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.0454-5648.20250811

基于集成学习与多目标优化的石灰石煅烧黏土水泥智能配比设计

杨嘉轩，王仕奇，卓峰谊，弓扶元

(浙江大学建筑工程学院，杭州 310058)

摘　要：石灰石煅烧黏土水泥(LC^3)以其低碳排放量和优异性能迅速成为新型胶凝材料的研究热点，但其性能常受到原料组分、处理工艺、制备方法等因素的影响，其中，基体配比是决定其性能的关键因素，基于此，本工作搜集整理了涵盖不同材料组分参数、制备条件、处理工艺的数据库，采用多种单一机器学习模型进行抗压强度预测，并通过贝叶斯优化优化各模型参数，提升模型性能，进而采用 Boosting 方法提升单一模型性能，得到性能良好的 LC^3 基体抗压强度预测模型，在此基础上，提出了结合抗压强度、成本与碳排放的多目标优化框架，并利用遗传算法获得 LC^3 基体的最优配合比。结果表明，该方法能够兼顾力学性能与可持续性，显著提升设计的科学性与工程适用性。本工作为新型低碳水泥材料的智能配比设计提供了有效思路。

关键词：石灰石煅烧黏土水泥；抗压强度；配合比设计；集成学习；多目标优化
中图分类号：TU528　　**文献标志码**：A　　**文章编号**：0454–5648(2026)03–0970–12
网络出版时间：2026–02–12



波特兰水泥是用于浇筑混凝土结构的关键基体，但其生产和应用会造成大量的碳污染和经济损耗[1]，故开发新型低碳水泥是实现碳中和可持续的重要技术途径。石灰石煅烧黏土水泥(LC^3)是一种可持续低碳复合型水泥，其采用石灰石和煅烧黏土取代部分波特兰水泥。石灰石中的钙元素可以激活煅烧黏土的活性，有效促进水泥的水化反应。除此之外，石灰石的煅烧极少产生 CO_2 且黏土的煅烧温度也低于波特兰水泥，可以降低能耗和 CO_2 排放量达 35%~40%[2]。相较硅酸盐水泥混凝土，LC^3 基混凝土在 LC^3 取代率不超过 50%时表现出相近的力学性能[3]。然而，LC^3 基混凝土在早龄期的强度发展趋势会低于硅酸盐水泥基混凝土，这是因为 LC^3 基混凝土的强度发展取决于 LC^3 中煅烧黏土的铝硅酸盐与石灰石中 $CaCO_3$ 等物质反应生成 C-S-H 凝胶的速度[4]。该过程受原料化学组成、颗粒分布及微观结构等多因素耦合影响，机理极为复杂。目前，针对 LC^3 组分配比的研究缺乏系统的理论指导与高效优化手段。 在人工智能迅猛发展的背景下，如何利用机器学习，深度挖掘组分–结构–性能间的内在规律，从而实现配比的精准、快速设计与性能预测，已成为突破 LC^3 材料研发瓶颈的关键问题。因此，有必要系统研究 LC^3 的材料组分配置比例。基于试验数据，学者们提出了基于机器学习的普通混凝土强度预测模型[5]，但这种基于部分参数回归的半经验半理论模型的适用性有限，无法充分评估多材料组分耦合作用机制对 LC^3 胶凝材料强度的影响。

机器学习方法因具备优异的非线性解耦和复杂系统评估能力，已被广泛应用于混凝土性能评估，涵盖抗压强度[6]、抗弯强度[7]、氯离子渗透[8]、碳化[9]等。然而，针对 LC^3 材料的性能预测尚处于起步阶段[10]，现有智能模型仅考虑预测精度，并未兼顾力学性能、碳排放与成本之间的博弈关系，无法充分发挥 LC^3 基混凝土的低碳优势。除此之外，单一机器学习模型在面对复杂数据集时存在精度有限等问题[11]，降低了模型的适用性。近年来，机器学习(ML)在混凝土配比设计中的应用已从单一的强度预测，发展到融合多种算法(如 BPNN、XGBoost)对工作性能、力学性能及耐久性的多目标精准预测(R^2 > 0.9)[12]。研究前沿进一步将 ML 预测模型与元启发式算法(如 NSGA-II、NSGA-III)相结合，以同时优化性能、成本及碳排放等多重目标[13]，显著提升了

收稿日期：2025–11–05。　修订日期：2025–11–28。
基金项目：宁波市“科创甬江 2035”重点研发计划(2024H019)。
第一作者：杨嘉轩(2001—)，男，博士研究生。
通信作者：弓扶元(1988—)，男，博士，教授。

Received date: 2025–11–05.　**Revised date:** 2025–11–28.
First author: YANG Jiaxuan (2001–), male, Doctoral candidate.
E-mail: yangjiaxuan@zju.edu.cn
Correspondent author: GONG Fuyuan (1988–), male, Ph.D., Professor.
E-mail: gongfy@zju.edu.cn

设计效率与材料的可持续性。然而，该领域的发展仍受限于高质量数据的稀缺、模型的可解释性不足以及复杂模型在工程实际中的泛化与应用壁垒。基于此，本工作提出了结合集成学习和多目标优化的 LC^3 水泥基材料性能预测和配合比优化设计模型，为该类低碳材料的高值化智能利用提供新见解。主要内容如下：1) 构建涵盖材料组分和养护条件的 LC^3 水泥基材料数据库，并采用相关系数法和协方差(VIF)法消除输入变量的多重共线性影响；2) 构建基于增强(Boosting)的集成学习模型以提高预测模型精度，并进行可解释性模型分析，分析变量的影响程度；3) 构建考虑抗压强度、成本及碳排放的 LC^3 混凝土多目标优化模型，采用遗传算法结合逼近理想解法确定最优的 LC^3 水泥设计方案。

1　集成学习算法与模型构建

1.1　多层感知机

多层感知机(MLP)是一种最基础的人工神经网络，被广泛应用于回归与分类任务。MLP 属于前馈型神经网络，由输入层、一个或多个隐藏层以及输出层构成，其核心机制是对输入加权求和后施加非线性激活函数，从而实现对任意复杂非线性映射的逼近。对于第 l 层的神经元 j，其激活值可表示为：

$$a_j^{(l)} = f\left(\sum_{i=1}^{n} w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}\right) \tag{1}$$

式中：$a_j^{(l)}$为前一层神经元的输出；$w_{ij}^{(l)}$为连接权重；$b_j^{(l)}$为偏置项；f为非线性激活函数。

1.2　随机森林

随机森林(RF)是典型的集成学习算法，其基本思想是通过 Bagging 框架将多个决策树模型进行组合，以降低方差并提升泛化性能。具体而言，随机森林在训练过程中对原始数据集进行自助采样，为每棵树构造不同的训练子集；同时，在每个节点分裂时仅随机选择部分特征参与决策，从而在样本和特征 2 个层面引入随机性。这种双重随机化机制保证了基学习器之间的差异性，是实现集成效能的关键。

在分类任务中，RF 通过多数投票确定最终结果；在回归任务中，则通过对各树的预测值取平均得到最终输出。依托这种机制，RF 相较于单一决策树具有更强的抗过拟合能力，并能在高维数据或含有噪声的数据集上保持稳定表现。此外，RF 还能提供特征重要性评估指标，为模型解释和变量筛选提供便利。其回归结果输出可以表示为：

$$H(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{2}$$

式中：$H(x)$为预测结果的平均值；$h_t(x)$为第 t 棵决策树的预测结果；T 为决策树的数量。

1.3　高斯过程回归

高斯过程回归(GPR)是一种非参数的贝叶斯学习方法，因其能够同时给出预测值与不确定性量化而被广泛应用于回归任务。高斯过程定义在函数空间上的概率分布，保证任意有限个函数取值均服从多元高斯分布。与假设固定参数化模型不同，GPR 将待拟合函数 $f(x)$ 建模为：

$$f(x) \sim GP\left(m(x), k\left(x, x'\right)\right) \tag{3}$$

式中：$m(x)$为均值函数，通常假设为零；$k(x,x')$为协方差函数(核函数)，用以刻画先验假设。

1.4　极端梯度提升

极端梯度提升(XGB)是基于梯度提升框架的高效集成学习算法，以其良好的计算效率、可扩展性和预测精度而被广泛应用。其基本思想是通过迭代拟合损失函数的负梯度不断生成新的回归树，以弥补前一阶段模型的不足，从而构建加性集成模型。对于给定的数据集 $D=\{(x_i,y_i)\}_{i=1}^{N}$，第 t 轮迭代的预测值为：

