

可解释性机器学习驱动的煤和污泥超临界水共气化产物预测研究

肖峰¹ 唐严² 马苗苗^{1,3} 安广晨¹ 张淑媛¹ 苏雷¹ 陈渝楠¹ 陈斌¹

(1. 西安交通大学绿色氢电全国重点实验室, 西安, 710049; 2. 国家电投集团四川电力有限公司, 成都, 610213; 3. 航空工业陕西航空电气有限责任公司, 西安, 710065)

摘要: 针对传统方法难以刻画小样本煤与污泥超临界水共气化实验数据中复杂非线性关系与多因素交互作用的问题, 本文构建了基于机器学习的产物预测与可解释性分析框架。采用留一法交叉验证评估随机森林 (RF)、极端梯度提升 (XGBoost)、梯度提升回归 (GBR) 和支持向量回归 (SVR) 对 H_2 、 CH_4 、 CO 和 CO_2 产率的预测性能, 并联合网格搜索和 Optuna 方法对模型超参数进行调优。结果表明, XGBoost 模型整体预测表现最优, 经 Optuna 优化后进一步提升 (平均 $R^2=0.828$, $E_{MAE}=0.614$, $E_{RMSE}=0.839$)。SHAP 特征重要性分析表明温度对气体产率总体贡献最大 (约 45%); 偏依赖图进一步揭示了实验参数对气化产物分布的非线性影响规律及其交互作用特征。本研究构建的机器学习框架适用于小样本超临界水气化实验数据, 可为煤与污泥共气化过程参数优化及产物调控提供数据驱动方法参考。

关键词: 煤与污泥; 超临界水共气化; 可解释性机器学习; 留一法交叉验证; 极端梯度提升

中图分类号: TK91 **文献标志码:** AA

文章编号: 0253-987X(XXXX)XX-0001-18

Prediction of Products from Supercritical Water Co-Gasification of Coal and Sludge Driven by Interpretable Machine Learning

Xiao Feng¹, Tang Yan², Ma Miaomiao^{1,3}, An Guangchen¹, Zhang Shuyuan¹, Su Lei¹, Chen Yunnan¹, Chen bin¹

(1. National Key Laboratory of Green Hydrogen and Electricity, Xi'an Jiaotong University, Xi'an 710049, China;

2. State Power Investment Group Sichuan Electric Power Co., Ltd., Chengdu 610213, China;

3. AVIC Shaanxi Aero Electric Co., Ltd., Xi'an 710049, China)

Abstract: To address the limitation of traditional methods in capturing complex nonlinear relationships and multi-factor interactions in small-sample experimental data from supercritical water co-gasification of coal and sludge, this study developed a machine learning-based framework for gasification product prediction and model interpretation. Leave-one-out cross-validation was employed to evaluate random forest (RF), extreme gradient boosting (XGBoost), gradient boosting regression (GBR), and support vector regression (SVR) models for predicting H_2 , CH_4 , CO , and CO_2 yields, while grid search and

收稿日期: XXXX-01-25。

Optuna were used for hyperparameter optimization. The results showed that XGBoost exhibited the best overall predictive performance, with an average R^2 of 0.828, MAE of 0.614, and RMSE of 0.839 after Optuna optimization. SHAP analysis identified reaction temperature as the most influential factor, contributing approximately 45% to the model output. Partial dependence plots further revealed the nonlinear effects of key experimental parameters on gasification product distribution and their interaction characteristics. The proposed framework is suitable for small-sample SCWG datasets and provides a data-driven approach for parameter optimization and product distribution regulation in coal and sludge co-gasification.

Keywords: coal and sludge; supercritical water co-gasification; interpretable machine learning; leave-one-out cross-validation; Extreme Gradient Boosting

在全球能源结构转型与“双碳”目标驱动的背景下,探索清洁、高效且可持续的氢能制备技术至关重要。国际能源署(IEA)报告显示,2024年全球氢能需求已接近1亿t,较2023年增长约2%^[1]。然而,目前超过99%的氢气生产仍依赖于天然气重整和煤气化等传统化石燃料工艺,存在碳排放高、能源转化效率低等问题,难以满足可持续发展需求^[2]。

超临界水气化(SCWG)技术以超临界水($T>374^{\circ}\text{C}$, $P>22.1\text{ MPa}$)为反应介质,能够将高分子有机质高效转化为富氢合成气^[3]。该技术具有传热传质速率快、有机物转化彻底、无需预干燥、无 NO_x/SO_x 排放、重金属固定效果好等显著优势^[4],特别适用于如农林生物质^[5]、微藻^[6]、塑料^[7]及污泥^[8]等多种有机废弃物的资源化处理,同时可实现废弃物处理和能源回收。

与单一原料相比,不同性质原料的共气化可能发挥协同效应,进一步提升气化效率、氢气产率并抑制焦油生成^[9]。Faki等^[10]在高粱与煤共气化研究

中发现,煤质量分数为25%时气化率和氢气产率均显著提高;Bian等^[11]在煤与塑料废物共气化中观察到,加入高氢碳比塑料可提升氢气和甲烷产率。基于共气化的协同潜力,Ma等^[12]将富含挥发分与碱金属的城市污泥与高热值的煤进行配对,探索了二者协同作用下的超临界水气化效果,并采用响应面法(RSM)工具建立预测模型。然而,受实验条件苛刻、成本高昂的限制,此类研究通常样本量有限,且参数间存在强非线性与交互效应,传统方法难以充分捕捉复杂关系,导致模型泛化能力和机理解释性不足^[13-14]。

近年来,机器学习(ML)方法在热化学转化领域迅速兴起,已成为处理预测产物分布并优化工艺参数的有效工具^[15]。相较于传统统计模型,ML算法在非线性和抗噪声方面表现出色。更值得关注的是,可解释性机器学习(Explainable ML)方法如SHAP、PDP等,能够打开模型“黑箱”,量化特征贡献、可视化边际效应、揭示参数交互,从而将数据驱动结论与反应机理相

