



基于集成学习的蔗渣灰混凝土抗压强度预测模型

林 星, 梁诗雪, 冯斯奕

(浙江理工大学建筑工程学院, 杭州 310018)

摘 要: 为了得到高效准确的基于集成学习的蔗渣灰混凝土抗压强度预测模型, 建立了 4 种集成学习模型, 即 eXtreme Gradient-Boosting (XGBoost)、Random Forest (RF)、Light Gradient-Boosting Machine (LightGBM) 和 Adaptive Boosting (AdaBoost); 通过模型的性能比较得到了预测能力最优的集成学习模型, 然后利用 SHAP (Shapley additive explanation) 值方法定量研究各输入变量对蔗渣灰混凝土抗压强度的影响。首先, 进行蔗渣灰混凝土抗压强度试验, 根据试验数据和文献数据构建了包含水泥含量、水灰比、蔗渣灰掺量和、细骨料含量、粗骨料含量等 5 个输入变量的集成学习数据库。然后, 采用决定系数、平均绝对误差、均方根误差、可靠性指数等 4 个评估指标来评估模型的预测能力。通过对比发现: XGBoost 模型的预测精度最高, 该模型训练集的评估指标决定系数、平均绝对误差、均方根误差、可靠性指数分别为 0.976、1.811、2.344、0.875。各输入变量对蔗渣灰混凝土抗压强度的影响从大到小排序为水泥含量、细骨料含量、粗骨料含量、蔗渣灰掺量和、水灰比; 水泥含量对混凝土抗压强度有正面影响, 蔗渣灰掺量低于 10% 时不会明显降低混凝土的抗压强度。该研究为蔗渣灰混凝土抗压强度的预测和影响因素解释提供了有益参考, 对于推动蔗渣灰混凝土等环保型材料的研究和应用具有一定价值。

关键词: 集成学习; 蔗渣灰混凝土; 抗压强度; SHAP 值方法; 预测模型

中图分类号: TU377.9

文献标志码: A

文章编号: 1673-3851 (2024) 04-0507-11

引文格式: 林星, 梁诗雪, 冯斯奕. 基于集成学习的蔗渣灰混凝土抗压强度预测模型[J]. 浙江理工大学学报(自然科学), 2024, 51(4): 507-517.

Reference Format: LIN Xing, LIANG Shixue, FENG Siyi. Models for predicting the compressive strength of bagasse ash concrete based on ensemble learning [J]. Journal of Zhejiang Sci-Tech University, 2024, 51(4): 507-517.

Models for predicting the compressive strength of bagasse ash concrete based on ensemble learning

LIN Xing, LIANG Shixue, FENG Siyi

(School of Civil Engineering and Architecture, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: To obtain an efficient and accurate ensemble learning model for predicting the compressive strength of sugarcane bagasse ash concrete, four ensemble learning models, namely eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Light Gradient Boosting Machine (LightGBM), and Adaptive Boosting (AdaBoost) were established. The predictive capabilities of these models were compared, and the ensemble learning model with the optimal predictive performance was identified. The Shapley Additive Explanation (SHAP) value method was employed to quantitatively study the impact of each input variable on the compressive strength of sugarcane bagasse ash concrete. Firstly, compressive strength experiments were conducted for sugarcane bagasse ash concrete. Based on experimental data and literature information, an ensemble learning database consisting of five input variables, namely cement content, water-to-cement ratio, sugarcane bagasse ash admixture content, fine aggregate content, and coarse aggregate content, was built. Subsequently, four evaluation metrics, namely determination

收稿日期: 2023-11-27 网络出版日期: 2024-04-11

基金项目: 国家自然科学基金项目(51808499); 浙江省自然科学基金项目(LY22E080016); 浙江理工大学科研业务费专项资金(24052126-Y)

作者简介: 林 星(2000—), 男, 浙江台州人, 硕士研究生, 主要从事新型混凝土结构方面的研究。

通信作者: 梁诗雪, E-mail: liangshixue0716@126.com

coefficient, mean absolute error, root mean square error, and reliability index, were used to assess the predictive capabilities of the models. In the performance comparison, it was observed that the XGBoost model exhibited the highest predictive accuracy. The evaluation metrics for the training set of the XGBoost model were determined as follows: determination coefficient of 0.976, mean absolute error of 1.811, root mean square error of 2.344, and reliability index of 0.875. The impact of each input variable on the compressive strength of sugarcane bagasse ash concrete was ranked from highest to lowest as follows: cement content, fine aggregate, coarse aggregate, sugarcane bagasse ash admixture content, and water-to-cement ratio. Cement content had a positive effect on concrete compressive strength, and the compressive strength of concrete was not significantly reduced when the sugarcane bagasse ash admixture content was below 10%. This study provides useful reference for predicting the compressive strength of sugarcane bagasse ash concrete and explaining influencing factors. It holds value in advancing research and applications of environmentally friendly materials such as sugarcane bagasse ash concrete.

Key words: ensemble learning; sugarcane bagasse ash concrete; compressive strength; SHAP value method; predictive model

0 引 言

随着社会的发展与人口的增加,农业生产在迅猛发展的同时也产生了大量的农业垃圾。农业垃圾包含大量生物质,为了充分利用农业垃圾,通常将其焚烧转化为能源,以满足工业发展的一定需要。农业垃圾焚烧后产生的生物质灰是农业生产链的最终废弃物,其产量呈现逐年增加的趋势。生物质灰富含二氧化硅(SiO_2),是建筑领域的重要材料,具备资源可持续利用的特点。目前关于将生物质灰应用在建筑领域的研究已成为研究热点^[1]。研究发现,生物质灰如稻壳灰、竹茎灰、麦秆灰、蔗渣灰等可用于制备混凝土、砖块等建筑材料^[2]。其中蔗渣灰是制糖业副产品甘蔗渣经焚烧后由除尘装置得到的灰分。我国拥有广阔的热带、亚热带农耕区域,蔗渣灰的来源稳定、基数大,其燃烧所释放的二氧化碳量远低于植物生长周期中排放的氧气量,是一种潜力巨大的绿色材料^[3]。自 1998 年 Martirena Hernández 等^[4]发现蔗渣灰具有一定的火山灰活性后,关于蔗渣灰的研究成果逐渐在建筑材料领域得到应用^[5-7]。在建筑材料中合理利用蔗渣灰,既能保证建筑材料的抗压强度、抗折强度等力学性能,又可以减少环境污染和资源浪费,符合绿色、可持续发展的要求,具有广阔的应用前景。

抗压强度是混凝土的力学性能中最受研究者关心的核心指标之一。混凝土抗压强度取决于多种因素,如水泥含量、粗骨料含量、细骨料含量等。一些研究者研究了蔗渣灰在混凝土中的应用。Xu 等^[8]使用蔗渣灰替代部分水泥,研究了蔗渣灰掺和量对

