0.09	0.28	0.19	0.03
SP	145.00	0.00	0.38	0.05	0.07
S-Fibre	145.00	0.00	2.56	0.57	0.94
3 d-h/(J·g^{-1})	145.00	42.59	345.57	179.07	52.10
MPD/nm	145.00	11.90	78.05	36.54	13.97
CS/MPa	145.00	66.68	201.47	115.84	28.03
FS/MPa	145.00	11.16	48.44	23.02	7.48
Workability/mm	145.00	112.71	314.00	206.25	51.54

SF—Silica fume; Th—Third type of powder; SA—Sand; w/b: Water-to-binder ratio; SP—Superplasticizer; S-Fibre—Steel fibre; 3d-h—3-day cumulative hydration heat; MPD—Median pore diameter; CS—28-day compressive strength; FS—Flexural strength.

1.2　数据库预处理

1.2.1　标准化处理　　在构建 UHPC 性能预测模型时，网络数据面临多源异构性挑战(如数据缺失、异常、量纲差异等问题)。这可能导致模型产生特征权重偏差与引发过拟合风险。为此，本工作采用标准化方法消除量纲影响，具体如下(如图 1a)：

$$X_{\text{normal}} = \frac{X - X_{\text{mean}}}{X_{\sigma}} \tag{1}$$

式中：X_{normal} 为 X 的标准化值；X_{mean} 和 X_{σ} 为训练集中每列的平均值和标准差。为了确保结果的一致性，测试数据也应使用上述 X_{mean} 和 X_{σ} 进行标准化。

1.2.2 缺失值处理 UHPC 数据库中，实验室与网络数据均存在波动性、缺失及质量不均问题。实验室数据因实验周期长、材料成分复杂导致一致性不足；网络数据因来源差异增加了不确定性。为此，本工作采用链式方程多重插补(MICE)方法(如图 1b)[31]，通过变量间相关性构建迭代预测模型，逐变量构建条件概率分布，经多轮迭代生成多个合理数据版本，动态填补缺失值并量化不确定性。该方法通过整合多数据集建模结果，可显著提升模型抗波动能力及预测稳定性，为 UHPC 性能预测模型建立高可靠性数据基础。

1.2.3 异常值处理 在 UHPC 材料性能测试中，多元组分原材料的非线性交互作用与实验测量系统的累积误差，导致数据集呈现显著的非高斯分布特性与多维异常扰动。为构建高置信度的材料性能数据库，本工作采用基于孤立森林(IF)的智能清洗框架(如图 1c)，可有效消除因仪器漂移或配比失误导致的离群点，并提升数据的鲁棒性[15]。

1.2.4 数据库划分 为了评估训练模型的准确性和泛化能力，数据集被划分为训练集和测试集，其中，70%的数据用于训练，30%的数据用于测试。此外，为了进一步提高模型的泛化能力并避免过度拟合，本工作采用了 5 折交叉验证方法替代简单的随机划分(如图 1d)。通过 5 折交叉验证，数据集被划分为 5 个子集，模型在每个子集上进行验证，从而更全面地评估模型在不同数据子集上的表现，以提高模型的鲁棒性与泛化能力。

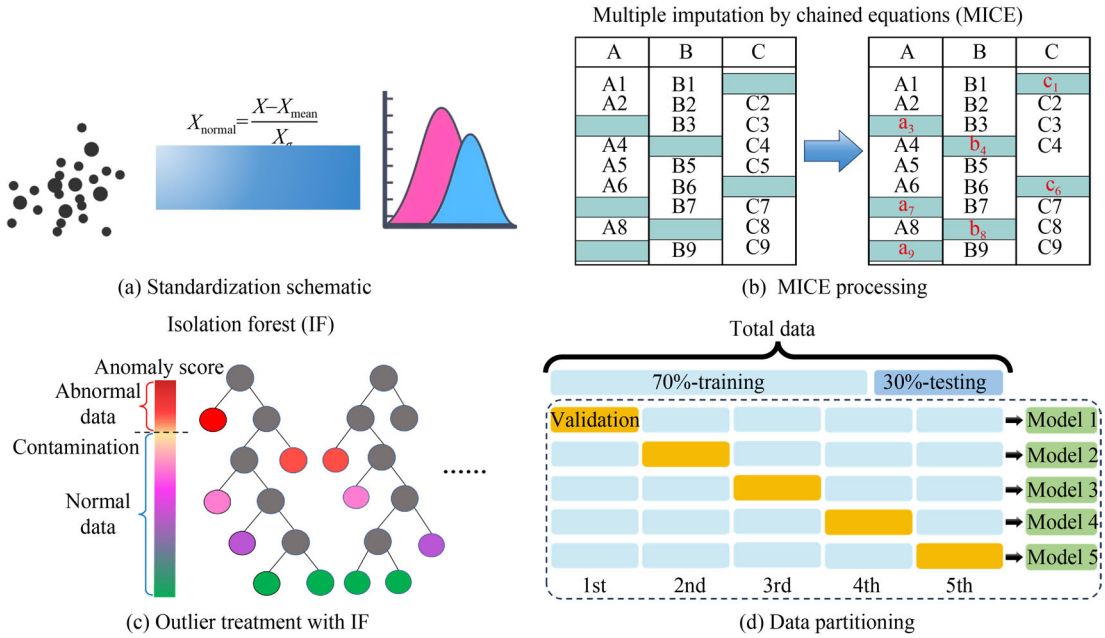


图 1 预处理过程图

Fig. 1 Preprocessing workflow

1.3 算法介绍

1.3.1 CGAN-MCMC 算法 CGAN-MCMC 是一种 CGAN 与马尔可夫链蒙特卡洛(MCMC)优化的混合生成模型，专为高维非独立同分布数据的概率密度逼近与统计特性保持而设计[29–30](如图 2)。其核心架构通过对抗训练生成初始分布与统计驱动的后验优化两阶段协同，突破传统生成模型在复杂分布建模中的局限性。该模型的关键技术可形式化表示为条件对抗生成阶段与 MCMC-KS 后验优化阶段。

1) 条件对抗生成阶段：生成器 G_θ 以噪声向量 $z\sim N(0, I)$和包含物理约束的条件变量 c (全变量约束)为输入，其中，物理约束综合考虑了材料组分与性能之间的相关性、数值关联以及组分与性能本身的合理取值范围。生成器通过深度映射学习目标分布的隐式表达，如式(2)所示。

$$\min_G \max_D V = \mathbb{E}_{x\sim p_{\text{data}}}(\ln D(x\,|\,c)) + \mathbb{E}_{z\sim p_z}(\ln(1-D(G(z\,|\,c))) \tag{2}$$

式中：判别器 D_ϕ 通过条件对抗机制引导生成样本 $\tilde{x}=G(z\,|\,c)$ 向真实数据流形 M_{real} 对齐，隐式实现特征间非线性依赖关系的建模。

2) MCMC-KS 后验优化阶段：针对生成样本 $\{\tilde{x}_i\}$，引入基于核密度估计(KDE)的 MCMC 转移核与统计量惩罚项，最小化生成分布 p_g 与真实分布 p_r

的差异：

$$\mathcal{L}_{\mathrm{MCMC}} = -\underbrace{\mathbb{E}_{\tilde{x}}\left[\ln p_{\mathrm{KDE}}(\tilde{x})\right]}_{\text{Local density matching}} + \lambda(\| \mathrm{Skew}(p_{\mathrm{g}}) - \mathrm{Skew}(p_{\mathrm{r}}) \|_2^2 + \| \mathrm{Kurt}(p_{\mathrm{g}}) - \mathrm{Kurt}(p_{\mathrm{r}}) \|_2^2) \tag{3}$$

通过 Metropolis-Hastings 准则实现样本迭代优化：

$$\alpha = \min\left(1, \frac{p_{\mathrm{KDE}}(x')\exp(-\mathcal{L}_{\mathrm{stats}}(x'))}{p_{\mathrm{KDE}}(x)\exp(-\mathcal{L}_{\mathrm{stats}}(x))} \cdot \frac{q(x \mid x')}{q(x' \mid x)}\right) \tag{4}$$