结合,增强模型的可信度与科学价值^[16]。例如,Yan等^[17]比较多种ML模型对煤、生物质与废塑料共气化的预测性能,并采用SHAP量化蒸汽/燃料比对氢产率的正向贡献;Gil等^[18]基于30组生物质样本构建高斯过程回归预测模型,并通过可解释性技术揭示原料特性对气体产物的差异化影响。

尽管如此,针对煤和污泥超临界水共气化的小样本可解释性机器学习研究仍较少,缺乏系统比较多种回归模型并结合多维度解释的完整框架。本研究基于Ma等^[12]的实验数据构建多输出回归模型,通过留一法交叉验证(LOOCV)评估了4个主流模型:随机森林(RF)、极端梯度提升(XGBoost)、梯度提升回归(GBR)和支持向量回归(SVR)的预测性能,并使用Optuna自动调参优化XGBoost模型。最终应用SHAP和PDP技术解析温度、停留时间、原料浓度和混合比4个重要因素对4种气体产率的贡献、方向性和交互效应。本研究建立了关于煤和污泥超临界水共气化实验的机器学习“建模-优化-解释”一体化分析框架。

1 研究方案

1.1 数据来源与预处理

本研究所用数据集来自Ma等^[12]关于煤和污泥的超临界水共气化实验研究,数据集总共包含28组实验。训练模型的输入特征共包含4个关键操作参数:温度、原料浓度、停留时间、以及

混合比例(煤和污泥的质量比),输出目标为4个气体 H_2 、 CH_4 、CO和 CO_2 的产率。相关实验参数,见表1。

表1 参数及其值范围

Table 1 The parameters and the range of values

	参数	范围
输入	反应温度/ $^{\circ}C$	450~700
	停留时间/min	10~60
	原料浓度/wt%	5~20
	混合比例/%	0~100
输出	H_2 产率	0.45~14.85
	CH_4 产率	0.15~5.45
	CO产率	0.09~2.39
	CO_2 产率	0.70~10.30

气体产率单位mol/kg基于干燥后的原料总质量(干基)计算。

在模型训练前,对数据进行预处理以确保数值稳定性并提升模型收敛性。使用“scikit-learn”中的“StandardScaler”对输入特征进行标准化处理,使其均值为0、方差为1,计算公式如下

$$X = \frac{x - \mu}{\sigma} \quad (1)$$

式中, X 和 x 分别指归一化值、真实值; μ 和 σ 则分别对应各特征的均值和标准差。

1.2 模型选择与优化

为全面评估不同机器学习算法对SCWG过程的建模能力,本研究选择了4类具有代表性的回归模型:随机森林(RF)、极端梯度提升(XGBoost)、梯度提升回归(GBR)和支持向量回

归 (SVR)。RF 是一种集成学习方法, 通过构建多棵决策树并结合它们的预测结果, 具有较好的抗过拟合能力和对非线性关系的捕捉能力。XGBoost 是一种高效的梯度提升树算法, 通过顺序构建弱学习器并最小化损失函数, 具有优异的预测精度和计算效率。GBR 是一种通过迭代训练弱学习器 (通常是决策树), 在每一步沿损失函数负梯度方向拟合残差, 从而将多个弱学习器组合成强学习器的集成方法。SVR 基于结构风险最小化原则, 通过核函数将数据映射到高维空间并寻找最优超平面。

在模型开发过程中, 为追求更强的鲁棒性和更好的预测能力, 需要通过适合的优化策略调整模型的超参数。本文的模型优化策略为: 首先对 4 个模型执行网格搜索 (GS) 优化^[19], 在合理的搜索空间 (表 2) 内确定较优超参数配置。然后基于网格搜索确定的最优配置, 采用 Optuna 方法对筛选出的初始最优模型进行超参数自动优化进一步提升模型预测能力。本研究采用的 Optuna 方法基于 Tree-structured Parzen Estimator (TPE) 算法, 通过迭代采样和评估候选超参数组合, 智能搜索最优配置^[20-21]。

1.3 模型评估

为全面评估四个机器学习模型的预测性能, 选用 3 个评价指标: 决定系数 (R^2)、平均绝对误差 (E_{MAE}) 和均方误差 (E_{RMSE}): R^2 用于衡量模型对目标变

量总方差的解释程度, 其值越接近 1 表明模型拟合度越高; E_{MAE} 衡量了预测误差的平均偏差, 但对异常值不敏感; E_{RMSE} 通过计算误差的平方根, 能有效反映模型预测的稳定性以及对异常值的敏感程度。相关计算公式如下:

表 2 网格搜索的超参数搜索空间

Table 2 Hyperparameter search space of grid search

模型	超参数	搜索空间
RF	树的数量	[100]
	最大树深	[5, None]
	节点分裂最小样本数	[2, 5]
	叶节点最小样本数	[1-2]
	最大特征数	['sqrt']
XGBoost	树的数量	[100]
	学习率	[0.1]
	最大树深	[2-3]
	最小叶子权重	[2-3]
	样本采样比例	[0.8]
	特征采样比例	[0.8]
	L2 正则化系数	[1-2]
	最小分裂损失	[0]
GBR	树的数量	[100]
	学习率	[0.1]
	最大树深	[2-3]
	节点分裂最小样本数	[2, 5]
	叶节点最小样本数	[1-2]
	样本采样比例	[0.8]
SVR	核函数	['rbf']
	惩罚系数	[10, 50]
	核函数系数	['scale']
	ϵ -不敏感损失函数参数	[0.1, 0.2]

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - y_i^{\text{pred}})^2}{\sum_{i=1}^m (y_i - y_i^{\text{avg}})^2} \quad \#(2)$$

$$E_{MAE} = \frac{1}{N} \sum_{i=1}^m |y_i - y_i^{\text{pred}}| \quad \#(3)$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^m (y_i - y_i^{\text{pred}})^2} \quad (4)$$

