

研究论文

DOI: 10.11949/0438-1157.

基于机器学习算法的酸催化蒸汽爆破生物质预处理产物产量预测研究

刘宸瑜^{1,2}, 关兆峰², 于梦², 何志鹏², 王远鹏³, 张雪英¹, 周俊^{1,2}

(¹ 南京工业大学环境科学与工程学院, 江苏 南京 211816; ² 南京工业大学生物与制药工程学院, 江苏 南京 211816; ³ 厦门大学化学化工学院, 福建 厦门, 361005)

摘要: 预处理可以有效提高木质纤维素中糖的释放量, 其中酸催化蒸汽爆破是一种广泛的预处理方法。通过整合文献与实验数据, 构建了包含生物质、酸性催化剂、蒸汽爆破条件和酶解条件等参数的数据集, 使用 MissForest 进行数据填充, 在此基础上利用贝叶斯优化和5折交叉验证对随机森林(Random Forest, RF)等四种模型进行调优构建了高精度的产量预测模型, 结果表明纤维素保留率、半纤维素去除率、糠醛、5-羟甲基糠醛(HMF)、乙酸和酶解葡萄糖产量预测的决定系数(R-Square, R^2)均超过了0.80。基于SHAP(Shapley Additive exPlanations)与偏依赖分析: 原料中低半纤维素(<18 wt%)利于提高固体残渣纤维素保留率, 高温预处理可弥补弱酸(高pKa)催化不足以促进半纤维素去除; 蒸汽爆破温度主导糠醛与HMF生成, 在低纤维素含量(<13 wt%)与高温(220 °C)下糠醛最少且HMF受到抑制; 乙酸受酸浓度和半纤维素共同影响; 延长酶解时间(>70 h)可最大化葡萄糖产量。此外, 本文开发了基于Web端的可执行文件, 将模型转化成为实际可视化和可操作的应用, 弥补了复杂数据分析之间的差距。本研究为生物质蒸汽爆破的智能化调控和应用提供了高效预测工具和理论基础。

关键词: 生物质; 酸催化; 蒸汽爆破; 机器学习; 特征重要性

中图分类号: TK 6; TP181

文献标志码: A

文章编号: 0438-1157 (XXXX) XX-0001-19

Prediction of acid-catalyzed steam explosion biomass product yield based on machine learning algorithms

LIU Chenyu^{1,2}, GUAN Zhaofeng², YU Meng², HE Zhipeng², WANG Yupeng³, ZHANG Xueying¹, ZHOU Jun^{1,2}

(¹ School of Environmental Science and Engineering, Nanjing Tech University, Nanjing 211816, Jiangsu, China; ² College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Nanjing 211816, Jiangsu, China; ³ College of Chemistry and Chemical Engineering, Xiamen University 361005, Fujian, China)

Abstract: Pretreatment can effectively enhance the release of sugars from lignocellulose, with acid-catalyzed steam explosion being a widely used pretreatment method. By integrating literature and experimental data, a dataset encompassing parameters such as biomass, acidic catalyst, steam explosion conditions, and enzymatic hydrolysis conditions was constructed. MissForest was employed for data imputation. Based on this, Bayesian optimization and

收稿日期: 2026-06-15 修回日期: 2026-07-09

通信作者: 周俊(1983-), 男, 博士, 教授, zhoujun@njtech.edu.cn; 张雪英(1976-), 女, 博士, 教授, xueyingzhang@njtech.edu.cn; 王远鹏(1979-), 男, 博士, 教授, wypp@xmu.edu.cn

第一作者: 刘宸瑜(2002-), 男, 硕士研究生, liuchenyu@njtech.edu.cn; 关兆峰(1998-), 男, 博士研究生, guan112600@njtech.edu.cn

基金项目: 国家自然科学基金项目(U24A20543)

引用本文: 刘宸瑜, 关兆峰, 于梦, 何志鹏, 王远鹏, 张雪英, 周俊. 基于机器学习算法的酸催化蒸汽爆破生物质预处理产物产量预测研究[J]. 化工学报, XXXX, XX(XX): 1-19