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k\left(x_i\right), \quad f_k \in F \tag{4}$$

式中：f_k 为第 k 棵回归树；F 为所有可能树的函数空间。

1.5　集成模型的构建与性能评估

基于上述 4 种基础模型，本节将构建集成学习模型，即基于 Boosting 的单一模型性能提升方法(图 1)，本方法通过将单一模型作为基模型并设置多个基模型对预测性能进行提升。本研究采用的集成方法属于一种误差迭代校正型 Boosting 框架，其思想与传统 Boosting(如 AdaBoost 或 Gradient Boosting)相似，均旨在通过多阶段模型的逐步优化减少总体预测误差。不同的是，本方法并未通过更新样本权重实现模型间的迭代，而是以先前模型的残差(误差项)作为后续模型的输入特征，实现预测偏差的逐步校正。因此，本方法在逻辑上是一种基于多层感知机(MLP)的串行 Boosting 集成方法。本方法的核心思想遵循 Boosting 范式，即通过序列化方式构建模型，后续模型致力于修正前序模型的预测误差。其与梯度提升机等经典 Boosting 方法的核心共性在于对“残差”的持续关注与修正，其主要区别在于基学习器的选择(使用 MLP 而非决策树)以及残差传递与利用方式

的具体实现。本工作将分别选择 MLP，RF，GPR 及 XGB 作为基础模型以探究单一性能提升的效果，将插补后的数据集输入到第 1 个基础模型中得到预测值，然后将实验值与预测值的误差作为下一个基础模型的输入。集成模型的输出预测是所有基础模型的预测值之和，这种方法可以有效地增强模型对累积误差的训练能力，因为后续模型会根据先前模型预测的真实值调整误差。

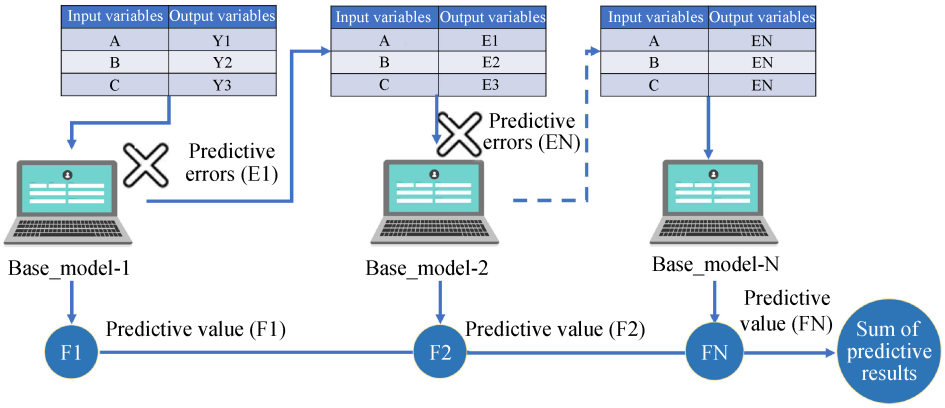


图 1 基于 boosting 的单一模型提升方法

Fig. 1 Single-model enhancement method based on boosting

2 数据库构建

2.1 变量选取

数据集的构建质量是控制机器学习模型预测性能的关键，故本工作在构建数据库时，选取的论文中都详细提供了 LC^3 水泥的材料组成，并包括用以确定材料质量的参数；选取的研究中所使用的煅烧黏土的化学组分均符合相关标准，且提供了黏土的煅烧条件等所需变量；砂浆抗压强度测试程序严格按照国际规范要求，不同尺寸试样抗压强度均按照相关规定进行折算；提供了试件水化和养护条件。基于此，本工作构建了 LC^3 水泥基材料的抗压强度数据集[4,14–23]，共包含 291 个数据样本，输入变量包括煅烧黏土含量(ω_{CL})；黏土煅烧温度(T_{CL})；颗粒比表面积(S_{CL})；煅烧时长(h)；石灰石含量(ω_{CC})；水泥含量(ω_{OPC})；水胶比(W/B)；养护温度(T)；养护湿度(R_H)；养护龄期(d)，输出变量为抗压强度(f_c)。同时，计算了各变量间的相关系数，如图 2 所示，皮尔逊相关系数是用于度量 2 个变量之间线性关系强度和方向的统计量。其值域在–1 到 1 之间，当皮尔逊相关系数为 1 时，表示 2 个变量完全正相关；当皮尔逊相关系数为–1 时，表示 2 个变量完全负相关；当皮尔逊相关系数为 0 时，表示 2 个变量没有线性关系。计算结果表明所选取的变量间独立性较好，只有配比关键参数间相关性较强，提高了模拟结果的可信度。表 1 为本工作所建数据库中各变量的分布指标。

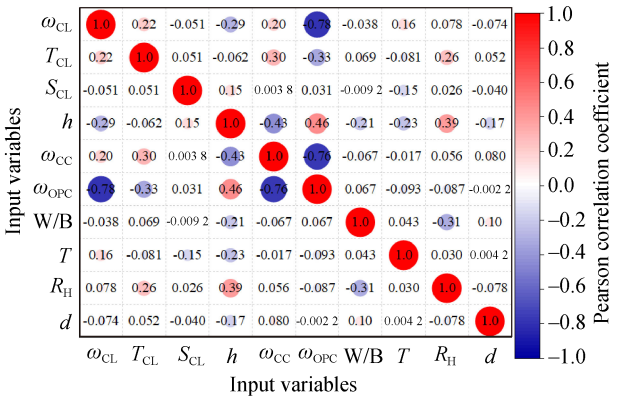


图 2 变量相关性系数热图

Fig. 2 Correlation matrix of variables

在机器学习建模过程中，多重共线性是影响模型稳定性和解释力的重要问题。当特征变量之间存在较强的线性相关性时，可能导致模型参数估计不稳定、特征重要性失真以及预测精度下降。为此，本工作引入方差膨胀因子(VIF)对数据进行处理。其基本原理是通过衡量某一特征对其他特征的可解释程度，进而评估该特征是否与其他变量存在显著多重共线性。VIF 的计算方法通常是将某一待检验特征作为因变量，其他特征作为自变量构建回归模型，根据回归的决定系数 R^2 计算得到：

$$\mathrm{VIF}_i = \frac{1}{1-R_i^2} \tag{5}$$

式中：R_i^2 为第 i 个特征对其余特征回归的拟合优度。

当 VIF>10 时，表明该变量存在较强共线性，应考虑剔除或进行降维处理[24]。计算得到各输入变

量的 VIF 值，如表 1 所示，各自变量的 VIF 值均小于 10，可用于机器学习模型训练。数据库中煅烧黏土含量(ω_{CL})、石灰石含量(ω_{CC})、水泥含量(ω_{OPC})的单位均为摩尔分数，为避免“定和约束”导致的强共线性，在计算 VIF 时，上述变量均按照绝对质量计算。同时，为了消除各输入变量间的量纲差异，本工作对数据集不同变量进行了归一化处理，按下式计算：

