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.0454-5648.20250734

基于生成对抗的钢纤维混凝土强度预测及沙普利加性解释分析

杨 哲[1]，张 悦[1]，徐 衍[2]，侯东帅[1]，王 攀[1]，王慕涵[1]，王鑫鹏[1]

(1. 青岛理工大学土木工程学院，山东 青岛 266520；2. 中国水利水电第四工程局有限公司，西宁 810000)

摘 要：为研究多因素耦合作用下钢纤维混凝土(SFRC)抗压性能的影响机制，本工作采用机器学习方法对其抗压强度进行预测。研究基于 141 组样本和 11 个关键参数，构建并评估了多种 ML 模型，结果表明，这些模型在强度预测中具有较高精度。为解决试验数据有限的问题，引入生成对抗网络(GAN)进行数据增强，并显著提升了模型的泛化性能。进一步的 SHAP 分析表明，水泥与硅灰含量对强度影响最大，其配比直接作用于最终强度的发展。通过将 ML 建模、GAN 增强与 SHAP 机理分析相结合，为 SFRC 的设计与工程应用提供了有效的预测工具和理论依据。

关键词：钢纤维混凝土；机器学习；随机森林；生成对抗网络；沙普利加性解释
中图分类号：TU528　　**文献标志码**：A　　**文章编号**：0454–5648(2026)03–0982–11
网络出版时间：2026–02–10

钢纤维混凝土(SFRC)[1–2]是在混凝土基体中掺入连续的长纤维或不连续的短纤维以增强抗裂性、韧性和耐久性等性能而制成的复合材料。

传统的实验方法通常通过控制变量法逐一研究各因素对 SFRC 性能的影响。例如，Wang 等研究了氯离子环境下纤维体积含量对聚丙烯纤维增强混凝土性能的影响，结果表明，纤维在提升耐久性方面作用显著，但最佳掺量仍需大量试验确定[3]。Lan 等考察了钢纤维、聚乙烯醇纤维与聚丙烯纤维在工程胶凝复合材料中的桥接效应，发现不同纤维在改善延展性与强度方面的作用差异明显，而同时兼顾两者仍具有挑战性[4]。

近年来，研究人员在混凝土抗压性能预测中逐渐采用数据驱动的机器学习(ML) [5]。例如，Dai 等采用 Bagging 技术预测 SFRC 抗压强度，结果显示其模型的决定系数(R^2)达到 0.94，并且 SHAP 分析揭示了固化时间与减水剂含量的重要作用[6]；李永健等利用 AdaBoost 预测 SFRC 28 d 抗压强度，取得了决定系数(R^2)=0.96，均方根误差(RMSE)=8 MPa 的结果，并确定硅灰、减水剂和水泥对强度影响最为显著[7]。上述研究表明，ML 在混凝土抗压性能预测中具有较高的应用潜力。

由于有限的数据集往往不足以支撑高性能 ML 模型的充分训练。Zhou 等[8]通过数据增强为深度学习分类器生成合成样本，从而在有限数据条件下实现对新兴技术的应用。Jiang 等[9]利用生成对抗网络(GAN)合成刀具磨损状态数据，显著提升了回归预测精度，结果表明，GAN 在材料与工程预测领域具有良好的应用潜力。

尽管 ML 在预测方面表现突出，但其“黑箱”特征导致的可解释性不足，限制了其在实际工程中的应用。研究人员提出 SHAP 方法，该方法是一种基于博弈论的特征贡献分析方法，能够量化各输入特征对模型预测结果的影响，揭示再生骨料混凝土强度的主控因素。因此，将 SHAP 引入 SFRC 性能预测可为机制理解提供理论依据。综上所述，传统实验方法在揭示多因素耦合作用下的 SFRC 抗压性能影响机制方面存在不足，而 ML 方法在预测中展现出较高精度，但受到数据规模和模型可解释性的限制。本工作通过将 GAN 数据增强方法与 SHAP 可解释性分析相结合，旨在提升模型的泛化能力，并揭示 SFRC 抗压强度的关键影响因素，从而为 SFRC 配合比设计与工程应用提供可靠工具与新思路。

收稿日期：2025–09–30。　修订日期：2025–11–13。
基金项目：国家自然科学基金(U24A20166)。
第一作者：杨 哲(1993)，男，博士研究生。
通信作者：侯东帅(1986)，男，博士，教授。

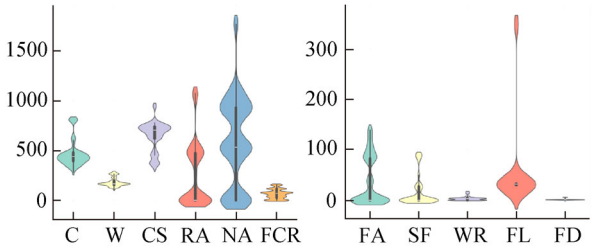
Received date: 2025–09–30.　**Revised date:** 2025–11–13.
First author: YANG Zhe (1993–), male, Doctoral candidate.
E-mail: yangzhe_x@163.com
Correspondent author: HOU Dongshuai (1986–), male, Ph.D., Professor.
E-mail: dshou@outlook.com

1　数据库建立和数据收集

1.1　数据收集与预处理

数据集来源为近 10 年发表的经同行评审的学术论文[10–21]。研究团队在数据筛选过程中重点关注混凝土制备工艺与配合比参数，包括水泥、粉煤灰等。所有不完整或不符合要求的样本均被剔除，最终获得 141 个有效数据条目。需要说明的是，尽管数据来源于不同文献，其原材料(如水泥强度等级、骨料来源等)存在一定差异，但本研究通过特征标准化处理和数据分布验证，确保了数据集的整体一致性。同时，后续采用的 GAN 增强方法进一步丰富了数据分布，提高了模型对原材料变化的适应能力。

图 1 为关键组分分布的小提琴图，小提琴的宽度越宽表示分布越集中，越窄则表示分布越稀疏。由图 1 可见，水泥(C)和天然骨料(NA)为主要成分，其占比范围为 30%～60%；水(W)和粗砂(CS)的占比分别为 7%～10%和 20%～30%；再生骨料(RA)的占比为 20%～40%，与 NA 用量接近，显示其在部分配方中具有替代潜力。粉煤灰(FA)、硅灰(SF)和减水剂(WR)的占比较低，一般在 1%～5%之间。纤维掺量比(FCR)均低于 6%，纤维长度(FL)与纤维直径(FD)的变化幅度较小，表明研究对象在这 2 方面的应用较为一致。该数据集较为全面地覆盖了 SFRC 的主要配合比范围，具有较强的代表性与可靠性。



C—Cement; W—Water; CS—Coarse Sand; RA—Recycled Aggregate; NA—Natural Aggregate; FCR—Fiber Content Ratio; FA—Fly Ash; SF—Silica Fume; WR—Water Reducer; FL—Fiber Length; and FD—Fiber Diameter. All abbreviations are used consistently throughout this paper.