Citation: LIU Chenyu, GUAN Zhaofeng, YU Meng, HE Zhipeng, WANG Yupeng, ZHANG Xueying, ZHOU Jun. Prediction of acid-catalyzed steam explosion biomass product yield based on machine learning algorithms[J]. CIESC Journal, XXXX, XX(XX): 1-19

5-fold cross-validation were utilized to fine-tune four models, including Random Forest (Random Forest, RF), to construct a high-precision yield prediction model. The results indicated that the determination coefficients (R^2) for cellulose retention rate, hemicellulose removal rate, furfural, 5-hydroxymethylfurfural (HMF), acetic acid, and enzymatic glucose yield prediction all exceeded 0.80. Based on SHAP(Shapley Additive exPlanations) and partial dependence analysis, a low hemicellulose content (<18 wt%) in the raw material is beneficial for enhancing the cellulose retention rate of solid residue, and high-temperature pretreatment can compensate for the insufficient promotion of hemicellulose removal by weak acid (high pKa) catalysis. Steam explosion temperature primarily determines the formation of furfural and HMF. At low cellulose content (<13 wt%) and high temperature (220 °C), furfural is minimized and HMF is suppressed. Acetic acid is jointly influenced by acid concentration and hemicellulose. Extending enzymatic hydrolysis time (>70 h) can maximize glucose yield. Furthermore, this study developed a web-based executable file to transform the model into a practical, visual, and operable application, bridging the gap between complex data analysis and practical use. This research provides an efficient prediction tool and theoretical foundation for the intelligent regulation and application of biomass steam explosion.

Keywords: biomass; acid catalysis; steam explosion; machine learning; feature importance

引 言

农业废弃物来源广泛、可再生性强,是生产燃料及高附加值化学品的重要生物质资源。典型的农业废弃物生物质主要组分为纤维素、半纤维素和木质素组成,各组分通过非共价作用及木质素-碳水化合物复合体(Lignin carbohydrate complexes, LCC)相互结合,形成具有抗降解性的细胞壁结构^[1],限制酶和化学试剂对多糖组分的有效接触。不同种类生物质在三组分含量、木质素结构及组分结合方式等方面存在差异,从而显著影响预处理过程中的细胞壁解构程度和多糖组分的分离效果。在生物炼制过程中,为解离木质素与碳水化合物之间的连接并提高多糖组分的可及性,开发了多种预处理技术,包括酸碱预处理、低共熔溶剂(Deep Eutectic Solvents, DES)预处理和亚临界水处理等^[2]。

蒸汽爆破(Steam Explosion, SE)作为一种高效的生物质预处理技术,利用高温高压水蒸气渗透生物质内部,通过瞬间释压产生的强大剪切力破坏木质纤维素的物理骨架,促进半纤维素水解和细胞壁解构,提高纤维素的酶可及性及后续酶解糖化效率^[3]。硫酸、盐酸、磷酸等酸性催化剂可用于强化蒸汽爆破预处理。研究表明酸性催化剂的加入可以显著降低汽爆时间和温度,并促进半纤维素水解与脱除,提高残渣纤维素酶解效果^[4]。然而,酸催化蒸汽爆破在促进木质纤维素解构的同时,也可能导致预处理浸出液中抑制性副产物的积累,主要包括糠醛、HMF、乙酸。这些副产物浓度的提升会抑制发

酵微生物生长和代谢,进而影响后续发酵过程^[5]。尽管蒸汽爆破预处理已得到广泛研究,但原料组成和工艺条件均会影响固体残渣组分、浸出液副产物及后续酶解产糖,且各因素之间相互关联。传统单因素试验通常只能考察少数变量,难以同时比较多种原料和工艺条件对预处理结果的影响。因此,需要采用更适合多变量数据的分析方法,进一步明确各因素与预处理效果之间的关系。