式中：Skew 为偏度衡量分布的高阶统计特性；Kurt 为峰度衡量分布的高阶统计特性；提议分布 q 采用高斯随机游走；λ 控制高阶统计量对齐强度。

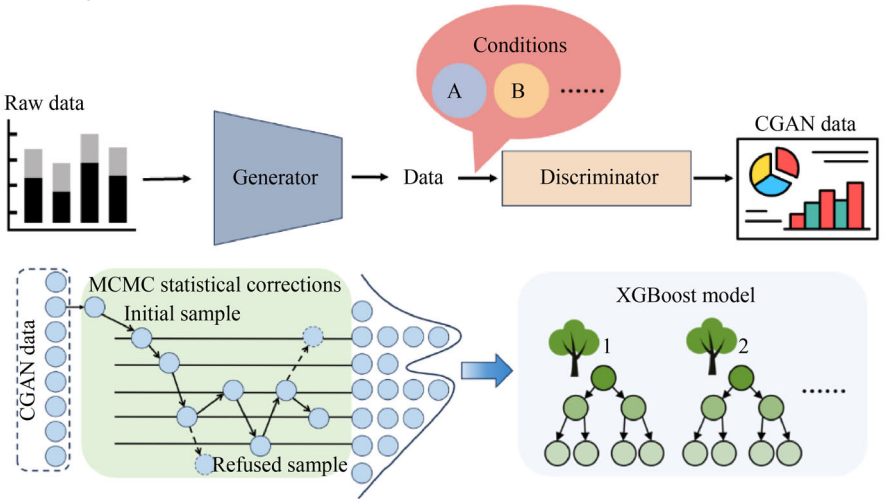


图 2　CGAN-MCMC 原理图

Fig. 2　CGAN-MCMC schematic

1.3.2　XGBoost 算法　　XGBoost 是一种基于梯度提升框架的机器学习算法，旨在通过集成弱学习器(决策树)的序列化训练，显著提升预测精度并控制模型复杂度[32]。其核心创新在于引入正则化项与动态优化策略，克服传统梯度提升树(如GBDT)的过拟合风险与计算效率限制。该模型通过最小化以下目标函数实现预测能力与可解释性的平衡：

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

式中：ℓ 为损失函数(本工作为回归预测任务，采用均方误差(MSE)作为损失函数)；$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \| w \|^2$ 为树模型的正则化项，通过惩罚叶子节点数量 T 与权重 Ω 的 L2 范值，实现双重正则化。

1.3.3　贝叶斯优化　　XGBoost 模型的预测性能在很大程度上依赖于超参数的合理设置。超参数直接决定了模型的复杂度、学习能力以及对未知数据的泛化能力。合适的超参数配置能够有效平衡偏差与方差，从而提升模型的整体表现。在未进行调优时，XGBoost 模型使用的是默认参数，尽管训练速度较快，但预测精度未必最优。为此，采用贝叶斯优化(BO)算法对 XGBoost 模型的关键超参数进行了自动化搜索与优化，以获取最适用于关键性能预测任务的超参数组合[33]。

1.3.4　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：

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。

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m} |f_i - y_i| \tag{6}$$

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。

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(f_i - y_i)^2} \tag{7}$$

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。

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(f_i - y_i)^2}{\sum_{i=1}^{m}(\bar{y}_i - y_i)^2} \tag{8}$$

式中：y_i 为实际值；f_i 为预测值；$\bar{y}_i$ 为实际值的平均值；m 为样本数量。

2　基于 CGAN-MCMC 的 UHPC 数据增强与评价

2.1　数据库预处理结果

UHPC 多源异构数据库存在数据缺失、异常值干扰等质量问题，本工作通过数据预处理的方法进行数据清洗，以构建满足机器学习需求的标准化数据集。通过 IF 异常检测算法对原始样本数据库进

行数据清洗，该算法基于树结构的隔离机制，通过量化样本在多维特征空间中的分布偏离程度识别潜在异常[16]。如图 3 所示，在数据预处理阶段，IF 模型通过自适应参数设定，从初始样本(N=157)中筛选出 12 个显著的异常数据点。清洗后的数据集(N=145)呈现出更稳定的分布特征，关键指标的离散程度明显降低，数据整体一致性显著增强。异常值剔除过程经可视化验证表明，被移除样本主要分布于特征空间的边缘区域或呈现与主体数据模式明显偏离的异常模式。该清洗操作有效提升了数据质量，为后续数据增强和建模分析提供了更可靠的基础，同时有助于降低异常干扰导致的模型偏差风险。

图 4 为预处理数据分布图。由图 4 可见，经异常值筛选后，所有的性能参数的数据频率分布呈现出改进的正态分布特性，偏度与峰度系数较原始数据显著改善，表明数据清洗有效提升了参数分布的规范性。然而，原始样本规模(N<200)仍难以满足高维特征空间建模需求，尤其在包含 10 余个工艺参数的复杂场景下，传统机器学习模型易受“维度灾难”制约，表现为模型对训练集噪声过度敏感，验证集准确率波动幅度超过 15%，存在明显的过拟合风险。为突破小样本数据约束，采用 CGAN-MCMC 模型进行数据增强，通过 CGAN 学习工艺参数与性能指

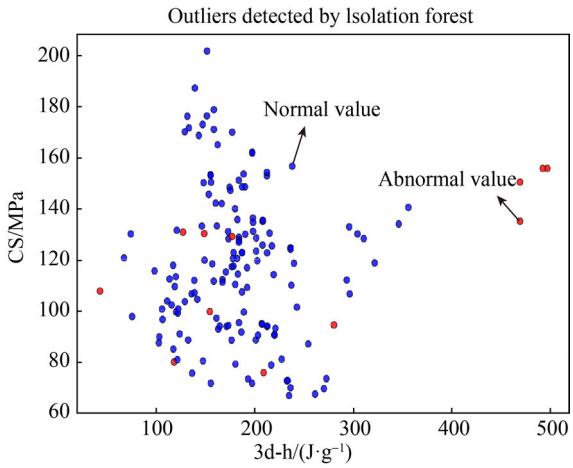


图 3 孤立森林筛选异常值

Fig. 3 Outliers detected by isolation forest

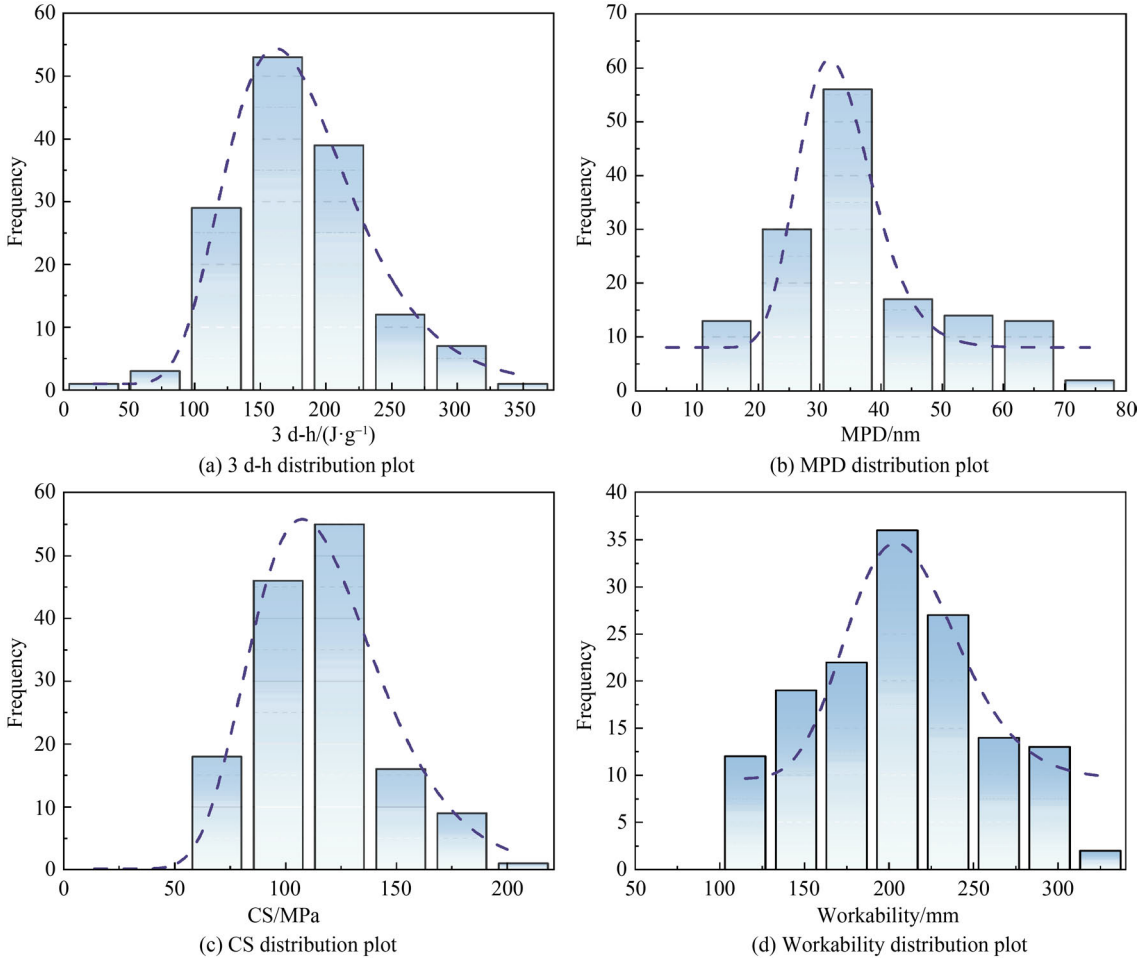


图 4 预处理数据分布图

Fig. 4 Distribution of pre-augmentation data

标间的复杂非线性映射关系，生成物理约束下的候选样本；继而利用 MCMC 算法对生成数据实施稳态检测与热力学平衡约束优化，通过接受–拒绝抽样机制确保合成数据与真实分布的兼容性，该混合策略在显著扩展数据规模的同时，严格维持关键参数的分布一致性，从而实现模型泛化能力的本质提升。