图 1　小提琴图显示了 SFRC 混合料中各组分的分布

Fig. 1　Violin plots showing the distribution of the components in SFRC mixtures

为了全面分析各参数的分布特征，对数据集进行了统计分析和正态性检验，其结果如表 1 所示。由表 1 可见，水泥 W 值为 0.769，其余参数亦均低于阈值，说明所有特征均不服从正态分布，这与图 1 的分布情况一致。鉴于此，基于正态分布假设的算法(如线性回归、高斯过程回归)并不适用。相反，决策树类方法(如随机森林、梯度增强树)对数据分布不敏感，更适合处理本研究所涉及的非正态分布特征，并能够保持较高的预测精度和模型鲁棒性。

表 1　SFRC 成分统计表和 Shapiro-Wilk 结果

Table 1　Summary statistics and shapiro-wilk normality test results for SFRC components

Parameter	Mean value	mid-value	Value range	Standard deviation	W value	α
Cement	467.070	440	280–810	132.382	0.769	0.05
Fly ash	39.377	0	0–140	48.173	0.762	0.05
Silica fume	16.702	0	0–90	29.866	0.598	0.05
Water	183.55	166	105.6–285.0	44.771	0.872	0.05
Coarse sand	644.7	705	317.8–948.0	145.917	0.879	0.05
Water reducer	3.543	2.96	0–17	4.110	0.759	0.05
Recycled aggregate	271.556	0	0–1073	338.117	0.758	0.05
Natural aggregate	593.206	536	0–1770	448.552	0.897	0.05
Fibre length	57.599	32	0–350	89.729	0.367	0.05
Fiber diameter	1.198	0.8	0–5.8	1.434	0.489	0.05
Fiber content ratio	65.629	70.2	0–156	47.937	0.923	0.05

Notes: α is significance level (typically set at 0.05 for a 95% confidence level).

1.2　基于关联的特征选择

为避免模型过拟合，使用 Pearson 相关系数矩阵分析输入特征的两两相关性。相关系数r计算公式如下：

$$r=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{1}$$

式中：X_i 和 Y_i 表示数据库中的 2 个特征；n 为数据

总量；r取值范围为[–1,1]，当r接近 1 或–1 时，说明变量之间存在较强的线性关系，此时，模型可能会过度依赖相关特征，从而影响泛化能力。

图 2 为数据集中各材料成分的关系热图，矩阵中的颜色和饼状图共同表示了各变量对之间的相关程度。由图 2 可见，所有变量的相关系数 $|r|<0.7$，不存在强相关性，多重共线性风险较低，无需对特征进行删除或合并处理。进一步显示，部分变量(如水泥与硅灰)呈中等程度的正相关($0.3\leq|r|<0.7$)，说明它们之间存在一定关联，但不足以影响模型稳定性，从而能确保各特征在预测中保持独立的贡献。

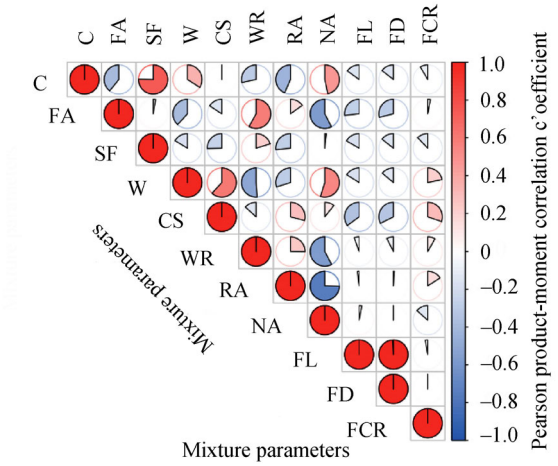


图 2 特征关系热图

Fig. 2 Characteristic relational heat map

2 模型性能评价

2.1 ML 模型的建立

图 3 为用于预测 SFRC 抗压强度的机器学习(ML)模型开发流程，包括数据收集、清洗与预处理等。结合 1.1 节的分析，选取了水泥、粉煤灰、硅灰、减水剂、再生骨料、天然骨料、纤维长度、纤维直径及纤维掺量比作为关键输入特征，以保证模型预测准确性。

在建模阶段，采用 XGBoost、Stacking、Random forest 和 AdaBoost 4 种 ML 算法构建预测模型，因其在处理高维、非线性数据方面具有优势，且能通过不同机制提升模型泛化能力，XGBoost 与 Random Forest 均为树模型，前者基于梯度提升，后者通过随机子空间法提升泛化性能；AdaBoost 用于弱学习器叠加增强预测；Stacking 综合多模型以进一步提升稳健性。本工作将数据集随机划分为 80%的训练集和 20%的测试集。训练集用于所有模型的拟合与参数学习；而测试集在整个训练过程中完全保持独立、不被使用，仅用于在模型训练完成后，最终一次性地评估其泛化能力和预测精度。最后，通过对预测结果进行分析，利用解释性方法揭示了关键因素对 SFRC 抗压强度的贡献，从而为混凝土抗压性能优化提供了一个数据驱动的研究框架。

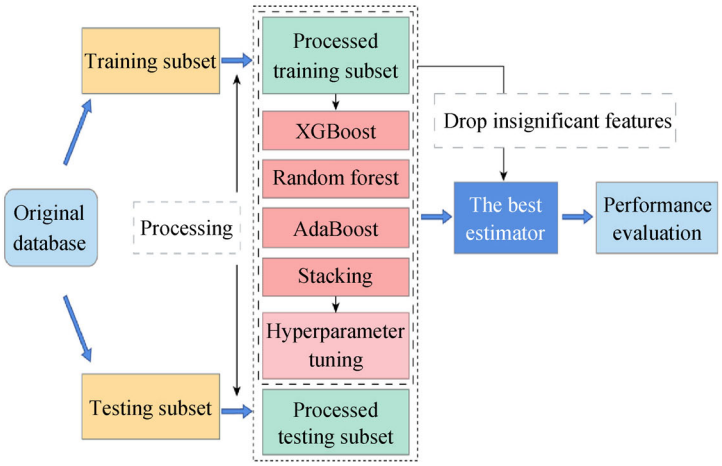


图 3 数据预处理和 ML 模型开发的工作流程

Fig. 3 Workflow of data preprocessing and ML model development

2.2 预测性能

图 4 为 4 种 ML 模型对 SFRC 抗压强度的预测性能，散点图显示预测值与实测值的关系，其中，45°参考线代表理想预测结果，数据点越接近该参考线，说明模型预测精度越高，为进一步量化模型性能，采用决定系数(R^2)和均方误差(MSE)作为评价指标。较高的 R^2 和较低的 MSE 值表明模型预测性能更优，如式(2)和式(3)所示：

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}\left|y_i - \bar{y}_i\right|} \tag{2}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \bar{y}_i\right| \tag{3}$$