式中, y_i 、 y_i^{pred} 和 y_i^{avg} 分别是第 i 个样本的实验值、预测值和对应实验值的平均值。

鉴于实验数据样本量有限,采用留一交叉验证 (LOOCV) 评估模型的预测能力。本文所用模型评估指标都表示为 R^2 (LOOCV)、 E_{MAE} (LOOCV) 和

E_{RMSE} (LOOCV)。图 1 为 LOOCV 评估方法示意图,以 28 次实验为例,LOOCV 每次使用 27 个样本训练模型,用剩下的 1 个样本测试,重复 28 次直至每个样本都被测试一次,得到的最终 LOOCV 性能指标是这些数值的平均值。虽然交叉验证方法的计算量大,但是能最大程度地利用有限数据,提供无偏的模型性能估计^[22]。

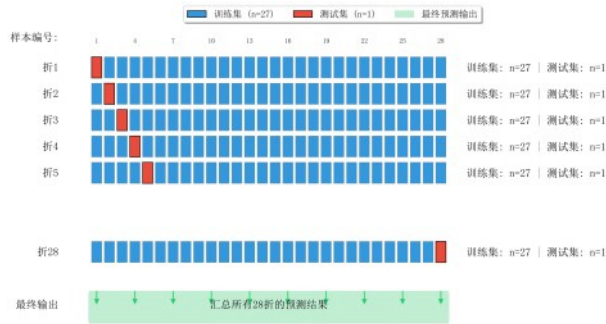


图 1 LOOCV 评估方法示意图

Fig. 1 Schematic diagram of LOOCV evaluation method

1.4 模型解释

可解释机器学习已成为机器学习研究的重要组成部分,特别是在科学和工程领域,理解模型决策过程与获得准确预测同等重要。本研究采用两种互补的可解释分析: □ SHAP (SHapley Additive exPlanations) 分析: SHAP 方法基于合作博弈论的 Shapley 值概念,提供了一种统一且理论完备的特征重要性量化方法。对于每个样本的每个特征,SHAP 值表示该特征对模型预测的贡献度^[23]; □ 偏依赖图 (Partial Dependence Plot,PDP) 分析: 部分依赖

图用于可视化一个或多个输入特征对预测结果的边际效应,适用于揭示特征与目标变量之间的非线性关系^[24]。这些方法共同确保模型预测与既定的气化热力学相符,弥补数据驱动代理模型与过程放大所必需的物理化学反应原理之间的差距。

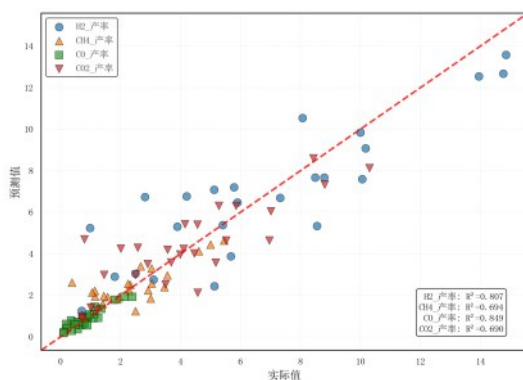
2 结果与讨论

2.1 模型性能比较

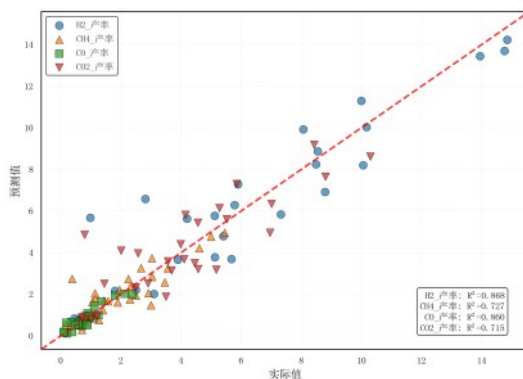
图 2 为 4 个模型气化产物预测产率与实际产率的对比图。可以看出,4 个

模型对各目标产率的预测点大多紧密聚集在45°线,这表明四个多输出回归模型都具备良好的预测能力。定量性能指标总结于表3,该表报告了每个模型-目标组合的 R^2 、 E_{MAE} 和 E_{RMSE} ,并附有各项指标的排名。XGBoost在多项指标上表现最优:对于 H_2 、 CO 、 CO_2 获得了最高的 R^2 ,分别为0.868、0.860和0.715,

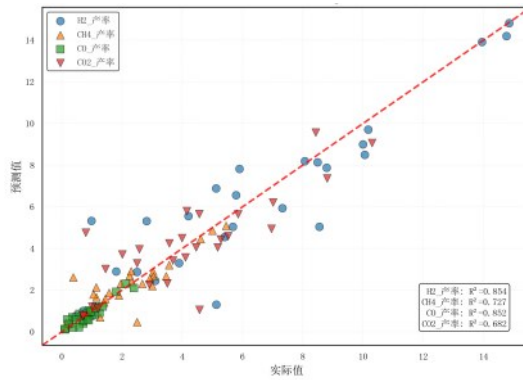
其平均 R^2 值较RF、GBR和SVR分别高出4.04%、1.64%和4.17%,预测性能表现最佳。并且XGBoost模型在三个预测目标中取得了最低的 E_{MAE} 和 E_{RMSE} ,其平均 E_{MAE} 和 E_{RMSE} 分别为0.700mol/kg和0.943mol/kg,这表明该模型的预测偏差更小。