2.2　数据相关性分析

对未增强 UHPC 网络数据库进行皮尔逊相关分析，结果如图 5 所示。由图 5 可以看出，钢纤维掺量与硅灰掺量对 UHPC 抗压强度呈现出显著正向影响，相关系数分别为 0.56 和 0.38。其中，钢纤维的强增韧效应主要源于其独特的应力传递机制，当混凝土基体产生微裂纹时，钢纤维通过桥接作用有效传递应力(桥联效应)，显著延缓裂纹扩展进程[34]。此外，钢纤维掺量与工作性能呈现负相关关系，这是由于钢纤维掺量的增大会形成搭接网络，阻碍了颗粒流动[35]。硅灰的增强作用则主要通过火山灰效应实现，可促进水化产物以及微观结构的致密化，并且其纳米级颗粒可填充水泥颗粒间空隙，从而优化基体颗粒堆积结构[36]。正如所预期的，水胶比与强度呈现负相关关系(–0.19～–0.22)，这是因为较低的水胶比降低了硬化后混凝土内部的孔隙率，同时提升了水泥水化产物的微观致密性，从而显著提升了基体力学性能[37]。

基于上述分析结果，由于不同特征对混凝土性能的影响程度与交互机制存在显著差异，建议在机器学习建模时采用分层特征处理架构：对钢纤维掺量等强相关因子($|\rho|>0.6$)构建主特征空间，通过 SHAP 值分析揭示其非线性作用规律；对硅灰掺量等中等相关变量($0.3<|\rho|<0.6$)，采用 Gram-Schmidt 正交化处理以消除与矿物掺合料的共线性干扰，可以使得模型在预测时表现出更好的鲁棒性[38]。

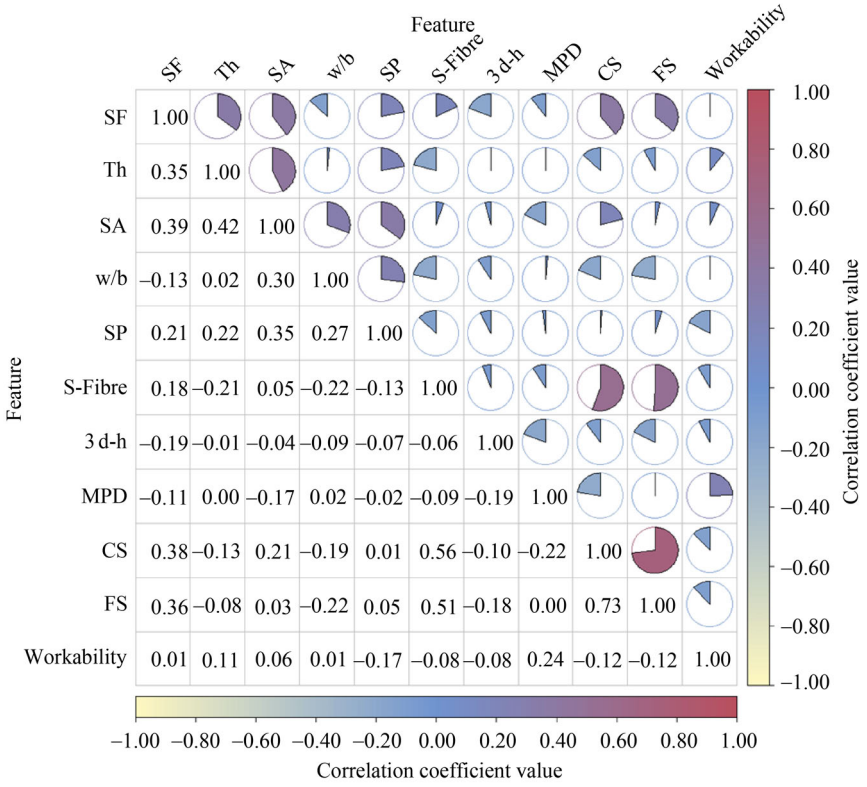


图 5　未增强的 UHPC 网络数据库的皮尔逊相关分析

Fig. 5　Pearson correlation analysis on the pre-augmentation database

2.3　CGAN-MCMC 增强数据库与可靠性评估

基于 CGAN-MCMC 框架实现数据增强后，UHPC 数据集规模扩展至 N=1500。如图 6 所示，增强数据集在抗压强度、工作性能等关键指标上均保持与原始分布的高度一致性，平均 Jensen Shannon(JS)距离<0.05，验证了物理信息约束生成机制对数据统计特性保真的有效性。

通过对比增强前后变量间的皮尔逊相关系数矩阵(如图 5 和图 7)，可以发现数据增强过程有效维持了原始数据库的统计关联特征。在关键材料参数维度，钢纤维掺量与力学性能的强正相关性保持稳定，其与 CS 的相关系数从增强前的 0.56 微调至 0.43 (降幅 23.2%)，但仍处于显著相关区间(p<0.01)。钢纤维掺量与工作性能的负向关联性

