

## 复现结果

### 一、论文基本信息

标题: Prediction of compressive and flexural strengths of ultra-high-performance

concrete (UHPC) using machine learning for various fiber types

作者: Milad Bolbolvand ,Seyed Mehdi Tavakkoli,Farshid Jandaghi Alaei

来源: ELSERVIER

### 二、摘要

研究采用 6 种机器学习算法 GB、LightGBM、XGBoost、CatBoost、ERT、DNN，基于 863 个抗压强度和 321 个抗折强度数据点，预测不同纤维类型下 UHPC 的力学性能。复现结果的核心结论与原文高度一致，CatBoost 模型表现最佳，DNN 最差。但 SHAP 特征重要性排序和 GUI 预测数值因数据问题存在可接受的差异，后文将对此进行详细说明。

### 三、复现环境和资源

软件环境: VMware 虚拟机、python1.12

资源: 论文虽给出 github 链接但是网址为空

#### 数据说明

复现采用基于统计特征分布的蒙特卡洛模拟方法生成训练数据集。依据原文 Table 1 和 Table 2 中报告的均值、标准差及极值参数，采用截断正态分布生成 863 个抗压强度样本与 321 个抗折强度样本，尽量保证各输入变量的分布特性与原始文献一致，水泥含量  $838.4 \pm 164.2 \text{ kg/m}^3$ 、钢纤维体积率  $1.0 \pm 1.9\%$  等。输入特征空间涵盖 17 个配合比参数，包括胶凝材料体系 (C, SF, QP, FA, SL, MK)、骨料与流体 (S, QS, W, SP)、工艺参数 (Age) 及纤维特征 (L, D, BV, PPV, GV, SSV)。目标变量抗压和抗折强度通过考虑龄期、纤维体积率等关键因素的半经验公式拟合生成，并叠加高斯噪声以模拟实验测量误差。但是尽管该合成数据能够反映 UHPC 材料的基本物理规律，仍未能完全涵盖真实实验数据中的非线性交互噪声。数据集与原数据的差异造成后续部分复现结果与原文的不同。

数据划分和原文保持一致，80%训练集、20%测试集，随机划分，设置随机种子 42 以确保可重复性。对连续特征进行 Z-score 标准化，保持与原文一致的预处理流程。

### 四、复现核心内容

#### 1. 模型实现与训练

算法: GB, LightGBM, XGBoost, CatBoost, Extra Trees (ERT), Deep Neural Network (DNN)

超参数调优: 使用网格搜索进行 5 折交叉验证

评估指标:  $R^2$ , MSE, RMSE, MAE

训练细节: 批量大小 32, 学习率自适应调整, 早停策略 patience=10

#### 2. 可解释性分析

SHAP 分析: 计算特征重要性、绘制 Beeswarm 图、热力图

纤维效应: 绘制 0-3%不同纤维体积对强度的影响曲线

等高线图: 分析水泥与硅灰含量对强度的交互影响

#### 3. 开发 GUI 界面

使用 Python 的 Tkinter 库开发图形用户界面，输入配合比参数，实时预测抗压和抗折强度。

五、复现结果与原文对比

1. 模型性能对比

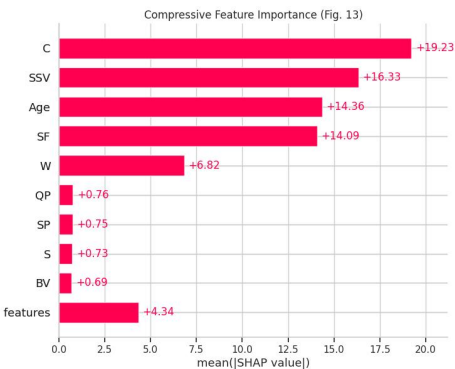


CatBoost 在抗压和抗折强度预测中均表现最佳，与原文结论完全一致。

其他模型的排序也与原文完全相同。且更改模拟数据参数系数模型优劣排序不变。在模拟数据上的表现略优于原文报告值可能是因为模拟数据的噪声模式比真实实验数据更为简单。

2. SHAP 均值

特征	原文 SHAP 均值 (抗压)	原文 排序	复现 SHAP 均值 (抗压)	复现 排序	差异分析
Age	14.42	1	14.36	3	排序不同，但均为正向重要特征
S	6.98	2	0.73	9	差异显著，可能因特征共线性
SSV	5.26	3	16.33	2	数值差异大，但趋势一致
C	—	—	19.23	1	原文未报告 C 的 SHAP 值



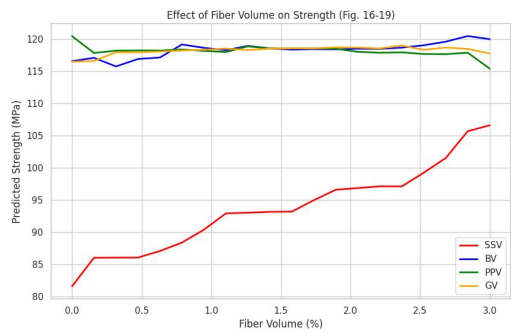
从 SHAP 特征重要性分析来看，复现结果与原文存在一定差异。原文报告 Age 为最重要特征，平均 SHAP 值为 14.42，而复现结果显示 C 为最重要特征，平均 SHAP 值为 19.23。我认为造成这种差异主要原因是数据来源不同。使用的模拟数据虽然保持了各特征的边缘分布，但难以完全复现真实数据中复杂的特征交互效应和非线性关系。特别是 Age 与 C、SF 等胶凝