混凝土抗收缩、抗侵蚀、抗压强度、抗折强度等力学性能的影响。Sales 等^[9]使用蔗渣灰替代部分天然砂,研究了蔗渣灰掺和量对混凝土的抗压强度、抗拉强度、弹性模量以及吸水性的影响。Katare 等^[10]进行了蔗渣灰混凝土单轴抗压试验,探讨了蔗渣灰掺和量、水灰比、骨料含量等因素对蔗渣灰混凝土抗压强度的影响。然而,蔗渣灰混凝土的抗压强度受到多种因素的共同作用,且添加了蔗渣灰这一成分复杂的生物质灰后,混凝土的力学性能更为复杂,单批试验难以反映其影响规律。因此,亟待建立预测精度较高的抗压强度预测模型。

近年来,越来越多的研究者们利用机器学习方法来解决各类工程问题。机器学习涉及统计学、概率论以及计算机科学等多个学科领域,其主要思想是通过对一部分数据进行学习,运用相关算法分析数据,发掘内在规律,从而对新数据作出预测。在土木工程领域,机器学习模型被广泛应用于研究结构(包括结构构件)承载能力预测、结构动力响应预测等问题^[11-13]。但目前鲜有研究者利用机器学习模型对蔗渣灰混凝土的抗压强度进行预测以及定量研究各种因素对蔗渣灰混凝土抗压强度的影响。

本文通过集成学习模型对蔗渣灰混凝土抗压强度进行预测,并且利用 SHAP(Shapley additive explanation)值方法定量研究各因素对蔗渣灰混凝土抗压强度的影响。首先,本文进行了蔗渣灰混凝土抗压强度试验,研究了蔗渣灰掺和量对混凝土抗压强度的影响,并将获得的试验数据与国内外文献中收集到的强度数据相结合,作为集成学习模型的训练和测试的数据库。然后,采用 4 种不同的集成

学习模型,即 eXtreme Gradient-Boosting (XGBoost)、Random Forest (RF)、Light Gradient-Boosting Machine (LightGBM)和 Adaptive Boosting (AdaBoost),对蔗渣灰混凝土的抗压强度进行预测,并得到预测精度最高的集成学习模型,从而建立高效准确的抗压强度预测模型,为实际工程中的材料选择和结构设计提供可靠的参考。最后,利用 SHAP 值方法定量分析水灰比、蔗渣灰掺和量、水泥含量、细骨料含量和粗骨料含量等 5 个输入变量对蔗渣灰混凝土抗压强度的影响,为混凝土配合比的设计提供科学依据,进而推进蔗渣灰混凝土等环保型材料的应用。

表 1 试验所用水泥的基本性能参数

| 强度等级 | 3 d 抗压强度/MPa | 3 d 抗折强度/MPa | 初凝时间/min | 终凝时间/min | 安定性 |
|------|--------------|--------------|----------|----------|-----|
| 42.5 | 27.20 | 5.50 | 172 | 234 | 合格 |

表 2 试验所用水泥的主要组分及含量

| 材料 | CaO | SiO ₂ | Al ₂ O ₃ | Fe ₂ O ₃ | MgO | SO ₃ | Cl | 其他 |
|----|-------|------------------|--------------------------------|--------------------------------|------|-----------------|------|------|
| 水泥 | 51.42 | 24.99 | 8.26 | 4.03 | 3.71 | 2.51 | 0.04 | 5.04 |

1.2.2 细骨料

砂子采用河砂,级配区间为Ⅱ区的中砂,细度模数为 2.6,表观密度为 2630 kg/m³,堆积密度为 1537 kg/m³,含泥量 0.8%。

1.2.3 粗骨料

石子采用碎石,最大粒径为 20 mm,表观密度为 2650 kg/m³,堆积密度为 1408 kg/m³,含泥量 0.4%。

1.2.4 蔗渣灰

蔗渣灰的制备过程涉及将甘蔗渣进行一系列处理,最终得到试验所使用的蔗渣灰,蔗渣灰制备过程中各阶段的示例照片见图 1。蔗渣灰的具体制备过程如下:a)收集甘蔗渣。收集蔗糖生产过程中剩余的甘蔗渣。b)清理和预处理。将收集的甘蔗渣去除杂质、清洗和干燥后进行初步研磨,得到粗磨甘蔗渣。c)研磨。对处理后的甘蔗渣进行细致研磨,获得更细的颗粒。d)灼烧。将研磨后的甘蔗渣置于锅炉中,用 500~600 ℃的高温灼烧。e)筛分。将灼烧后的蔗渣灰通过筛子进行筛分,得到的灰烬小于 75 μm。

1.3 混凝土配合比设计

本次试验选定混凝土设计强度为 C30,按照 JGJ 55—2019《普通混凝土配合比设计规程》进行混凝土配合比设计。试验设置了对照组 S0,该组的蔗渣灰掺和量为 0,并且设置试验组 S1、S2、S3,这 3 组的蔗渣灰掺和量分别为 10.00%、20.00%、30.00%,其具体配合比见表 3。

1 蔗渣灰混凝土抗压强度试验

1.1 试验方案

本文通过试验分析不同的蔗渣灰掺和量对混凝土抗压强度的影响。采用了取代率不同的蔗渣灰替代水泥,制备混凝土试件;并根据 GB/T 50081—2019《混凝土物理力学性能试验方法标准》进行混凝土抗压强度试验。

1.2 原材料

1.2.1 水泥

水泥采用 P·O42.5 级普通硅酸盐水泥,其基本性能参数及主要组成分别见表 1—表 2。



(a) 收集的甘蔗渣



(b) 清理和预处理后的甘蔗渣



(c) 研磨后的甘蔗渣



(d) 灼烧和筛分后的蔗渣灰

图 1 蔗渣灰制备过程中各阶段的示例照片

1.4 试验方法

首先,按照表 3 的配合比称取原材料,接着依次向搅拌机内投入水泥、骨料、水以及蔗渣灰,在确保搅拌均匀后排出混凝土料。然后,制备尺寸为 150 mm×150 mm×150 mm 的混凝土试件,共 4 组,每组 3 个试块。5 h 后进行收面处理,在相对湿度为 60%、温度为(20±1)℃的环境中养护 1 d 后脱模。随后将试件放置于相对湿度为 95%以上、温度为(20±2)℃的标准养护室内养护至 28 d。最后使用 TYE-2000B 型压力试验机进行抗压强度试验。

表 3 蔗渣灰混凝土配合比

| 编号 | 水灰比 | 蔗渣灰掺和量/% | 水泥含量/(kg·m ⁻³) | 细骨料含量/(kg·m ⁻³) | 粗骨料含量/(kg·m ⁻³) |
|----|------|----------|----------------------------|-----------------------------|-----------------------------|
| S0 | 0.38 | 0 | 450.00 | 715.00 | 937.00 |
| S1 | 0.42 | 10.00 | 406.00 | 715.00 | 937.00 |
| S2 | 0.48 | 20.00 | 360.00 | 715.00 | 937.00 |
| S3 | 0.55 | 30.00 | 315.00 | 715.00 | 937.00 |

