https://doi.org/10.3799/dqkx.2026.261

岩石强度的可解释性预测：基于COA-LightGBM-SHAP框架

张 岩[1]，杨奇来[1]，罗 璟[1*]，刘亚斌[2]，简怀磊[2]，曾 鹏[1]，李天斌[1]

1. 成都理工大学地质灾害防治与地质环境保护全国重点实验室，四川成都 610059
2. 陕西兴通监理咨询有限公司，陕西西安 710061

摘 要：岩石单轴抗压强度(UCS)是表征岩体力学特性的核心参数，在工程稳定性分析中具有重要作用.然而，由于岩石材料的复杂性，高效且精确地确定UCS仍是一项重大挑战.本研究通过将Light Gradient Boosting Machine(LightGBM)与Coati优化算法(COA)相结合，开发了一个稳健的预测框架.基于包含273组岩石样本的综合数据库，利用COA系统地优化了LightGBM模型的超参数.研究结果表明，COA-LightGBM模型在岩石单轴抗压强度预测中表现卓越，其RMSE、R^2和MAE值分别为20.48、0.83和15.24.对比试验结果表明，COA-LightGBM模型在岩石单轴抗压强度预测上的性能显著优于最近邻(KNN)、决策树(DT)、随机森林(RF)等对比模型.此外，引入了Shapley Additive Explanations(SHAP)方法来阐明内部决策机制，并弥合“黑箱”模型与物理可解释性之间的差距.特征分析表明，孔隙率(n)是控制岩石强度的最关键因素.这项研究不仅证实了COA优越的优化能力，还为工程实践中岩石力学参数的快速评估提供了科学且可解释的解决方案.

关键词：岩石单轴抗压强度；机器学习；LightGBM模型；COA优化算法；SHAP特征分析；工程地质.

中图分类号：P642 **文章编号：**1000－2383(2026)09－3433－12 **收稿日期：**2026－01－26

Interpretable Prediction of Rock Strength: A Study Based on the COA-LightGBM-SHAP Framework

Zhang Yan[1], Yang Qilai[1], Luo Jing[1*], Liu Yabin[2], Jian Huailei[2], Zeng Peng[1], Li Tianbin[1]

1. *State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu* 610059, *China*
2. *Shaanxi Xingtong Supervision and Consulting Co., Ltd., Xi'an* 710061, *China*

Abstract: Uniaxial compressive strength (UCS) is a fundamental parameter for characterizing the mechanical behavior of rock masses which is essential for geotechnical stability analysis. However, the efficient and precise determination of UCS remains a significant challenge due to the complexity of rock materials. This study develops a robust predictive framework by integrating the Light Gradient Boosting Machine (LightGBM) with the Coati Optimization Algorithm (COA). Based on a comprehensive database of 273 rock samples, the COA is utilized to systematically optimize the hyperparameters of the LightGBM model. The results indicate that the proposed COA-LightGBM model exhibits exceptional predictive accuracy, achieving an RMSE of

基金项目：国家自然科学基金资助项目(No.42130719)；四川省自然科学基金资助项目(No.2025ZNSFSC0097).
作者简介：张岩(1988—)，男，研究员，主要从事地下工程地质灾害机理与防控研究.ORCID：0000-0002-4791-7081. E-mail：zhylr1988@126.com
*通讯作者：罗璟，ORCID：0000-0001-6746-1520. E-mail：luoj@cdut.edu.cn

引用格式：张岩，杨奇来，罗璟，刘亚斌，简怀磊，曾鹏，李天斌，2026.岩石强度的可解释性预测：基于COA-LightGBM-SHAP框架.地球科学，51(9)：3433－3444.

Citation: Zhang Yan, Yang Qilai, Luo Jing, Liu Yabin, Jian Huailei, Zeng Peng, Li Tianbin, 2026. Interpretable Prediction of Rock Strength: A Study Based on the COA-LightGBM-SHAP Framework. *Earth Science*, 51(9): 3433－3444.

20.48, an R^2 of 0.83, and an MAE of 15.24. Comparative analyses demonstrate that the COA-LightGBM model significantly outperforms conventional models such as K-Nearest Neighbors, Decision Trees, and Random Forests. Furthermore, the Shapley Additive Explanations (SHAP) method is introduced to elucidate the internal decision-making mechanism and bridge the gap between "black-box" models and physical interpretability. Feature analysis reveals that porosity (n) is the most critical factor governing rock strength. This research not only validates the superior optimization capability of COA but also provides a scientific and interpretable solution for the rapid assessment of rock mechanical parameters in engineering practice.

Key words: uniaxial compressive strength of rock; machine learning; LightGBM model; COA optimization algorithm; SHAP feature analysis; engineering geology.

0 引言

单轴抗压强度(UCS)是衡量岩石抗压性能的重要指标之一,通常被用来评估岩石在受到单向应力压缩时所能承受的最大应力,在岩土工程施工的稳定性分析中,UCS的运用极为广泛(李文和谭卓英,2016; Zhang *et al.*, 2026a).通过UCS,工程人员能够更为准确地判断岩石的工程性质,从而做出合理的决策(方灯明等,2001;李中等,2019).

在实际工程中,岩石的UCS通常是由单轴压缩试验来获取(Armaghani *et al.*, 2016; Zhang *et al.*, 2025a, 2025b).虽然通过试验可以高精度地测得UCS,但是试验的过程繁琐、测试周期长,不适用于大规模的工程现场(Wei *et al.*, 2022; Zhang *et al.*, 2026c).因此,工程技术人员需要便捷且高效的UCS获取方法.

人工智能技术的发展大力推动了机器学习在岩石力学研究中的应用,并且仍在不断深入(邓子昊等,2024;李海波等,2024; Zhang *et al.*, 2026b, 2026d).早期研究通过引入基础的机器学习模型(如人工神经网络(ANN)和最小二乘支持向量机(LSSVM)等)来拓展UCS的预测领域(Li and Tan, 2017).后续的研究中,基于组合模型发展而来的集成学习逐渐成为UCS预测的关键技术.该方法通过整合多个基模型的预测结果,来显著提升模型的整体性能.比如吴顺川等(2022)构建的融合了多项式回归、岭回归等七种基学习器的Stacking集成模型,结果表明其预测精度高,且与UCS实测值吻合度高.吴禄源等(2023)对多种机器学习方法进行对比研究后,发现LightGBM在UCS预测中泛化性能优秀且计算效率极高.

集成学习的UCS预测能力相较于传统机器学习有了显著提升,然而,其参数配置过程较为复杂,并且模型的参数优化难度会随数据规模的增大而急剧增加.为此,研究人员引入了优化算法来高效优化模型的超参数(Daviran and Ghezelbash, 2026).例如,Lei *et al.*(2022)采用群智能算法优化了BP神经网络的参数,其在提升模型预测精度的同时还加速了训练过程;Xie *et al.*(2024)利用麻雀搜索算法(SSA)优化了XGBoost的超参数配置,有效增强了模型的预测能力.此类优化算法可高效定位全局最优参数,使模型的综合性能相较于未调整的状态显著提升.

针对目前岩石UCS预测中存在的预测精度瓶颈、集成学习参数寻优困难以及模型黑盒特性导致的可解释性缺失等问题,本文的创新性工作主要体现在以下三个方面:(1)算法框架创新:首次将生物启发式的长鼻浣熊优化算法(Coati Optimization Algorithm, COA)引入LightGBM模型,利用COA独特的"全局捕食-局部逃逸"双阶段搜索机制,克服了传统优化算法易陷入局部最优的缺陷,实现了超参数的高效全局搜索;(2)预测鲁棒性提升:基于包含273组多源岩石样本的数据集,构建了具备强泛化能力的COA-LightGBM预测方法,并通过独立工程实例验证了其在复杂地质条件下的可靠性;(3)物理机制深度解析:突破了单纯的重要性排序,引入SHAP依赖图揭示了孔隙率等关键特征与强度的非线性响应规律,从岩石力学物理机制层面论证了数据驱动模型的合理性.