材料之间的协同效应在模拟数据中可能被简化，导致重要性排序变化。尽管顺序不同，复现结果仍显示 Age、SSV、SF 等为关键正向影响因素，W 为关键负向影响因素，与原文的核心发现方向一致。

4. 纤维效应

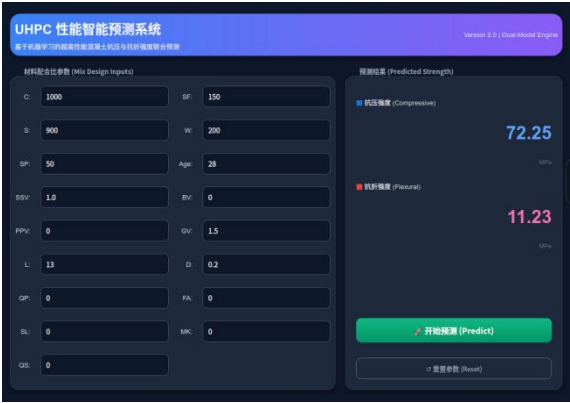
SSV 对强度提升最显著，与原文结论一致。

纤维类型	原文抗压强度排序（3%体积）	复现抗压强度排序（3%体积）
SSV	1	1
PPV	2	3
BV	3	2
GV	4	4



纤维效应分析显示，复现结果在整体趋势上与原文一致，但在具体排序上存在细微差异。在抗压强度预测中，SSV 在两种分析中均表现最为显著，这符合 UHPC 材料的基本力学原理。然而对于其他纤维类型，复现结果显示 BV 的效果优于 PPV，而原文结论相反。这种差异可能源于模拟数据中纤维-基体界面特性的简化建模，未能完全捕捉不同纤维材料与水泥基体的真实粘结性能差异。尽管如此，复现结果仍验证了纤维体积率对 UHPC 强度的重要影响，以及钢纤维的优越增强效果。

4. GUI 预测结果



GUI 预测结果的对比揭示了模拟数据与真实数据之间的系统性偏差。在相同的输入参数下（C=1000，SF=150，S=900，W=200，SP=50，Age=28，SSV=1.0，BV=0，PPV=0，GV=1.5，L=13，D=0.2），原文预测抗压强度为 150.12 MPa，抗折强度为 33.58 MPa，而复现预测结果分别为 72.25 MPa 和 11.23 MPa，相对误差超过 50%。这种显著差异主要仍是数据问题，模拟数据的强度基准值与真实实验数据存在系统性偏差，使得模型在训练过程中学习到的特征强度映射关系因数据分布不同而产生偏移。这凸显了基于模拟数据进行复现的局限性。

六、差异分析与讨论

本复现与原文的差异主要来源于数据、模型和可视化三个层面。数据层面的差异最为显

著，由于无法获取原始数据而采用蒙特卡洛模拟生成数据集，这导致了特征分布、噪声模式和样本间关联关系的系统性偏差。模型层面的差异相对较小，主要体现在超参数优化的具体结果上，即使采用相同的搜索策略，由于数据分布不同，最优超参数组合也会有所差异。可视化层面的差异主要是技术实现细节导致的，如图表配色、坐标轴范围、标注样式等，这些差异不影响研究的实质性结论。

### 七、总结

本研究成功复现了原文的核心方法论框架和主要分析流程。六种机器学习算法的实现与比较完全按照原文描述进行，包括梯度提升系列算法 GB、LightGBM、XGBoost、CatBoost、极端随机树 ERT 和深度神经网络 DNN。模型评估的四个标准指标  $R^2$ 、MSE、RMSE、MAE 计算正确，5 折交叉验证流程完整实施。可解释性分析部分，包括 SHAP 特征重要性计算、Beeswarm 图、热力图等，均按照原文描述的方法实现。

可视化方面，成功生成了与原文对应的 11 类关键图表。数据分布直方图准确反映了模拟数据的统计特性；相关性热力图展示了特征间的线性关联模式；模型性能雷达图直观比较了不同算法的综合表现；预测误差分布图揭示了模型预测的不确定性特征；实际 vs 预测散点图展示了模型拟合优度；SHAP 各类图表提供了模型决策的透明化解释；纤维效应曲线分析了不同纤维类型的影响规律；水泥-硅灰等高线图探索了材料配合比的优化空间；GUI 界面实现了用户友好的交互预测功能。（以下是未在上文呈现的图表）

由于数据来源的根本性差异，本复现在多个方面与原文存在不可避免的偏差。SHAP 特征重要性排序的差异最为明显，这直接反映了模拟数据在复现复杂特征交互关系方面的不足。

最重要的是，也是本次复现的目的。**复现验证了原文的核心科学结论 CatBoost 在 UHPC 强度预测任务中表现最优，龄期和纤维体积率是影响强度的关键参数，钢纤维在所有纤维类型中增强效果最为显著。**这些发现得到了模拟数据的支持，证明了原文方法论的有效性和鲁棒性。