机器学习(Machine Learning, ML)技术在处理生物质转化过程中的多变量、非线性及高维度特征映射方面展现出显著优势。Xie等人通过比较多种机器学习模型,发现支持向量回归(Support Vector Regression, SVR)对木质纤维素的酶解糖产量($R^2=0.99$),并揭示了溶液pH值是影响预处理效果及糖产率的核心因素^[6],Zhang等人研究对比了多种机器学习模型对木质纤维素葡萄糖产率的预测能力,结果表明卷积神经网络(Convolutional Neural Network, CNN)表现最优(R^2 达0.95),并指出木质素去除率和纤维袋/木质素比值是影响产量的关键因素^[7],Huang等利用人工神经网络(Artificial Neural Network, ANN)构建了酸催化蒸汽爆破(Acid Catalysis Steam Explosion, ACSE)优化框架,通过权衡葡萄糖产率与抑制效应之间的竞争关系,实现了预处理条件的智能化寻优^[8]。Zhang等(2023)开发了一款基于Python的交互式图形用户界面(Graphical User Interface, GUI),通过可视化的方式实现了对生物炭强化厌氧消化性能的快速预测及其关键影响因子的直观解析^[9]。Liu等(2025)基于Python的Tkinter库同样开

发了一款GUI,实现了根据生物质特性和热解参数对生物炭产量及其氮磷钾(NPK)含量的即时、可视化预测^[10]。Chen等(2026)开发了一款基于Python的GUI,通过集成的极度梯度提升(eXtreme Gradient Boosting,XGBoost)模型实现了从原料组分到生物炭产率、理化品质及吸附性能的全流程即时预测,为生物炭的定向制备与环境应用提供了直观的可视化决策工具^[11]。然而,现有多针对特定转化环节或单一类型的输出变量,对原料组成、蒸汽爆破条件、固体组分转化、液相副产物生成及后续酶解性能之间关系仍缺乏统一建模与定量分析。

针对蒸汽爆破预处理过程中原料组成与工艺参数关系复杂,以及固相组分转化、液相副产物生成和酶解糖化效果难以协同定量分析等问题,构建了一种“原料特性-蒸汽爆破工艺参数-目标产物”的机器学习预测模型,比较RF、XGBoost、ANN和SVR四种算法的预测性能,通过SHAP分析和偏依赖探究原料组成与工艺条件对残渣纤维素保留率、半纤维素去除率、糠醛、HMF和乙酸生成量以及酶解葡萄糖得率的影响规律。为蒸汽爆破过程的性能预测和工艺优化提供参考。此外,本平台基于研究得出的最佳模型构建,通过数字化交互界面根据原料特性和蒸汽爆破条件快速预测固相组分转化、液相副产物生成及酶解糖化结果。该交互平台可