1 模型原理

1.1 LightGBM模型

LightGBM (Light Gradient Boosting Machine)是微软在GBDT基础上开发的高效梯度提升算法框架,其核心思想是利用加法模型和前向分步算法,通过迭代训练多个弱学习器(通常为决策树)并进行线性组合,从而构建出性能更强的预测模型(Chen *et al.*, 2022; Wang *et al.*, 2026).LightGBM的两项关键技术包括:叶子优先的树生长策略以及基于直方图的特征分裂算法.前者通过设置叶节点

最小样本数与最大深度,有效降低了过拟合风险;后者利用特征离散化与直方图统计,大幅减少了特征的遍历成本,使其在大规模数据环境中有更高的训练效率与更低的内存开销(Daniel *et al.*, 2025;张越等,2026).

LightGBM的目标函数定义如下:

$$\mathrm{Obj}=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k), \tag{1}$$

$$\Omega(f_k)=\gamma T+\frac{1}{2}\lambda /\!/ \omega /\!/ ^2, \tag{2}$$

其中,$l(y_i,\hat{y}_i)$为样本损失,f_k为第k棵树,γ和λ分别为叶节点数量与叶权重的正则化系数,T与ω为叶节点数和叶权重.该正则化项用于控制模型复杂度,从而提升泛化能力.

1.2　长鼻浣熊优化算法(COA)

COA是Dehghani *et al.*(2023)提出的群智能优化算法.其灵感来源于浣熊的捕食和逃离行为,特点是参数调整少且结构简洁,具有良好的搜索平衡性和较强的优化效率.其主要分为全局勘探和局部开发两个阶段.

在全局勘探阶段,COA模拟长鼻浣熊的捕食行为,并假定当前最优解为鬣鳞蜥的位置.种群前一半的个体模拟攀爬并发动攻击的长鼻浣熊,其位置不断向最优解移动,从而实现全局搜索:

$$x_i^{t+1}(j)=x_i^t(j)+r\cdot(x_{i_{\mathrm{best}}}^t(j)-I\cdot x_i^t(j)),i=1,2,\cdots,\frac{N}{2}, \tag{3}$$

其中,$x_i^{t+1}(j)$是第i只长鼻浣熊计算的新位置,$x_i^t(j)$是其的第j维,r是区间[0,1]内的随机实数,$x_{i_{\mathrm{best}}}^t(j)$是最优位置,$I$是从集合{1,2}中随机选取的整数.

另一半个体则在鬣鳞蜥落地后依据其随机落点来更新位置,并根据适应度进行方向选择,使搜索具有跳跃性和竞争性.这一机制避免了算法陷入局部极值.

局部开发阶段模拟的是长鼻浣熊逃离捕食者的行为.在该过程中,每个个体在其附近的局部区域生成候选位置,局部边界随迭代次数的增加逐渐收缩.位置更新方式如下:

$$lb_i^{\mathrm{local}}=\frac{lb_i}{t},ub_i^{\mathrm{local}}=\frac{ub_i}{t},\ t=1,2,\cdots,T, \tag{4}$$

$$X_i^{t+1}(j)=X_i^t(j)+(1-2r)\cdot(lb_i^{\mathrm{local}}+r\cdot(ub_i^{\mathrm{local}}-lb_i^{\mathrm{local}}))i=1,2,\cdots,N, \tag{5}$$

其中,$X_i^{t+1}(j)$是基于COA局部开发阶段计算出的第i只浣熊的新位置,$X_i^t(j)$是其第j维,r是区间[0,1]内的随机实数,t是迭代计数器,lb^{local}和ub^{local}分别是第j个决策变量的局部下界和局部上界,lb_i和ub_i分别是第j个决策变量的下界和上界.该机制可确保算法在后期执行更细致的局部搜索,加速收敛.

在两个阶段中,COA均采用贪婪选择机制确保每个个体只接受能使适应度改善的位置.总体而言,COA通过"全局攻击-局部逃逸"的双阶段策略有效平衡了探索能力与开发能力,从而获得稳定且高质量的优化性能.

相较于传统的群智能算法,COA的优势主要体现在其独特的"双阶段寻优机制".第一阶段模拟合作狩猎,保证了模型在复杂高维超参数空间(如LightGBM的8维参数空间)中的全局勘探能力,能够有效跳出局部最优解;第二阶段模拟逃离捕食者,通过逐渐收敛的局部边界实现精确开发.与PSO或GA相比,COA的控制参数更少,降低了调优的复杂性,且在处理非线性、非凸的UCS预测问题时表现出更强的稳健性和更快的收敛速度.

2　数据分析

岩石的单轴抗压强度通常受到多种因素的影响.然而,在工程应用中,受限于时间、环境与设备等约束,这些参数的获取往往依赖复杂的实验过程.因此,基于易获取参数的UCS预测方法在此类工程中的运用越发频繁.

基于上述背景,本文致力于构建一种基于易获取参数的UCS预测模型.为了确保模型训练的科学性和数据的代表性,本文收集并整理了273组岩石的UCS实验数据(Kılıç and Teymen, 2008; Dehghan *et al.*, 2010; Mishra and Basu, 2013; Mahmoodzadeh *et al.*, 2021).所有数据均来源于公开发表的学术研究,具有较高的可靠性和适用性.根据UCS的影响因素,输入特征包括点荷载指数($I_s(50)$)、施密特回弹值(SHR)、纵波速度(V_p)以及孔隙率(n)这四种便于获取且与UCS有较强相关性的岩石物理性质指标(Matin *et al.*, 2018).

表1详细列出了各输入指标的统计特征,其中包括平均值、标准差和各指标的数据按照升序排列时其位于最小值位置、位于25%数据量位置、中值位置、位于75%数据量位置和位于最大值位置的数据值,从数值层面上清晰地展示了数据的集中趋势、离散程度和范围.该数据集中沉积岩是主要岩石类型,其中有石灰岩(占数据集总量的22.0%,60个样本)和砂

表 1 数据集统计特征

Table 1 Statistical features of the dataset

类型	$I_s(50)$ (MPa)	SHR	V_p (km/s)	n (%)	UCS (MPa)
平均值	4.36	43.78	5.24	3.76	87.20
标准差	2.62	12.54	1.07	5.50	53.06
最小值	0.86	17.00	1.47	0.06	6.20
25% 数据值	2.86	30.30	4.78	0.31	37.89
50% 数据值	3.39	46.00	5.45	0.50	80.00
75% 数据值	5.59	54.18	5.92	6.66	131.00
最大值	14.13	67.07	7.94	37.81	239.00

岩(占数据集总量的19.0%,52个样本)等岩石类型,其次是花岗岩(占数据集总量的10.3%,28个样本).同时还包括板岩、白云岩和泥灰岩等少数岩性,确保模型能够反映各种不同的地质条件.该数据集包含6.20~239.00 MPa的岩石强度数据.每个特征属性覆盖了广泛的数值区间,体现出数据集中多样的岩石类型.

图1展示了数据集中各输入特征之间的相关性以及各输入特征的分布直方图.图中表明数据整体分布集中,大多数样本密集落在部分区间内.由于岩石种类的不同,部分特征的分布略有偏差,但这些偏差仍在研究允许的误差范围之内,并不会显著影响预测结果的准确性和泛化性.