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{6}$$

式中：x_i 为归一化前的输入变量值；x_i^*为归一化之后的值；$x_{\max}$ 和 $x_{\min}$ 分别为输入变量的最大值和最小值。

表 1　数据库各变量的统计分布

Table 1　Descriptive statistics of database variables

Symbol	Unit	Num	Max	Min	Mean	SD	VIF
ω_{CL}	%	291	40.0	10.0	23.50	7.90	1.228
T_{CL}	℃	291	925.0	600.0	760.70	84.06	1.271
S_{CL}	m^2/g	291	45.7	2.5	18.50	7.30	1.047
h	h	291	3.0	0.2	1.27	0.75	1.946
ω_{CC}	%	291	31.1	0.0	7.50	7.70	2.581
ω_{OPC}	%	291	90.0	37.6	68.90	12.13	2.784
W/B		291	0.9	0.1	0.50	0.09	1.108
T	℃	291	50.0	5.0	22.60	6.00	1.148
R_H	%	291	100.0	80.0	92.30	5.30	1.519
d	d	291	270.0	1.0	30.70	41.30	1.057
f_c	MPa	291	75.0	5.0	39.20	16.60	

2.2　模型训练和超参数优化

将上述数据库按照 80%与 20%的比例划分为训练集和数据集，针对本工作数据量，按此比例划分能够避免欠拟合和泛化评估不可靠[25]。使用 MATLAB 软件训练模型，训练过程中对模型超参数进行优化，以提升模型性能。因此，调参算法成为优化过程中的关键环节。常见的调参方法包括：网格搜索，通过穷举预设的参数组合进行性能比较，适用于低维参数空间但计算代价较高；随机搜索，在参数空间内随机采样，能够在较低计算成本下获得接近最优的结果；进化算法(如遗传算法、粒子群算法等)，通过模拟自然选择过程在复杂空间中迭代寻优，适用于大规模非凸优化问题。相比之下，贝叶斯优化通过构建代理模型(如高斯过程)对目标函数进行建模，并基于采集函数平衡探索与利用，能够在有限计算资源下高效找到接近最优的超参数组合。

在本工作中，采用贝叶斯优化分别对多层感知机(MLP)、随机森林(RF)、高斯过程回归(GPR) 以及极端梯度提升(XGB) 的关键超参数进行自动化寻优。表 2 给出了各基模型和元模型的寻优超参数。XGB 的寻优参数中，Colsample_bytree 代表每棵树的特征子集采样比率，该参数决定了各学习器对数据选的比例；Gamma 定义了一个节点分裂所需的最小损失函数减少量，当模型的复杂度较高时，可以通过增加 Gamma 值来降低节点分裂的复杂度；Max_depth 代表学习器的最长路径，当 Max_depth 较大时会导致模型发生过拟合；N_estimators 代表学习器的数量，与 Max_depth 相似，N_estimators 越大，模型所需要迭代的次数就越多，模型的复杂度也就越高，进而导致模型过拟合；Subsample 代表学习器在迭代过程中选择样本的比例。RF 模型的寻优参数中，Min_samples_leaf 为每个数据数据节点所需的最小样本数，Min_samples_split 为每个数据树在进行分裂时所需的最小样本数。增大该值可以使数据树的深度变浅，复杂度降低，有助于减少过拟合。但是，如果设置过大，可能会导致模型无法捕捉到数据中的复杂模式，降低模型精度。N_estimators 是 RF 中数据树的数量。它决定了模型中包含多少棵决策树。每一棵树都会对样本数据进行独立的预测，最终的预测结果是所有树预测结果的平均值。MLP 的寻优参数中，Activation 为 MLP 所使用的激活函数，该函数将决定网络训练精度与效率；Alpha 是 L2 正则化项的系数，控制网络的权重衰减，可避免过拟合；Hidden_layer_sizes 为隐藏层数，控制模型的复杂程度，适度的隐藏层数将提高模型的预测精度。GPR 的寻优参数为 Kernel，参数范围在 Rbf、Rqk、Maternal 中选择。通过贝叶斯优化，不仅显著降低了超参数搜索的计算成本，同时提升了模型在预测任务中的稳定性与精度。

表 2 各机器学习模型的寻优参数

Table 2 Optimal parameters of different machine learning models

Algorithm name	Optimization parameters	Parameter range	Optimum values
XGB	Colsample_bytree	[0.5, 1.0]	0.6
	Gamma	[0, 5]	1
	Max_depth	[3, 10]	5
	N_estimators	[100, 1000]	100
	Subsample	[0.5, 1.0]	0.5
RF	Max_depth	[5, 100]	100
	Min_samples_leaf	[1, 20]	1
	Min_samples_split	[2, 20]	10
	N_estimators	[100, 200]	100
MLP	Activation	[ReLU, Sigmoid, Tanh]	Tanh
	Alpha	[1.0×10^{-6}, 0.1]	0.1
	Hidden_layer_sizes	[1, 100]	1
GPR	Kernel	[Rbf，Rqk，Maternal]	Rbf

2.3 模型性能评估

采用决定系数(R^2)、均方根误差(RMSE)、均方误差(MSE)及平均绝对误差(MAE)衡量 4 种单一模型的预测性能，各参数指标具体计算公式如式(7)~式(10)所示：

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2} \tag{7}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \tag{8}$$

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 \tag{9}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{10}$$

式中：y_i为数据集中第 i 个试验结果；$\hat{y}_i$为第 i 个预测结果；$\bar{y}$为所有试验数据的平均值；n 为样本数量。

考虑到集成学习模型可以组合不同的单独机器学习算法，本节将其作为构建元模型的依据。元模型的预测性能是决定集成模型预测性能的关键，本工作采用 MLP 作为元模型，在元模型的选择上，采用 MLP 而非线性回归模型，主要基于以下考虑。1)非线性映射能力。基学习器输出值与真实值之间往往存在复杂的非线性关系。MLP 具有较强的特征组合与非线性建模能力，能够有效学习基学习器间的隐含交互关系，从而提升最终预测精度；2)高维特征融合能力。在多模型集成中，各基学习器可能捕捉到不同层次的特征信息(如输入敏感性、局部规律、数据残差结构)。MLP 能够在隐层中自动学习这些特征的加权关系，实现信息的高维融合；3)稳定性保证措施。针对 MLP 训练可能出现的收敛不稳定问题，本工作通过以下方式提高其鲁棒性：采用早停机制防止过拟合，使用 Adam 优化器加速梯度收敛，并在输入层与隐层间引入 Dropout 正则化，以抑制噪声干扰。综合考虑模型性能与训练可控性，MLP 作为元学习器能够更好地捕捉复杂的误差修正规律，优于线性或基于树的简单学习器。这是因为 Boosting 的工作机制更容易对预测性能较弱的学习器进行强化。采用 MLP 作为基模型。在此基础上，将训练集带入基模型进行预测，并将预测结果作为输入变量代入元模型中进行训练，通过将数据集进行多指标的对比，可以确保集成模型的预测性能。