减少重复试验,提高数据分析与工艺参数优化的效率,为生物质预处理过程的数字化研究提供工具支持。

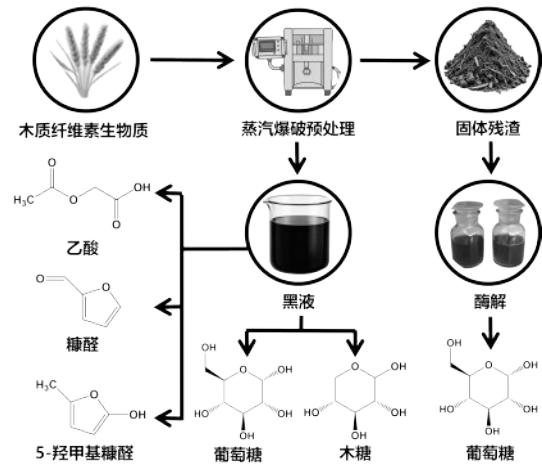


图1 蒸汽爆破反应路径图

Fig.1 Steam explosion reaction pathway diagram

1 材料与方法

本研究流程图如图2所示,包括四个阶段:(1)数据集收集和预处理;(2)模型构建与开发;(3)模型可解释性分析;(4)基于Web可视化页面以及可执行文件。

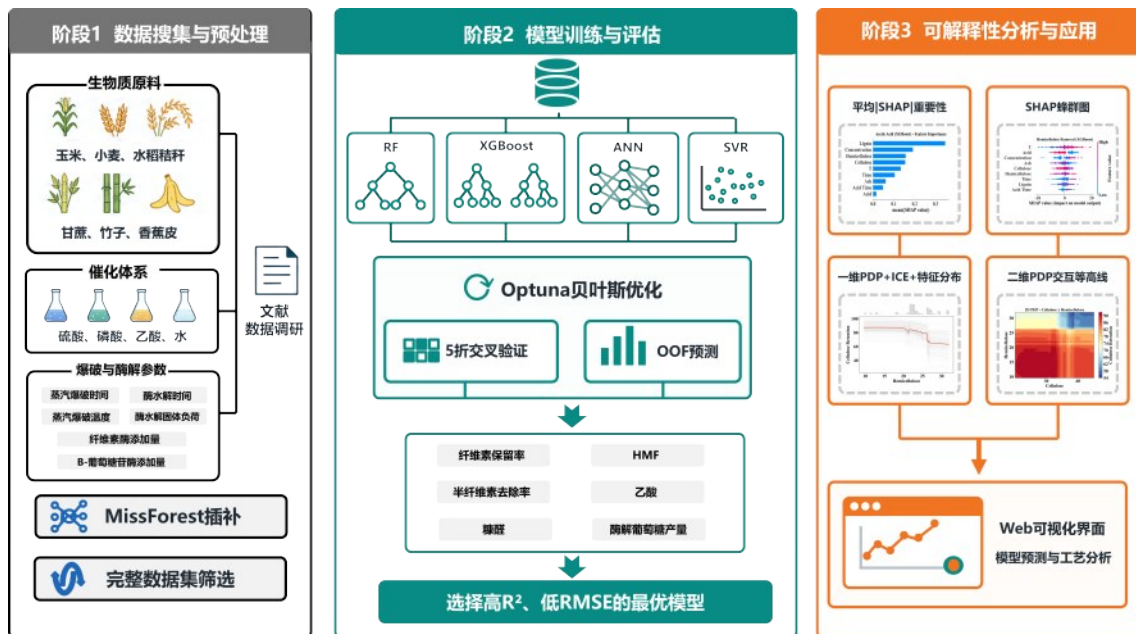


图2 本文流程框架

Fig.2 The process framework of this study

1.1 数据获取与筛选

基于谷歌学术、web of Science 和中国知网等数据库,通过国内大语言模型 deepseek-v4-pro 结合“机器学习”、“蒸汽爆破”等关键词进行文献检索和筛选,最终选取了 31 篇符合标准的文献作为文献数据来源,其中图数据使用 Getdata 软件获取,并结合实验数据补充文献数据空白,本文构建了结构化的蒸汽爆破数据集,并将输入特征按照功能划分为四个类别:生物质特性、酸性催化剂性质、汽爆条件、酶解条件等参数(见表1)。其中生物质种类包括了玉米秸秆^[12-23]、水稻秸秆^[24-29]、小麦秸秆^[30-34]、香蕉^[4]、甘蔗^[35-38]、大麦^[39]、竹子^[40,41]等产物,酸性催化剂种类涵盖了硫酸($\text{pKa}=-3$)、磷酸($\text{pKa}=2.15$)、乙酸($\text{pKa}=$

4.75)和水($\text{pKa}=15.70$)。文献调研数据中,蒸汽爆破温度范围:120–280 °C;时间范围:0–60 min。酶解条件中纤维素酶添加量范围:3–70 FPU/g,酶水解时间范围:6–96 h,固体负荷范围:2–48 %;最终构建可用数据 403 条。为方便模型训练,将浸液时间过夜假设为 12h,酸性催化剂种类使用酸解离常数(pKa)代替,其余特征单位统一。

本文中所涉及相关实验数据由南京工业大学生物能源研究所实验室提供,采用汽爆设备 QBS-80(分析型汽爆工艺试验台),内部容积 405 mL。该设备具有蒸汽发生与汽爆系统一体化功能,设备自动化程度高,操作方便。