3 模型预测与结果分析

3.1 COA-LightGBM模型的构建

为了解决集成学习模型的超参数优化困难的问题,本文选用LightGBM模型作为基准预测模型,并引入COA算法对其进行超参数优化.随机打乱数据集,并按8∶2的比例分割数据集,最终获得218组训练样本与55组测试样本.

首先,将218组训练集数据作为初始样本输入LightGBM模型,随后采用COA对模型关键超参数进行系统搜索.该算法通过长鼻浣熊合作狩猎鬣鳞蜥来进行模型超参数的全局搜索,通过长鼻浣熊逃离捕食者的行为来实现模型超参数局部优化.并基于不同参数配置下的模型性能动态评估,持续逼近最优超参数组合.图2展示了COA优化LightGBM模型的完整流程.

COA算法在规定范围内对参数组合进行系统搜索,每个超参数的搜索范围以及优化完成后获得的最优超参数如表2所示.其中的超参数主要分为两类,一类是模型的核心架构参数,包括树的数量(n_estimators)、树的最大深度(max_depth)、叶节点数量(num_leaves)和学习率(learning_rate),它们构成模型基础预测能力;另一类是模型的抗过拟合和

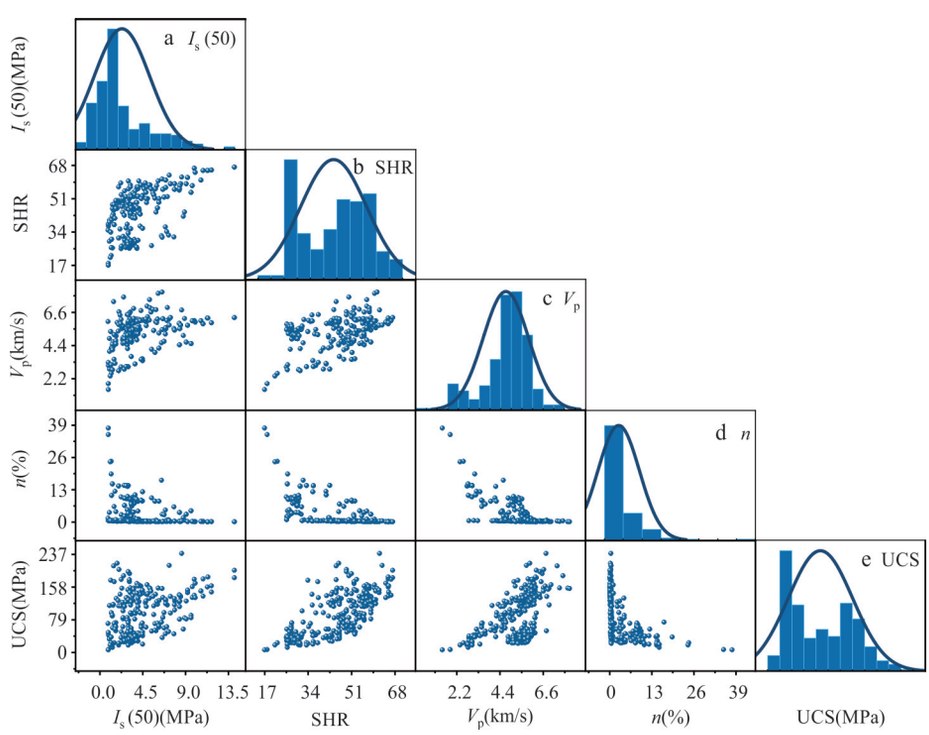


图1 数据的相关性散点矩阵及直方分布图

Fig.1 Scatterplot matrix and histograms of the dataset

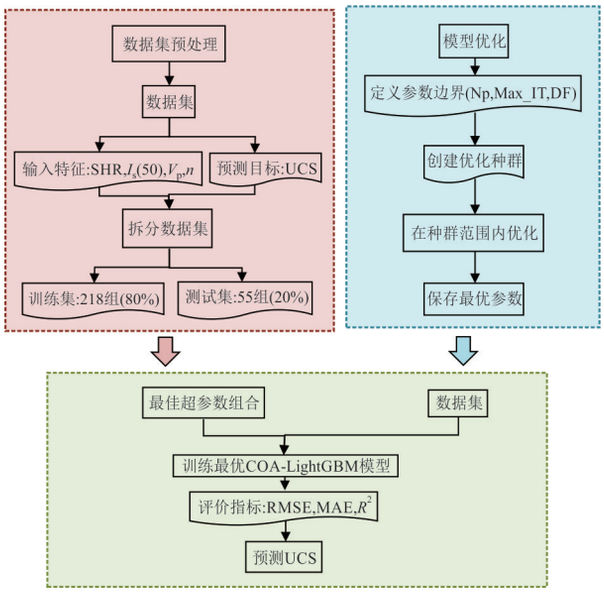


图2　模型优化及模型预测流程

Fig.2　Flowchart of model optimization and prediction

表2　超参数的范围和优化结果

Table 2　Range of hyperparameters and optimization results

超参数类型	数值优化范围	优化结果
n_estimators	(200,3 000)	219
max_depth	(3,15)	3
num_leaves	(20,300)	212
learning_rate	(0.001,0.15)	0.062 4
feature_fraction	(0.5,1.0)	0.713 9
bagging_fraction	(0.5,1.0)	0.716 5
lambda_l2	(0,50)	13.906 2
min_child_samples	(5,50)	18

随机化的参数,分别为特征采样率(feature_fraction)、数据采样率(bagging_fraction)、L2正则化项(lambda_l2)和叶节点最小样本数(min_child_samples),它们主要控制模型各种学习和分裂策略,保证模型的泛化性和过滤掉数据中噪声的能力,防止模型过拟合.

在模型优化过程中,COA算法的运行参数设定如下:种群规模(SearchAgents_no)设为20,最大迭代次数(Max_iteration)设为500.这些参数的选取旨在平衡计算成本与寻优深度,确保算法能够充分探索8维超参数空间.表2中所列出的超参数搜索范围(如n_estimators为200～3 000,learning_rate为0.001～0.150)主要参考了LightGBM的官方文档建议以及相关文献中关于中小型岩石数据集预测的经验法则,旨在确保模型在具备足够学习能力的同时,通过合理的正则化约束防止过拟合现象的发生.

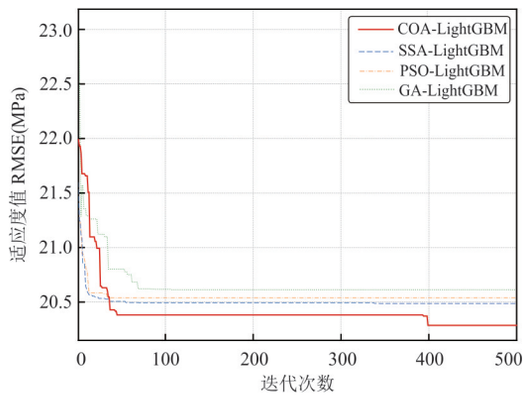


图3　优化算法适应度对比曲线

Fig.3　Fitness comparison curves for optimization algorithms

同时,本文还将优化完成后所得到的参数组合应用于LightGBM模型中,构建出优化后的COA-LightGBM预测模型,并对该模型进行保存,以便后续的研究及实验调用.

3.2　结果分析

为了客观评估COA算法的优越性,本研究引入了PSO、GA和SSA三种算法在相同种群规模(N=20)和迭代次数(T=500)下对LightGBM进行优化对比(图3).从收敛曲线可以看出,SSA虽然在初期下降迅速,但在约50次迭代后进入平台期,陷入局部最优.而COA算法虽然在初始阶段下降稍缓,但在50～100次迭代期间表现出极强的跳出局部最优的能力,其最终的RMSE稳定在所有对比算法中的最低水平.这证明了COA在平衡全局探索与局部开发方面的卓越性能,能够为岩石强度预测提供精度更高的参数组合.