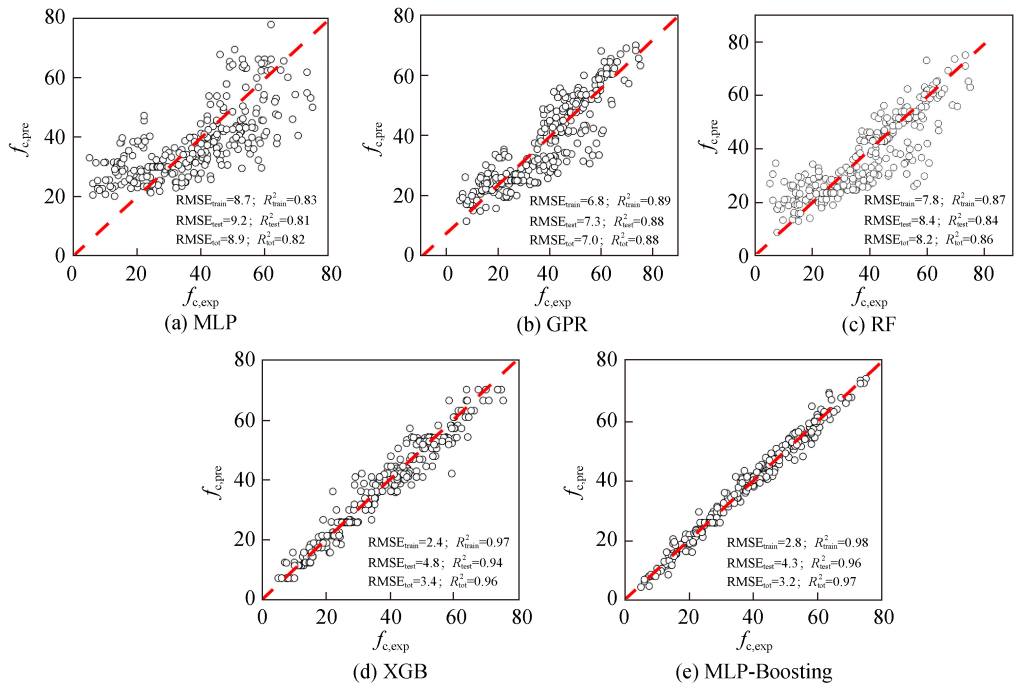
本工作在 MLP-Boosting 框架中引入了模型层与残差层的双重动态加权机制，以保持传统 Boosting 算法的自适应学习特征。对于模型层，根据各基学习器在验证集上的预测性能，为其分配动态权重系数，使得精度更高的基模型在最终融合中占比更大。对于残差层，在每一迭代轮次中，根据前一轮残差的绝对值大小为样本分配权重，使后续模型更关注误差较大的样本区域，从而在结构上实现对样本重要度的自适应调整。该双层加权机制在保持残差学习思路的同时，继承了 Boosting 的动态优化特征，提升了模型的稳定性与泛化性能。

各模型训练集、测试集及总数据集的训练结果如图 3 所示。由图 3 可见，梯度提升树模型在预测时表现最好，模型决定系数为 0.96，均方根误差为 1.84；多层感知机模型预测性能最差，但经过以其为元模型的 Boosting 性能提升之后，其预测性能超过梯度提升树单一模型，模型决定系数为 0.97，均方根误差为 1.76。

各个基模型的参数指标评估如表 3 所示。其中，XGB 表现最佳，说明其在处理非线性关系与复杂特征交互方面具有优势；GPR 与 RF 次之，显示了深度结构在非线性拟合中的有效性；而 MLP 则在整体性能上略逊一筹，但依然保持了较高的稳定性。

2.4 参数分析

基于上述过程，针对训练结果进行参数分析，分析过程采用 SHAP (SHapley Additive exPlanations) 方法，该方法通过计算每个特征对预测结果的贡献



$f_{c,exp}$ denotes the experimental compressive strength (MPa), $f_{c,pre}$ denotes the compressive strength predicted by the model (MPa).

图 3　MLP、GPR、RF、XGB、MLP-Boosting 的预测性能

Fig. 3　Prediction performance of MLP, GPR, RF, XGB, and MLP-Boosting

表 3　预测模型统计指标评估

Table 3　Evaluation of model performance metrics

Model	R^2	RMSE	MSE	MAE
MLP	0.82	2.99	8.9	2.73
RF	0.86	2.86	8.2	2.67
GPR	0.88	2.65	7.0	2.52
XGB	0.96	1.84	3.4	1.72
MLP-Boosting	0.97	1.76	3.1	1.68

度，可以忽略模型影响，衡量单次预测结果中的特征贡献，进而聚合局部结果成为对模型的整体解释。

通过分析所有样本的 SHAP 值与特征取值的关系，绘制依赖图(图 4a)，揭示模型对该特征的响应趋势。计算各特征 SHAP 值的绝对值的平均值，量化特征在所有样本中的平均贡献大小，通过排序可识别模型中最关键的驱动因素，得到特征重要性图(图 4b)。对于单个预测样本(图 4c 和图 4d)，SHAP 分解模型输出为各特征的贡献值之和,直观显示各特征对该样本预测结果的正负影响。另外，为了探明不同样本间的差异与不同输入变量间的相关性对输出结果的影响，对不同变量间展开依赖性分析，如图 5 所示。

2.4.1　输入变量对抗压强度预测的影响分析(全局依赖性分析)　从整体依赖性分析结果来看，变量 ω_{CL} (煅烧黏土含量)与 ω_{OPC} (普通水泥含量) 对模型预测输出的影响均呈现出显著的规律性趋势。ω_{CL} 在低取值区间对抗压强度表现为正向贡献，而在高取值区间逐渐转为负向作用，这是由于煅烧黏土含量低时，其提供的活性铝相与石灰石生成碳铝酸盐，有益于抗压强度；含量高时，水泥含量被过度稀释，导致硅酸盐相水化产物不足，抗压强度下降。因此，ω_{OPC} 呈现出相反的规律，低取值时多为负向影响，而在较高水平下贡献持续增强。这表明 2 类变量在全局层面上对预测结果存在明显的非线性关系和阈值效应。

与之相比，单样本解释结果进一步揭示了在特定预测案例下各变量的实际贡献情况。对于该样本，S_{CL} (比表面积)和 W/B(水灰比)是最主要的驱动因素，前者的细度特征抑制了预测值，而后者则显著提高了预测值。与此同时，T 和 T_{CL} 的正向贡献也较为突出，说明养护条件和胶凝材料总量对早期强度的提升具有重要作用。其余变量如 d (龄期)、ω_{CC} (石灰石)、ω_{OPC} (普通水泥含量)等则表现为次要影响。