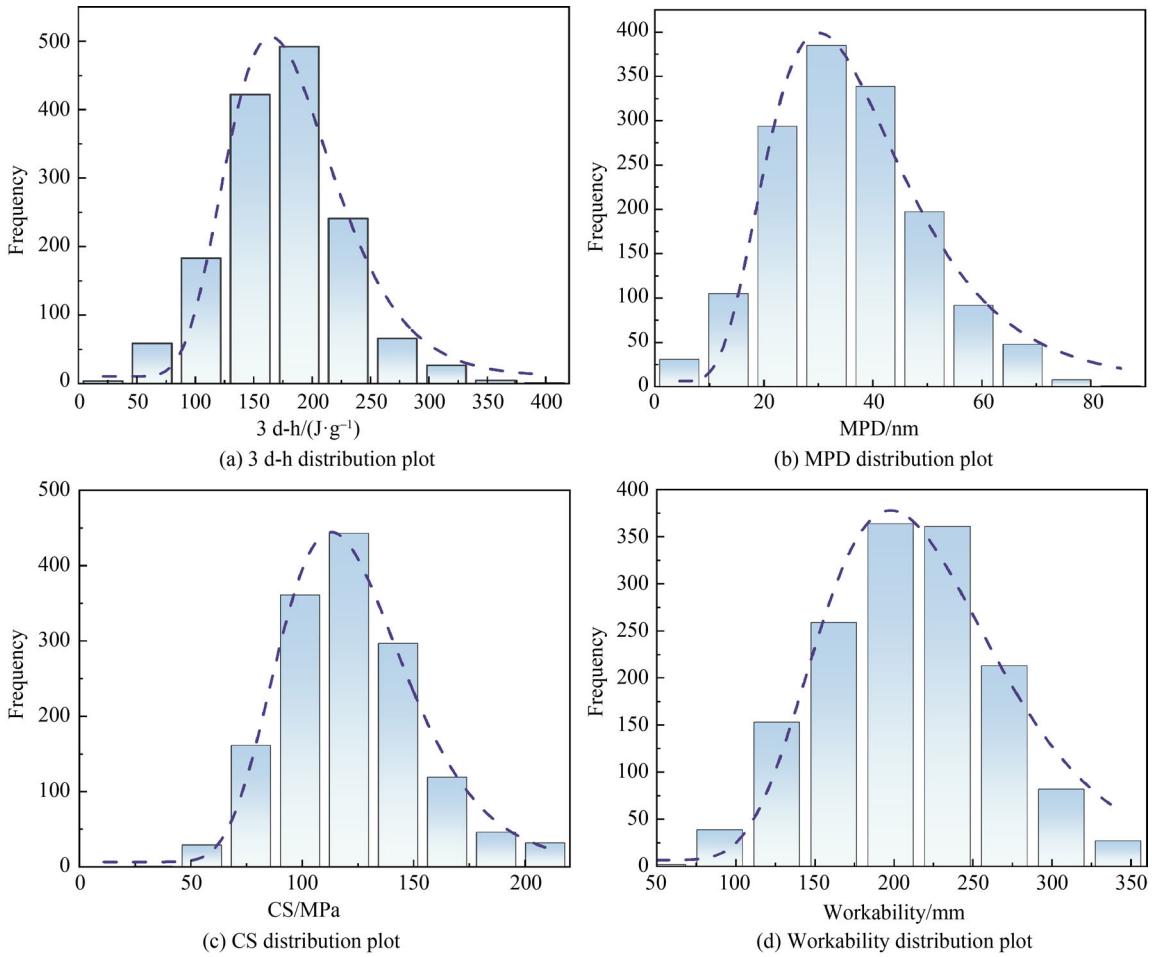


图 6 增强数据分布图

Fig. 6 Distribution of augmented data

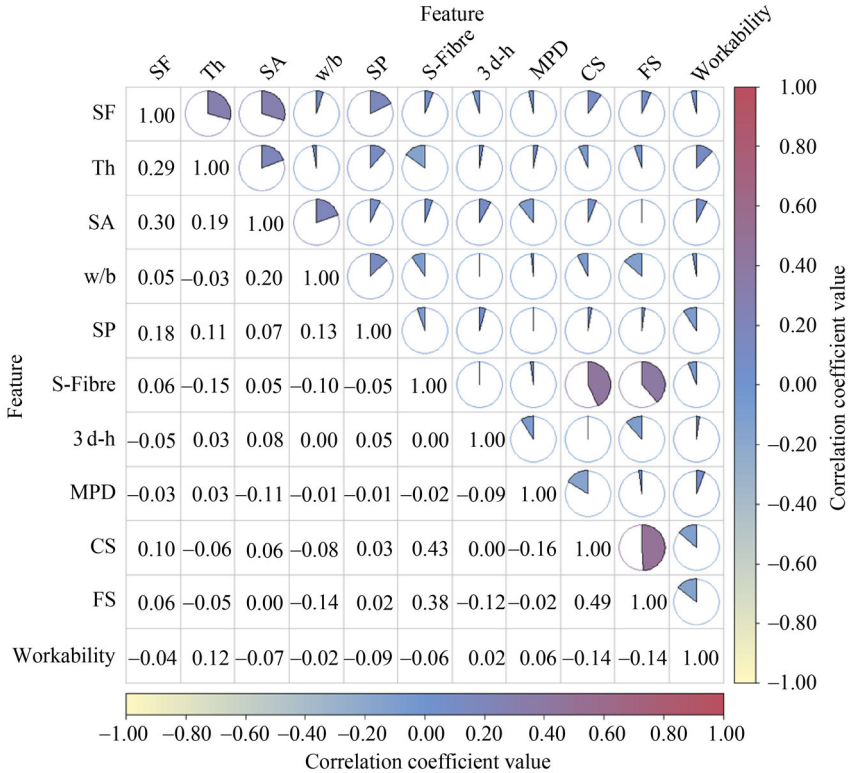


图 7 增强数据后的皮尔逊相关系数

Fig. 7 Pearson correlation coefficient of augmented data

(原始数据 r= –0.08)在增强后略微变化至 r= –0.06，波动幅度仅为 0.02，表明合成数据生成过程未“破坏纤维网络效应–浆体流变性”的物理耦合结果[39]。此外，w/b、SP 对力学性能的影响模式保持高度稳定，如水胶比(w/b)与抗压强度(CS)的负向关联性从 –0.19 变为–0.08，其统计学显著性水平(p=0.032→0.041)未发生本质改变。深层分析表明，约 82%的变量对相关系数波动幅度控制在±0.15 以内，其中 54%的关联性变化绝对值小于 0.05。这种统计特性的高度保真，证实了 CGAN-MCMC 生成器通过物理约束模块的有效性，实现了数据扩充与机理传承的平衡。

在初步验证原始数据与增强数据的分布一致性后，本工作采用 Kolmogorov-Smirnov (K-S)假设检验体系，系统论证 2 组数据的分布同源性。图 8 为经验累积分布函数(ECDF)结果，发现所有材料参数的完整曲线均展现出高度重叠特征(显著性水平 α=0.05)。具体表现为：抗压强度(图 8a，p=0.089)、工作性能(图 8b，p=0.751)、第三元粉料(图 8c，p=0.126)、中值孔径(图 8d，p=0.332)、3 d 累计水化放热量(图 8e，p=0.917)以及砂(图 8f，p=0.304)的 ECDF 分布。所有参数的统计检验结果均显著高于拒绝域阈值(p>0.05)，从数理统计角度判定了数据增强过程的分布不变性。