为验证所构建的最优模型在预测UCS时的准确性,本文将测试集输入到COA-LightGBM模型中进行预测.同时还分别训练并测试了K最近邻(KNN)、决策树(DT)、随机森林(RF)和未优化的LightGBM四种常用机器学习模型,并与COA-LightGBM模型进行对比验证.模型的预测性能通过均方根误差(RMSE)、平均绝对误差(MAE)以及决定系数(R^2)三项常用的回归类评价指标进行评估.模型评价指标公式如下:

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}\ , \tag{6}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|\ , \tag{7}$$

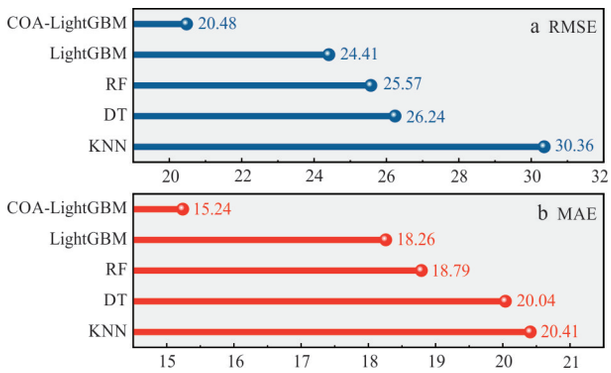


图 4 各模型的 RMSE、MAE 对比

Fig.4 Comparison of RMSE and MAE for different models

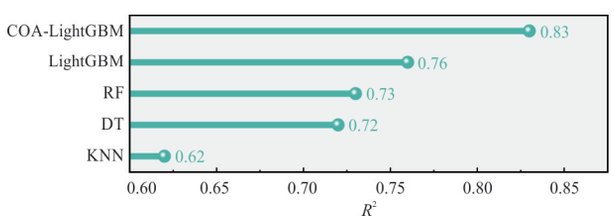


图 5 各模型的 R^2 对比

Fig.5 Comparison of R^2 for different models

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2}, \tag{8}$$

其中，n 是预测样本总数，y_i 是真实值，$\hat{y}_i$ 是预测值，$\bar{y}_i$ 是平均真实值.

图 4 和图 5 分别展示了各对比模型与优化后的 COA-LightGBM 模型在预测 UCS 时的误差评价指标情况.其中，指标 RMSE 和 MAE 表示模型预测的误差，其数值越小表明模型误差越小，R^2 表示模型拟合度，数值范围为 0～1，越接近 1 表明模型拟合度越高.

在使用 LightGBM 模型时，预测结果的 RMSE 为 24.41，MAE 为 18.26，R^2 为 0.76，表明存在一定程度的预测误差.而在使用 COA-LightGBM 模型时，预测结果的 RMSE 显著降低至 20.48，MAE 降至 15.24，R^2 则提高至 0.83.相较于原始模型，COA-LightGBM 模型在三个评价指标上均有所改善：RMSE 降低了 16.10%，MAE 减少了 16.53%，R^2 提升了 9.21%.这些变化表明，经过超参数优化后的模型在准确率更高的同时，还具有更高的拟合精度和更强的泛化能力.

最终的预测性能可视化对比结果汇总于图 6，可直观展示各模型在 UCS 预测中的表现.与图 5 所示的多模型预测结果对比不难看出，KNN 和 DT 模型预测能力明显不足，在图中体现为预测散点偏离中间拟合线较远.其次，集成学习算法 RF 和未优化调整的 LightGBM 算法在预测能力上相较于 KNN 和 DT 已经有了一定的提升，虽然还是一定程度上偏离拟合线，但偏离程度较小.而经过 COA 调整过超参数的 COA-LightGBM 模型的散点图更贴合拟合线，表明模型在本研究数据集上的表现显著优于 KNN、DT、RF 和 LightGBM 模型.

如图 6 所示，模型对极个别极高强度样本（UCS＞180 MPa）的预测存在一定滞后（预测值偏低）.这主要是由于：（1）岩石试验本身存在的离散性，极高强度样本通常具有复杂的微观胶结结构，其破坏模式具有极强的非线性；（2）样本集中极高强度样本占比相对较少，导致模型对该区间的演化规律捕捉不够充分.这为后续通过扩充极值数据集进一步优化模型指明了方向.

为了更全面、深入地评估所构建模型的预测性能与稳健性，本文在常规全局评价指标（RMSE、MAE 和 R^2）的基础之上，进一步引入了残差分析（Residual Analysis）.残差分析能够直观地揭示模型在不同预测区间的误差分布特征，是检验模型是否存在系统性偏差的重要手段.图 7 展示了 COA-LightGBM 模型在测试集上的残差综合分析结果.

图 7a 为模型的预测值与残差分布散点图.由图可知，测试集样本的残差整体呈现出随机且均匀分布在零误差线（红色水平虚线）两侧的特征，并未表现出明显的“喇叭状”发散或特定的非线性趋势.这一结果表明，COA-LightGBM 模型的预测误差不存在显著的异方差性（Heteroscedasticity）.换言之，无论是面对低强度（如 40 MPa 以下）还是高强度（如 140 MPa 以上）的岩石样本，该模型均能保持相对稳定的预测方差，未出现系统性的高估或低估现象，证明了其在全强度区间内的泛化可靠性.

图 7b 进一步展示了预测误差（残差）的频数直方图及其核密度估计（KDE）曲线.从图中可以清晰地观察到，误差的整体分布呈现出以 0 为中心、高度集中的近似正态分布（Normal Distribution）特征.绝大多数测试样本的预测误差被严格控制在［－20，20］MPa 的区间内.核密度曲线的峰谷极度贴近零误差线（红色垂直虚线），充分证明了经过 COA 优化的 LightGBM 模型具有优异的无偏估计能力.尽管数据集中存在极少数残差较大的离群点（如绝对误差超过 40 MPa），但考虑到岩石材料本身固有的强非均质性与地质成因的复杂性，这种程度的极值