1.5 试验结果

蔗渣灰混凝土抗压强度试验数据见图 2,从图中可以发现:当蔗渣灰掺和量为 0 时,混凝土抗压强度为 33.76 MPa;而在掺和量为 10.00%时,混凝土抗压强度略微下降至 33.61 MPa;当蔗渣灰掺和量从 10.00%增加至 20.00%时,混凝土抗压强度快速降至 26.93 MPa;随着蔗渣灰掺和量超过 20.00%,混凝土抗压强度的下降幅度减小;当掺和量达到 30.00%时,混凝土抗压强度为 24.00 MPa。

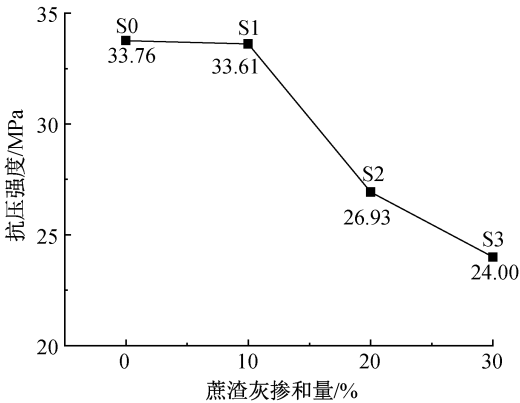


图 2 蔗渣灰混凝土抗压强度与蔗渣灰在掺和量间的关系曲线

综上所述,蔗渣灰在掺和量低于 10.00%时对混凝土抗压强度无明显负面影响。这可能是因为蔗渣灰颗粒粒径较小,在混凝土中发挥了填充的作用,填补了水泥浆体与骨料之间的空隙,使得混凝土强度没有明显下降^[14]。随着掺和量的增加,负面影响增大。这可能是由于蔗渣灰颗粒之间的粘合力低,并且水泥含量减少导致混凝土内部的粘合力下降,混凝土强度明显下降^[14]。

2 集成学习模型

本文选取 4 种集成学习模型来预测蔗渣灰混凝土的抗压强度,即 eXtreme Gradient-Boosting (XGBoost)、Random Forest (RF)、Light Gradient-Boosting Machine (LightGBM) 和 Adaptive Boosting (AdaBoost)。集成学习模型是一种机器学习方法,其通过整合多个弱学习器的预测结果,以弥补单个弱学习器的不足,从而获得具有更高预测能力

的强学习器。集成学习模型通常在性能方面表现出卓越的优势,能够克服单个学习器的局限性,提高模型的整体性能。

2.1 eXtreme Gradient-Boosting (XGBoost)

XGBoost 是一种高效且灵活的集成学习模型,属于梯度提升框架(Gradient boosting framework, GBF)中的一种。XGBoost 通过迭代的方式构建决策树,每一轮迭代都生成一棵决策树。在每轮迭代中,首先模型会计算当前模型的残差,接着构建一棵新的决策树来减小残差。在构建决策树的过程中,XGBoost 会通过最小化损失函数来确定决策树的结构,然后引入正则化项以避免过拟合^[15]。构建完成后,XGBoost 会得到 n 棵决策树。当对新样本进行预测时,每个新样本会被分配到决策树上对应的叶节点,每个叶节点有一个关联的分数,最终的输出结果是所有决策树的叶节点分数的累加和^[16]。图 3 展示了 XGBoost 的算法原理。

2.2 Random Forest (RF)

RF 属于集成学习模型中的一种组合分类算法模型。RF 通过构建多个决策树,将所有决策树的输出结果结合起来进行最终的预测。在构建决策树的过程中,RF 首先利用随机采样法从数据集集中有放回或无放回地抽取多个样本,使得每棵决策树的训练数据集都略有不同,以避免过拟合。然后用抽取的样本对决策树进行训练,最后把训练后的 n 棵决策树组合在一起,通过投票得出最终的预测结果^[17]。图 4 展示了 RF 的算法原理。

2.3 Light Gradient-Boosting Machine (LightGBM)

LightGBM 是一种基于 GBF 的集成学习模型,原理上与 XGBoost 相似,能够在不降低准确率的前提下加快模型的训练速度。图 5 展示了 LightGBM 的优化原理。LightGBM 相对于 XGBoost 在一些方面进行了优化:

a)采用带深度限制的节点展开(Leaf-wise)算法。大多数梯度提升决策树工具使用低效的按层生长的 Level-wise 决策树生长策略,不加区分地对待同一层的叶节点,探索 and 分裂了很多分裂效益较低的叶节点。LightGBM 使用了带有深度限制的节点