表1 输入特征类别
Table 1 Input feature category

类别	特征	单位	翻译(缩写)
生物质特性	纤维素含量	wt%	Cellulose
	半纤维素含量	wt%	Hemicellulose
	木质素含量	wt%	Lignin
	灰分含量	wt%	Ash
酸催化性质	酸解离常数	–	pKa(Acid)
	酸浓度	wt%	Acid Concentration(Concentration)
	浸泡时间	min	Acid Soaking time(Acid Time)
蒸汽爆破参数	蒸汽爆破温度	°C	Steam explosion Temperature(<i>T</i>)
	蒸汽爆破时间	min	Steam explosion time(Time)
酶解参数	纤维素酶添加量	FPU/g	Cellulase loading
	β -葡萄糖苷酶添加量	IU/g	β -Glucosidase
	酶水解时间	h	Enzymatic hydrolysis time
	固体负荷	%	Solid loading
输出特征	固体残渣纤维素保留率	%	Cellulose recovery
	固体残渣半纤维素去除率	%	Hemicellulose removal
	浸出液糠醛含量	wt%	Furfural
	浸出液 HMF 含量	wt%	HMF
	浸出液乙酸含量	wt%	Acetic acid
	固体残渣酶解葡萄糖产量	g/L	glucose yield from enzymatic hydrolysis(Glucose)

1.2 数据预处理

对搜集的原始数据进行系统的预处理,以确保数据质量和模型训练的有效性。进行数据清洗,去除异常值和不合理的数据点,确保数据的准确性和一致性。

缺失值的插补采用 MissForest 算法^[42],相对于传统的插补方法, MissForest 算法具有以下优势:该算法采用随机森林算法来生成预测模型,具有高效的计算速度和可扩展性。该算法能够自动学习数据

的特征和模式,并通过迭代的方式提高插补准确性,具有高准确性和精度。该算法能够适应不同类型和结构的数据,包括数值型数据、分类数据和文本数据。首先以各变量的中位数对缺失值进行初始填充,随后按照缺失值由低到高对变量进行排序,对每个变量以其余变量为特征训练随机森林回归器,预测并更新其缺失值,循环迭代直至收敛。参数设置为:决策树数量 $n_{\text{estimators}}=200$,最大迭代次数 $\text{max_iter}=15$,分裂特征树 $\text{max_features}=\text{sqrt}$,

启用袋外评分(oob_score=True),随机种子random_state=42以保证结果可复现,收敛阈值设为相邻两轮插补值的最大绝对变化量小于 10^{-6} 。为评估插补质量,随机遮住数据集中10%的已知值作为验证集,插补后与真实值对比计算归一化均方根误差(NRMSE)和决定系数(R^2)。验证结果表明,全部变量的平均袋外 R^2 达到了0.96,所有变量的OOB的 R^2 均高于0.80,人工缺失验证的平均NRMSE为0.06(<0.10),平均 R^2 为0.84(>0.80),表明插补精度优异,数据质量满足后续建模要求。

1.3 模型构建与超参数调优

机器学习模型由python 3.12中的sklearn包构建。对机器学习模型中的超参数进行微调可以提高模型的精度,避免过拟合,因此对每个机器学习模型的超参数进行优化。与随机搜索和网格搜索相比,贝叶斯优化方法在超参数调优方面效率更高。

由于RF与XGBoost等树模型基于特征切分,对输入特征量纲不敏感;而SVR与ANN分别依赖于距离度量与梯度下降优化,其收敛速度与模型精度易受到特征数值范围影响。因此,本文仅针对SVR和ANN进行了数据标准化处理。

RF模型,“n_estimators”、“max_depth”、“min_samples_leaf”和“max_features”是重要参数。“n_estimators”决定了模型进行迭代的次数,在提升法(Boosting)中,每一棵树都在尝试修复前一棵树的残差。“max_depth”决定了每一颗基础决策树的最大层数。“min_samples_leaf”代表着叶子节点所需的最小样本数。“max_features”代表着分裂考虑的最大特征数。