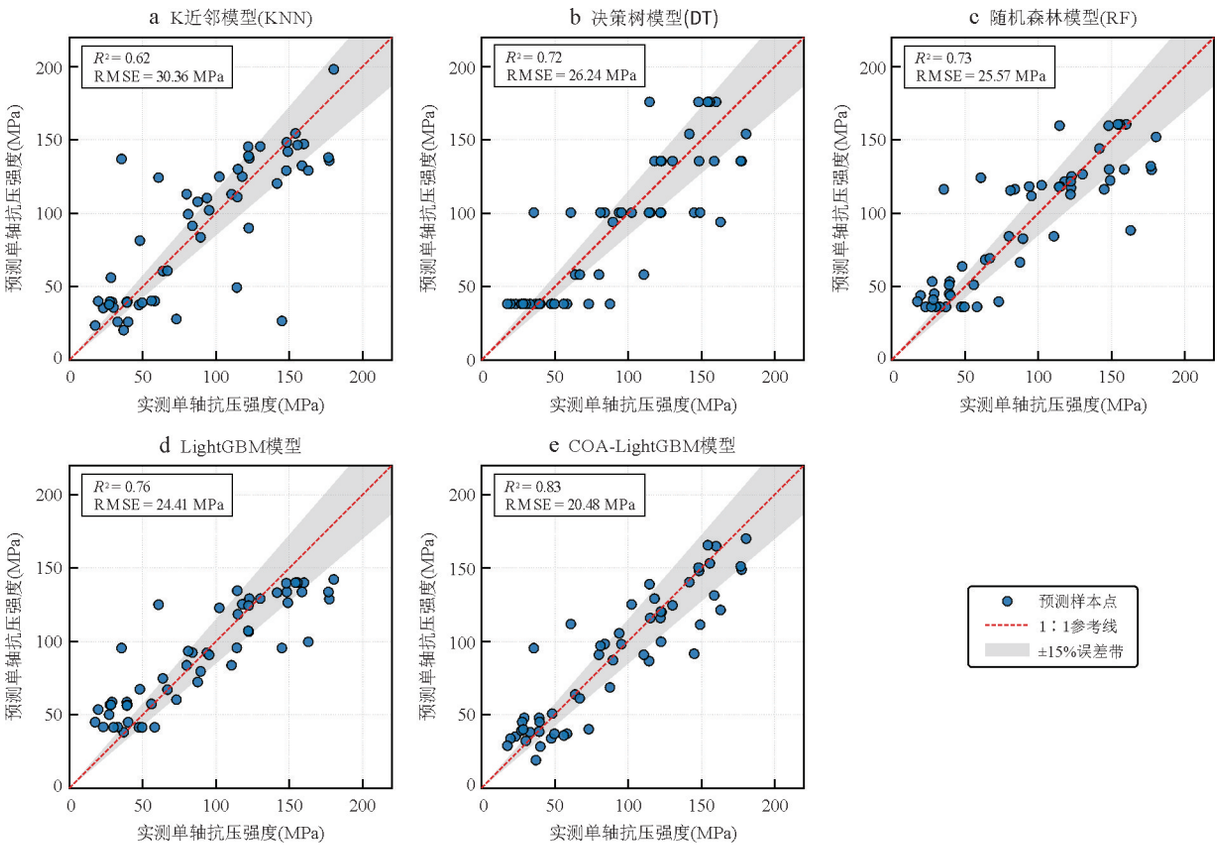


图 6　各模型预测散点对比

Fig.6　Scatter plots of predicted versus actual values for different models

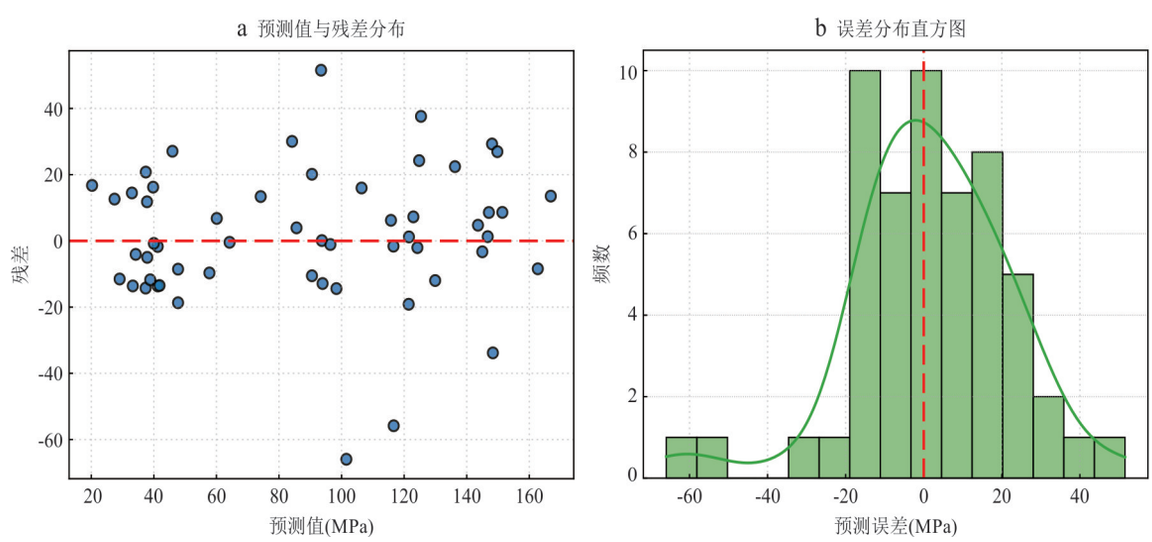


图 7　COA-LightGBM模型的残差分布图和误差直方图

Fig.7　Residual distribution plot and error histogram of the COA-LightGBM model

波动在岩石力学工程预测的合理容差范围之内.

综合上述残差分析结果,COA-LightGBM模型不仅在宏观统计指标上取得了最低的误差,其微观误差分布同样具备极高的稳健性,验证了模型内部决策逻辑的科学性.

以上结果表明,COA算法在超参数优化过程中有效提升了LightGBM模型在岩石单轴抗压强度预测方面的性能.该结果不仅验证了COA算法在岩石强度类回归任务中的实用性和适应性,也为类似工程地质问题的预测建模提供了一定的参考价值.

4 工程验证

为进一步验证所提出的COA-LightGBM模型在实际工程场景中的应用效果与泛化能力,本文引入了Ng *et al.*(2015)收集的花岗岩UCS数据集进行独立测试.该数据集共包含115组完整样本,涵盖了岩石在不同物理条件下的力学特性.其具体数值分布见表3.

数据集经标准化、随机打乱等预处理后,选择其中23组数据作为验证测试集.将验证测试集输入已保存的模型进行训练,相关的预测结果见图8和图9.

模型在该工程数据集上的预测结果为:RMSE=8.55,MAE=7.18,R^2=0.76.从评价指标及预测结果图像可以看出,COA-LightGBM模型在该外部验证数据集上表现出较强的预测准确性和稳定性,R^2达到0.76,表明模型能够较好地学习输入变量与UCS之间的非线性关系,具备较强的拟合能力.同时,RMSE与MAE的数值明显低于原始数据的波动范围,表明预测误差较低.验证结果体现了COA-LightGBM模型在实际岩石工程应用中具有优良的预测性能和良好的实用性,并具有一定的泛化能力,为该模型在更广泛地质条件下的应用提供了理论支撑与实践依据.

表3 验证数据样本的统计描述

Table 3 Statistical description of the validation data sample

类型	$I_s(50)$ (MPa)	SHR	V_p (km/s)	n(%)	UCS (MPa)
平均值	4.47	42.3	4.15	2.27	49.65
标准差	1.63	7.54	0.85	1.53	15.71
最小值	1.08	16.8	1.16	0.64	20.3
25%数据值	3.25	37.4	3.68	1.15	38.3
50%数据值	4.5	42.9	4.29	1.7	48.9
75%数据值	5.61	48.5	4.77	2.81	61.3
最大值	8.19	56.5	5.66	7.23	91

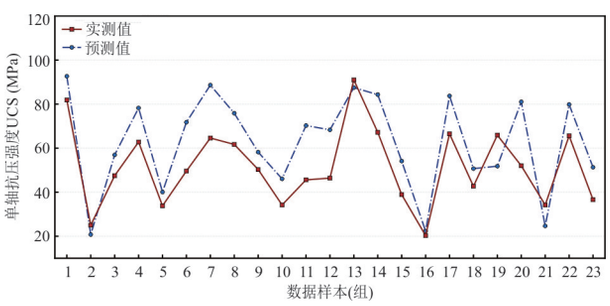


图8 预测值与真实值的折线对比图

Fig.8 Line chart comparing predicted and actual values

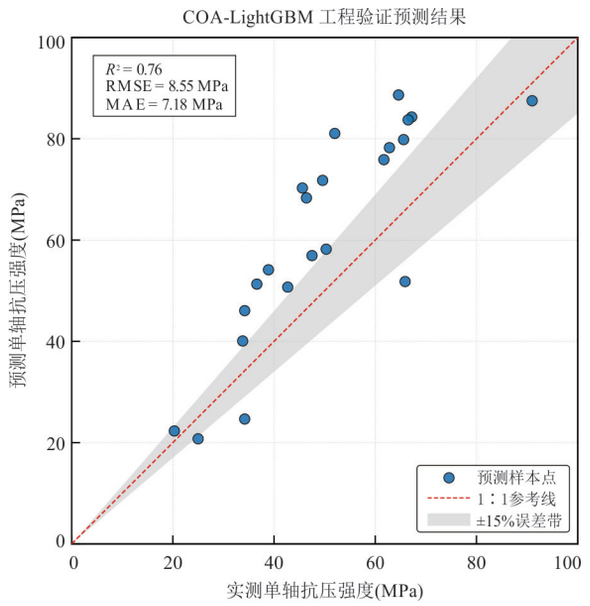


图9 工程验证数据的模型预测散点图

Fig.9 Scatter plots of predicted versus measured values for engineering validation data