2.4.2　输入变量之间的依赖性分析(变量间交互分析)　图 5 为各输入变量间依赖性分析。由图 5a 可见，ω_{CL} 与模型输出之间存在明显的非线性关系。

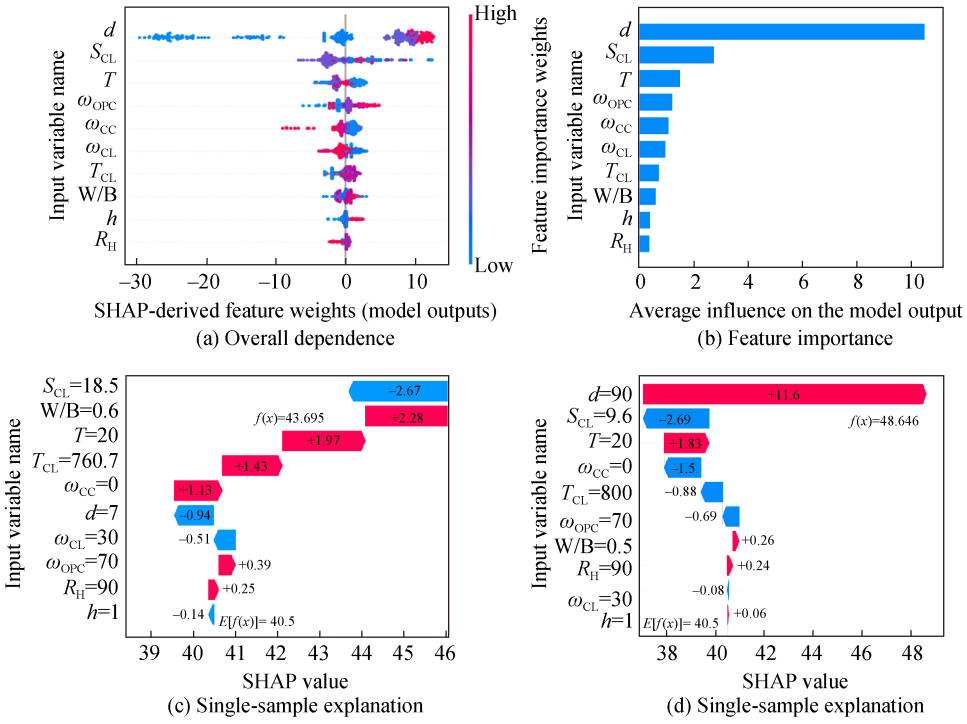
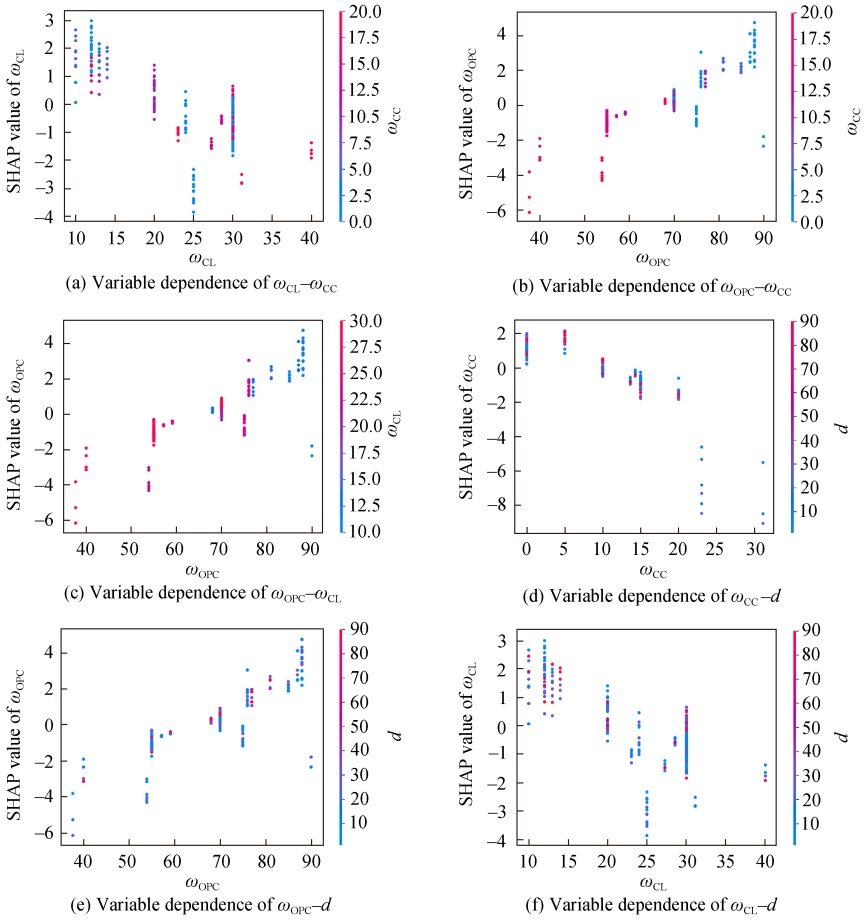


图 4 各输入变量对抗压强度预测结果影响程度

Fig. 4 Relative importance of input variables for compressive strength prediction results



In Fig. 5(a), the horizontal axis represents the values of *CL*, the vertical axis denotes the corresponding predicted compressive strength, and the color scale indicates the variation in *CC* (limestone content).

图 5 各输入变量间依赖性分析

Fig. 5 Dependency analysis among input variables

当 ω_{CL} 在 10~15 区间时，其 SHAP 值主要为正，说明此时 ω_{CL} 对抗压强度预测具有积极作用。相反，当 ω_{CL} 大于 20 时，其 SHAP 值多为负，表明其对预测抗压强度产生不利影响。此外，颜色分布表明 ω_{CC} 与 ω_{CL} 存在交互作用，高 ω_{CC} 取值会加剧 ω_{CL} 的负向影响。

从多变量 SHAP 依赖分析结果可以看出，输入变量在全局与交互层面均表现出显著的差异性作用。首先，d 与预测结果的关系具有明显的阶段性特征：早期龄期(0~28 d)对应的 SHAP 值大多为负，说明早期龄期对强度预测有削弱作用；而在中后期(>50 d)，SHAP 值逐渐转正，表明龄期增长显著增强了预测强度。此外，d 与 ω_{CL}、ω_{OPC} 的颜色分布呈现协同效应，即较高的胶凝材料用量能够进一步放大龄期的正向贡献。

其次，ω_{CC} 对预测结果总体上表现为负向作用，随着掺量的增加，SHAP 值逐渐下降，尤其在超过 20 时负效应更为显著，这是因为超过了与活性铝相反应生成碳铝酸盐的化学计量比，多余的石灰石仅作为惰性填料，体现为稀释效应，从而导致抗压强度增长停滞甚至下降。同时，ω_{CL} 与 ω_{OPC} 的分布色阶表明，它们能够调节 ω_{CC} 的影响程度：在较高 ω_{CL} 或 ω_{OPC} 含量下，ω_{CC} 的负面效应更加突出。

对于 ω_{CL} 与 ω_{OPC}，两者分别展现出相反的趋势：ω_{CL} 在低含量区间对抗压强度有正向贡献，但在高含量时转为负向；而 ω_{OPC} 则表现为随用量增加而 SHAP 值持续上升。这说明 2 类胶凝材料在预测结果中的作用存在互补性。进一步对比色阶分布可见，d 与 ω_{CC} 均能显著改变 ω_{CL} 与 ω_{OPC} 的贡献幅度，体现了变量之间的强交互性。

综上，全局依赖性与变量间交互分析共同揭示了混凝土配合比设计中的关键规律：龄期与水泥用量是主导性正向因素，而掺合料 ω_{CL} 与 ω_{CC} 的过量使用可能引入负向作用；同时，不同变量之间存在明显的协同或抑制效应，这些发现为配合比优化与性能预测提供了可解释性依据。

为了进一步揭示各输入变量对预测结果的影响规律及其与材料性能演化之间的内在联系，本研究在分析图 5 所示的特征重要性结果基础上，结合既有文献[26]中关于 LC^3 体系的材料学与化学演变规律进行深入讨论。LC^3(水泥–石灰石–煅烧黏土复合体系)的性能提升并非仅源于数据层面的变量相关性，而体现为材料组分间的多重协同效应，其核心机制包括：石灰石粉提供的成核与填充效应、煅烧黏土中的活性铝参与的火山灰反应、以及石灰石与铝相反应形成的碳铝酸盐相等。适量的石灰石可显著促进早期水化并改善孔结构，但过高的取代率会产生稀释效应，从而降低后期强度；该趋势与模型在高石灰石掺量区域出现的负向特征贡献高度一致。

煅烧黏土的反应性则主导了体系中 C-A-S-H 凝胶的生成与孔隙结构的细化过程，高活性黏土可持续消耗 $Ca(OH)_2$ 并形成高铝硅比的水化产物，促进致密化发展，从而在模型特征分析中表现出随龄期增长而增强的正向贡献。

除主要组分因素外，龄期、煅烧黏土颗粒比表面积以及养护温度对 LC^3 体系性能的发展亦具有显著影响。随着龄期增长，体系中 $Ca(OH)_2$ 的持续消耗与 C-A-S-H 凝胶的进一步生成，使孔隙结构逐步细化，强度及致密性不断提高。模型特征分析显示，龄期变量的影响权重在整体中占据主导地位，且其边际贡献曲线呈现出由快速增长到趋于平稳的阶段性特征，这与 LC^3 体系“早期成核加速—中期火山灰反应—后期致密化”三阶段水化规律一致。