由图 4 可见，4 种模型均能合理预测 SFRC 抗压强度，大多数预测点分布在参考线附近，误差在±15%以内。Random forest 和 XGBoost 的点最集中，表现出较强一致性。Random forest 通过套袋方法减轻过拟合，在高维复杂特征及噪声数据下表现出较好的鲁棒性；XGBoost 利用梯度增强机制捕捉非线性交互特征。Stacking 整合多个模型，提高了泛化能力，但受限于训练复杂性和有限数据量，性能略低于前两者；AdaBoost 对噪声敏感，导致 R^2最低、MSE 最高，预测精度较差。

柱状图显示，Random forest R^2 最高，为 0.93，MSE 最低，为 15.6，XGBoost 紧随其后(R^2=0.92，MSE=16.6)。Stacking 和 AdaBoost 的 R^2 分别为 0.91 和 0.89，MSE 分别为 17.8 和 20.8。这些结果表明，Random forest 和 XGBoost 的预测精度更高，平均误差更小。Stacking 模型整合了多个基模型的预测结果，因此，R^2 达到 0.91，显示出较强的泛化能力。相比之下，AdaBoost 的 MSE 最高，R^2 最低，其预测精度较低。

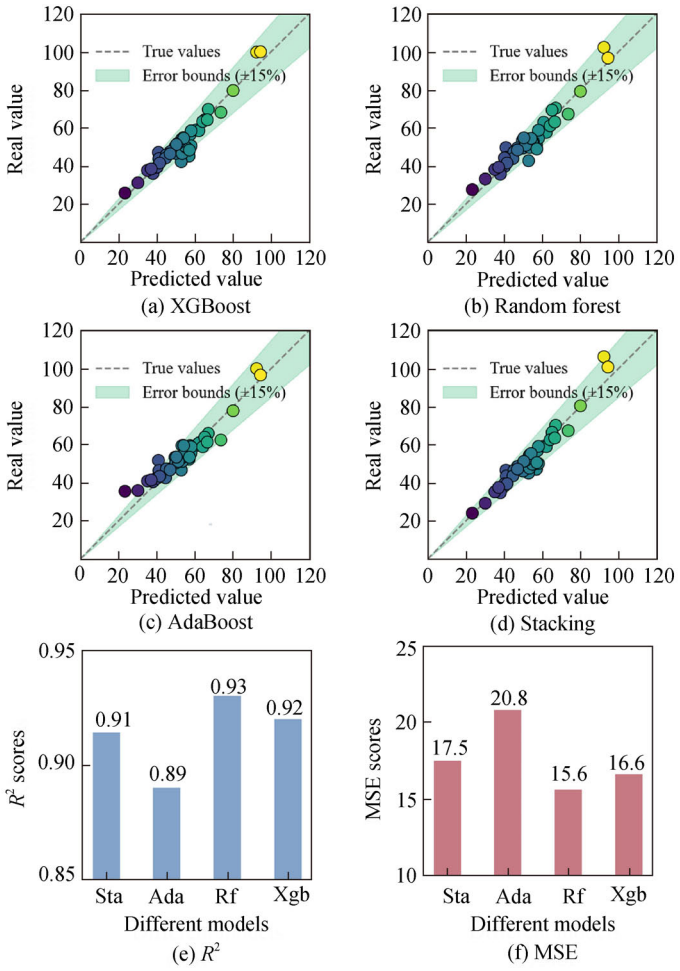


图 4　在预测抗压强度和试验强度之间进行测试

Fig. 4　Test between predicted compressive strength and test strength

2.3　数据增强预测

2.3.1　GAN 增强数据　　考虑到数据集包含 11 个特征和 141 个样本，这些样本无法充分捕捉特征间的复杂交互，容易导致模型过拟合或泛化能力受限。为了提升模型性能，本工作引入生成对抗网络(GAN)进行数据增强[22]。通过生成与原始数据统计分布相似的合成样本，从而可以评估模型的稳健性和预测精度，并揭示混凝土配合比与抗压强度之间的关系。

图 5 为使用的 GAN，由生成器 G 和鉴别器 D 组成。由图 5 可见，生成器接收随机噪声，并学习

将其映射到高维空间，以复制目标分布的统计特性，通过对抗性训练，生成器不断提高生成样本的逼真度，而鉴别器则提升分类准确率，实现数据增强和特征分布保真。

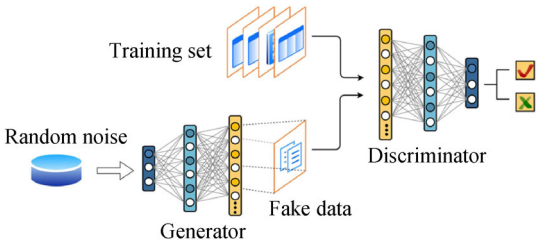


图 5 数据增强流程图

Fig. 5 Data enhancement flow chart

基于 141 个原始样本，通过 GAN 生成了 282 个新样本，生成器采用 3 层全连接网络，输入为随机噪声向量，输出为与原始数据同维度的合成样本；判别器采用类似结构，通过交替训练使生成样本分布逼近真实数据分布，其特征与原始数据一致。生成器和鉴别器的结构设计参考了 Random forest，通过对抗性训练生成逼真数据。随后，本工作使用这些增强样本训练随机森林模型，用于预测 SFRC 抗压强度。

为评估生成数据的质量与多样性，使用 *t*-分布随机邻居嵌入(*t*-SNE)对原始数据和生成数据在特征空间的分布进行可视化。最终，*t*-SNE 通过最小化 KL 散度优化低维点的位置，尽可能保留高维数据的局部结构，从而直观呈现生成样本与原始数据之间的分布关系。式(4)~式(6)给出了其数学表述。

$$q_{j|i}=\frac{\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k\neq i}\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma_i^2}\right)} \tag{4}$$

$$q_{j|i}=\frac{(1+\|y_i-y_j\|^2)^{-1}}{\sum_{k\neq i}(1+\|y_i-y_j\|^2)^{-1}} \tag{5}$$

$$C=\sum_i\sum_{j\neq i}p_{j|i}\log\frac{p_{j|i}}{q_{j|i}} \tag{6}$$

图 6 为原始和 GAN 数据的 *t*-SNE 可视化图。由图 6 可见，GAN 生成的样本有效捕获了原始数据的分布特征，红蓝点高度重叠，表明 GAN 成功保留了数据的局部结构和簇特征。密度轮廓进一步显示，生成数据不仅与原始数据的整体模式一致，而且保持了簇结构特征，这一结果验证了 GAN 能够再现原始数据的固有特征，支持其在数据增强和合成样本生成中的潜在应用价值。