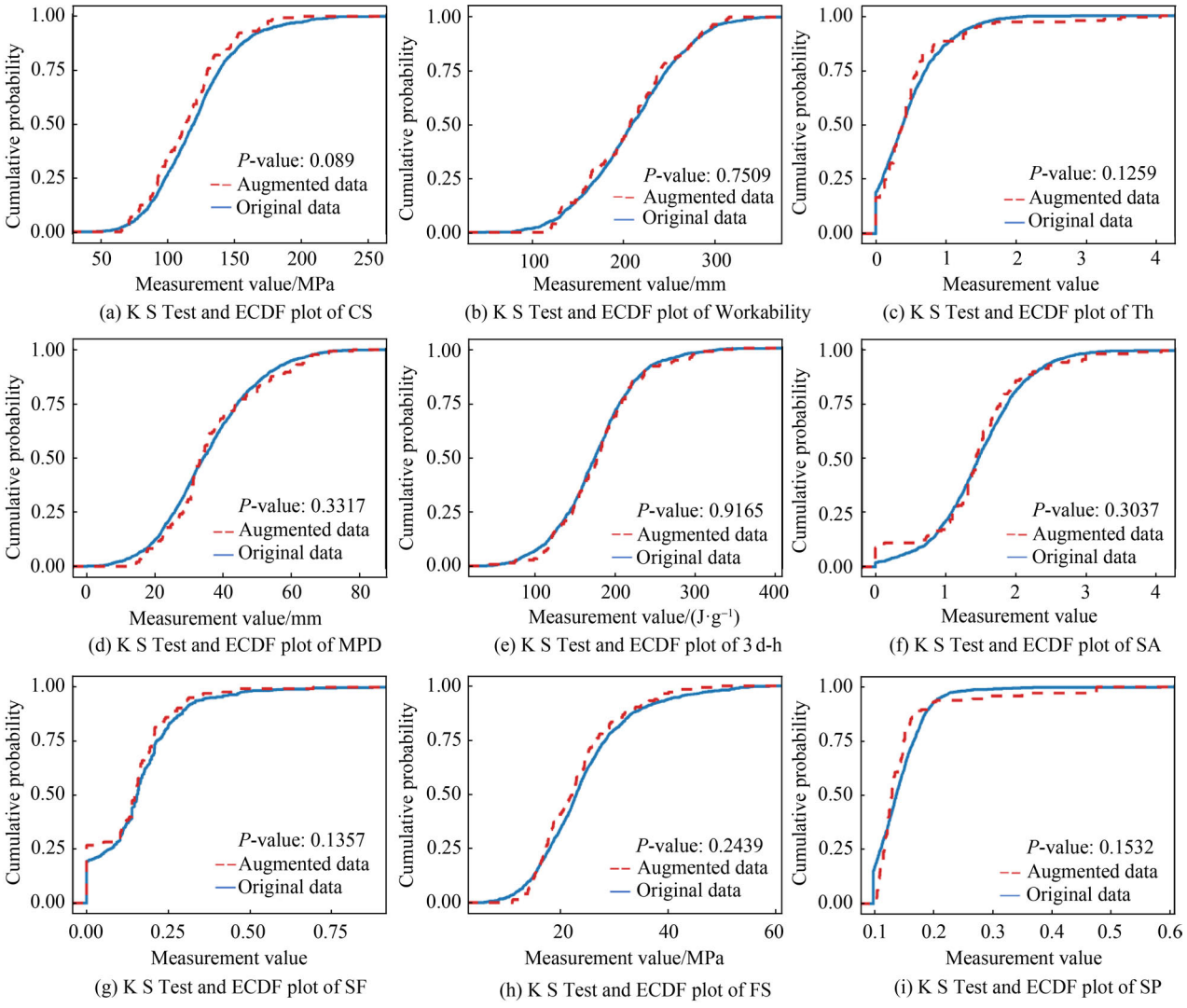


图 8　K-S 检验与 ECDF 图

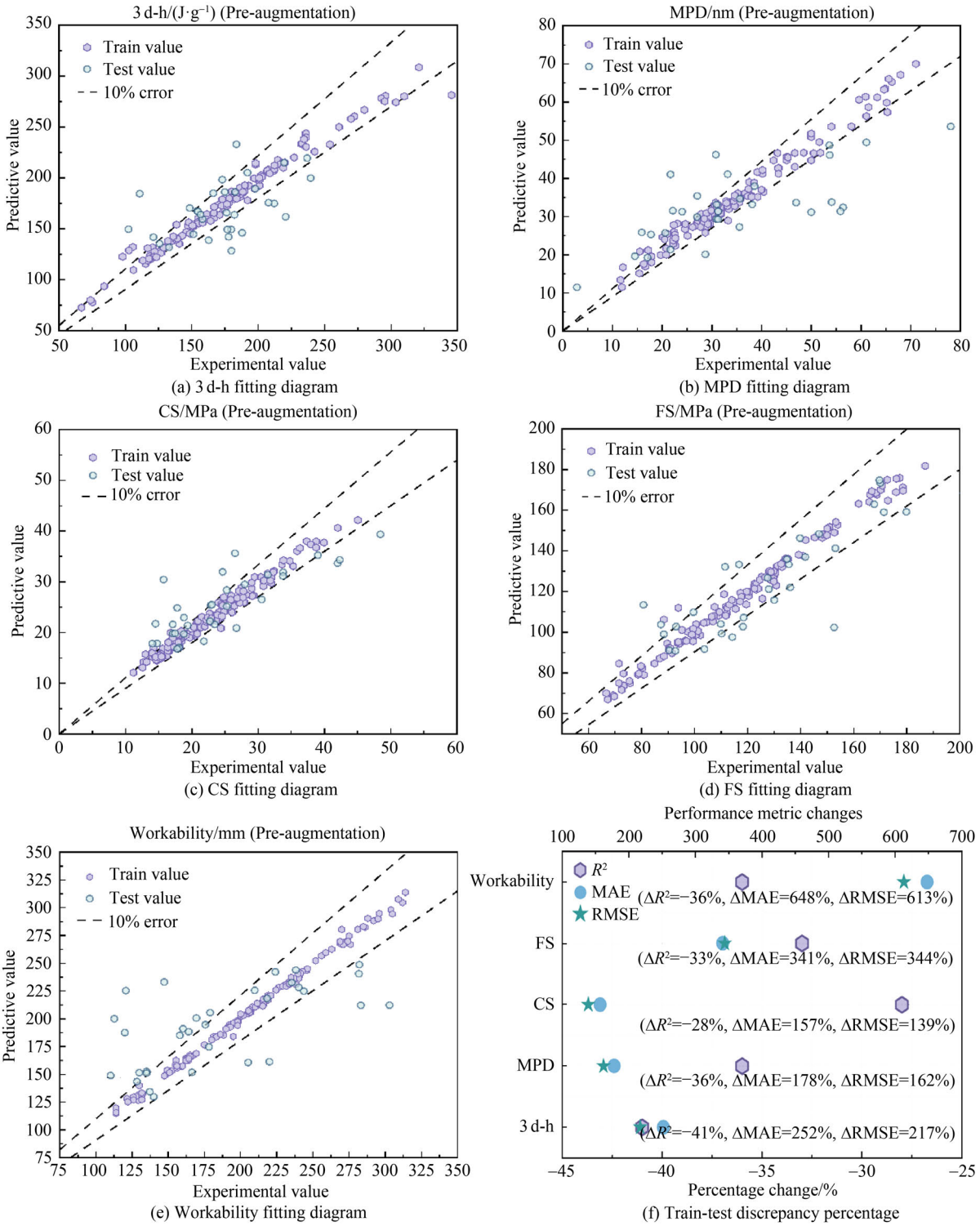
Fig. 8　K-S test validation with ECDF plot

定量分析表明，92.7%的合成数据点落入原始数据 95%置信区间，同时实现样本量 1,035%的跨数量级扩展。这一双重突破证明 CGAN-MCMC 生成器不仅完整保持了源数据的本征统计特性(峰度差异<0.15，偏度偏移量<0.08)，更通过马尔可夫链蒙特卡洛的动态平衡机制，在高维特征空间内实现了拓扑结构的精准复现。由此，通过构建“保真度–多样性”双目标优化验证框架，本方法在数学意义上保

证了增强数据集在提升机器学习模型鲁棒性时的零分布偏移特性，为小样本材料信息学中的可信数据增强提供了理论完备的解决方案。

2.4 人工智能算法评估增强数据库

通过 BO 优化后的 XGBoost 模型对增强前后的数据库进行性能预测，以验证 CGAN-MCMC 模型增强 UHPC 数据的可靠性。图 9 和表 2 分别为模型评估结果。由图 9 和表 2 可以看出，测试集结果大多分布在 10%误差线以外，所有性能指标均呈现显著的“高训练精度–低测试精度”特征，3 d 累计水化放热量从 0.98 下降至 0.58 (降幅 41%)、中值孔径从 0.97 下降至 0.62 (降幅 36%)、抗压强度从 0.98 下降



“Δ” represents the rate of change of the indicator, and a positive value indicates an increase while a negative value indicates a decrease.

图 9 XGBoost 模型预测增强前数据库结果

Fig. 9 XGBoost (Pre-augmentation) Performance

表 2　XGBoost 模型在未增强数据库上预测的评估结果

Table 2　Evaluation Results of XGBoost Model Predictions on the pre-augmentation database

Pre-augmentation	Training data			Testing data		
	R^2	MAE	RMSE	R^2	MAE	RMSE
3d-h	0.98	5.21	7.89	0.58	18.34	24.67
MPD	0.97	3.12	4.55	0.62	8.67	11.93
CS	0.98	4.82	6.93	0.71	12.37	16.55
FS	0.97	0.92	1.17	0.65	4.06	5.20
Workability	0.99	4.71	6.58	0.63	35.21	46.93

至 0.71 (降幅 28%)、抗折强度从 0.97 下降至 0.65 (降幅 33%)、工作性能从 0.99 下降至 0.63 (降幅 36%)，跨尺度性能预测中，测试集决定系数较训练集平均下降 34%。这种系统性泛化失效在误差指标中更为直观，所有性能的测试集 MAE 与 RMSE 均达到训练集的 3.1~7.5 倍(如工作性能 MAE:4.71→35.21，RMSE:6.58→46.93)，表明模型未能有效捕捉流变行为的本质规律。可见当面对样本量有限且数据质量欠佳的原始数据库时，未经数据增强处理的机器学习模型容易过度拟合训练数据中的噪声特征，导致其在新数据上的预测性能显著下降，表明模型仅记忆了训练数据的表层特征，未能有效捕捉 UHPC 性能演化的内在规律。

深层机理分析发现，原始数据库的固有缺陷是导致模型失效的主因，有限样本量(N<200)难以支撑 XGBoost 复杂树结构的可靠训练，特别是在 3 d 累计水化放热量(测试集 R^2=0.58)等涉及时序演变的预测任务中，模型易受实验噪声干扰，这种全性能域的过拟合现象本质上暴露了小样本条件下机器学习模型的根本局限：当采用 XGBoost 等复杂算法时，即便通过贝叶斯优化精细调参(最大深度=9，学习率=0.03)，模型仍会优先拟合数据统计特征而非物理规律。

经过 CGAN-MCMC 数据增强后，XGBoost 模型在 UHPC 多尺度性能预测中呈现出系统性优化特征，研究结果体现了 CGAN-MCMC 方法在数据增强方面的有效性。如表 2、表 3 和图 10 所示，数据增强后模型在测试集的拟合优度(R^2)显著提升：3d 累计水化放热拟合优度 R^2 从 0.58 提升至 0.94 (升幅 62%)，中值孔径从 0.62 提升至 0.96 (升幅 55%)，抗压强度从 0.71 提升至 0.84 (升幅 18%)，抗折强度从 0.65 提升至 0.95 (升幅 46%)，工作性能从 0.63 提升至 0.85 (升幅 35%)，关键性能指标平均提升幅度达 43%。值得注意的是，涉及时间依赖特性的水化放热预测改善最为显著，表明 CGAN-MCMC 算法成功捕获了材料时变特性的动力学机制。误差分析显示，测试集平均绝对误差(MAE)与均方根误差(RMSE)均降低了 65%，其中工作性能的 MAE 从 35.21 降至 11.13 (降幅 68%)，RMSE 从 46.93 降至 15.40 (降幅 67%)。