煅烧黏土颗粒的比表面积直接决定其反应活性与离子传输速率。高比表面积的细颗粒可提供更多活性位点，加速铝相溶出与硅铝酸盐凝胶生成，从而增强早期强度与孔隙结构的致密化效应。然而，过高的比表面积也会增加需水量与黏度，会在拌和过程中产生分散性降低与气泡不稳定等不利影响。模型的特征依赖结果表明，比表面积存在一个最佳区间：在该区间内，反应促进效应与施工性能保持平衡，预测性能达到最优，这与关于高反应性煅烧黏土活性窗口的实验结论相吻合。

养护温度则通过影响水化反应速率与产物形貌，对 LC^3 体系的结构演化产生深远作用。适度升温(如 35~45 ℃)可显著促进碳铝酸盐及 C-A-S-H 凝胶的形成，使体系早期强度显著提高；但在较高温度条件下(>50 ℃)，过快的水化反应可能导致凝胶结构缺陷增加与后期收缩风险上升。模型结果显示，温度变量的 SHAP 贡献呈现明显的“正向—饱和—反转”特征，表明温度对性能的促进具有阈值效应，这一趋势与文献中关于 LC^3 热养护与后期致密性变化的实验结果保持一致。

3　LC^3 基体配合比优化设计

为实现 LC^3 基体配合比的综合性能优化，本工作将抗压强度、材料成本与碳排放量作为主要目标，构建多目标优化框架。抗压强度基于本文所建

立模型的预测结果，各材料成本、密度与碳排放量如表 4 所示[27]。研究流程包括目标函数定义、约束条件设定、最优解搜索及结果验证 4 个步骤，具体如下。

表 4 材料密度、成本与碳排放量

Table 4 Material density, cost, and carbon emissions

Material	Density/(kg·m^{-3})	Cost/(¥·t^{-1})	Carbon emissions/(kg CO_2·t^{-1})
Portland cement (42.5)	3100	500	912.00
Calcined clay	2600	600	230.00
Limestone powder	2700	100	15.77
Water	1000	3	0.00

1) 目标函数的建立

首先，定义决策变量 $\boldsymbol{x}$：

$$\boldsymbol{x}=\{r_{\mathrm{OPC}},r_{\mathrm{CL}},r_{\mathrm{CC}},\mathrm{w/b},B\} \tag{11}$$

式中：r_{OPC}, r_{CL}, r_{CC} 分别为水泥含量(质量分数)、煅烧黏土含量、石灰石粉含量；w/b 为水胶比；B 为总胶凝材料用量(kg/m^3)。

在此基础上分别建立以下 3 个目标函数：

$$\max F_1(\boldsymbol{x})=\hat{f}_c(\boldsymbol{x}) \tag{12}$$

式中：$\hat{f}_c(\mathbf{x})$ 为基于前序 MLP-Boosting 模型得到的抗压强度。

$$\min F_2(\boldsymbol{x})=\sum_i p_i q_i(\boldsymbol{x}) \tag{13}$$

式中：p_i 为第 i 种原材料的单价；$q_i(x)$为该材料在单位体积混凝土中的用量。

$$\min F_3(\boldsymbol{x})=\sum_i e_i q_i(\boldsymbol{x}) \tag{14}$$

式中：e_i 为第 i 种原材料的碳排放量；$q_i(x)$为该材料在单位体积混凝土中的用量。

上述 3 个目标共同构成多目标优化问题。

2) 构建约束条件

在本工作的优化框架中，决策变量聚焦于胶凝材料的核心组分，即水泥(ω_{OPC})、煅烧黏土(ω_{CL})和石灰石(ω_{CC})的配比。具体变量可按实际工程情况取值，本工作取煅烧粘土颗粒比表面积 18.5 m^2/g，煅烧温度 600℃，胶凝材料用量 380 kg/m^3。固定上述变量取值旨在构建一个清晰且可解的初始优化问题。若将所有变量同时作为决策变量，将使优化问题的维度和复杂性急剧增加，导致帕累托前沿难以收敛，且最优解集过于分散。其次，所固定的参数数值均代表当前工程实践与科学研究中的典型条件。例如，选定的煅烧温度(600 ℃)和煅烧粘土比表面积(18.5 m^2/g)是大多数研究中公认的、能够有效激活黏土火山灰活性并保证经济性的工艺窗口[26]。其余条件均按标准取值。同时为保证配合比的可行性和工程适用性，对材料组分与水胶比设置边界条件：

$$r_{\mathrm{OPC}}+r_{\mathrm{CL}}+r_{\mathrm{CC}}=1 \tag{15}$$

$$0.28\leqslant \mathrm{w/b}\leqslant 0.55 \tag{16}$$

3) 优化求解过程

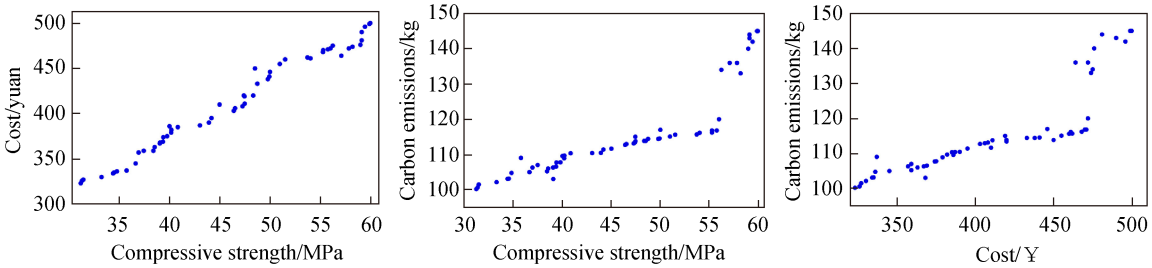
定义目标与约束后，本工作采用 NSGA-II 算法进行求解，具体过程包括：在可行域内利用拉丁超立方抽样生成初始种群；计算每个个体的抗压强度、成本与碳排放量；采用模拟二进制交叉与多项式变异进行个体进化；通过非支配排序与拥挤度比较确定种群迭代方向；以帕累托前沿稳定或迭代代数达到设定值为终止条件；得到强度优先、低碳优先、低成本优先及折中解等多类代表性最优解。

4) 结果验证

通过多目标优化分析，获得了若干帕累托最优解集。由图 6 可以看出，帕累托前沿分布较为均匀，随着混凝土强度的增大，成本和碳排放量相应增加。针对不同目标偏好，优化方案不同。帕累托前沿分类图如图 7 所示，若实际工程对经济性和环保性能要求较高，则选择 Ⅰ 区域非劣解集；若对抗压性能要求较高，则选择Ⅲ区域非劣解集；若均衡考虑抗压、成本及环保性能,则选择 Ⅱ 区域非劣解集。为了获得唯一的最优解，通常采用理想点法，即从帕累托最优解集中挑选一个平衡解，理想点由抗压强度、成本及碳排放量所对应的最优值构成。在确定理想点后，将最优帕累托前沿中的所有解代入理想点公式，获取到理想点距离最短的点,从而得到混凝土多目标优化的最优解。

在保证 28 d 抗压强度不低于设计要求的前提下，兼顾成本与碳排放量的综合最优解配合比如表 5 所示，该配合比的预测 28 d 抗压强度为 51.15 MPa。该预测结果与文献中的实验数据[28]能够较好吻合，如图 8 所示，其 28 d 龄期抗压强度值相对误差仅为 0.99%。相较于普通硅酸盐水泥降低了约 22%的材料

成本，同时减少了约 38%的单位体积碳排放。该结果表明，通过合理控制熟料比例并引入煅烧黏土与石灰石的协同效应，可以在保持较高力学性能的同时显著提升经济性与环境效益。



(a) Pareto frontier of compressive strength–cost　(b) Pareto frontier of compressive strength–carbon emissions　(c) Pareto frontier of cost–carbon emissions