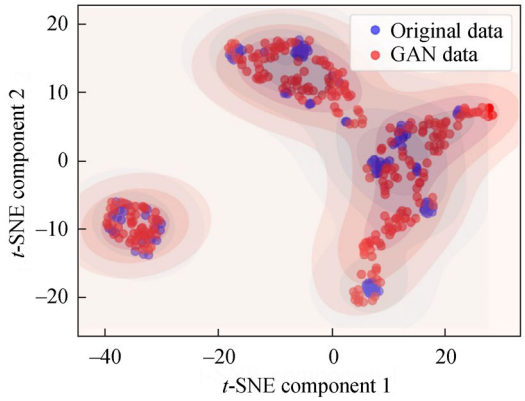


图 6 原始和 GAN 数据的 *t*-SNE 可视化图

Fig. 6 *t*-SNE Visualization of Original and GAN Data

经过 GAN 增强后，统计量分析(图 7)表明，GAN 生成数据在整体分布和特征尺度上与原始数据具有高度一致性。为进一步严格验证生成样本的物理合理性，分析发现，在低掺量区间，强度提升有限；随着掺量增至中等水平(约 2%~4%)，纤维的桥接效应充分发挥，强度显著增长；而当掺量过高时，因纤维团聚和工作性下降，强度增长趋于平缓，这一趋势与文献[3–4]及经典理论完全吻合。GAN 不仅学习了数据的统计特征，更捕捉并内化了支配 SFRC 强度的物理机制，从而保证了生成数据用于模型增强的可靠性。

2.3.2 GAN 增强数据建模性能 图 8 为预测抗压强度与试验强度的对比。由图 8a~图 8d 可见，使用 GAN 生成数据进行抗压强度预测时，预测值与实际值的关系中，大多数数据点位于绿色阴影区域内，说明模型预测误差处于可接受范围内。随机森林和 Stacking 模型的预测精度相近，误差极少超过 15%，而 AdaBoost 的预测效果仍然较差。这主要是因为 AdaBoost 对错误样本过于敏感，并且依赖单个弱分类器的加权组合，难以像随机森林或 Stacking 那样有效处理噪声和异常值。由图 8e~图 8f 可见，使用 GAN 增强的数据显著提升了各模型的拟合性能。所有模型的 R^2值均明显提高，其中 Stacking 模型 R^2 达到 0.99；同时，所有模型的均方误差(MSE)均下降超过 50%，Stacking 模型的 MSE 由 17.5 降至 4.2，表现出卓越性能。Random Forest 和 XGBoost 的 MSE 也下降约 50%。GAN 生成的数据优化了样本分布，为模型提供了更全面、均衡的数据集，从而增强了模型对特征与目标值映射关系的捕捉能力，有效提

升了预测精度和泛化能力。考虑到机器学习模型对训练数据分布高度敏感，稀疏或不均匀的数据常导致过拟合或欠拟合，因此 GAN 数据的引入对提升模型稳健性和性能具有重要作用。

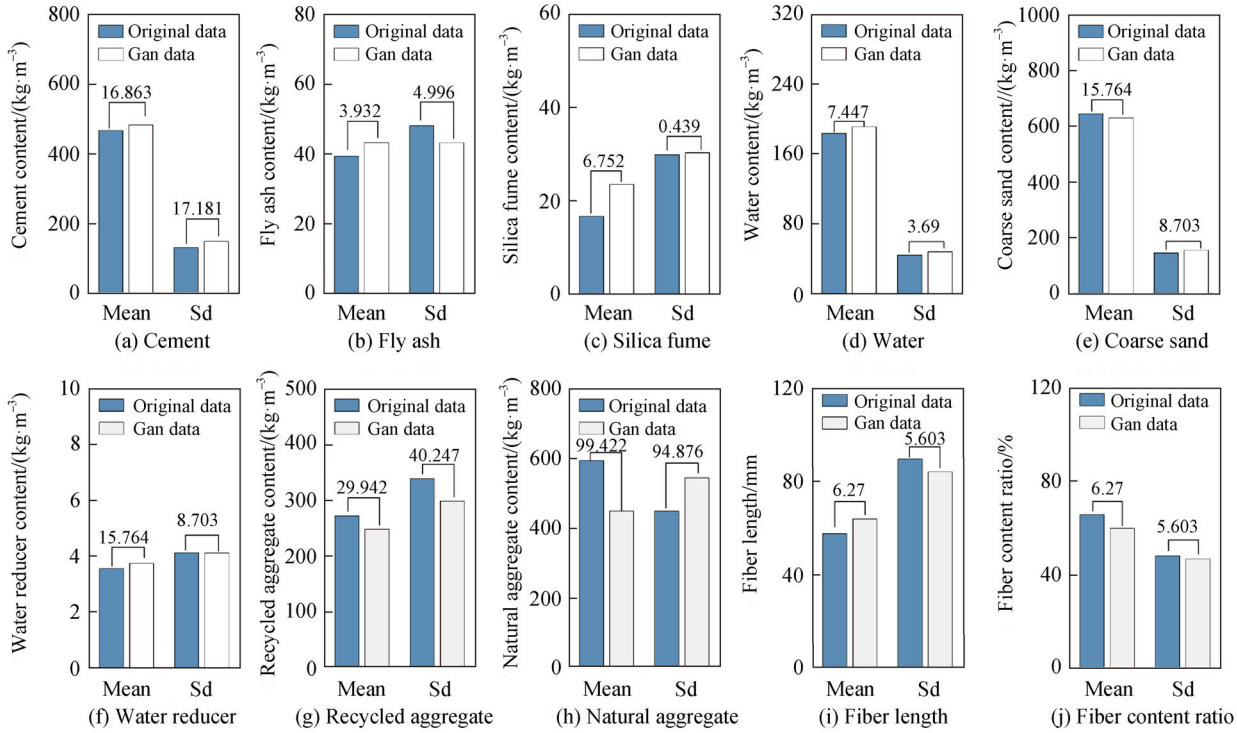


图 7　原始数据和 GAN 数据的均值和标准差

Fig. 7　Mean and standard deviation of the original data and GAN data

GAN 生成的数据优化了样本分布，为模型提供了更全面、均衡的数据集，从而增强了模型对特征与目标值映射关系的捕捉能力，有效提升了预测精度和泛化能力。尽管如此，基础数据集规模仍属有限，且 GAN 增强并不能完全替代真实试验数据。因此，构建的模型其最终实用性与可靠性，仍有待于在未来通过补充系统的试验数据进行进一步的验证与完善。