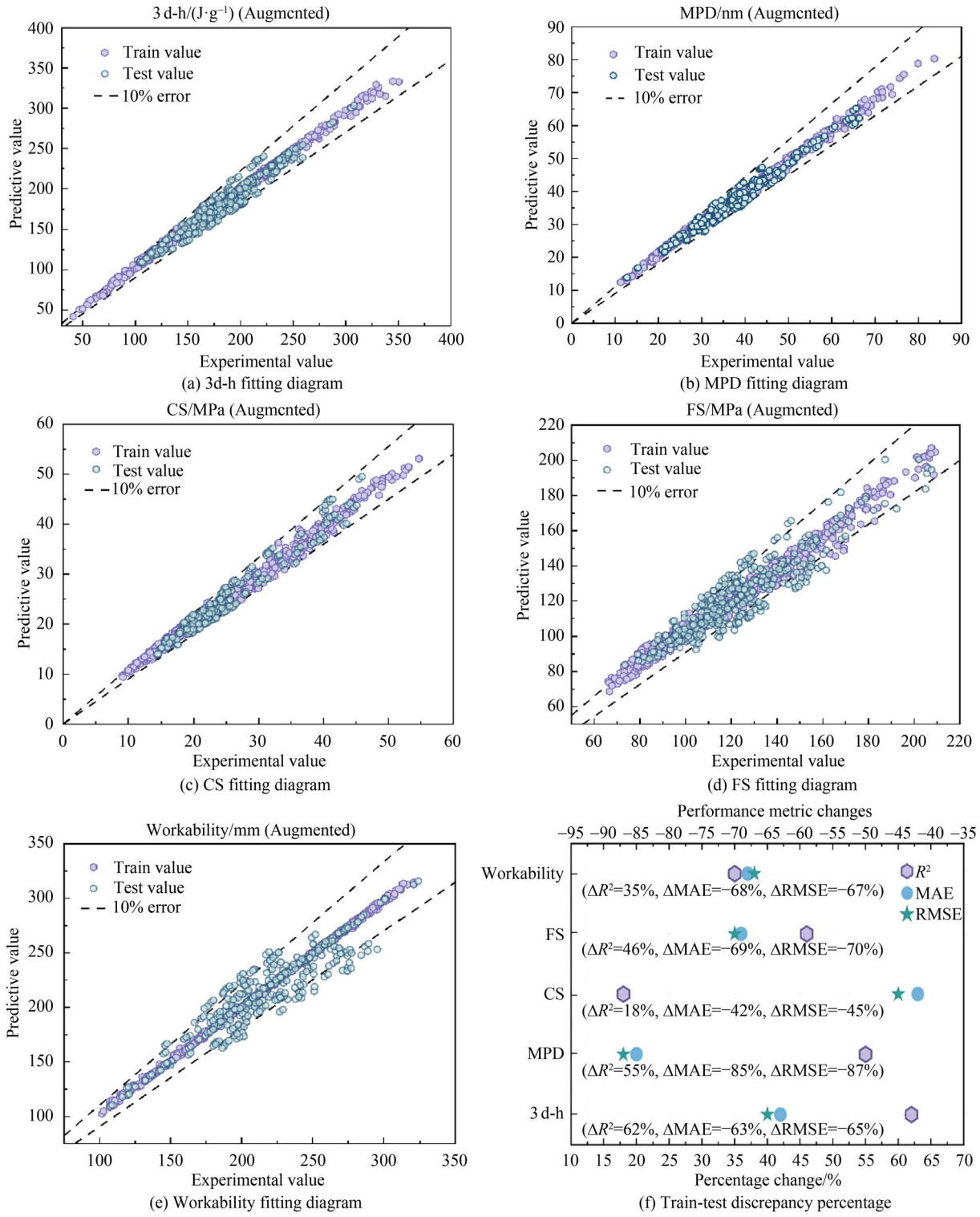
增强后模型训练集与测试集性能差距显著缩小，说明模型从“记忆噪声”向“学习规律”发生根本转变。CGAN-MCMC 增强策略成功解决了小样本条件下机器学习模型的根本性局限。通过构建具有物理一致性的合成数据，模型不仅突破了“高训练精度–低测试精度”的泛化困境，更在跨尺度性能预测中实现了工程级精度(测试集平均 R^2=0.91)，为 UHPC 材料智能化设计提供了可靠的数据驱动解决方案。

2.5　实际工程预测验证

本工作依托广东省某 UHPC 漂浮平台工程的材料设计数据进行实际工程验证。如表 4 所示，工程实验数据库(N=36)涵盖了水泥(P·Ⅱ 52.5 型，C)、硅灰(SF)、粉煤灰(FA)、双级配河砂(0~0.60 mm 与 0.60~1.25 mm，Sand1 & Sand2)、直钢纤维(L 13 mm×ϕ0.2 mm，Fibre)、偏高岭土(MK)、减水剂(SP)原材料参数及抗压强度(CS)。

基于 CGAN-MCMC 框架对实验室原始数据进行增强处理，得到 1000 组增强数据。如图 11 所示，采用数据增强前后的 XGBoost 模型预测了强度性能，可以发现在实验同源测试集上，模型测试集拟合优度 R^2 从 0.65 提升至 0.92，同时训练集与测试集的平均绝对误差(MAE)分别下降 35%与 50%，均方根误差(RMSE)分别下降 39%与 48%，这些指标的系统性改善，不仅反映了模型整体预测精度的跃升，更深刻揭示了增强数据在优化模型非线性映射能力与有效抑制过拟合倾向方面的双重优势。因此，上述结果证明 CGAN-MCMC 有效消除了非物理震荡信号，构建了符合实验规律的特征关联，使模型在训练与测试误差收敛趋势上呈现更高一致性。这不

仅从量化指标上证实了方法的有效性，更从工程可解释性的深层次角度，验证了 CGAN-MCMC 所生成增强数据的高度合理性与可靠性，显著提升了模型的泛化能力和实际应用价值。



"Δ" represents the rate of change of the indicator, and a positive value indicates an increase while a negative value indicates a decrease.

图 10 XGBoost 模型预测增强后数据库结果

Fig. 10 XGBoost (Augmented) Performance

2.6 不足与展望

尽管本工作证明了 CGAN-MCMC 模型是增强 UHPC 数据库的有效手段，但其在实际应用中仍存在一些不足：生成数据的有效性高度受制于原始数

据集的分布真实性，一旦初始数据存在系统性偏差，便可能放大预测误差；模型训练过程中需进行 10^5 级迭代，计算耗时约达传统 GAN 模型的 2.2 倍，对计算资源要求较高；在面对更小初始数据集时，生成样本的多样性与准确性均可能出现下滑。基于此，未来研究方向可聚焦于：1) 将 Fick 扩散定律等材料科学基础方程嵌入生成器损失函数，构建物理信息增强的机器学习模型，增强输出结果的物理可解释性；2) 积极拓展模型在冻融循环、氯盐侵蚀等极端环境复杂工况下的应用；3) 同时引入迁移学习等前沿技术降低对大规模数据的依赖，并通过优化网络架构削减计算成本，以此全方位提升模型的鲁棒性与工程适用性，为 UHPC 领域的智能化发展注入新动力。

表 3　XGBoost 模型在增强数据库上预测的评估结果

Table 3　Evaluation results of XGBoost model predictions on the augmented database

Augmented	Training data			Testing data		
	R^2	MAE	RMSE	R^2	MAE	RMSE
3d-h	0.99	2.39	2.83	0.94	6.84	8.72
MPD	0.99	0.73	0.75	0.96	1.34	1.50
CS	0.98	2.89	3.55	0.84	7.16	9.16
FS	0.99	0.67	0.89	0.95	1.27	1.55
Workability	0.99	0.93	1.17	0.85	11.13	15.40

表 4　工程配合比数据分布

Table 4　Distribution of engineering mixtures

Materials	Unit	N	Min	Max	Mean	Std
C	ρ/(kg·m^{-3})	36	450.00	950.00	722.20	102.59
FA	ρ/(kg·m^{-3})	36	0	306.00	127.63	100.23
SF	ρ/(kg·m^{-3})	36	0	165.00	113.77	44.31
MK	ρ/(kg·m^{-3})	36	0	300.00	55.87	84.54
Sand1	ρ/(kg·m^{-3})	36	0	990.00	591.40	199.82
Sand2	ρ/(kg·m^{-3})	36	0	742.5.00	355.37	183.36
Fibre	ρ/(kg·m^{-3})	36	39.00	234.00	141.97	65.59
Water	ρ/(kg·m^{-3})	36	169.00	198.00	185.20	8.45
SP	ρ/(kg·m^{-3})	36	30.00	35.00	31.80	1.80
CS	MPa	36	116.10	180.10	149.42	15.16