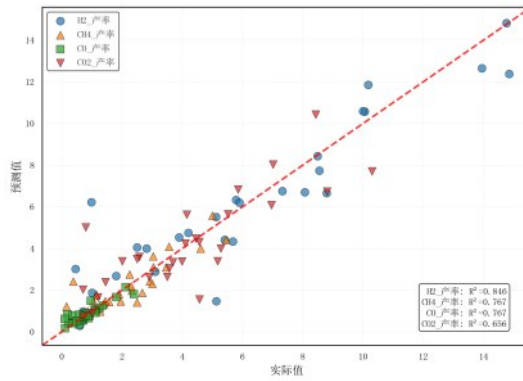
(a) RF



(b) XGBoost



(c) GBR



(d) SVR

图2 LOOCV评估四种机器学习模型气化产物预测产率与实际产率对比图

Fig. 2 Actual vs. predicted gas yields from four models evaluated by LOOCV

此外,为提供跨目标的综合排名,表4计算了每个模型-产物组合的分项排名之和。由表可知,三个模型对 H_2 的预测性能排名为 $XGBoost > GBR > SVR > RF$, CH_4 的排名为 $SVR > GBR > XGBoost > RF$, CO 的排名为 $XGBoost > GBR > RF > SVR$, CO_2 的排名为 $XGBoost > RF > GBR > SVR$ 。基于上述分析,四个模型对气化产物产率的综合预测性能排序为 $XGBoost > GBR > SVR > RF$ 。该结果与各模型的内在机

制及数据特征紧密相关。对于本研究这类高度非线性且特征间存在强交互的小样本SCWG过程数据, $XGBoost$ 和 GBR 表现更优的核心原因是顺序纠正残差的加法学习方式——每一棵新树都致力于弥补前一模型的不足,从而在小样本条件下反复挖掘数据中的有效信息,在训练过程中逐步提升模型的拟合能力^[23,25]。 $XGBoost$ 是 GBR 的优化算法,具有正则化项与二阶梯度优化,其表现更优调参也更复杂^[26]。相比之下, RF 需要足够多的树才能稳定,小样本

下每棵树看到的样本有限，可能无法充分学习特征间的复杂交互。SVR的核方法非线性建模能力强，但对参数调优和尺度敏感，因此在小样本下泛化能力可能较弱。

表 3 模型表现和排名
Table 3 Model performance and rankings

ML 模型	输出	R^2	R^2 排名	E_{MAE}	E_{MAE} 排名	E_{RMSE}	E_{RMSE} 排名
RF	H ₂	0.807	4	1.442	4	1.840	4
	CH ₄	0.694	3	0.587	4	0.753	4
	CO	0.848	3	0.167	2	0.213	3
	CO ₂	0.690	2	1.049	2	1.391	2
	平均	0.760	3	0.811	4	1.049	4
XGBoost	H ₂	0.868	1	1.081	1	1.522	1
	CH ₄	0.727	2	0.539	3	0.712	3
	CO	0.860	1	0.162	1	0.205	1
	CO ₂	0.715	1	1.018	1	1.334	1
	平均	0.792	1	0.700	1	0.943	1
GBR	H ₂	0.854	2	1.129	2	1.600	2
	CH ₄	0.727	2	0.488	1	0.711	2
	CO	0.852	2	0.172	3	0.211	2
	CO ₂	0.682	3	1.073	3	1.408	3
	平均	0.779	2	0.715	2	0.982	2
SVR	H ₂	0.846	3	1.182	3	1.645	3
	CH ₄	0.767	1	0.501	2	0.658	1
	CO	0.767	4	0.193	4	0.264	4
	CO ₂	0.656	4	1.098	4	1.466	4
	平均	0.759	4	0.743	3	1.008	3

2.2 模型优化

本节基于表现较优的 XGBoost 模型，利用 Optuna 进行超参数自动优化，以进一步提升模型的预测性能。优化目标为最大化在 LOOCV 评估下气化产物的平均 R^2 ，搜索过程中的内部 LOOCV 评估确保了稳健的泛化能力评估。综合考虑探索充分性与计算成本，共进行 100 次试验。针对 XGBoost 的 9 个关键

超参数（树的数量、学习率、最大树深度、最小叶子权重、样本采样比例、特征采样比例、L1 正则化、L2 正则化、最小分裂损失），为每个参数设定合理的搜索范围以平衡模型复杂度与泛化性能^[20]。

图 3 展示了 XGBoost 模型的超参数优化过程。图 3（a）为 Optuna 优化过程中的收敛曲线。由图可知，模型性能在优化初期提升较快，说明 Optuna 算

法能够在较少试验次数内快速定位较优的超参数区域。试验次数达 20 次后,最优 R^2 提升幅度逐渐减小并趋于稳定,表明优化过程已较充分地探索了预设搜索空间,所得最优超参数组合具有较好的稳定性和可靠性。

表 4 各模型对气体产率的预测力排名
Table 4 Rankings of model predictive performance for gas yield

ML 模型	输出	排名之和	总排名
RF	H ₂	12	4
	CH ₄	11	4
	CO	8	3
	CO ₂	6	2
XGBoost	H ₂	3	1
	CH ₄	8	3
	CO	3	1
	CO ₂	3	1
GBR	H ₂	6	2
	CH ₄	5	2
	CO	7	2
	CO ₂	9	3
SVR	H ₂	9	3
	CH ₄	4	1
	CO	12	4
	CO ₂	12	4

图 3 (b) 展示了最终获得的最优超参数组合。由于各超参数的量纲和取值范围不同,图中采用归一化搜索位置表示各最优参数在其搜索区间中的相对位置,并在条形末端标注对应的真实最优取值。结果表明,优化后的 XGBoost 模型整体表现出较浅树结构、适中学习率和弱正则化的特征。较浅的树深有助于降低模型复杂度,避免小样本条件下出现过拟合;适中的学习率则有利于在

模型收敛速度和预测稳定性之间取得平衡;较低的正则化与分裂损失参数说明,在当前数据规模和特征条件下,过强的约束可能会削弱模型对复杂非线性关系的学习能力。

性能提升的定量总结见表 5,优化后的 XGBoost 模型平均 R^2 达到 0.828,平均 E_{MAE} 降低了 12.3%,平均 E_{RMSE} 降低了 11.0%。性能提升在四个气体产率目标上保持一致,其中 H₂ 的提升最为显著,CH₄ 次之,CO 和 CO₂ 则得到较小的提升。图 4 展示了优化后 XGBoost 模型的预测产率与实际产率对比图,四种产物的预测数据点都紧密地聚集在 45° 线周围,与基线模型结果相比提升较小所以直观差距不大。