5 讨论

5.1 SHAP特征重要性分析

本文采用的COA-LightGBM模型在预测精度方面表现优越,但受限于其"黑盒"模型的本质,其内部决策逻辑较难解释,无法直接反映输入特征对输出结果的影响机制.为了解决这一问题,本文引入了SHAP(SHapley Additive Explanations)解释方法(Chen and Seo, 2023).SHAP源于Shapley值理论,能够为复杂模型提供准确、可解释性更高的特征贡献度评估(陈忠元等,2025; Yu *et al.*, 2026).

图10展示了COA-LightGBM模型中各输入特征的SHAP值整体排序.从图中可以看出,n对预测结果的影响最大,属于模型中最关键的输入特征.具体而言,n数值越小,SHAP值越大,说明n数值越小,其对UCS的影响越大且其对UCS预测的贡献为正;反之,n数值越大,SHAP值越小,说明n数值越大,其对UCS的影响越小且其对UCS预测的贡献为负.这一结果反映了孔隙越少、岩石越致密,其强度越大的工程特性.相比之下,SHR和V_p对预测结果也有较为显著的影响,且二者数值越高,岩石强度越大.$I_s(50)$虽然在多数研究中被广泛采用,但在本文数据集中,其对UCS的相对贡献度略低,预测变化相对稳定,说明其在当前样本结构中的解释力有限,但仍具有一定的参考价值.

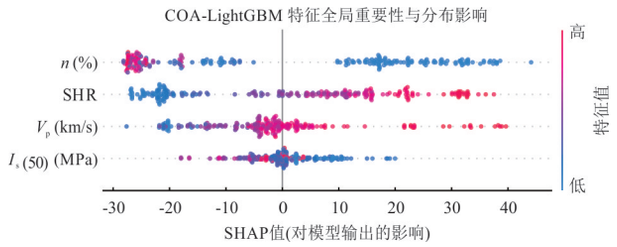


图 10　各输入特征的SHAP值排序

Fig.10　Feature importance ranking based on SHAP values

为了进一步揭示COA-LightGBM模型内部的决策逻辑,并验证其是否符合岩石力学基本物理规律,本文引入了SHAP特征依赖图对各输入特征进行深度物理机制剖析,如图11所示.

孔隙率(n)的微观损伤力学解释:如图11d所示,孔隙率n对UCS的预测具有决定性的非线性负向影响.从微观损伤力学角度来看,孔隙率代表了岩石内部原生微裂隙和孔洞等缺陷的分布密度.当n<2.5%时,岩石趋于极度致密,微裂纹较少,SHAP值呈现急剧上升的正向增益(最高可达30以上);而当孔隙率越过该阈值后,岩石的有效承载截面积减小,且在外部荷载作用下,孔隙和微裂隙尖端会产生强烈的应力集中效应(Stress Concentration),促使裂纹提前萌生、扩展与贯通.模型敏锐地捕捉到了这一力学衰减机制,表现为n>2.5%后SHAP值迅速下降并维持在负值区间,这与岩石强度劣化的物理常识完美契合.

回弹值(SHR)的表面硬度映射:图11b展示了施密特回弹值SHR与SHAP值的严格正相关线性趋势.SHR本质上反映了岩石基质的表面硬度与弹性反弹能力.图中显示,随着SHR增大,模型赋予其极强的正向贡献.值得注意的是,图中的颜色交互表明,处于高SHR区间的岩石样本往往伴随着极低的孔隙率(深蓝色散点),这证实了高强度岩石在宏观上表现为高表面硬度,在微观上表现为低孔隙特征的耦合物理属性.

纵波速度(V_p)的声学-力学特征耦合解释:图11c反映了纵波速度V_p对岩石强度的正向促进作用.在岩石弹性波理论中,声波在介质中的传播速

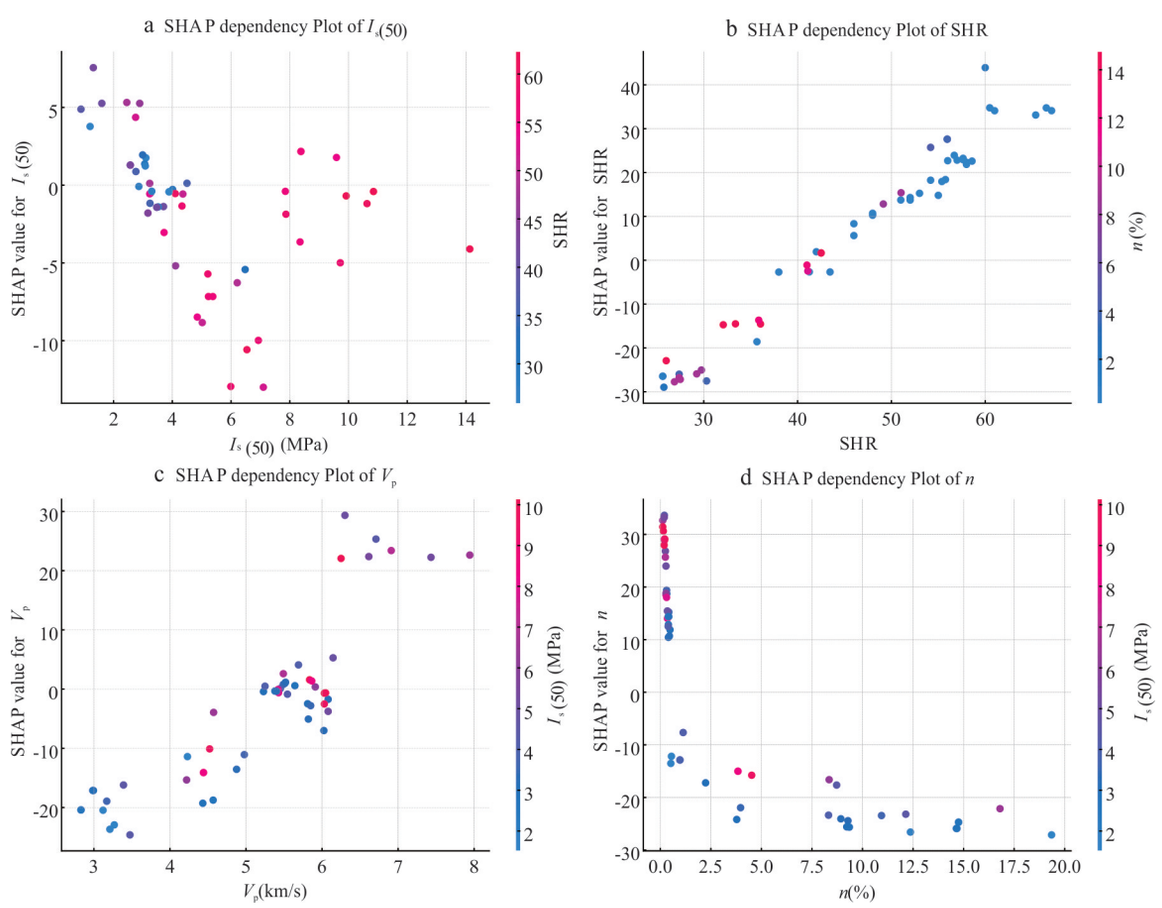


图 11　各输入特征的SHAP特征依赖

Fig.11　SHAP feature dependence plots for each input feature

度直接受控于岩体的固体介质连续性：岩体越完整、矿物颗粒胶结越致密，弹性波能量衰减越小，V_p越高．图中显示，当V_p>5.5 km/s时，其对强度的正向贡献出现显著跃升，这对应了完整坚硬岩石的高强度声学特征．