XGBoost模型,“n_estimators”、“max_depth”、“learning_rate”和“colsample_bytree”是重要参数。“learning_rate”通过缩减每一步的权重,防止模型过快地收敛到局部最优解。“colsample_bytree”在构建每一棵树时,对特征进行随机采样地比例。因此,“n_estimators”、“max_depth”、“learning_rate”和“colsample_bytree”分别在150-250、5-10、0.01-0.20和0.60-1之间进行了优化。

SVR模型对参数敏感,所以在此进行了数据标准化操作。SVR模型通过使用“C”(0-5)和“epsilon”(0.05-0.20)进行了优化。其中,“C”用来平衡“对超出误差的容忍度”,“epsilon”定义了预测线周围的一个免惩罚区域。

对于ANN模型,通过数据标准化消除量纲差异以加速收敛,采用ReLU激活函数引入非线性特征并缓解梯度消失问题,利用Adam优化器在500个迭代周期(epochs)中高效更新网络权重,并以均方根误差(RMSE)作为损失函数来不断缩小预测值与实际值间的偏差,从而确保模型在处理复杂的生物质转换数据时具备优异的拟合精度与泛化性能。

1.4 模型评估指标

回归模型用于以下评估,来确定模型的性能优异。数据库被分为训练集和测试集,分别用于模型训练和评估。测试集的作用是评估模型的泛化能力及预测新数据的能力。本文采用了5倍交叉验证搜索最佳参数,在5倍交叉验证中,数据集被分为5个大小相同的子集,模型由4个子集迭代训练,并由剩下一个子集进行模型评估。因此,对于每个超参数组合,对模型进行5次训练和评估。使用平均 R^2 来评估80%的数据集。较高的 R^2 值(接近于1)表明性能优于较低的平均 R^2 值。

本文所建立的模型为非线性回归模型,因此采用均方根误差(Root Mean Square Error, RMSE)和 R^2 对模型进行评价。更精确的模型对应更低的RMSE值, R^2 通常用于评价回归模型在实验数据和预测数据之间的相关性。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{\text{exp}} - X_{\text{pred}})^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (X_{\text{exp}} - X_{\text{pred}})^2}{\sum_{i=1}^n (X_{\text{exp}} - \bar{X}_{\text{pred}})^2} \quad (2)$$

式中,n是数据总量; X_{exp} ,实验数据; $\bar{X}_{\text{exp}}$,平均实验数据; Y_{pred} ,预测数据

1.5 模型可视化

本文使用vue3+Fastapi构建前后端使用界面,使研究人员能够直接访问本研究的模型。可以在Github(https://github.com/kongyushengcao/straw_ml)平台上找到该页面的源代码,包含最佳模型,按照README.md说明可以获取相关页面。另外,还使用pyinstaller库创建exe文件,方便下载直接使用。

2 结果与讨论

2.1 皮尔逊相关系数

皮尔逊相关系数如图3所示,展示了生物质预处理、酶解过程和结果产物中19个特征之间的Pearson相关系数矩阵,颜色从蓝色($r=-1$)经白色($r=0$)到红色($r=1$)。171对特征组合中包含了3对强相关($|r|\geq 0.60$),17对中等相关($0.40\leq |r|\leq 0.60$),其余特征之间相关性较弱,表明所选特征具有较低的冗余性,有利于模型关注到不同维度的信息。不同种类原料中纤维素与半纤维素含量成较强正相关($r=0.61$),两种组分与灰分含量分别成负相关($r=-0.55$; $r=-0.58$),反映了不同种类生物质组成差异主要存在于结构性碳水化合物与无机组分之间。固体残渣纤维素保留率与生物质半纤维含量呈较强负相关($r=0.62$),表明汽爆预处理后纤维素的保留率随着不同生物质半纤维素含量增加而降低。原因可能与生物质中各组分通过不同方式形成了精密的交联结构有关,其中半纤维素与纤维素主要通过氢键和范德华力连接,同时借助LCC连接键与木质素发生交联形成“抗降解屏障”限制三素预处理的择性分离^[1]。生物质半纤维素含量越高则会提高解聚难度,在预处理促进半纤维素溶出的同时增加纤维素损失的可能性^[43]。相应地,半纤维素脱除率与酸浓度、酸浸时间和温度均呈正相关($r=0.30$ 、 0.23 和 0.25),说明预处理强度的提高会促进半纤维素糖苷键及部分对酸敏感的连接键断裂,进而增加纤维素表面的可及性^[44, 45]。