3　基于 SHAP 的随机森林模型可解释性分析

3.1　基于 SHAP 蜂窝图的特征分析

为深入探讨各配合比特征对混凝土抗压强度的影响及其在模型预测中的作用，本工作基于随机森林模型进行了 SHAP 解释性分析。式(7)和式(8)为 SHAP 值的数学定义与计算原理。

$$\phi_i = \sum_{S\subseteq N\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}\left[f(S\cup\{i\}-f(S)\right] \quad (7)$$

$$x_i \to \phi_i \text{ with color encoding for } x_i \quad (8)$$

SHAP 分析能够全面揭示预测模型中各特征的相对重要性及其对输出结果的影响，如图 9 所示。由图 9 可见，水泥和水是影响最大的变量，这与水泥和水分在材料性能中的基础性作用高度一致，相比之下，硅灰和粗砂的贡献虽不如水泥和水显著，但仍表现出可观的 SHAP 值，进一步验证了补充胶凝材料(如硅灰)在混凝土抗压性能优化中的重要性，尤其是在与水泥共同作用时发挥次要但不可忽视的作用。与纤维几何特性相关的变量呈现出较为集中的 SHAP 值分布。这表明它们对预测结果的贡献相对细微，纤维的微观结构特性虽不是主要驱动因素，却能够在特定条件下对整体性能产生影响，这与纤维在混凝土中的主要作用机制相符，纤维主要通过桥接裂缝改善材料的韧性和抗裂性能，而对基体抗压强度的提升作用较为有限。

在 SHAP 汇总图中，颜色梯度直观反映了特征值大小与预测结果之间的复杂关系。图 10 为双组分对混凝土抗压强度的影响。由图 10a~图 10c 可见，天然粗骨料的 SHAP 值大多分布在 0 附近，整体较为分散，说明其对抗压强度的贡献较小。同时，粗砂、纤维直径和纤维含量比也表现出类似特征，表明它们的影响相对较弱，且缺乏明显规律。相比之下，减水剂含量展现出更为清晰的模式(图 10d)。

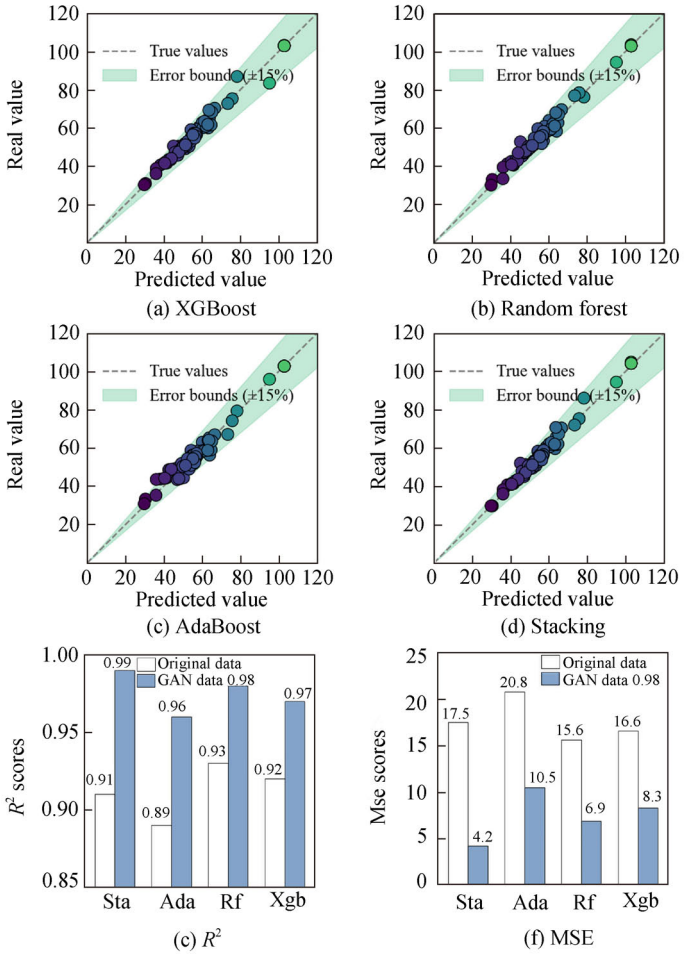


图 8 预测抗压强度与试验强度的对比

Fig. 8 Comparison between predicted and experimental compressive strength using GAN-augmented data

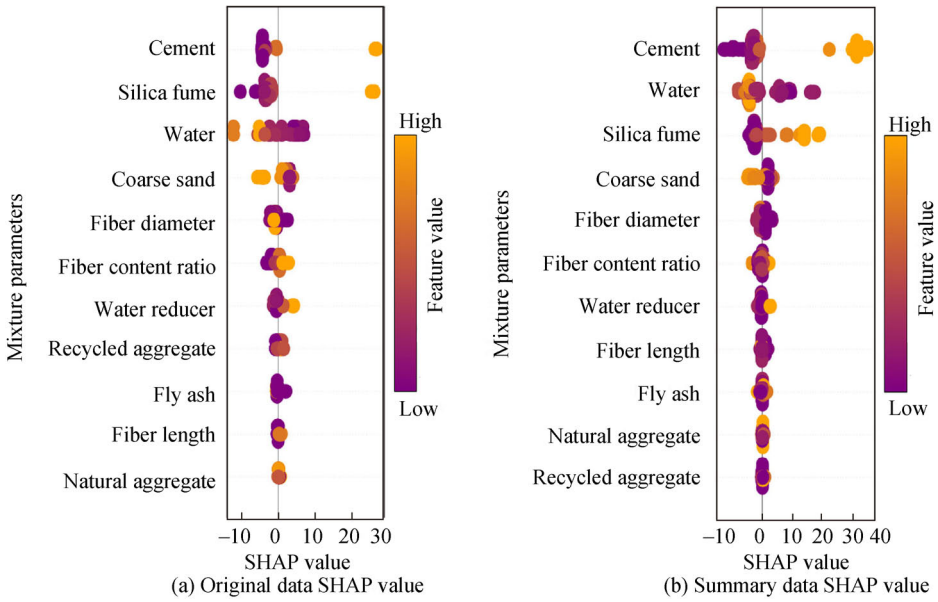


图 9 SHAP 值依赖关系图

Fig. 9 SHAP value dependency diagram

当减水剂含量处于较低水平(0~5)时，其 SHAP 值主要分布在负区域，表明减水剂不足会导致抗压强度下降；当减水剂含量超过 5 时，其 SHAP 值迅速转向正区域并显著上升，显示减水剂对抗压强度提升具

有积极贡献。此外，水与减水剂之间的交互效应在图中也有所体现：当含水量较高(橙色)时，部分 SHAP 值落在较低区域，说明过量用水会削弱减水剂的作用，使抗压强度略有下降；当含水量较低(紫色)时，减水剂的正向贡献更加稳定且显著，强化了水灰比与外加剂用量之间的耦合效应。