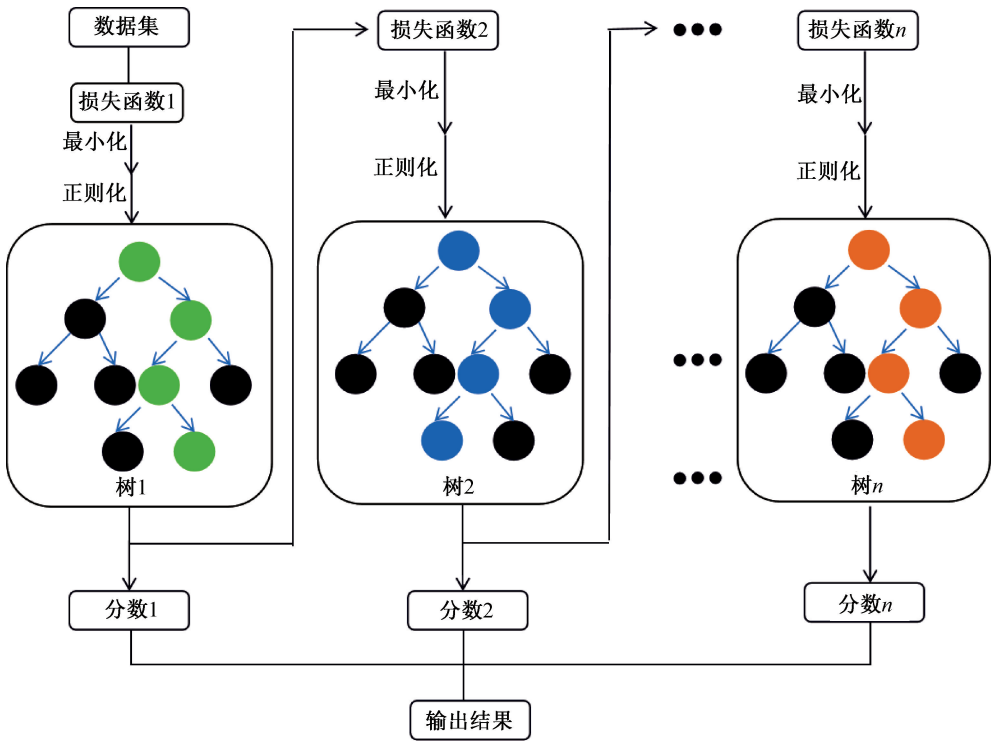


图 3 XGBoost 算法原理示意图

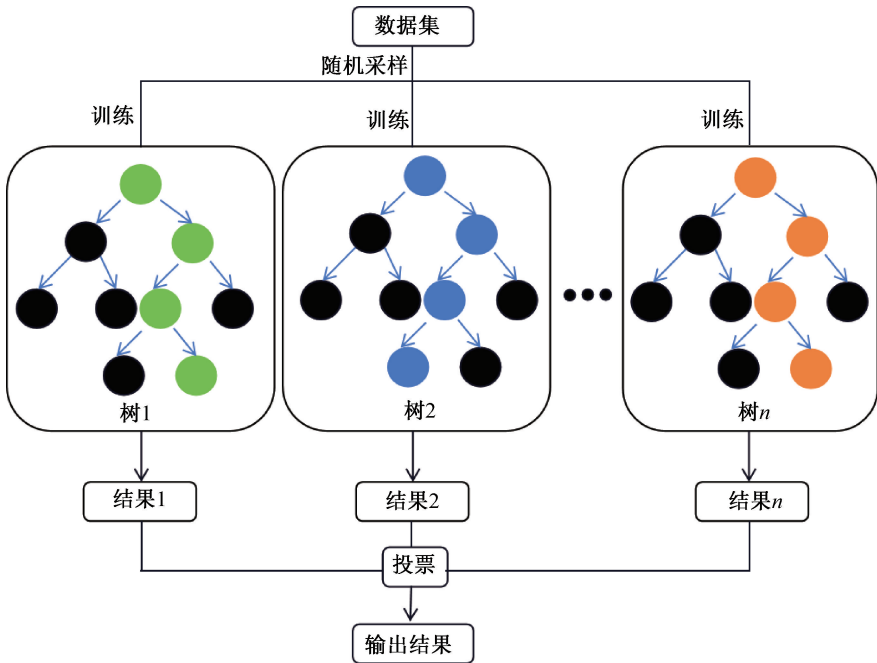


图 4 RF 算法原理示意图

展开的 Leaf-wise 算法^[18],决策树的层级不会无限制地增长,这有助于控制模型的复杂度,防止过拟合。

b)引入直方图算法。为了加速梯度计算,首先 LightGBM 会将连续的特征值划分为多个离散区间,然后在训练过程中构建直方图,最后使用直方图差来近似直方图的梯度信息,以降低梯度的计算复杂度。

2.4 Adaptive Boosting (AdaBoost)

AdaBoost 是由 Yoav Freund 和 Robert E.

Schapire 在 1995 年提出的一种集成学习模型^[19],它通过迭代训练一系列弱学习器,然后将它们组合成一个强大的分类器。AdaBoost 的核心思想是通过调整训练样本的权重,使得前一个弱学习器分错的样本在后续的学习中得到更多的关注,从而纠正错误,提高模型的整体性能。AdaBoost 将多个弱学习器的结果进行加权组合,形成一个在整体上具有更好的泛化能力的强学习器。图 6 展示了 AdaBoost 的算法原理。

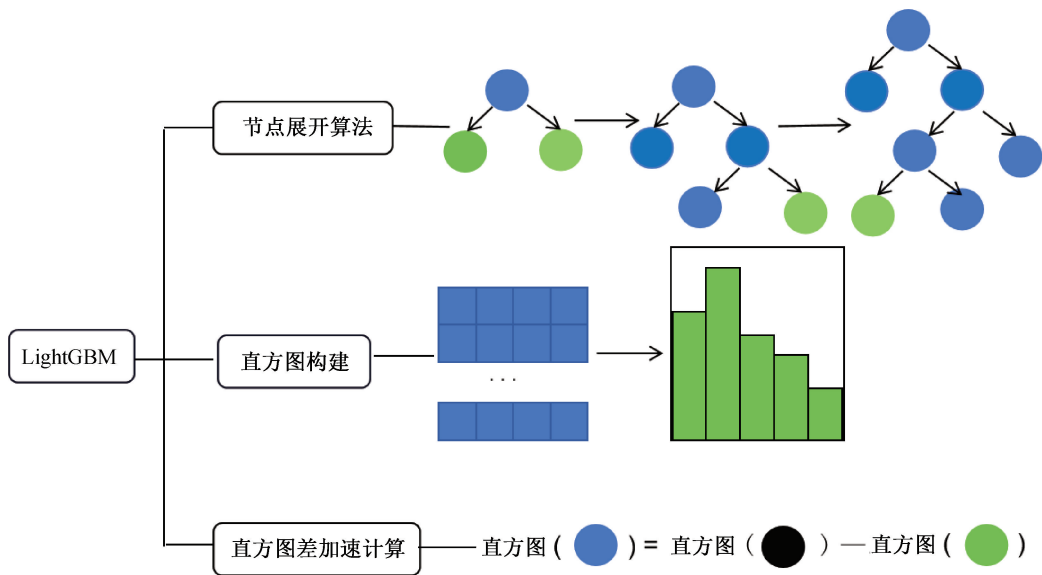


图 5 LightGBM 优化原理示意图

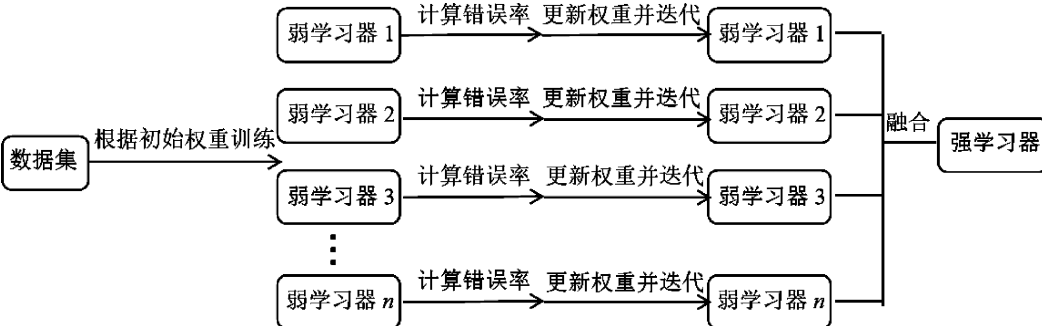


图 6 AdaBoost 算法原理示意图

3 蔗渣灰混凝土强度预测

3.1 蔗渣灰混凝土抗压强度数据库

结合本文开展的试验以及文献[20-21],共收集了 68 组蔗渣灰混凝土抗压强度的数据作为数据集(见附录 A),并且将数据集按照 8:2 的比例划分为训练集和测试集,其中训练集为 55 组,测试集为 13 组。本文根据收集的数据集,选取了 5 个变量作为输入变量,它们分别为水灰比、蔗渣灰掺和量、水泥含量、细骨料含量、粗骨料含量;将蔗渣灰混凝土的抗压强度作为输出变量。集成学习建模可接受的数据量应大于输入变量个数的 10 倍^[22],因此数据量应大于 50 组,本文数据集的数据量符合要求。