图 6　帕累托前沿图

Fig. 6　Pareto frontier diagrams

表 5　最优配合比

Table 5　Optimal mix proportion

Portland cement	Calcined clay	Limestone	W/B	Compressive strength / MPa	Cost / ¥	Carbon emissions / kg
40%	40%	20%	0.4	51.15	390	565.44

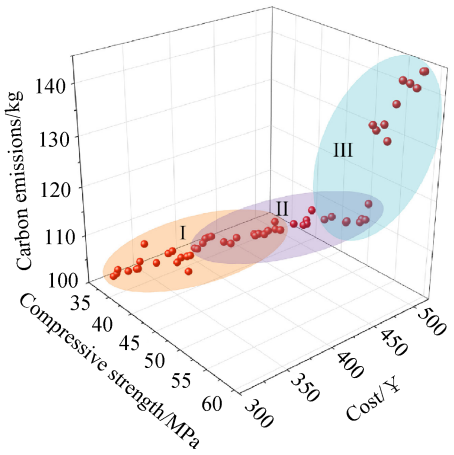


图 7　帕累托前沿分类图

Fig. 7　Pareto frontier classification diagram

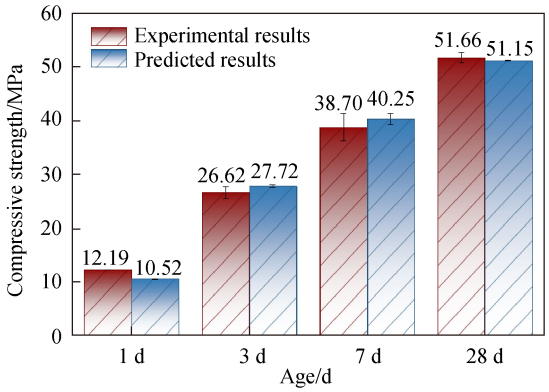


图 8　最优配比抗压强度结果与实验数据对比

Fig. 8　Comparison of optimal mix compressive strength results with experimental data

4　结论与展望

4.1　结论

1) 本工作建立的集成学习模型可以较好地预测 LC³ 基体的抗压强度，梯度提升树模型在预测时表现最好，模型决定系数为 0.96，均方根误差为 1.84；多层感知机模型预测时性能最差，但经过以其为元模型的 Boosting 性能提升之后，其预测性能超过梯度提升树单一模型，模型决定系数为 0.97，均方根误差为 1.76。

2) 变量标准化回归系数分析表明，龄期是影响 LC³ 基体的关键参数，其次是煅烧黏土颗粒比表面积和养护温度；从配比的角度而言，水泥用量、煅烧黏土用量及石灰石用量是影响 LC³ 基体抗压强度的最主要参数，梯度提升树模型和 MLP-Boosting 模型针对关键变量波动的变化趋势与已有试验结果是一致的，验证了 LC³ 基体强度预测机器学习模型的准确性。

3) LC³ 基体的最优配合比为：水泥用量 40%、煅烧黏土用量 40%、石灰石粉用量 20%、抗压强度达到 51.15 MPa，与实验数据相匹配，使用机器学习方法得出的最优配合比有较大可信度。

4) 结合 LC³ 预测机器学习模型和 NSGA-II多目标优化算法提出了 LC³ 基体配合比优化设计方法。该设计方法优化结果符合工程实际要求，显示出良好的工程应用潜力，可用于指导新型低碳黏土

基胶凝材料配合比设计。

4.2 展望

1) 数据样本量的限制可能对模型的泛化能力造成一定影响。因此，未来工作需通过多源数据集扩充与迁移学习策略提升模型的稳健性与适用范围。

2) 输入变量的构成存在一定简化。本工作主要以宏观配比参数与养护条件作为输入，而未纳入原料的化学与矿物学组成特征，后续研究可结合XRD、热重分析及化学滴定数据，以提升模型的物理可解释性。

3) 本工作聚焦于抗压强度的预测与短期性能解释，模型优化框架尚未考虑长期耐久性指标。

综上所述，尽管本研究基于可解释性集成学习方法揭示了 LC^3 体系性能预测的关键影响因素及其机理一致性，但仍存在若干局限性，有待在后续工作中进一步完善。

参考文献:

[1] 秦继辉, 钱觉时, 宋庆, 等. 磷酸镁水泥的研究进展与应用[J]. 硅酸盐学报, 2022, 50(6): 1592–1606.
QIN Jihui, QIAN Jueshi, SONG Qing, et al. J Chin Ceram Soc, 2022, 50(6): 1592–1606.

[2] HOSEN K, CHEN B. Limestone calcined clay cement (LC3): A review of materials, properties, production and environmental impact[J]. J Build Eng, 2025, 112: 113672.

[3] SHEKARCHI M, BONAKDAR A, BAKHSHI M, et al. Transport properties in metakaolin blended concrete[J]. Constr Build Mater, 2010, 24(11): 2217–2223.

[4] DHANDAPANI Y, SANTHANAM M. Assessment of pore structure evolution in the limestone calcined clay cementitious system and its implications for performance[J]. Cem Concr Compos, 2017, 84: 36–47.

[5] NGUYEN H, VU T, VO T P, et al. Efficient machine learning models for prediction of concrete strengths[J]. Constr Build Mater, 2021, 266: 120950.

[6] ANTONI M, ROSSEN J, MARTIRENA F, et al. Cement substitution by a combination of metakaolin and limestone[J]. Cem Concr Res, 2012, 42(12): 1579–1589.

[7] FERNANDEZ R, MARTIRENA F, SCRIVENER K L. The origin of the pozzolanic activity of calcined clay minerals: A comparison between kaolinite, illite and montmorillonite[J]. Cem Concr Res, 2011, 41(1): 113–122.

[8] MARAGHECHI H, AVET F, WONG H, et al. Performance of Limestone Calcined Clay Cement (LC3) with various kaolinite contents with respect to chloride transport[J]. Mater Struct, 2018, 51(5): 125.

[9] NUNEZ I, NEHDI M L. Machine learning prediction of carbonation depth in recycled aggregate concrete incorporating SCMs[J]. Constr Build Mater, 2021, 287: 123027.

[10] LAPEYRE J, HAN T H, WILES B, et al. Machine learning enables prompt prediction of hydration kinetics of multicomponent cementitious systems[J]. Sci Rep, 2021, 11(1): 3922.

[11] CANBEK O, XU Q Z, MEI Y J, et al. Predicting the rheology of limestone calcined clay cements (LC3): Linking composition and hydration kinetics to yield stress through Machine Learning[J]. Cem Concr Res, 2022, 160: 106925.

[12] 胡定宇. 基于机器学习和优化算法的混凝土性能预测及配合比设计[D]. 重庆: 重庆交通大学, 2022.
HU Dingyu. Performance prediction and mix design of concrete based on machine learning and optimization algorithm [D]. Chongqing: Chongqing Jiaotong University, 2022.

[13] 李鹏远, 牛豪爽, 刘毅豪. 基于深度学习和多目标优化的混凝土配合比设计及性能预测研究[J]. 水利水电技术(中英文), 2025, 56(4): 194–210.
LI Pengyuan, NIU Haoshuang, LIU Yihao. Water Resour Hydropower Eng, 2025, 56(4): 194–210.

[14] AKINDAHUNSI A A, AVET F, SCRIVENER K. The Influence of some calcined clays from Nigeria as clinker substitute in cementitious systems[J]. Case Stud Constr Mater, 2020, 13: e00443.

[15] LIN R S, LEE H S, HAN Y, et al. Experimental studies on hydration–strength–durability of limestone-cement-calcined Hwangtoh clay ternary composite[J]. Constr Build Mater, 2021, 269: 121290.

[16] DIXIT A, DU H J, PANG S D. Performance of mortar incorporating calcined marine clays with varying kaolinite content[J]. J Clean Prod, 2021, 282: 124513.