点荷载指数($I_s(50)$)的非线性分布特征：如图11a所示，相较于其他三个特征，$I_s(50)$的依赖图呈现出更为分散和复杂的非线性分布形态．这说明在当前包含多种岩性(沉积岩、花岗岩等)的复杂数据集中，单一的点荷载指数在表征全岩UCS时存在局限性，其转化系数受到其他变量的强烈干扰．这也从侧面解释了在图10的全局重要性排序中，$I_s(50)$的相对贡献度偏低的原因．

综合上述分析，COA-LightGBM模型并非单纯的数学拟合，其从数据中自主学习到的决策边界高度遵循了经典的岩石微观损伤与弹性力学规律，这赋予了该模型极强的工程可解释性与可靠性．

5.2 局限性

值得注意的是，本研究模型的优异性能建立在特定的数据集和实验条件下．实际工程中复杂多变的地质环境及岩石性质的不确定性，可能会影响模型的泛化能力．因此，在不同工程场景应用中，需对模型进行进一步验证与调整，以适应特定岩石样本，确保预测准确性．

在数据规模方面，本研究使用的273组样本虽然已经具有了一定的规模，并且一定程度上提升了模型的稳定性与泛化能力，但为了进一步提高模型的适用范围、稳健性和工程实用性，未来的研究仍需纳入更多种类、更广泛的工况、更大规模的数据集．同时，包含多种岩石类型的数据集又会使得模型的学习变得困难，模型在预测时可能会出现部分数据点误差较大的情况，要解决这类问题，就要求模型拥有更高的性能．为此，未来研究也应当考虑提升模型的性能从而提升预测能力，比如引入物理约束等．

COA-LightGBM模型在岩石UCS预测方面展现出良好的应用潜力．其进一步的优化与推广有赖于更广泛的数据积累与更高性能的机器学习模型．后续研究应聚焦于扩展数据集的规模和提升模型的预测性能，以推动岩石力学智能预测方法的发展．

6 结论

岩石的单轴抗压强度(UCS)是广泛应用于岩土工程中的表征岩体物理力学性能的重要指标，在工程稳定性分析中具有重要作用．然而，如何在保障精度的前提下便捷、高效地获取UCS值，一直是工程实践和学术研究中的重点难题．本文提出采用COA优化算法对梯度提升树模型(LightGBM)进行超参数的优化，构建出一套高效的UCS预测模型．研究得出以下结论：

(1)通过COA算法对LightGBM模型的超参数进行调整后，模型的RMSE降低了16.10%，MAE减少了16.53%，R^2提升了9.21%，模型性能显著提高．

(2)在KNN、DT、RF、LightGBM和COA-LightGBM模型上的测试结果表明，COA-LightGBM模型性能最佳，其评价指标为：RMSE=20.48，MAE=15.24，R^2=0.83，体现出较高的预测精度与稳定性．

(3)运用23组岩石强度数据对模型进行了工程验证，结果显示COA-LightGBM模型在实际岩石工程应用中预测性能优秀并且有一定的泛化能力．

(4)引入SHAP方法对模型进行了特征指标对UCS预测的影响程度分析．结果显示，n对UCS预测值的影响最大，SHR与V_p次之，$I_s(50)$影响最小但稳定．

本文所提出的COA-LightGBM模型不仅在预测精度上优于传统模型，还通过SHAP方法实现了模型结果的可解释性分析，为岩石强度的高效预测与工程设计提供了一种科学、可行的解决方案．同时，该模型在实际工程中具有较高的应用潜力，但在泛化能力、数据样本拓展等方面仍需进一步研究与优化．

References

Armaghani, D. J., Tonnizam Mohamad, E., Momeni, E., et al., 2016. Prediction of the Strength and Elasticity Modulus of Granite through an Expert Artificial Neural Network. *Arabian Journal of Geosciences*, 9(1): 48. https://doi.org/10.1007/s12517-015-2057-3

Chen, C., Seo, H., 2023. Prediction of Rock Mass Class Ahead of TBM Excavation Face by ML and DL Algorithms with Bayesian TPE Optimization and SHAP Feature Analysis. *Acta Geotechnica*, 18(7): 3825−3848. https://doi.org/10.1007/s11440-022-01779-z

Chen, J., Tang, H., Chen, W. K., et al., 2022. A Prediction Method of Ground Motion for Regions without Available Observation Data (LGB-FS) and Its Application to both Yangbi and Maduo Earthquakes in 2021. *Journal of Earth Science*, 33(4): 869−884. https://doi.org/10.1007/s12583-021-1560-6

Chen, Z. Y., Ren, T., Zhao, D., 2025. TPE-SVM Model and SHAP Analysis to Identify PB-ZN Deposit Types

Based on Sphalerite Trace Elements. *Earth Science*, 50(11): 4355－4369 (in Chinese with English abstract).

Daniel, C., Gao, F., Yin, X., et al., 2025. Exploring the Interpretability of Machine Learning Approaches in Modelling the Uniaxial Compressive Strength of Rocks. *Mining, Metallurgy & Exploration*, 42(4): 2471－2497. https://doi.org/10.1007/s42461-025-01282-5

Daviran, M., Ghezelbash, R., 2026. Detecting Low-Risk Exploration Targets from Automated Tuned Machine Learning Algorithms: Hybrid Particle Swarm Optimization, Grid Search and Genetic Optimization Algorithms with Support Vector Machine. *Journal of Earth Science*, 37(2): 843－869. https://doi. org/10.1007/s12583-025-0276-4

Dehghan, S., Sattari, G., CHEHREH CHELGANI, S., et al., 2010. Prediction of Uniaxial Compressive Strength and Modulus of Elasticity for Travertine Samples Using Regression and Artificial Neural Networks. *Mining Science and Technology* (*China*), 20(1): 41－46. https://doi.org/10.1016/S1674-5264(09)60158-7

Dehghani, M., Montazeri, Z., Trojovská, E., et al., 2023. Coati Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Optimization Problems. *Knowledge-Based Systems*, 259: 110011. https://doi.org/10.1016/j.knosys.2022.110011

Deng, Z. H., Zhang, L. M., Xu, X. H., et al., 2024. Evaluation for Broken Line Slope Stability Based on Ensemble Learning and LEM. *Earth Science*, 49(11): 4216－4224 (in Chinese with English abstract).

Fang, D. M., 2001. Main Problems and Determination of Foundation Strength of Pier of the Second Nanjing Bridge. *Earth Science*, 26(4): 406－409 (in Chinese with English abstract).

Kılıç, A., Teymen, A., 2008. Determination of Mechanical Properties of Rocks Using Simple Methods. *Bulletin of Engineering Geology and the Environment*, 67(2): 237－244. https://doi.org/10.1007/s10064-008-0128-3

Lei, Y., Zhou, S. T., Luo, X. D., et al., 2022. A Comparative Study of Six Hybrid Prediction Models for Uniaxial Compressive Strength of Rock Based on Swarm Intelligence Optimization Algorithms. *Frontiers in Earth Science*, 10: 930130. https://doi. org/10.3389/feart.2022.930130

Li, H. B., Li, X., Wang, S. J., et al., 2024. Cross Project Conversion Relationship of Key Parameters of TBM Rock Breaking. *Earth Science*, 49(5): 1722－1735 (in Chinese with English abstract).