3.2 超参数取值

为了提高集成学习模型的预测准确性和泛化能力,需要进行超参数调优。超参数是在训练集成学习模型之前预设的参数,其值不能由模型从数据中自动学习得出。超参数优化的常见方法包括网格搜索、随机搜索和贝叶斯优化等,本文采用了网格搜索结合 K-fold 交叉验证的方法来优化超参数。

网格搜索通过事先定义一组可能的超参数值,然后尝试所有可能的超参数组合,以遍历模型的各种超参数组合^[23]。而 K-fold 交叉验证是一种数据拆分技术,可以让使用者在未曾见过的数据上评估模型的性能^[24]。首先,它将数据集分成 K 个子集,然后多次训练模型,每次使用不同的子集作为验证集,其余的作为训练集。最后,将这些验证集上的性能指标的平均值作为最终性能的评估指标^[18]。

本文采用了 10 折交叉验证。首先,将所有样本均匀分成 10 个子集。然后,依次遍历这 10 个子集,每次将 1 个子集作为验证集,其余 9 个子集作为训练集用于模型的训练。最后,取这 10 次评估的平均值作为最终评估指标。

表 4 列出了 4 种集成学习模型在网格搜索中需要调优的主要超参数,表 5 总结了每个模型的最佳超参数组合以及相应的交叉验证性能评分。

3.3 模型预测结果

为了准确比较 4 种集成学习模型的预测能力,这 4 种模型均使用相同的训练集和测试集,最终通过 4 个评估指标(决定系数、平均绝对误差、均方根

表 4 集成学习模型超参数组合

| 集成学习模型 | 弱学习器数量 | 学习率 | 学习深度 |
|----------|----------------|-----------------|------------|
| XGBoost | 50,60,70,⋯,200 | 0.1,0.2,0.3,⋯,1 | 1,2,3,⋯,10 |
| RF | 50,60,70,⋯,200 | | 1,2,3,⋯,20 |
| AdaBoost | 50,60,70,⋯,200 | 0.1,0.2,0.3,⋯,1 | |
| LightGBM | 50,60,70,⋯,200 | 0.1,0.2,0.3,⋯,1 | 1,2,3,⋯,20 |

表 5 集成学习模型最优超参数

| 集成学习模型 | 弱学习器数量 | 学习率 | 学习深度 | 平均输出结果 |
|----------|--------|-----|------|--------|
| XGBoost | 200 | 0.1 | 2 | 0.76 |
| RF | 50 | | 11 | 0.73 |
| AdaBoost | 160 | 0.7 | | 0.56 |
| LightGBM | 60 | 0.4 | 10 | 0.30 |

误差、可靠性指数)评价模型的预测能力。4 个评估指标的计算公式为式(1)–(4):

$$R=1-\frac{\sum_{i=1}^m(f(x_i)-y_i)^2}{\sum_{i=1}^m(\overline{y_i}-y_i)^2}$$

(1)

$$M=\frac{1}{m}\sum_{i=1}^m|f(x_i)-y_i|$$

(2)

$$r=\sqrt{\frac{1}{m}\sum_{i=1}^m(f(x_i)-y_i)^2}$$

(3)

$$a_{10}=\frac{m_{10}}{m}$$

(4)

其中: R 为决定系数; M 为平均绝对误差; r 为均方根误差; a_{10} 为可靠性指数; $f(x_i)$ 为模型对蔗渣灰混凝土抗压强度的预测值,MPa; y_i 为蔗渣灰混凝土抗压强度的实验值,MPa; $\overline{y_i}$ 为蔗渣灰混凝土抗压强度实验值的平均值,MPa; m 为数据集样本的数量; m_{10} 为实验值与预测值之比在 0.9 和 1.1 之间的数据集样本数量。

决定系数是一种用于评估模型拟合程度的统计指标,表示因变量(输出变量)的变异性能够被自变量(输入变量)解释的比例,它的取值范围在 0 到 1 之间,其值越接近于 1,说明模型对数据拟合效果越好;平均绝对误差用于评估预测值和实验值之间的接近程度,其值越小说明拟合效果越好;均方根误差是一种用于衡量预测值和实验值之间的偏差的度量,它对一组数据中非常大或者非常小的误差反应非常敏感,其值越小说明模型拟合效果越好;可靠性指数通过显示预测值与实验值偏差在±10%之间的样本数量占全部样本数量的比例来评估模型的可靠性,其值越靠近 1 说明模型可靠性越高,模型拟合效果越好。

图 7(a)–(d)展示的预测结果可以作为评估 4 个集成学习模型预测能力的参考依据。图中的蓝色圆点代表训练集结果,橙色方点代表测试集结果。4 种模型预测结果的数据点散布在基线($y=x$)周围。观察图 7(a)–(d)可以发现,图 7(a)中预测值与实验值之间的差异较小,表明集成学习模型 XGBoost 的预测能力较好。

为了比较 4 种模型的预测能力,表 6 列出了 4 种模型在训练集和测试集上的评估指标具体数值。在训练集上,XGBoost 模型表现出色,其 R 为 0.976, M 为 1.811, r 为 2.344, a_{10} 为 0.875,优于其他 3 个模型。LightGBM 模型的表现较差,4 个评估指标的数值分别为 0.529、7.959、10.490、0.271,准确性最低。

在测试集上,XGBoost 模型展现出良好的泛化能力,其 R 为 0.833, M 为 6.229, r 为 11.898, a_{10} 为 0.333。LightGBM 模型的表现最差,其 R 为 0.384, M 为 12.722, r 为 16.601, a_{10} 为 0.250,表明其在未见数据上的适应能力较差。综合考虑 4 个评估指标,4 个模型的预测能力由高到低依次为 XGBoost、RF、AdaBoost、LightGBM。

3.4 SHAP 影响因素分析

基于以上结果分析,选用 XGBoost 模型进行影响因素分析。由于集成学习模型只能得到最终的预测结果,无法解释特征值对结果的影响。因此本文通过 SHAP 值方法定量解释每个输入变量对于预测结果的影响。

集成学习模型的解释方法包括 Partial dependence plot(PDP)、Individual conditional expectation(ICE)、Permuted feature importance(PFI)、Global surrogate(GS)、Local surrogate(LS)、Shapley value(SHAP)等,其中 SHAP 值方法在集成学习模型的解释中应用最为广泛,因为它既能反映每个样本所有特征的影响,又能显示影响的正负性^[25]。