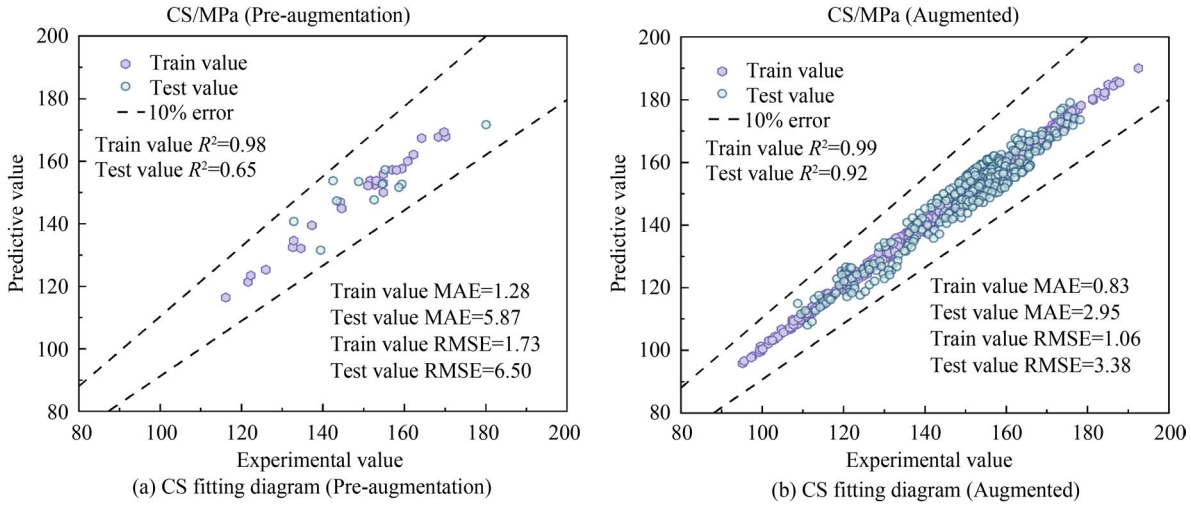


图 11　XGBoost 模型预测增强前后的实验室抗压强度

Fig. 11　Predicting compressive strength of laboratory-tested data before and after enhancement using XGBoost model

3 结论

1) 采用 MICE 和 IF 算法联用，实现了缺失值的填补和异常数据的精准剔除；随后基于 CGAN-MCMC 算法构建了 1500 组 UHPC 增强数据库。增强前后数据库呈现高度分布一致性，其 JS 散度<0.05，K-S 检验 p 值均>0.05，且输入输出变量间皮尔逊相关系数差异<0.13，验证了数据增强过程在统计学意义上有效保持了原始特征关联性及分布特性。

2) 基于 BO 优化的 XGBoost 模型，对数据增强前后的 UHPC 网络数据库进行性能预测，其中 3 d 累计水化放热量、中值孔径、抗压强度、抗折强度及工作性能的测试集的 R^2 分别提升了 62%、55%、18%、46%和 35%，有效验证了 CGAN-MCMC 模型数据增强的可靠性。

3) 基于广东省某 UHPC 实际工程，对增强前后的数据库进行性能预测，测试集的 R^2 从 0.65 提升至 0.92，同时训练集与测试集的 MAE 分别降低 35%与 50%，RMSE 分别下降 39%与 48%，从实际工程上验证了 CGAN-MCMC 模型数据增强的可行性。

4) CGAN-MCMC 数据增强框架为 UHPC 的智能化发展提供了理论创新与方法突破，该模型有效解决了 UHPC 领域实验成本高昂、多尺度数据稀缺导致的分析瓶颈，显著提升了材料性能预测的可靠性与泛化能力，为建筑工程材料数字化及智能基础设施建设提供了可迁移的技术路径。

参考文献:

[1] 元强, 张胶玲, 张苏辉, 等. CO_2 矿化再生骨料物理性能影响因素与机理[J]. 建筑材料学报, 2024, 27(10): 895–903.
YUAN Qiang, ZHANG Jiaoling, ZHANG Suhui, et al. J Build Mater, 2024, 27(10): 895–903.

[2] ZHAO D Q, SHEN W G, BIRHANU A H, et al. A novel sintering technique to control cement's water-soluble chromium (VI) release: Steel slag fusion adhering calcined gradient high-ferrite clinker[J]. Constr Build Mater, 2025, 466: 140126.

[3] 刘剑辉, 王猛生, 杨玉柱, 等. 混凝土抗冻融破坏性能研究进展[J]. 硅酸盐学报, 2025, 53(1): 190–211.
LIU Jianhui, WANG Mengsheng, YANG Yuzhu, et al. J Chin Ceram Soc, 2025, 53(1): 190–211.

[4] 张鹏, 董必钦, 元强. 先进海洋材料服务海洋强国建设[J]. 硅酸盐学报, 2023(11): 2739–2740.
ZHANG Peng, DONG Biqin, YUAN Qiang. J Chin Ceram Soc, 2023(11): 2739–2740.

[5] FAN D Q, CHEN Z A, CAO Y, et al. Intelligent predicting and monitoring of ultra-high-performance fiber reinforced concrete composites − A review[J]. Compos Part A Appl Sci Manuf, 2025, 188: 108555.

[6] XU Z S, SUN W, MIAO S L, et al. Effect of container geometry induced slippage on rheological measurement of fresh cement-based materials using coaxial cylinder rheometer[J]. Constr Build Mater, 2025, 471: 140790.

[7] 冯甘霖, 罗盛禹, 罗启灵, 等. 氧化石墨烯涂覆砂增强水泥基材料抗氯离子侵蚀性能[J]. 硅酸盐学报, 2025, 53(3): 505–518.
FENG Ganlin, LUO Shengyu, LUO Qiling, et al. J Chin Ceram Soc, 2025, 53(3): 505–518.

[8] FANG L, FU D B, YUAN Q, et al. Combined effects of low-heat cement, expansive agent and shrinkage-reducing admixture on drying shrinkage and cracking of concrete[J]. Case Stud Constr Mater, 2025, 22: e04344.

[9] XIANG Y, ZHAO H T, LIU J P, et al. Shrinkage cracking law and anti-crack inverse design in early-age concrete: A novel perspective on the development of crack resistance properties[J]. Eng Fract Mech, 2024, 310: 110395.

[10] WANG H, LIU Y P, HU Z L, et al. Influencing aspects and mechanisms of steel bar reinforcement on shrinkage and cracking of cement-based materials: A review[J]. J Build Eng, 2023, 77: 107476.

[11] TENG L, KHAYAT K H, LIU J P. Recent advances in the rheological properties of ultra-high-performance concrete: A critical review[J]. Engineering, 2025

[12] CHEN Y J, XU Y, YANG G, et al. Research on creep behaviour of UHPC based on experiments and viscoelastic modelling[J]. J Build Eng, 2024, 84: 108585.

[13] ZHANG X H, WU Z M, XIE J, et al. Trends toward lower-carbon ultra-high performance concrete (UHPC)–A review[J]. Constr Build Mater, 2024, 420: 135602.

[14] JIA M J, CHEN X Y, XU Z C, et al. A new approach for constructing UHPC conductive pathways: Oriented deposition of conductive hydration products[J]. Cem Concr Res, 2025, 191: 107827.

[15] 余睿, 陈飞翔, 范定强, 等. 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[J]. 硅酸盐学报, 2025, 53(5): 1165–1176.
YU Rui, CHEN Feixiang, FAN Dingqiang, et al. J Chin Ceram Soc, 2025, 53(5): 1165–1176.

[16] 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.

[17] GAO C, LUO C Y, ZHANG B, et al. Creep behavior evaluation of recycled aggregate concrete using machine learning technology[J]. J Build Eng, 2025, 99: 111538.

[18] 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.

[19] YU F, CAI H, ZHANG H, et al. Automatic identification method for three-phase structure of pervious concrete based on deep learning network of Mask R-CNN[J]. Constr Build Mater, 2024, 420: 135534.

[20] RANJBAR I, TOUFIGH V. Deep long short-term memory (LSTM) networks for ultrasonic-based distributed damage assessment in concrete[J]. Cem Concr Res, 2022, 162: 107003.

[21] 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.

[22] ZHOU X X, DUAN F, LIU W H, et al. Research on prediction of rheological properties of steel-polypropylene fiber UHPC mortar based on ensemble machine learning[J]. Constr Build Mater, 2025, 472: 140833.

[23] LI Z Z, YOON J, ZHANG R, et al. Machine learning in concrete science: Applications, challenges, and best practices[J]. NPJ Comput Mater, 2022, 8: 127.