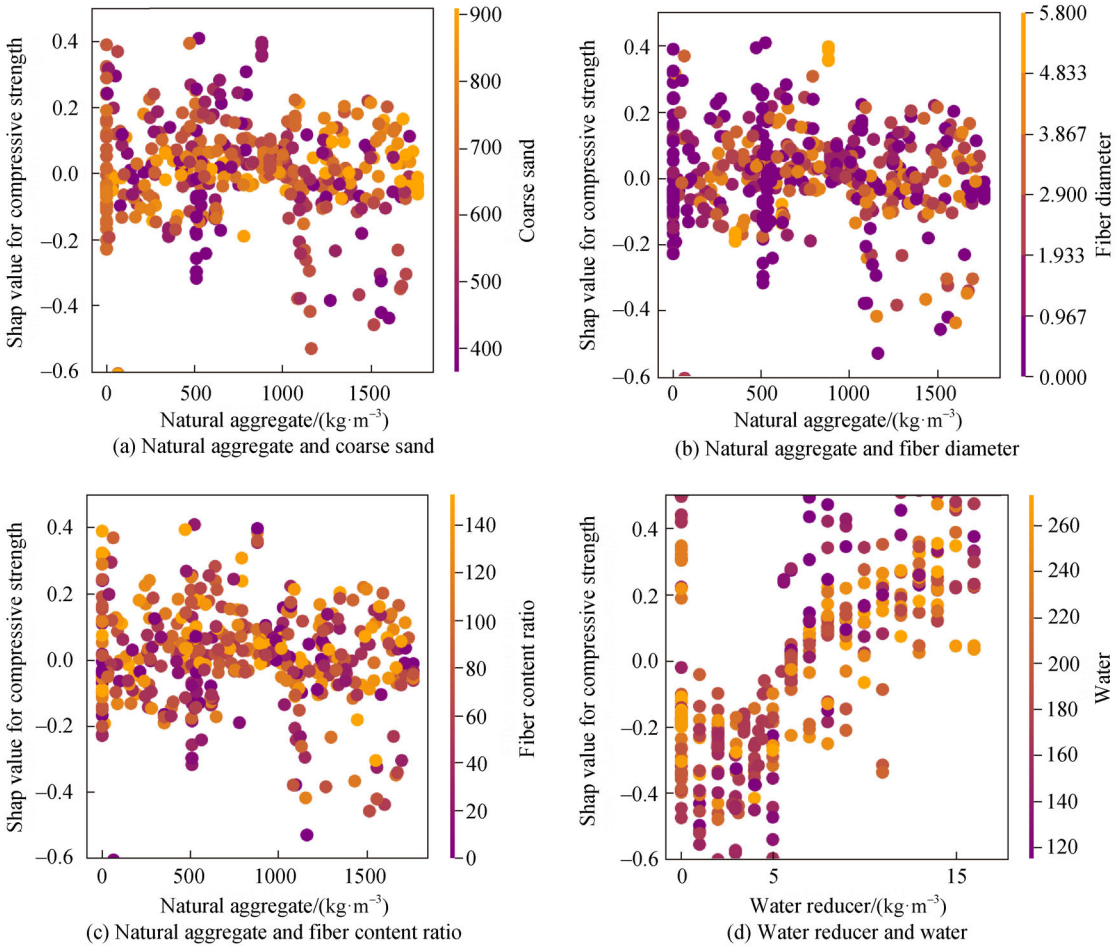


图 10　双组分对混凝土抗压强度的影响

Fig. 10　Influence of two-component combinations on compressive strength

为了进一步解释特定预测结果中单个特征的贡献，本工作使用 SHAP 力图对原始数据集和 GAN 增强数据集的最小与最大抗压强度进行分析，如图 11 所示。

在原始数据集中，最小抗压强度(27.36 MPa)主要受到水泥(C=340.0)、水(W=200.0)和粗砂(CS=780.0)低值的共同影响，这反映了配合比不合理导致的抗压强度不足。相反，最大抗压强度(133.5 MPa)则由水泥(C=810.0)和硅粉(SF=90.0)高值主导，突出显示了它们在提升抗压强度方面的关键作用。

在 GAN 增强数据集中，整体特征贡献趋势与原始数据保持一致，水泥和硅粉依然是最大强度预测的核心驱动因素。然而，GAN 增强数据使模型能够识别更多微妙的交互效应。例如，对于最小抗压强度(27.84 MPa)，纤维含量比(FCR=7.0)被引入为新的贡献变量，这说明 GAN 增强扩展了模型对特征空间的理解，捕捉到了原始数据中未显现的潜在规律。此外，在最大抗压强度(130.3 MPa)的预测中，SHAP 值的分布更为平滑且均衡，表明数据空间的扩展提升了模型的鲁棒性和稳定性。

图 12 为基于 SHAP 的 Sankey 图。由图 12 可见，水泥和硅灰是驱动高抗压强度(σ_c)的主要因素，它们通过促进致密、连续的水化产物形成，显著增强了基体的整体结构完整性。水的影响呈现明显的非线性效应：适量水有助于充分水化反应，从而提高混凝土强度；而过量水会增加孔隙率，导致基体密实性下降，从而削弱σ_c。粗砂的高掺量与较低的σ_c相关，这是因为粗砂削弱了胶凝材料的连续性，

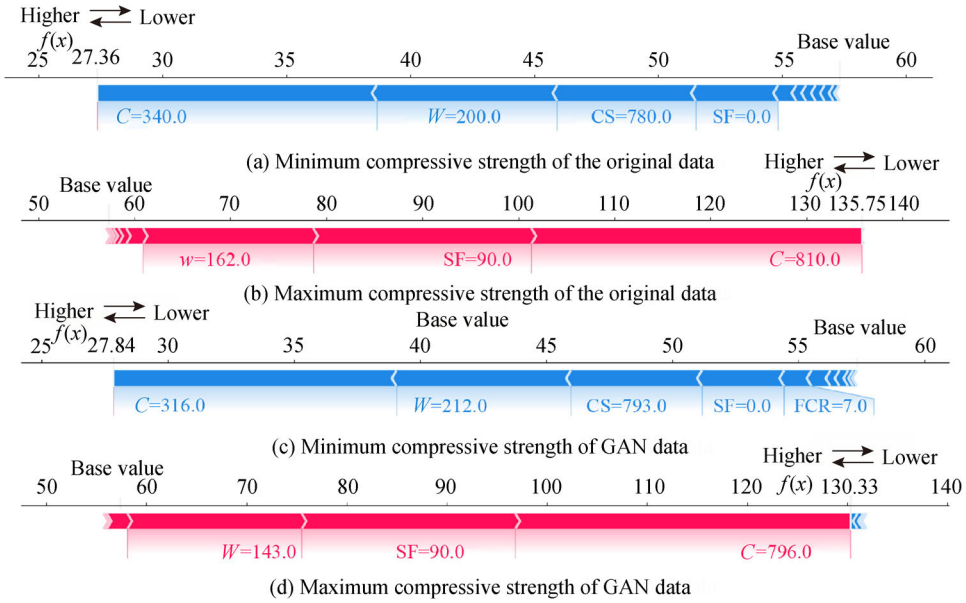


图 11 SHAP 力图比较原始数据和 gan 增强数据中最小和最大抗压强度的特征贡献

Fig. 11 SHAP force plots comparing feature contributions between original and GAN-augmented data