根据得到的 SHAP 值创建 XGBoost 模型的全局特征重要性条形图和全局特征重要性蜂窝图,如图 8—图 9 所示。观察图 8 可以看出,各输入变量对抗压强度的影响从大到小排序为:水泥含量、细骨

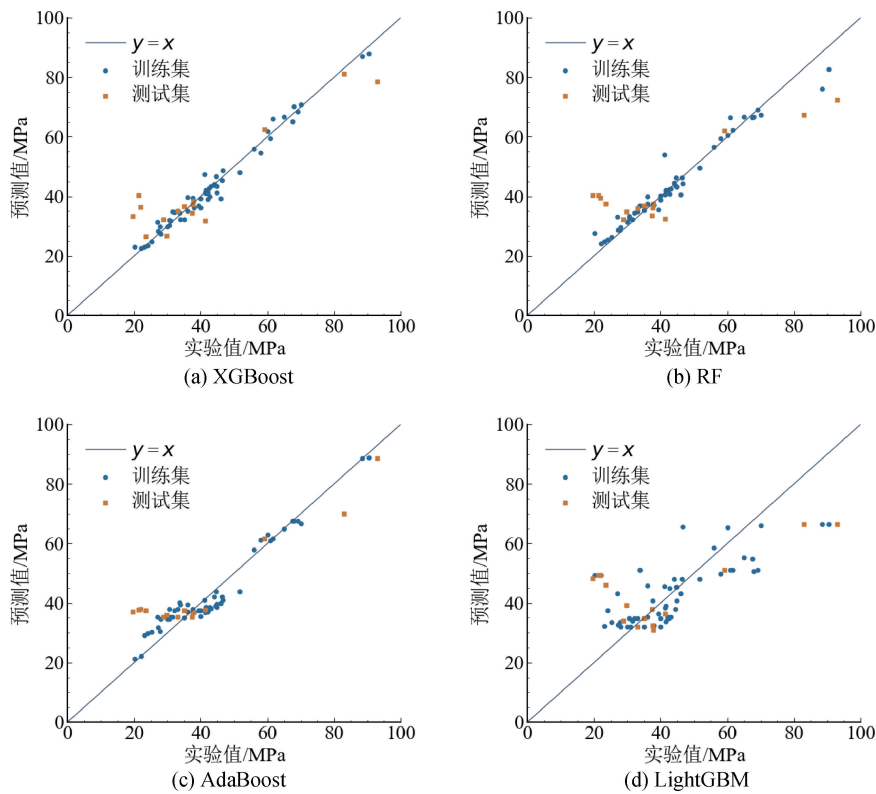


图 7 集成学习模型预测结果散点图

表 6 集成学习模型综合性能

| 集成学习 模型 | 训练集 | | | | 测试集 | | | |
|------------|---------|-------|----------|----------|---------|--------|----------|----------|
| | XGBoost | RF | AdaBoost | LightGBM | XGBoost | RF | AdaBoost | LightGBM |
| R | 0.976 | 0.949 | 0.926 | 0.529 | 0.833 | 0.684 | 0.802 | 0.384 |
| M | 1.811 | 2.205 | 3.483 | 7.959 | 6.229 | 9.206 | 7.238 | 12.722 |
| r | 2.344 | 3.441 | 4.167 | 10.490 | 11.898 | 11.898 | 9.418 | 16.601 |
| a_{10} | 0.875 | 0.875 | 0.417 | 0.271 | 0.333 | 0.250 | 0.417 | 0.250 |

料含量、粗骨料含量、蔗渣灰掺和量、水灰比。由此可以得到结论:水泥含量对抗压强度的影响是最大的,而蔗渣灰掺和量的影响相对较小。

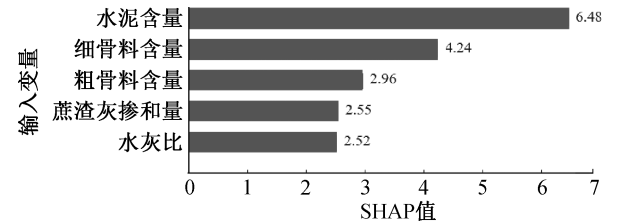


图 8 XGBoost 全局特征重要性条形图

XGBoost 模型全局特征重要性蜂窝图如图 9 所示,其中每个点的颜色从蓝色过渡到红色表示相应特征值由小变大。观察图 9 可以得到以下结论:

- a)水泥含量对蔗渣灰混凝土抗压强度的影响。水泥含量的特征值与 SHAP 值呈现正相关,这表明水泥含量对抗压强度有正面的影响。
- b)细骨料含量和蔗渣灰掺和量对蔗渣灰混凝土

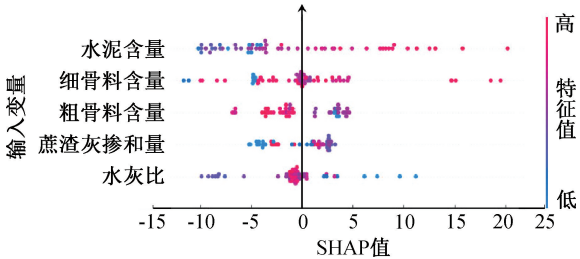


图 9 XGBoost 模型全局特征重要性蜂窝图

土抗压强度的影响。随着细骨料含量和蔗渣灰掺和量的增加,SHAP 值呈现先增加后减小的趋势,这说明过多的细骨料含量和蔗渣灰掺和量对混凝土抗压强度有负面影响。

c)水灰比对蔗渣灰混凝土抗压强度的影响。水灰比的值与 SHAP 值呈现负相关,这表明水灰比对抗压强度具有负面影响。

综合图 8 和图 9 的信息,可以得出结论:蔗渣灰掺和量较低时并不会明显降低混凝土的抗压强度。

4 结 论

本文构建了基于集成学习的蔗渣灰混凝土抗压强度预测模型,首先通过试验探究了蔗渣灰掺和量对混凝土抗压强度的影响,然后挑选出了预测能力最佳的集成学习模型,最后定量分析了输入变量对蔗渣灰混凝土抗压强度的影响。所得主要研究结论有:

a)蔗渣灰混凝土抗压强度试验结果显示,蔗渣灰掺和量在 10.00% 以下时对混凝土抗压强度没有明显的负面影响,蔗渣灰掺和量超过 10.00% 时对混凝土抗压强度表现出了负面影响。

b)XGBoost 模型在 4 个模型中预测效果最好,它在训练集的 R 、 M 、 r 、 a_{10} 分别为 0.976、1.811、2.344、0.875,在测试集中的 R 、 M 、 r 、 a_{10} 分别为 0.833、6.229、11.898、0.333。

c)蔗渣灰混凝土抗压强度的影响因素分析结果显示,水泥含量对抗压强度影响最大,对抗压强度有正面影响;蔗渣灰掺和量影响相对较小,并且蔗渣灰掺和量过多会有负面影响,该结果与本文进行的试验结果基本符合。

本文发现,在混凝土中使用替代率不超过 10.00% 的蔗渣灰来替代水泥,不仅能够确保混凝土的性能,还有助于减少环境污染,这为混凝土配合比设计提供了有益的指导。XGBoost 模型的应用不仅能够支持混凝土质量控制方面的工程实践,还突显了在可持续性设计中合理使用蔗渣灰的重要性。

参考文献:

[1] 黄萍,徐青,吴能森,等. 蔗渣灰特性及其在建筑材料的应用研究综述[J]. 武汉工程大学学报, 2022, 44(6): 599-607.