模型优化的结果显示,虽然在性能提升绝对值上较为有限,但考虑到数据集规模小且基线性能已经较强,这些提升仍具有意义。最终优化的 XGBoost 模型为后续应用可解释分析进行机理解释奠定了坚实基础,这是揭示特征重要性和交互效应反映真实物理化学关系而非数据噪声的关键。

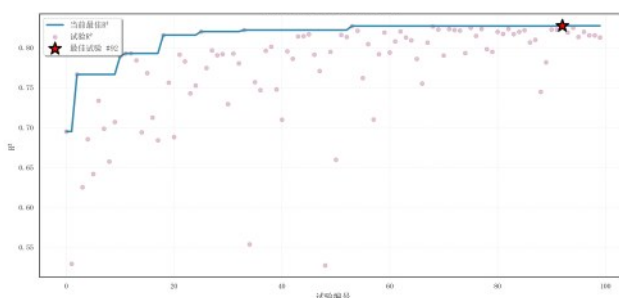
2.3 可解释性分析

2.3.1 基于 SHAP 值的特征重要性分析

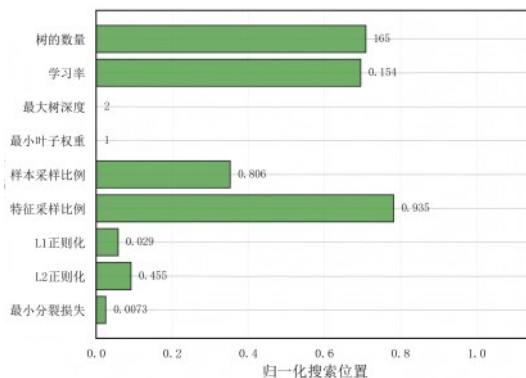
为理解特征对模型预测气体产率的贡献程度与方式,采用 SHAP 方法进行归因分析。SHAP 值通过“TreeExplainer”计算,确保了高效且精确的贡献分配。由堆叠式 SHAP 特征

重要性图(图5)可知,温度在四个目标产物(H_2 、 CH_4 、 CO 和 CO_2)的预测中贡献占比均为最高,总体占比约45%。这表明温度是模型决策的最主导因素。高温能促进超临界水中的自由基反应,加速重整、裂解和水气变换反应,显著提升气体产率并调控产物分布^[27]。停留时间和原料浓度的贡献次之,二者较为均衡(约25%),反映出

传质限制和反应动力学的辅助调控作用。相比之下,混合比的整体贡献最小(约6%),这表明相较于温度、停留时间等参数,该特征对气体产物分布的影响相对有限。此外,不同特征的重要性在各目标产物间也存在差异。例如,对于 CH_4 产率,混合比的相对贡献要显著高于其对其他气体的贡献。



(a) 优化收敛曲线



(c) 最优超参数组合

图3 XGBoost模型超参数优化过程

Fig. 3 Hyperparameter optimization process of the XGBoost model

图6的蜂群图进一步揭示了特征效应的方向性与非线性。图中每个点代表一个样本,颜色表示特征值的高低,其在x轴(SHAP值)上的位置表示该特

征对该样本预测的贡献方向(右为正,左为负)。由图可知,除了原料浓度外其他特征的高值红点几乎都集中在右侧,这表明这些特征对各气体产率均为

正向作用。其中，温度对四个气体产物尤其是H₂具有强烈的正向作用（SHAP >+3）。而混合比例的红点紧挨着SHAP值为0的轴线，这是因为其正向作用表现较弱。而原料浓度的高值样本整体偏向负贡献，表明高原料浓度对气体产率有一定抑制作用。

表5 XGBoost超参数优化前后性能对比

Table 5 Performance comparison of XGBoost before and after hyperparameter optimization

目标	优化前			优化后		
	R^2	E_{MAE}	E_{RMSE}	R^2	E_{MAE}	E_{RMSE}
H ₂	0.868	1.081	1.522	0.910	0.886	1.255
CH ₄	0.727	0.539	0.712	0.791	0.437	0.622
CO	0.860	0.162	0.205	0.873	0.151	0.195
CO ₂	0.715	1.018	1.334	0.736	0.981	1.284
平均	0.792	0.700	0.943	0.828	0.614	0.839