[17] KRISHNAN S, KANAUJIA S K, MITHIA S, et al. Hydration kinetics and mechanisms of carbonates from stone wastes in ternary blends with calcined clay[J]. Constr Build Mater, 2018, 164: 265–274.

[18] MISHRA G, EMMANUEL A C, BISHNOI S. Influence of temperature on hydration and microstructure properties of limestone-calcined clay blended cement[J]. Mater Struct, 2019, 52(5): 91.

[19] AVET F, SNELLINGS R, ALUJAS DIAZ A, et al. Development of a new rapid, relevant and reliable (R3) test method to evaluate the pozzolanic reactivity of calcined kaolinitic clays[J]. Cem Concr Res, 2016, 85: 1–11.

[20] MSINJILI N S, GLUTH G J G, STURM P, et al. Comparison of calcined illitic clays (brick clays) and low-grade kaolinitic clays as supplementary cementitious materials[J]. Mater Struct, 2019, 52(5): 94.

[21] ALUJAS A, FERNÁNDEZ R, QUINTANA R, et al. Pozzolanic reactivity of low grade kaolinitic clays: Influence of calcination temperature and impact of calcination products on OPC hydration[J]. Appl Clay Sci, 2015, 108: 94–101.

[22] LORENTZ B, ZHU H, STETSKO Y, et al. Feasibility study for calcined clay use in the southeast USA[C]// Calcined Clays for Sustainable Concrete. Singapore: Springer Singapore, 2020: 27–36.

[23] MACHNER A, ZAJAC M, BEN HAHA M, et al. Portland metakaolin cement containing dolomite or limestone–Similarities and differences in phase assemblage and compressive strength[J]. Constr Build Mater, 2017, 157: 214–225.

[24] 刘凯华, 郑佳凯, 谢维力, 等. 基于机器学习的再生混凝土配合比设计方法[J]. 湖南大学学报(自然科学版), 2023, 50(9): 88–96.
LIU Kaihua, ZHENG Jiakai, XIE Weili, et al. J Hunan Univ:Nat Sci, 2023, 50(9): 88–96.

[25] BICHRI H, CHERGUI A, HAIN M. Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets[J]. Int J Adv Comput Sci Appl, 2024, 15(2): 331–339.

[26] 董烨民, 钱雄, 胡传林, 等. 新型胶凝材料: 石灰石煅烧黏土水泥研究进展[J]. 硅酸盐学报, 2023, 51(9): 2446–2464.
DONG Yemin, QIAN Xiong, HU Chuanlin, et al. J Chin Ceram Soc, 2023, 51(9): 2446–2464.

[27] CHEN X, CHEN W Y, LI Z A, et al. A hybrid prediction and multi-objective optimization framework for limestone calcined clay cement concrete mixture design[J]. Sci Rep, 2025, 15(1): 22120.

[28] RUAN Y X, JAMIL T, HU C L, et al. Microstructure and mechanical properties of sustainable cementitious materials with ultra-high substitution level of calcined clay and limestone powder[J]. Constr Build Mater, 2022, 314: 125416.

Intelligent Mixture Design of Limestone Calcined Clay Cement Based on Ensemble Learning and Multi-Objective Optimization

YANG Jiaxuan, *WANG Shiqi*, *ZHUO Fengyi*, *GONG Fuyuan*
(College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058)

Extended Abstract

Introduction Limestone Calcined Clay Cement (LC^3) has attracted attention as a promising low-carbon cementitious material due to its reduced CO_2 emissions and satisfactory mechanical performance. However, its properties are strongly affected by raw-material composition, processing conditions, and mixture proportions. In this study, a database encompassing diverse material parameters and preparation processes was developed, and multiple machine-learning models were applied to predict the compressive strength of LC^3 systems. The predictive accuracy of these models was further enhanced through the Bayesian optimization and the Boosting-based ensemble strategies. Based on the optimized predictive models, a multi-objective optimization framework was proposed to integrate compressive strength, materials cost, and embodied carbon emissions. A genetic algorithm was subsequently applied to identify optimal mixture proportions for LC^3 matrices. The results showed that the proposed approach could achieve a balanced trade-off between mechanical performance and sustainability, thereby improving both the scientific robustness and practical applicability of LC^3 mixture design. This work could provide a data-driven pathway for performance-oriented optimization of emerging low-carbon cementitious materials.

Methods A database of 291 LC^3 mixtures was compiled, covering a wide range of compositions and curing conditions. Input variables included clinker, calcined clay, limestone, gypsum, water-to-binder ratio, calcination parameters, curing conditions, and curing age. The compressive strength served as an output. Correlation analysis and variance inflation factor (VIF) evaluation were applied to reduce multicollinearity. Four machine learning models (*i.e.*, Multilayer Perceptron (MLP), Random Forest (RF), Gaussian Process Regression (GPR), and Extreme Gradient Boosting (XGB)) were trained and tuned *via* the Bayesian optimization. An ensemble framework was then constructed with MLP as the meta-learner and XGB, RF, and GPR as base models within a Boosting strategy to enhance accuracy.

In parallel, a multi-objective optimization model was formulated, targeting compressive strength, material cost, and embodied carbon emissions. Cost and emission factors were drawn from market and life cycle databases. Practical constraints on binder fractions, water-to-binder ratio, and curing requirements ensured a feasibility. The non-dominated Sorting Genetic Algorithm II (NSGA-II) was used to search for Pareto-optimal mixtures that reflect different trade-offs among strength, cost, and carbon footprint.

Results and Discussion The ensemble learning notably improves prediction compared with individual models. The XGB provides the strongest performance among single algorithms, while GPR and RF also yield the good results. The MLP alone is weaker, but the predictive accuracy is further enhanced when integrating as a meta-learner in the Boosting ensemble. This underscores the advantage of combining diverse learners to model a complex LC^3 behavior.

The variable importance analysis identifies curing age as a dominant positive factor, followed by calcined clay fineness and curing temperature. In contrast, the excessive replacement of calcined clay and limestone has a negative effect, revealing threshold behaviors. The interactions among clinker, calcined clay, and limestone confirm synergistic benefits in moderate levels, but dilution effects at a higher dosage. These findings clarify the nonlinear mechanisms affecting LC^3 strength and validate the interpretability of the model.

The NSGA-II optimization generates a Pareto front of candidate solutions. A strength-prioritized mixture achieves mechanical properties, comparable to OPC concrete with a reduced carbon impact. A carbon-prioritized mixture minimizes emissions, while maintaining an acceptable strength, making it suitable for sustainability-driven projects. A balanced solution achieves compromise among all three objectives, representing a practical option for general engineering applications. The verification against experimental results confirms the reliability of predictions, demonstrating that the proposed framework can simultaneously address the performance, economic, and environmental targets.

Conclusions This study introduced an integrated framework that could combine ensemble learning with a multi-objective optimization for LC^3 mixture design. The main contributions involved, 1) Boosting-based ensemble learning significantly enhanced the accuracy of compressive strength prediction; 2) Curing age, calcined clay fineness, and curing temperature were identified as key positive drivers, while excessive calcined clay and limestone replacement exerted negative influences; 3) NSGA-II effectively produced Pareto-optimal mixtures balancing strength, cost, and carbon emissions, offering flexible solutions under different priorities; and 4) The experimental validation confirmed that optimized LC^3 mixtures could maintain competitive strength, while substantially reducing both cost and carbon footprint.

The integration of predictive modeling and optimization provided a powerful pathway for advancing LC^3 research and practice. It could contribute a methodological innovation and demonstrate a clear potential for sustainable, design of low-carbon cementitious materials in the construction industry.

Keywords limestone calcined clay cement; compressive strength; mixture design; ensemble learning; multi-objective optimization