[2] 王卓琳,黎庆涛,田潇凌,等. 甘蔗糖厂废弃物蔗渣灰综合利用的研究进展和展望[J]. 轻工科技, 2018, 34(2): 74-76.

[3] Andreão P V, Suleiman A R, Cordeiro G C, et al. Beneficiation of sugarcane bagasse ash: Pozzolanic activity and leaching behavior[J]. Waste and Biomass Valorization, 2020, 11(8): 4393-4402.

[4] Martirena Hernández J F, Middendorf B, Gehrke M, et al. Use of wastes of the sugar industry as pozzolana in lime-pozzolana binders: Study of the reaction [J]. Cement and Concrete Research, 1998, 28(11): 1525-1536.

[5] Ambedkar B, Alex J, Dhanalakshmi J. Enhancement of

mechanical properties and durability of the cement concrete by RHA as cement replacement: Experiments and modeling[J]. Construction and Building Materials, 2017, 148: 167-175.

[6] Rodier L, Bilba K, Onésippe C, et al. Study of pozzolanic activity of bamboo stem ashes for use as partial replacement of cement [J]. Materials and Structures, 2016, 50(1): 87.

[7] Al-Akhras N M, Abu-Alfoul B A. Effect of wheat straw ash on mechanical properties of autoclaved mortar[J]. Cement and Concrete Research, 2002, 32(6): 859-863.

[8] Xu Q, Ji T, Gao S J, et al. Characteristics and applications of sugar cane bagasse ash waste in cementitious materials[J]. Materials, 2018, 12(1): 39.

[9] Sales A, Lima S A. Use of Brazilian sugarcane bagasse ash in concrete as sand replacement [J]. Waste Management, 2010, 30(6): 1114-1122.

[10] Katare V D, Madurwar M V. Experimental characterization of sugarcane biomass ash: A review [J]. Construction and Building Materials, 2017, 152: 1-15.

[11] 马财龙,王文虎,侯宪龙,等. 基于可解释机器学习方法的 RC 深弯构件开裂剪力预测[J]. 新疆大学学报(自然科学版)(中英文), 2023, 40(5): 621-629.

[12] 李萌,武海军,董恒,等. 基于机器学习的混凝土侵蚀深度预测模型[J]. 兵工学报, 2023, 44(12): 3771-3782.

[13] 李宏伟,王文武,贾冯睿,等. 钢筋混凝土黏结-滑移行为敏感性分析及机器学习模型[J/OL]. 辽宁石油化工大学学报: 1-9 (2023-06-26) [2024-01-26]. <http://kns.cnki.net/kcms/detail/21.1504.TE.20230625.1448.008.html>.

[14] 李腾宇,梁晋鹏,陀广泉,等. 蔗渣灰对砂浆性能及孔结构的影响研究[J]. 混凝土与水泥制品, 2022, (10): 92-95.

[15] Sagi O, Rokach L. Approximating XGBoost with an interpretable decision tree [J]. Information Sciences, 2021, 572: 522-542.

[16] 张驰,郭媛,黎明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57-69.

[17] 李慧,李正,余堃. 一种基于综合不放回抽样的随机森林算法改进[J]. 计算机工程与科学, 2015, 37(7): 1233-1238.

[18] Guo R H, Fu D L, Sollazzo G. An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree [J]. International Journal of Pavement Engineering, 2022, 23(10): 3633-3646.

[19] Freund Y, Schapire R E. A desicion-theoretic generalization of on-line learning and an application to boosting[C]//European Conference on Computational Learning Theory, 1995. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995: 23-37.

[20] Shah M I, Javed M F, Aslam F, et al. Machine learning modeling integrating experimental analysis for predicting the properties of sugarcane bagasse ash concrete[J]. Construction and Building Materials, 2022,314:125634.

[21] Sai Vamsi Gangadhar M, Absar W, Siddharth A, et al. Comparative study on sugarcane bagasse ash and banana peel powder on concrete and its properties[J/OL]. Materials Today: Proceedings. (2023-01-24) [2023-11-27]. <https://doi.org/10.1016/j.matpr.2023.04.041>.

[22] Wang H Q, Yi W, Liu Y. An innovative approach of determining the sample data size for machine learning models; a case study on health and safety management for infrastructure workers[J]. Electronic Research Archive, 2022, 30(9): 3452-3462.

[23] 王健峰, 张磊, 陈国兴, 等. 基于改进的网格搜索法的 SVM 参数优化[J]. 应用科技, 2012, 39(3): 28-31.

[24] Naderpour H, Mirrashid M. Estimating the compressive strength of eco-friendly concrete incorporating recycled coarse aggregate using neuro-fuzzy approach[J]. Journal of Cleaner Production, 2020, 265: 121886.

[25] Mi J X, Li A D, Zhou L F. Review study of interpretation methods for future interpretable machine learning[J]. IEEE Access, 2020, 8: 191969-191985.