酶解后葡萄糖产量是检验预处理效果的关键指标,图中酶解后葡萄糖产量与半纤维素脱除率呈正相关($r=0.33$),与生物质半纤维素含量和木质素含量呈负相关($r=-0.14$),这与半纤维素的脱除改善了固体残渣中纤维素可及性。但酶解后葡萄糖含量与温度呈正相关($r=0.30$),与酸浓度和酸浸时间分别呈负相关($r=-0.27$ 和 -0.26),说明适度升温有利于后续酶解效果,但是体系中过多酸性催化剂的加入以及酸浸时间的增长会使预处理后水解液中的抑制物产量增加^[46]。综上所述,汽爆预处理主要通过改变生物质中三素成分的比例,促进半纤维素的脱除,从而影响最终酶解葡萄糖的产量,但相关系数关联性并未达到较高水平,说明最终酶解效果是由不同生物质原料中三素比例,汽爆预处理强度以及酶解条件协同决定。

2.2 模型拟合与性能评估

为了评估四种机器学习算法(RF、XGBoost、ANN、SVR)的预测性能,本文对比了不同预测模型在不同目标变量上的RMSE和 R^2 (图4)。生成的副产物主要在液相中,而酶解对副产物通常没有影响。因此,仅固体残渣酶解葡萄糖产量预测需要酶解操作变量数据。在 R^2 方面,固体残渣纤维素保留率预测结果在所有模型中较为稳定, R^2 均高于0.81,RF和XGBoost两种模型 R^2 均为0.87。相同预测模型对于不同目标变量的预测结果差异较大,例如XGBoost模型预测半纤维素去除率 R^2 为0.90,预测HMF的 R^2 仅为0.82。在RMSE方面,各个模型在半纤维素去除率预测上表现出最大的误差波动,其中ANN模型在半纤维素去除率上的RMSE达到10.10。浸出液糠醛和浸出液HMF的预测误差较低,RMSE均小于0.25,但由于不同目标变量的量纲和数值范围不同,RMSE不宜直接进行跨目标比较。XGBoost目标变量较多的情况下,泛化差距较小且预测精度较高,说明模型具备较好泛化能力。综合来看,XGBoost模型在保持较高 R^2 的同时,表现出在复杂生物化学组分预测中更优的预测精度与泛化能力。

通过分析 R^2 与RMSE评估数据,确定了针对不同产物的最优预测模型。XGBoost凭借更强大的非线性捕捉高维能力,在目标变量较多的情况下通常具备更优的预测性能。评估模型数据表明,RF模型在固体残渣纤维素保留率预测上具有较高优势,其在保证高拟合度($R^2=0.87$)的同时,展现了比XGBoost(RMSE=7.09)算法更低的预测误差(RMSE=6.94),因此被选定为固体残渣纤维素保留率预测的最优模型。但XGBoost在半纤维素去除率、糠醛、HMF、乙酸和葡萄糖的预测中具备较高的 R^2 值与更低的RMSE。

图5展示了在最佳模型上通过 R^2 和RMSE比较了不同模型的实际拟合图。红色虚线表示等值线($y=x$),纤维素保留率、半纤维素去除率、糠醛、HMF、乙酸和葡萄糖的 R^2 分别是0.87、0.90、0.84、0.82、0.85和0.84;RMSE分别是6.94、6.52、0.20、0.10、0.38和2.88。其中半纤维素去除率拟合效果最佳($R^2=0.90$),HMF相对较低($R^2=0.82$)。RMSE最小的是HMF(RMSE=0.10),最大的是纤维素保留率(RMSE=6.94)。整体来看,散点紧密分布于拟合直线周围, R^2 接近1且RMSE较低,表明模型具有较高的预测精度和良好的泛化能力。