并引入了额外孔隙，从而降低了力学性能。相比之下，其他特征对混凝土强度分级的贡献相对有限，仅在局部范围内表现出次要作用。

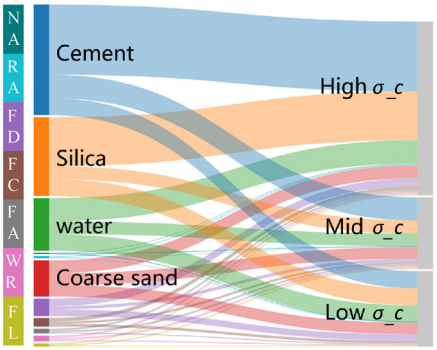


图 12 基于 SHAP 的 Sankey 图

Fig. 12 Sankey diagram of SHAP-based feature contribution paths for SFRC compressive strength

Sankey 图揭示了各混合成分在不同强度区间中的作用机制，为研究者理解混凝土抗压性能形成机理并优化配合比设计提供了直观证据。

4 结论

1) 本工作基于 141 个样本和 11 个特征参数，构建了 XGBoost、Stacking、AdaBoost 和 Random Forest 4 种机器学习模型，用于预测 SFRC 抗压强度。结果显示，随机森林模型的预测精度最高，R^2达到 0.93，能够较好地捕捉特征间的非线性关系。

2) 研究采用 GAN 方法将原始数据集扩展至 242 个样本，有效提升了模型性能。扩展后，随机森林模型的 R^2 由 0.93 提高至 0.98，MSE 由 15.6 下降至 6.9，降幅约 56%，表明 GAN 数据增强能够显著改善模型的泛化能力。

3) SHAP 分析结果表明，水泥、硅灰和水是影响 SFRC 抗压强度的主要因素。其中，水泥和硅灰通过形成致密基体提高混凝土强度，而水表现出非线性效应；适量减水剂能够增强抗压强度，但高含水量会削弱减水剂的作用。粗砂对抗压强度具有负面影响，而纤维相关参数(长度、直径和含量)的贡献相对较小。

4) SHAP 分析表明，水泥、硅灰与水是影响 SFRC 抗压强度的 3 个关键因素。该结论与材料学原理一致：水泥与硅灰形成强度骨架，而水则通过控制水灰比直接决定基体密实度。此发现与文献[7]中“水泥、硅灰、减水剂为主因”的结论相互支持，并从机理上明确了“水”的核心作用，因为减水剂的功能本质上是为优化水灰比。纤维参数对抗压强度贡献相对有限，与其主要增强韧性的机制相符。

5) 本工作的预测模型基于现有的文献配合比数据构建，尚未考虑纤维取向与分布等细观结构因素的影响。后续研究可通过引入图像识别与数值仿真数据来丰富特征维度，以期在更广泛的力学性能预测中取得突破。

参考文献:

[1] ZHANG Y F, CHEN J J, XIA J. Compressive strength and chloride

resistance of slag/metakaolin-based ultra-high-performance geopolymer concrete[J]. Materials, 2023, 16(1): 181.

[2] IOANA A, PAUNESCU L, CONSTANTIN N, et al. High-strength and heat-insulating cellular building concrete based on calcined gypsum[J]. Materials, 2023, 16(1): 118.

[3] WANG C F, GUO Z X, NIU D T. Influence of the fiber volume content on the durability-related properties of polypropylene-fiber-reinforced concrete[J]. Sustainability, 2020, 12(2): 549.

[4] LAN M Z, ZHOU J, XU M F. Effect of fibre types on the tensile behaviour of engineered cementitious composites[J]. Front Mater, 2021, 8: 775188.

[5] DASH P K, PARHI S K, PATRO S K, et al. Influence of chemical constituents of binder and activator in predicting compressive strength of fly ash-based geopolymer concrete using firefly-optimized hybrid ensemble machine learning model[J]. Mater Today Commun, 2023, 37: 107485.

[6] DAI L, WU X, ZHOU M R, et al. Using machine learning algorithms to estimate the compressive property of high strength fiber reinforced concrete[J]. Materials, 2022, 15(13): 4450.

[7] LI Y J, ZHANG Q Z, KAMIŃSKI P, et al. Compressive strength of steel fiber-reinforced concrete employing supervised machine learning techniques[J]. Materials, 2022, 15(12): 4209.

[8] ZHOU Y, DONG F, LIU Y F, et al. Forecasting emerging technologies using data augmentation and deep learning[J]. Scientometrics, 2020, 123(1): 1–29.

[9] JIANG Y C, DRESCHER B, YUAN G G. A GAN-based multi-sensor data augmentation technique for CNC machine tool wear prediction[J]. IEEE Access, 2023, 11: 95782–95795.

[10] HE W C, KONG X Q, FU Y, et al. Experimental investigation on the mechanical properties and microstructure of hybrid fiber reinforced recycled aggregate concrete[J]. Constr Build Mater, 2020, 261: 120488.

[11] 蒋毓晋. 钢纤维再生混凝土力学性能及声发射特征试验研究[D]. 邯郸: 河北工程大学, 2022.
JANG Yujin. Experimental Study on Mechanical Properties and Acoustic Emission Characteristics of Steel Fiber Recycled Concrete [D]. Handan: Hebei University of Engineering, 2022.

[12] 贾星亮. 钢纤维-玄武岩纤维再生混凝土力学性能与冻融腐蚀耦合研究[D]. 西安: 西京学院, 2021.
JIA Xinglian. STUDY ON MECHANICAL PROPERTIES AND FREEZ-THAW CORROSION OF STEEL-BASALT HYBRIDFIBER RECYCLED AGGREGATE CONCRETE [D]. Xian: Xijing University, 2021.

[13] 张志银. 钢纤维高强轻骨料混凝土力学性能及韧性研究[D]. 西安: 长安大学, 2019.
ZHANG Zhiyi. Study on Mechanical Properties and Toughness of Steel Fiber High Strength Lightweight Aggregate Concrete [D]. Xian: Chang'an University, 2019.

[14] 商冉. 钢纤维陶粒混凝土力学性能与耐久性研究[D]. 北京: 中国地质大学(北京), 2018.
SHANG Ran. Research on Mechanics Performance and Durability of Steel Fiber Reinforced Ceramisite Concrete [D]. Beijing: China University of Geosciences(Beijing), 2018.

[15] WANG J. Study on the influence of different steel fiber content on the mechanical properties of concrete[J]. Jiangxi Build Mater, 2021, 40: 35–36.

[16] 于剑桥. 西部环境下混杂纤维混凝土力学性能与耐久性能研究[D]. 兰州: 兰州理工大学, 2022.
YU Jianqiao. Research on the Mechanical Properties and Durability of Hybrid Fiber Reinforced Concrete under Western Environment [D]. Lanzhou: Lanzhou University of Technology, 2022.

[17] ZHANG, Y. , B. CHEN, AND Y. WANG. "Analysis on the mechanical properties of steel fiber reinforced concrete." Concrete 4 (2020): 74–77.