附录 A

表 A.1 蔗渣灰混凝土试验数据

| 来源 | 编号 | 水灰比 | 蔗渣灰 掺和量/% | 水泥含量/ (kg·m ⁻³) | 细骨料含量/ (kg·m ⁻³) | 粗骨料含量/ (kg·m ⁻³) | 抗压强度/MPa |
|------------------------|----|------|--------------|--------------------------------|---------------------------------|---------------------------------|----------|
| Shah 等 ^[20] | 1 | 0.55 | 0.00 | 347.00 | 548.00 | 1260.00 | 30.50 |
| | 2 | 0.55 | 5.00 | 330.00 | 548.00 | 1260.00 | 33.00 |
| | 3 | 0.55 | 10.00 | 312.00 | 548.00 | 1260.00 | 37.50 |
| | 4 | 0.55 | 15.00 | 295.00 | 548.00 | 1260.00 | 38.00 |
| | 5 | 0.55 | 20.00 | 278.00 | 548.00 | 1260.00 | 35.00 |
| | 6 | 0.55 | 25.00 | 260.00 | 548.00 | 1260.00 | 30.00 |
| | 7 | 0.55 | 30.00 | 243.00 | 548.00 | 1260.00 | 28.00 |
| | 8 | 0.45 | 0.00 | 450.00 | 750.00 | 955.00 | 41.10 |
| | 9 | 0.45 | 10.00 | 405.00 | 740.00 | 945.00 | 46.50 |
| | 10 | 0.45 | 20.00 | 360.00 | 735.00 | 935.00 | 46.00 |
| | 11 | 0.45 | 30.00 | 315.00 | 730.00 | 925.00 | 41.20 |
| | 12 | 0.45 | 40.00 | 270.00 | 720.00 | 915.00 | 39.40 |
| | 13 | 0.45 | 50.00 | 225.00 | 710.00 | 905.00 | 37.60 |
| | 14 | 0.30 | 0.00 | 560.00 | 710.00 | 924.00 | 65.00 |
| | 15 | 0.30 | 10.00 | 504.00 | 717.00 | 896.00 | 69.00 |
| | 16 | 0.30 | 20.00 | 448.00 | 711.00 | 888.00 | 68.00 |
| | 17 | 0.30 | 30.00 | 392.00 | 708.00 | 885.00 | 67.50 |
| | 18 | 0.35 | 0.00 | 478.00 | 865.00 | 905.30 | 60.86 |
| | 19 | 0.35 | 10.00 | 430.20 | 865.00 | 905.30 | 61.56 |
| | 20 | 0.35 | 15.00 | 406.30 | 865.00 | 905.30 | 59.01 |
| | 21 | 0.35 | 20.00 | 382.40 | 865.00 | 905.80 | 57.83 |
| | 22 | 0.45 | 0.00 | 165.70 | 240.00 | 490.00 | 36.00 |
| | 23 | 0.45 | 5.00 | 157.40 | 240.00 | 490.00 | 42.00 |

续表 A. 1

| 来源 | 编号 | 水灰比 | 蔗渣灰 掺和量/% | 水泥含量/ (kg•m ⁻³) | 细骨料含量/ (kg•m ⁻³) | 粗骨料含量/ (kg•m ⁻³) | 抗压强度/MPa |
|--|----|------|--------------|--------------------------------|---------------------------------|---------------------------------|----------|
| Shah 等 ^[20] | 24 | 0.45 | 10.00 | 149.10 | 240.00 | 490.00 | 43.00 |
| | 25 | 0.45 | 15.00 | 140.80 | 240.00 | 490.00 | 42.00 |
| | 26 | 0.45 | 20.00 | 132.60 | 240.00 | 490.00 | 40.00 |
| | 27 | 0.45 | 25.00 | 124.30 | 240.00 | 490.00 | 35.00 |
| | 28 | 0.45 | 30.00 | 116.00 | 240.00 | 490.00 | 32.00 |
| | 29 | 0.55 | 0.00 | 350.00 | 650.00 | 1205.00 | 42.67 |
| | 30 | 0.56 | 5.00 | 332.50 | 650.00 | 1205.00 | 44.80 |
| | 31 | 0.58 | 15.00 | 297.50 | 650.00 | 1205.00 | 41.22 |
| | 32 | 0.60 | 25.00 | 262.50 | 650.00 | 1205.00 | 37.29 |
| | 33 | 0.50 | 0.00 | 500.00 | 745.00 | 945.00 | 62.00 |
| | 34 | 0.50 | 5.00 | 475.00 | 745.00 | 945.00 | 93.00 |
| | 35 | 0.50 | 10.00 | 450.00 | 745.00 | 945.00 | 88.50 |
| | 36 | 0.50 | 15.00 | 425.00 | 745.00 | 945.00 | 83.00 |
| | 37 | 0.50 | 20.00 | 400.00 | 745.00 | 945.00 | 70.00 |
| | 38 | 0.50 | 25.00 | 375.00 | 745.00 | 945.00 | 60.00 |
| | 39 | 0.50 | 30.00 | 350.00 | 745.00 | 945.00 | 56.00 |
| | 40 | 0.50 | 0.00 | 383.00 | 727.00 | 1103.00 | 23.43 |
| | 41 | 0.50 | 5.00 | 363.85 | 727.00 | 1103.00 | 29.75 |
| | 42 | 0.50 | 10.00 | 344.70 | 727.00 | 1103.00 | 30.57 |
| | 43 | 0.50 | 15.00 | 325.55 | 727.00 | 1103.00 | 27.22 |
| | 44 | 0.50 | 20.00 | 306.40 | 727.00 | 1103.00 | 23.13 |
| | 45 | 0.53 | 0.00 | 383.00 | 575.00 | 1150.00 | 36.00 |
| | 46 | 0.53 | 5.00 | 364.00 | 575.00 | 1150.00 | 41.50 |
| | 47 | 0.53 | 10.00 | 345.00 | 575.00 | 1150.00 | 42.50 |
| | 48 | 0.53 | 15.00 | 326.00 | 575.00 | 1150.00 | 42.00 |
| | 49 | 0.53 | 20.00 | 306.00 | 575.00 | 1150.00 | 40.00 |
| | 50 | 0.53 | 25.00 | 287.00 | 575.00 | 1150.00 | 33.00 |
| | 51 | 0.53 | 30.00 | 268.00 | 575.00 | 1150.00 | 31.00 |
| | 52 | 0.40 | 0.00 | 480.00 | 335.00 | 865.00 | 22.04 |
| | 53 | 0.40 | 5.00 | 456.00 | 335.00 | 865.00 | 20.13 |
| | 54 | 0.40 | 10.00 | 432.00 | 335.00 | 865.00 | 21.25 |
| | 55 | 0.40 | 15.00 | 408.00 | 335.00 | 865.00 | 21.94 |
| | 56 | 0.40 | 20.00 | 384.00 | 335.00 | 865.00 | 19.56 |
| | 57 | 0.48 | 0.00 | 399.12 | 516.13 | 1218.17 | 44.89 |
| | 58 | 0.48 | 10.00 | 359.20 | 516.13 | 1218.17 | 44.44 |
| | 59 | 0.48 | 15.00 | 339.20 | 516.13 | 1218.17 | 41.56 |
| | 60 | 0.48 | 20.00 | 319.20 | 516.13 | 1218.17 | 37.78 |
| Sai Vamsi Gangadhar 等 ^[21] | 61 | 0.40 | 5.00 | 450.00 | 625.38 | 1168.58 | 46.33 |
| | 62 | 0.40 | 10.00 | 450.00 | 614.53 | 1148.31 | 51.70 |
| | 63 | 0.40 | 15.00 | 428.45 | 611.57 | 1142.78 | 44.61 |
| | 64 | 0.40 | 0.00 | 450.00 | 635.24 | 1142.78 | 44.00 |
| 本文 | 65 | 0.38 | 0.00 | 450.00 | 715.00 | 937.00 | 33.76 |
| | 66 | 0.42 | 10.00 | 406.00 | 715.00 | 937.00 | 33.61 |
| | 67 | 0.48 | 20.00 | 360.00 | 715.00 | 937.00 | 26.93 |
| | 68 | 0.55 | 30.00 | 315.00 | 715.00 | 937.00 | 24.00 |