Li, W., Tan, Z. Y., 2016. Comparison on Rock Strength Prediction Models Based on MLR and LS-SVM. *Mining Research and Development*, 36(11): 36－40 (in Chinese with English abstract).

Li, W., Tan, Z. Y., 2017. Research on Rock Strength Prediction Based on Least Squares Support Vector Machine. *Geotechnical and Geological Engineering*, 35(1): 385－393. https://doi.org/10.1007/s10706-016-0114-7

Li, Z., Zhang, Z. X., Yang, J., et al., 2019. Research and Application of Formation Pressure Monitoring while Drilling in Deepwater with High Temperature and High Pressure. *Earth Science*, 44(8): 2597－2602 (in Chinese with English abstract).

Mahmoodzadeh, A., Mohammadi, M., Hashim Ibrahim, H., et al., 2021. Artificial Intelligence Forecasting Models of Uniaxial Compressive Strength. *Transportation Geotechnics*, 27: 100499. https://doi. org/10.1016/j. trgeo.2020.100499

Matin, S. S., Farahzadi, L., Makaremi, S., et al., 2018. Variable Selection and Prediction of Uniaxial Compressive Strength and Modulus of Elasticity by Random Forest. *Applied Soft Computing*, 70: 980－987. https://doi.org/10.1016/j.asoc.2017.06.030

Mishra, D. A., Basu, A., 2013. Estimation of Uniaxial Compressive Strength of Rock Materials by Index Tests Using Regression Analysis and Fuzzy Inference System. *Engineering Geology*, 160: 54－68. https://doi. org/10.1016/j.enggeo.2013.04.004

Ng, I. T., Yuen, K. V., Lau, C. H., 2015. Predictive Model for Uniaxial Compressive Strength for Grade III Granitic Rocks from Macao. *Engineering Geology*, 199: 28－37. https://doi.org/10.1016/j.enggeo.2015.10.008

Wang, W., Li, Y. L., Wu, N. Y., 2026. Urgent Need of Submarine Landslide Risk Monitoring after 2025 Mega-Quake off Coast of Kamchatka. *Journal of Earth Science*, 37(3): 1513－1518. https://doi. org/10.1007/s12583-026-0047-x

Wei, M. D., Meng, W. Z., Dai, F., et al., 2022. Application of Machine Learning in Predicting the Rate-Dependent Compressive Strength of Rocks. *Journal of Rock Mechanics and Geotechnical Engineering*, 14(5): 1356－1365. https://doi.org/10.1016/j.jrmge.2022.01.008

Wu, L. Y., Li, J. H., Ma, D., et al., 2023. Prediction for Rock Compressive Strength Based on Ensemble Learning and Bayesian Optimization. *Earth Science*, 48(5): 1686－1695 (in Chinese with English abstract).

Wu, S. C., Wang, Y. C., Zhang, H. J., 2022. Study on Prediction Method of Uniaxial Compressive Strength of Rocks Based on Stacking Ensemble Algorithm. *Mining*

Research and Development, 42(6): 105－111 (in Chinese with English abstract).

Xie, H. H., Lin, P., Kang, J. T., et al., 2024. Prediction Method of Rock Uniaxial Compressive Strength Based on Feature Optimization and SSA-XGBoost. *Sustainability*, 16(19): 8460. https://doi.org/10.3390/su16198460

Yu, Y. L., Cai, H. C., Chen, C. W., et al., 2026. Reconstruction Modeling-Driven Analysis of Shale Oil Production Dynamics Using Explainable Machine Learning CatBoost-SHAP: A Case Study from the Bohai Bay Basin. *Journal of Earth Science*, 37(2): 923－927. https://doi.org/10.1007/s12583-026-0606-1

Zhang, Y., Ma, C. C., Zou, M. L., et al., 2025a. Thermo-mechanical Failure Analysis of Sandstone Subjected to High Ground Temperature. *Journal of Rock Mechanics and Geotechnical Engineering*, 17(6): 3524－3545. https://doi.org/10.1016/j.jrmge.2025.04.004

Zhang, Y., Meng, K., Fan, X. M., et al., 2025b. Advancements in Laboratory Studies of Layered Rock Masses for Deep Engineering: Insights and Future Perspectives. *Journal of Earth Science*, 36(3): 1334－1340. https://doi.org/10.1007/s12583-025-2032-1

Zhang, Y., Meng, K., Zhang, Z. L., et al., 2026a. Evolution of Characteristic Stresses in Deep Hard Rocks under True Triaxial Stresses: The TT-CVS Identification Approach. *Rock Mechanics and Rock Engineering*, Online. https://doi.org/10.1007/s00603-026-05364-8

Zhang, Y., Qiao, J., Li, T. B., et al., 2026b. An Improved LightGBM Model with ADASYN and Whale Optimization Algorithm for Rockburst Intelligent Prediction. *Rock Mechanics and Rock Engineering*, Online. https://doi.org/10.1007/s00603-026-05437-8

Zhang, Y., Zheng, X. S., Li, S. J., et al., 2026c. Instability Mechanisms of Deep Tunnels Induced by Stiff Discontinuities: Insights from Laboratory Comparative Tests and Numerical Simulation Analysis. *Tunnelling and Underground Space Technology*, 172: 107554. https://doi.org/10.1016/j.tust.2026.107554

Zhang, Y., Zhu, G. Q., Qiao, J., et al., 2026d. Dual-Modality Fusion with HRNet for High-Accuracy Surrounding Rock Classification in TBM Tunneling: Integrating Geological and Operational Parameters. *Tunnelling and Underground Space Technology*, 173: 107619. https://doi.org/10.1016/j.tust.2026.107619

Zhang, Y., Zhou, B. F., Guo, W. X., et al., 2026. Spike Waveform Recognition for Strong-Motion Records Based on LightGBM-SVM Stacking Algorithm. *Earth Science*, 51(1): 185－198 (in Chinese with English abstract).

中文参考文献

陈忠元, 任涛, 赵冻, 2025. 基于TPE-SVM模型和SHAP解释的闪锌矿微量元素特征识别铅锌矿床类型. 地球科学, 50(11): 4355－4369.

邓子昊, 张黎明, 徐兴华, 等, 2024. 基于集成学习和LEM的折线型边坡稳定性评价方法. 地球科学, 49(11): 4216－4224.

方灯明, 2001. 南京长江二桥南汊桥塔基主要工程地质问题与力学参数的确定. 地球科学, 26(4): 406－409.

李海波, 李旭, 王双敬, 等, 2024. TBM破岩关键参数跨工程转换关系. 地球科学, 49(5): 1722－1735.

李文, 谭卓英, 2016. 基于MLR与LS-SVM的岩石强度预测模型比较. 矿业研究与开发, 36(11): 36－40.

李中, 张祯祥, 杨进, 等, 2019. 地层压力随钻监测方法在深水高温高压井中的研究与应用. 地球科学, 44(8): 2597－2602.

吴禄源, 李建会, 马丹, 等, 2023. 基于集成学习与贝叶斯优化的岩石抗压强度预测. 地球科学, 48(5): 1686－1695.

吴顺川, 王艳超, 张化进, 2022. 基于Stacking集成算法的岩石单轴抗压强度预测方法研究. 矿业研究与开发, 42(6): 105－111.

张越, 周宝峰, 郭文轩, 等, 2026. 基于LightGBM-SVM堆叠算法的强震动记录尖刺波形识别. 地球科学, 51(1): 185－198.