[18] 范存园. 混杂钢纤维混凝土力学性能研究[D]. 郑州: 郑州大学, 2017.
FAN Cunyuan. Study on Mechanical Properties of Hybrid Steel Fiber Reinforced Concrete [D]. Zhengzhou: Zhengzhou University, 2017.

[19] 陈图真. 钢纤维和聚丙烯粗纤维改性再生混凝土力学性能试验研究[D]. 广州: 广东工业大学, 2015.
CHEN Tuzhen. Experimental study on mechanical properties of steel fiber and Macro polypropylene fiber reinforced recycled concrete[D]. Guangzhou: Guangdong University of Technology, 2015.

[20] 徐朦. 多种钢纤维对超高性能混凝土力学性能影响的比较研究[D]. 北京: 北京交通大学, 2014.
XU Meng. Comparative study on the influence of various steel fibers on mechanical properties of ultra-high performance concrete(UHPC)[D]. Beijing: Beijing Jiaotong University, 2014.

[21] 芦静云. 钢纤维再生混凝土配合比及基本力学性能[D]. 郑州: 郑州大学, 2014.
LU Jingyun. Experimental Research on Fracture Properties and Mixed Proportion of Steel Fiber Reinforced Recycled Concrete [D]. Zhengzhou: Zhengzhou University, 2014.

[22] MOU X Q, CHEN X L, GUAN J, et al. Sea clutter suppression for radar PPI images based on SCS-GAN[J]. IEEE Geosci Remote Sens Lett, 2021, 18(11): 1886–1890.

Strength Prediction and SHapley Additive Explanations Interpretability Analysis of Steel Fiber Reinforced Concrete Based on Generative Adversarial Networks

YANG Zhe[1], *ZHANG Yue*[1], *XU Yan*[2], *HOU Dongshuai*[1], *WANG Pan*[1], *WANG Muhan*[1], *WANG Xinpeng*[1]
(1. School of Civil Engineering, Qingdao University of Technology, Qingdao 266033, Shandong, China;
2. China Water Resources and Hydropower Fourth Engineering Bureau Co., Ltd., Xining 810000, China)

Extended Abstract

Introduction Predicting the compressive strength of Steel Fiber Reinforced Concrete (SFRC) is challenging due to the complex, coupled influences of numerous mix design parameters. Traditional experimental methods are often costly and time-consuming for exploring this high-dimensional factor space. While machine learning (ML) models offer a powerful alternative, their practical

application is hindered by two major limitations: the scarcity of comprehensive, high-quality experimental data which restricts model performance and generalizability, and the inherent "black-box" nature of advanced ML algorithms which obscures the underlying mechanisms linking mix proportions to mechanical properties. This study aims to address these critical gaps by developing a hybrid framework that integrates ML-based prediction, data augmentation via Generative Adversarial Networks (GAN), and post-hoc model interpretability using SHapley Additive exPlanations (SHAP). The objective is not only to achieve high-precision strength prediction but also to enhance model robustness with limited data and, crucially, to uncover the influential mechanisms of key components on SFRC compressive strength.

Methods A database of 141 mix designs with 11 input features—Cement (C), Fly Ash (FA), Silica Fume (SF), Water (W), Coarse Sand (CS), Water Reducer (WR), Recycled Aggregate (RA), Natural Aggregate (NA), Fiber Length (FL), Fiber Diameter (FD), and Fiber Content Ratio (FCR)—and the corresponding compressive strength was compiled from peer-reviewed literature. Four tree-based ML algorithms (Random Forest, XGBoost, Stacking, and AdaBoost) were employed for regression modeling. To mitigate the data scarcity issue, a GAN architecture was implemented to generate 282 synthetic but physically plausible samples, expanding the training set. The quality of the GAN-generated data was validated using *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) visualization and statistical moment analysis. The model trained on the augmented dataset was evaluated using the coefficient of determination (R^2) and Mean Squared Error (MSE). Finally, the optimal model (Random Forest) was interpreted using the SHAP framework to quantify and visualize the contribution and interaction effects of each input feature on the model's predictions.

Results and discussion Using the original dataset, all ML models demonstrated competent predictive capability. The Random Forest model achieved the best performance with an R^2 of 0.93 and an MSE of 15.6 MPa^2, followed closely by XGBoost (R^2=0.92, MSE=16.6 MPa^2). The introduction of GAN-generated data led to a remarkable and consistent improvement in all models' performance. For Random Forest, the R^2 increased to 0.98 and the MSE decreased by approximately 56% to 6.9 MPa^2. The Stacking model achieved an R^2 of 0.99. This significant enhancement confirms that GAN-based data augmentation effectively enriches the feature space and improves the models' ability to generalize from limited experimental data.

The SHAP analysis provided profound insights into the governing factors of compressive strength. The summary plot identified cement, water, and silica fume as the three most influential features. Cement and silica fume exhibited strong positive SHAP values, highlighting their fundamental role in forming a dense cementitious matrix. The impact of water displayed a distinct non-linear relationship: optimal content promoted hydration and strength, while excess water reduced strength, likely by increasing porosity. The interaction SHAP dependence plots revealed nuanced coupling effects. For instance, the positive contribution of a high WR dosage was more pronounced and stable under conditions of lower water content, emphasizing the critical coupling between water and chemical admixtures in determining the water-to-cementitious materials ratio. In contrast, parameters related to fiber geometry (FL, FD) and content (FCR) showed relatively low and concentrated SHAP values near zero, indicating their secondary role in influencing compressive strength, which aligns with their primary function of improving toughness rather than pure compressive capacity. The SHAP force plots and Sankey diagram further illustrated how specific combinations of high cement/silica fume and optimal water content drive predictions toward high strength, whereas high coarse sand content or inappropriate water levels contribute to lower strength outcomes.

Conclusions This work successfully developed an integrated framework combining ML, GAN, and SHAP for predicting and interpreting the compressive strength of SFRC. The main conclusions are as follows:1) Among the evaluated ML models trained on the original dataset, Random Forest provided the most accurate predictions for SFRC compressive strength.2) GAN-based data augmentation proved highly effective, substantially improving the predictive accuracy and generalization ability of all models, with performance metrics (R^2 and MSE) showing dramatic improvement.3) SHAP interpretability analysis quantitatively identified cement, silica fume, and water as the most critical factors affecting compressive strength, with water exhibiting a non-linear effect. The analysis also elucidated important feature interactions, such as that between water and water reducer. Fiber parameters were found to have a minor influence on compressive strength prediction.4) The proposed framework offers a powerful, data-driven tool that moves beyond mere "black-box" prediction. It enhances prediction robustness from small datasets and provides actionable, mechanism-based insights for optimizing SFRC mix design, thereby bridging the gap between empirical data mining and fundamental materials science understanding.

Keywords steel fiber reinforced concrete; machine learning; random forest; generative adversarial network; shapley additive explanations