[24] FORD E, MANEPARAMBIL K, KUMAR A, et al. Transfer (machine) learning approaches coupled with target data augmentation to predict the mechanical properties of concrete[J]. Mach Learn Appl, 2022, 8: 100271.

[25] 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.

[26] MIYAN N, ANOOP KRISHNAN N M, 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.

[27] CHEN N, ZHAO S B, GAO Z W, et al. Virtual mix design: Prediction of compressive strength of concrete with industrial wastes using deep data augmentation[J]. Constr Build Mater, 2022, 323: 126580.

[28] WANG X M, XIAO Y J, LI W Q, et al. Kriging-based surrogate data-enriching artificial neural network prediction of strength and permeability of permeable cement-stabilized base[J]. Nat Commun, 2024, 15(1): 4891.

[29] DING X, WANG Y W, XU Z H, et al. Distilling and transferring knowledge *via* cGAN-generated samples for image classification and regression[J]. Expert Syst Appl, 2023, 213: 119060.

[30] DE OLIVEIRA J T, DE CARVALHO COSTA L R, CARDOSO ESTUMANO D, et al. Applying Bayesian statistics and MCMC to ozone reaction kinetics: Implications for water treatment models[J]. Chemosphere, 2025, 373: 144164.

[31] LYNGDOH G A, ZAKI M, ANOOP KRISHNAN N M, et al. Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning[J]. Cem Concr Compos, 2022, 128: 104414.

[32] DONG W, HUANG Y M, LEHANE B, et al. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring[J]. Autom Constr, 2020, 114: 103155.

[33] WANG S Q, XIA P, GONG F Y, et al. A Bayesian-physical informed conditional tabular generative adversarial network framework for low-carbon concrete data augmentation and hyperparameter optimization[J]. Eng Appl Artif Intell, 2025, 152: 110811.

[34] CUNHA V M C F, BARROS J A O, SENA-CRUZ J M. An integrated approach for modelling the tensile behaviour of steel fibre reinforced self-compacting concrete[J]. Cem Concr Res, 2011, 41(1): 64–76.

[35] ZHANG H M, CAO L N, DUAN Y F, et al. High-flowable and high-performance steel fiber reinforced concrete adapted by fly ash and silica fume[J]. Case Stud Constr Mater, 2024, 20: e02796.

[36] WU Z M, KHAYAT K H, SHI C J. Changes in rheology and mechanical properties of ultra-high performance concrete with silica fume content[J]. Cem Concr Res, 2019, 123: 105786.

[37] SU W W, LIU J H, LIU L P, et al. Internal curing of fine coral aggregate in cement mortars with low water-to-cement ratio: Difference in freshwater and seawater[J]. Cem Concr Compos, 2024, 150: 105572.

[38] XU W Y, ZHANG L Y, FAN D Q, et al. AI-infused characteristics prediction and multi-objective design of ultra-high performance concrete (UHPC): From pore structures to macro-performance[J]. J Build Eng, 2024, 98: 111170.

[39] JI D D, WU Z R, GAO D Y, et al. Analysis on rheological and workability properties of steel-polypropylene fibers reinforced self-compacting concrete with recycled coarse and fine aggregates[J]. J Build Eng, 2023, 79: 107931.

A Novel Approach for Ultra-High Performance Concrete Data Augmentation and Intelligent Prediction Driven by Conditional Generative Adversarial Network-Markov Chain Monte Carlo Model

YU Rui[1], *LI Wang*[1], *FAN Dingqiang*[2,3]

(1. State Key Laboratory of Silicate Materials for Architectures, Wuhan University of Technology, Wuhan 430070, China; 2. Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong 999077, China; 3. Building Materials Engineering Laboratory, Department of Architecture, The University of Tokyo, Tokyo 113-8654, Japan)

Extended Abstract

Introduction Under the dual pressures of global extreme climates and the demand for long service lives in major engineering projects, conventional concrete materials increasingly struggle to meet performance requirements for critical infrastructures like cross-sea bridges and nuclear facilities, highlighting a urgent need for advanced construction materials. Ultra-high performance concrete (UHPC) emerges as a transformative solution for modern engineering challenges due to its exceptional compressive strength (>120 MPa), durability, and lightweight design advantages. However, UHPC development faces critical barriers, *i.e.*, the intricate nonlinear relationships among its composition, microstructure, and properties undermine conventional modeling approaches, while the scarcity of high-quality datasets (typically <200 samples) with inconsistent experimental protocols severely limits the reliability of data-driven models. This “algorithm-data mismatch” dilemma necessitates innovative database enhancement strategies.

The existing efforts in transfer learning and data augmentation, including generative adversarial networks (GANs) and Kriging interpolation, show a promise but encounter three fundamental limitations, *i.e.*, physical inconsistency in synthetic data (*e.g.*, illogical porosity-strength correlations), high-dimensionality failures in interpolation methods, and domain-shift distortions during cross-material knowledge transfer. To address these challenges, this study was to propose a dual-driven data enhancement framework

integrating conditional GAN (CGAN) and Markov Chain Monte Carlo (MCMC) techniques. The methodology systematically resolved dimensionality issues through preprocessing, generated physically constrained virtual samples *via* CGAN under constitutive material relationships, and ensures topological consistency using MCMC statistical correction. A Bayesian-optimized XGBoost model could validate the enhanced dataset, with experimental verification from marine floating platform UHPC applications, demonstrating the framework efficacy in overcoming multi-dimensional data generation constraints. This approach was to establish a robust paradigm for intelligent performance prediction under small-sample conditions, advancing digital design capabilities for concrete structures in complex marine environments.

Methods This study developed a CGAN-MCMC model to address small database augmentation and intelligent prediction for UHPC. The process commenced with standardized preprocessing of the original UHPC database, where missing values and outliers were systematically handled through the multiple imputation by chained equation (MICE) method and isolation forest algorithm, respectively, and followed by K-fold cross-validation partitioning. The CGAN-MCMC model was then employed to enhance the UHPC database as distribution comparisons and Kolmogorov-Smirnov (KS) tests conducted to verify the homogeneity between the original and augmented databases. Subsequently, a Bayesian optimization (BO)-tuned XGBoost model was implemented to evaluate performance prediction capabilities by both pre- and post-augmentation datasets. The model validity was further confirmed through validation against an actual UHPC engineering dataset from a floating platform project in Guangdong Province, China, demonstrating its practical applicability and robustness in engineering scenarios.

Results and discussion The data standardization effectively eliminates feature scale discrepancies, mitigating overfitting risks caused by parameter weight bias. The MICE method significantly enhances model noise resistance and prediction stability, establishing a high-reliability data foundation for UHPC performance prediction models. The isolation forest (IF) algorithm effectively removes outliers induced by instrument drift or mixing errors, while improving dataset robustness. The K-fold cross-validation further strengthenes model generalization *via* comprehensively evaluating performance across diverse data subsets. The optimized dataset demonstrates marked improvements in prediction accuracy and consistency across all performance metrics.

The CGAN-MCMC model is applied to augment the preprocessed UHPC database (N=145), generating an enhanced dataset (N=1,500). A statistical homogeneity between original and augmented data is confirmed through comparative analysis of distribution patterns, Pearson correlation matrices, and Kolmogorov-Smirnov (KS) tests (p-values >0.05). A Bayesian optimization (BO)-tuned XGBoost model reveals substantial improvements in test-set R^2 values for 3-day hydration heat (62%), median pore size (55%), compressive strength (18%), flexural strength (46%) and workability (35%) when using augmented data, quantitatively validating the CGAN-MCMC enhancement methodology.

The validation with engineering data from a floating platform project in Guangdong Province, China, demonstrates an enhanced model performance (*i.e.*, test-set R^2 increases from 0.65 to 0.92, with mean absolute error (MAE) reductions of 35% (training) and 50% (testing), alongside root mean squared error (RMSE) decreases of 39% (training) and 48% (testing)). These results empirically confirm an engineering feasibility of the proposed framework. The proposed methodology advances UHPC intelligent design through improved data quality and model generalization, providing an innovative solution for performance optimization.

Conclusions This study demonstrated a transformative potential of machine learning in overcoming data limitations and enhancing UHPC performance prediction. The study established a paradigm-shifting approach for UHPC design optimization *via* integrating advanced data augmentation with optimized ML modeling. The CGAN-MCMC framework could resolve critical challenges in material informatics and provide actionable insights for developing sustainable, high-performance concrete systems, accelerating the digital evolution of construction materials engineering.

Keywords artificial intelligence; ultra-high performance concrete; data augmentation; machine learning; intelligent prediction