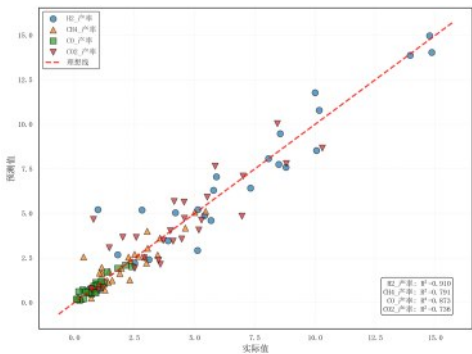


图4 -优化后的XGBoost模型气化产物预测产率与实际产率对比图

Fig. 4 Predicted vs actual gas yields of the optimized XGBoost model

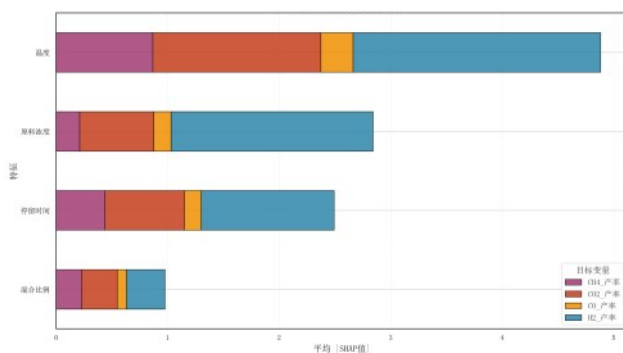
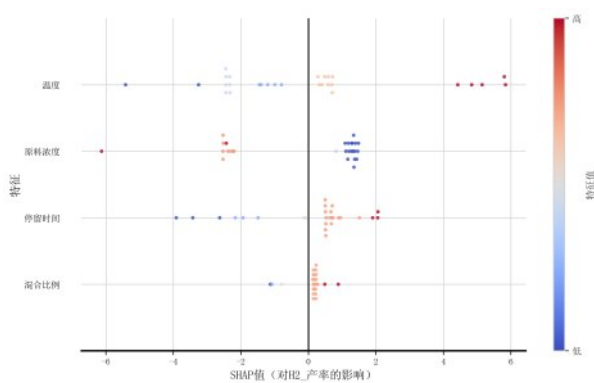
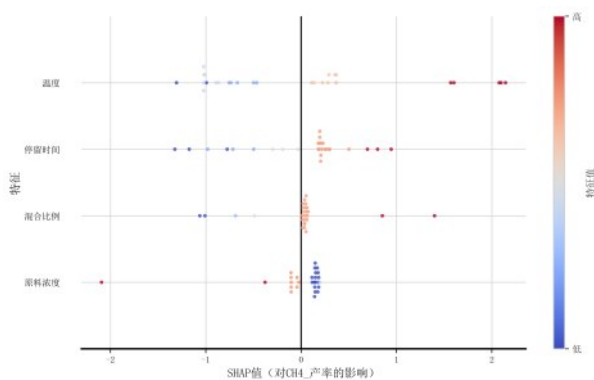


图5 堆叠式SHAP特征重要性图

Fig. 5 Stacked SHAP feature importance plot

(a) H₂(b) CH₄

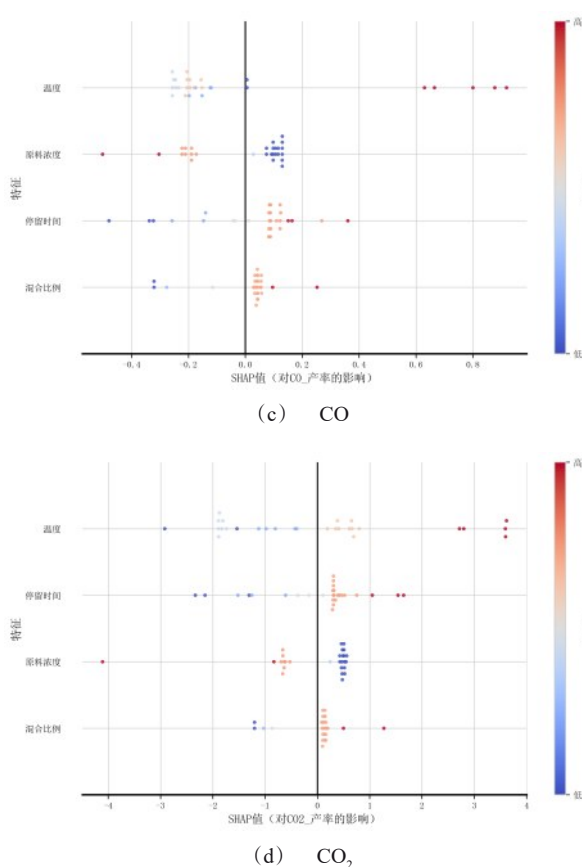


图6 SHAP值蜂群图

Fig. 6 SHAP Beeswarm Summary Plot

2.3.2 基于偏依赖图 (PDPs) 的特征效应分析

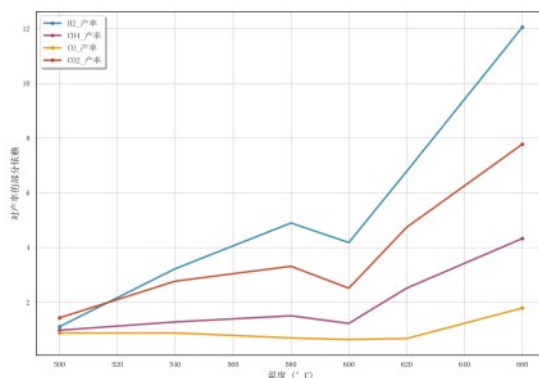
图7展示了单一特征对XGBoost模型预测结果的边际影响。结果显示,各气体产率都随温度的升高而增加,600℃后各曲线斜率剧增,其中H₂产率增速最快,CO垫底(图8a)。温度升高使超临界水介电常数与密度降低,能强化有机质及中间产物的溶解与传质过程;同时促进·H、·OH等自由基生成,加速有机大分子裂解及二次气化反

应^[28]。停留时间对气体产率均呈正向影响,20min后稍显平缓但未达到稳定状态(图8b)。这说明体系中仍存在难裂解有机结构及中间产物的持续转化过程,反应尚未达到完全转化。原料浓度则表现出明确的抑制效应,气体产率都随浓度增加单调下降,且中高浓度(12wt%)后趋势更明显(图8c),体现了高固体负载下传质阻力与稀释效应的负面作用^[29]。混合比对气体产率的影响趋势表现为气体产率在0.4-0.6区间有较大增长率,区间前后略为平缓

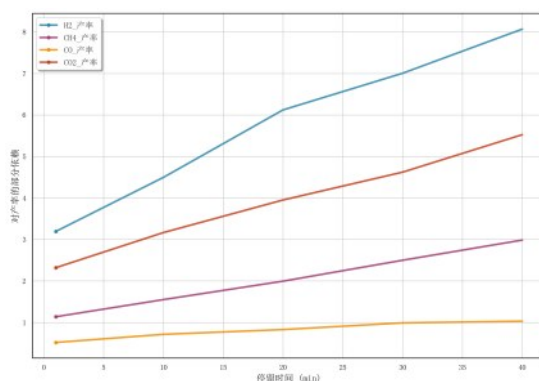
(图8d)。

图8为特征交互对 H_2 产率的二维偏依赖等高线图。高产率区域主要集中于高温、低浓度、较长停留时间和高混合比窗口。温度与停留时间、温度与原料浓度、温度与混合比等交互曲线均显示

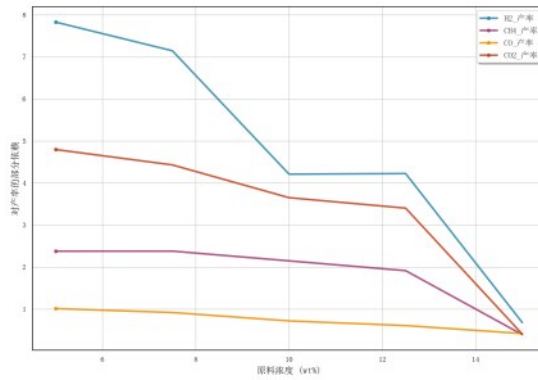
出明显的非线性协同或拮抗效应,例如高温下延长停留时间或保持低浓度可显著放大 H_2 产率,而高浓度区域即使在高温下产率仍受抑制。综上所述,特征对模型预测存在显著的非线性与交互性影响。



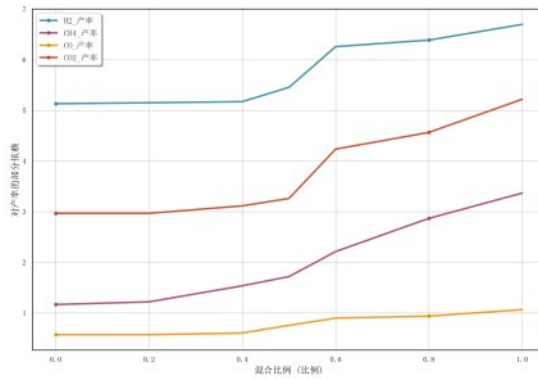
(a) 温度的影响



(b) 停留时间的影响



(c) 原料浓度的影响



(d) 混合比的影响

图7 各特征对气体产率的一维偏依赖图

Fig. 7 1D PDPs of features on gas yield

其 E_{MAE} 和 E_{RMSE} 分别降低约 12.3% 和 11.0%。

3 结论

(1) 基于 28 组实验数据, 比较了 RF、XGBoost、GBR 和 SVR 四个模型对 H_2 、 CH_4 、CO 和 CO_2 产率的预测性能, XGBoost 在本次模型对比中展现了相对最佳的预测潜力, 平均 R^2 为 0.792。

(2) Optuna 自动调参方法成功提升了模型的预测能力, 平均 R^2 提升至 0.828。同时还得到更小的预测偏差,

(3) SHAP 分析结果表明, 温度对气体产率的总体贡献最为显著 (约 45%), 是模型决策的主导因素; 停留时间和原料浓度的贡献次之 (约 25%); 混合比的整体贡献最小 (约 6%)。除原料浓度外, 其余特征对各气体产率均呈正向效应。

(4) 一维 PDP 图量化了各特征独立的主效应方向与强度; 二维 PDP 图则揭示了实现高 H_2 产率目标的高温、

低浓度、长停留时间和高混合比协同参数窗口。

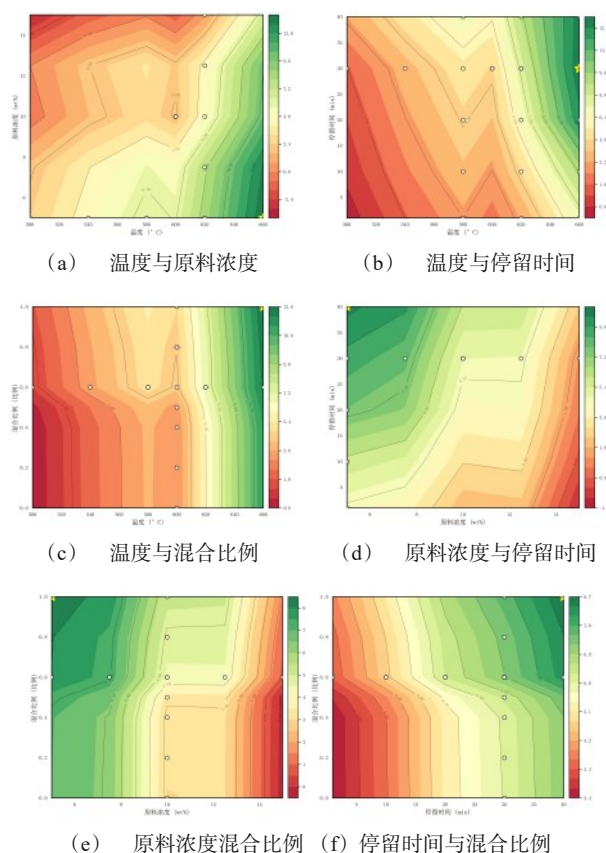


图8 特征交互对 H_2 产率影响的二维偏依赖等高线图

Fig. 8 2D PDPs of feature interactions on H_2 yield

参考文献

- [1] International Energy Agency (IEA). Global Hydrogen Review 2025 [R]. Paris: IEA, 2025.
- [2] Hu Y, Liu M, Li H, et al. Recent research progress in supercritical water gasification of biomass for hydrogen production[J]. Chemical Communications, 2025, 61(96): 19004-19019.
- [3] Bai J, Wang Y, Liu S, et al. Sustainable hydrogen production from coal tar residue via supercritical water gasification[J]. Journal of Cleaner Production, 2025, 519: 145967.
- [4] 郭栋, 周臣臣, 葛志伟, 等. 煤炭超临界水气化技术研究进展[J]. 现代化工, 2024, 44(3): 43-47.
- GUO D, ZHOU C, GE Zi, et al. Advances on coal supercritical water gasification[J]. Modern Chemical Industry, 2024, 44(03): 43-47.
- [5] 郭烈锦, 王乐, 黄勇, 等. 农林业固废超临

- 界水热化学制氢进展——反应机理[J]. 西安交通大学学报, 2023, 57(1): 1-14.
- GUO L, WANG L, HUANG Y, et al. Review of thermochemical conversion of agriculture and forestry waste for hydrogen production in supercritical water——mechanism analysis[J]. Journal of Xi'an Jiaotong University, 2023, 57(01): 1-14.
- [6] Heeley K, Orozco R L, Macaskie L E, et al. Supercritical water gasification of microalgal biomass for hydrogen production-A review[J]. International Journal of Hydrogen Energy, 2024, 49: 310-336.
- [7] Fang H, Fan D, Yao B, et al. Supercritical water gasification of plastic waste for hydrogen production: A review[J]. Journal of the Energy Institute, 2025, 123: 102239.
- [8] 王玉龙, 王树众, 蒋华义, 等. 市政污泥超临界水气化制氢技术研究进展[J]. 石油学报(石油加工), 2023, 39(4): 931-940.
- WANG Y, WANG S, JIANG H, et al. Research progress of hydrogen production from municipal sludge gasification in supercritical water[J]. Acta Petrolei Sinica (Petroleum Processing Section), 2023, 39(04): 931-940.
- [9] Wei J, Gong Y, Guo Q, et al. A mechanism investigation of synergy behaviour variations during blended char co-gasification of biomass and different rank coals[J]. Renewable Energy, 2019, 131: 597-605.
- [10] Faki E, Üzden Ş T, Seçer A, et al. Hydrogen production from low temperature supercritical water Co-Gasification of low rank lignites with biomass[J]. International Journal of Hydrogen Energy, 2022, 47(12): 7682-7692.
- [11] Bian C, Zhang R, Dong L, et al. Hydrogen/Methane Production from Supercritical Water Gasification of Lignite Coal with Plastic Waste Blends[J]. Energy & Fuels, 2020, 34(9): 11165-11174.
- [12] Ma M, Zhang S, Chen Y, et al. Optimization of hydrogen-rich syngas from coal and sewage sludge co-gasification in supercritical water[J]. Chemical Engineering Journal, 2024, 497: 154792.
- [13] Okolie J A, Epelle E I, Nanda S, et al. Modeling and process optimization of hydrothermal gasification for hydrogen production: A comprehensive review[J]. The Journal of Supercritical Fluids, 2021, 173: 105199.
- [14] Khandelwal K, Dalai A K. Prediction of Individual Gas Yields of Supercritical Water Gasification of Lignocellulosic Biomass by Machine Learning Models[J]. Molecules, 2024, 29(10): 2337.
- [15] Wang M, Chen L, Chen D, et al. Modeling study on biomass gasification for H₂-rich syngas production based on machine learning: A comprehensive review[J]. Renewable and Sustainable Energy Reviews, 2026, 226: 116223.
- [16] Zhao S, Li J, Chen C, et al. Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass[J]. Journal of Cleaner Production, 2021, 316: 128244.
- [17] Yan S, Mao J, Wei X, et al. Co-gasification of coal, biomass with waste plastic in a fluidized bed: Machine learning model optimized and gasification scheme[J]. Journal of Environmental Chemical Engineering, 2025, 13(6): 119916.
- [18] Gil M V, Jablonka K M, Garcia S, et al.

- Biomass to energy: a machine learning model for optimum gasification pathways [J]. *Digital Discovery*, 2023, 2(4): 929-940.
- [19] Liu S, Yang Y, Yu L, et al. Predicting gas production by supercritical water gasification of coal using machine learning[J]. *Fuel*, 2022, 329: 125478.
- [20] Al-Salih O G, Guangjian D, Al-Mudhafar W J, et al. Using extreme gradient boosting with Optuna hyperparameter tuning for efficient lost circulation prediction[J]. *Energy Geoscience*, 2026, 7(2): 100540.
- [21] Almarzooq H, Bin Waheed U. Automating hyperparameter optimization in geophysics with Optuna: A comparative study[J]. *Geophysical Prospecting*, 2024, 72(5): 1778-1788.
- [22] Devasahayam S, Albijanic B. Predicting hydrogen production from co-gasification of biomass and plastics using tree based machine learning algorithms[J]. *Renewable Energy*, 2024, 222: 119883.
- [23] Tian J, Dong R, Jia H, et al. Interpretable machine learning for predicting and evaluating hydrogen production from supercritical water gasification of coal[J]. *Fuel*, 2026, 404: 136173.
- [24] Li J, Li L, Tong Y W, et al. Understanding and optimizing the gasification of biomass waste with machine learning[J]. *Green Chemical Engineering*, 2023, 4(1): 123-133.
- [25] Abioye S O, Babatunde Y O, Abikoye O A, et al. Optimized machine learning algorithms with SHAP analysis for predicting compressive strength in high-performance concrete[J]. *AI in Civil Engineering*, 2025, 4(1): 16-32.
- [26] Khandelwal K, Nanda S, Dalai A K. Machine learning modeling of supercritical water gasification for predictive hydrogen production from waste biomass[J]. *Biomass and Bioenergy*, 2025, 197: 107816.
- [27] 徐功迅, 陈伟, 方真. 生物质超临界水制氢研究进展[J]. *农业工程学报*, 2023, 39(7): 24-35.
- XU G, CHEN W, FANG Z. Research progress of supercritical water hydrogen production from biomass [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(7): 24-35.
- [28] Chen J, Wang Q, Xu Z, et al. Process in supercritical water gasification of coal: A review of fundamentals, mechanisms, catalysts and element transformation[J]. *Fuel Processing Technology*, 2022, 237: 107456.
- [29] 戚新刚, 路利波, 陈渝楠, 等. 造纸黑液超临界水气化制氢与高附加值化学品回收研究进展 [J]. *化工学报*, 2022, 73(8): 3338-3354.
- QI X, LU L, CHEN Y, et al. Review of black liquor supercritical water gasification for hydrogen production with high value-added chemicals recovery [J]. *CIESC Journal*, 2022, 73(8): 3338-3354.