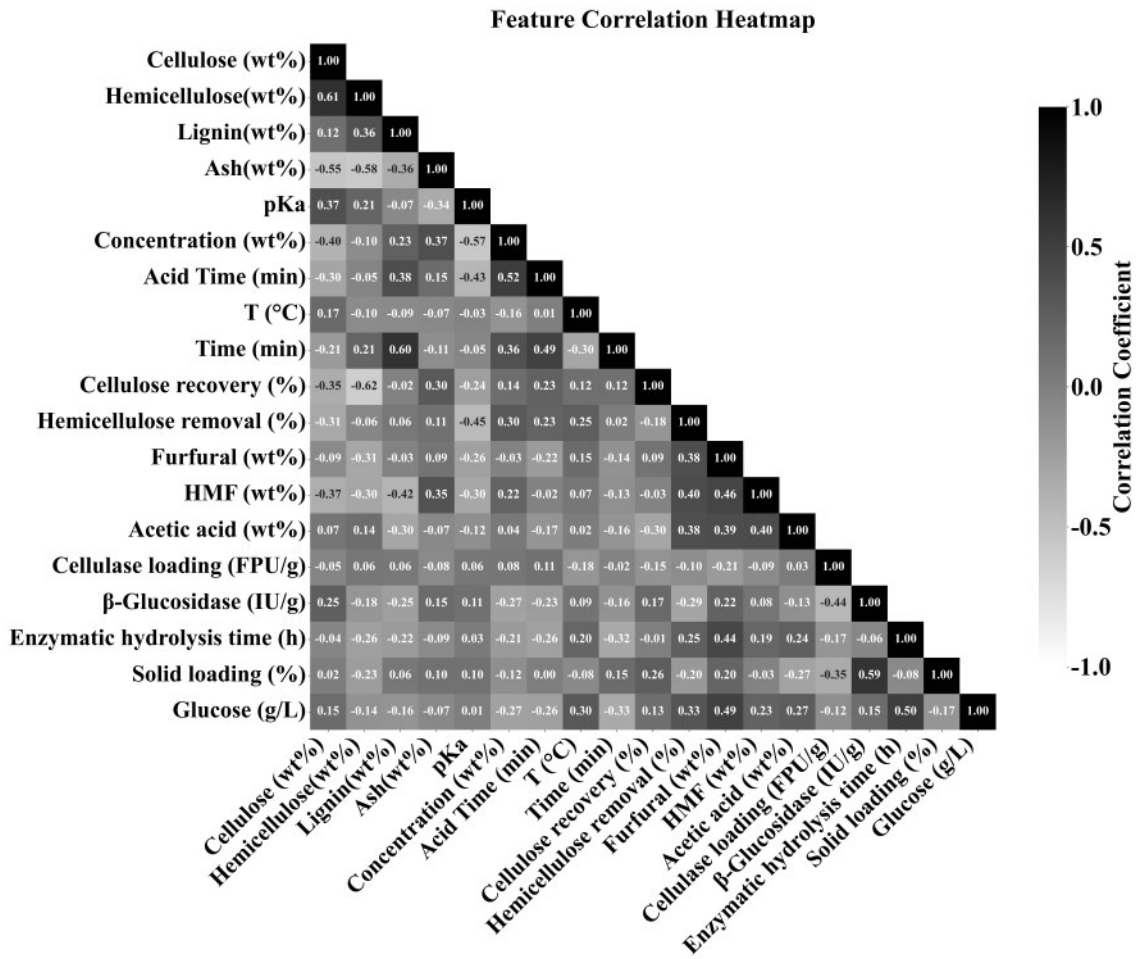
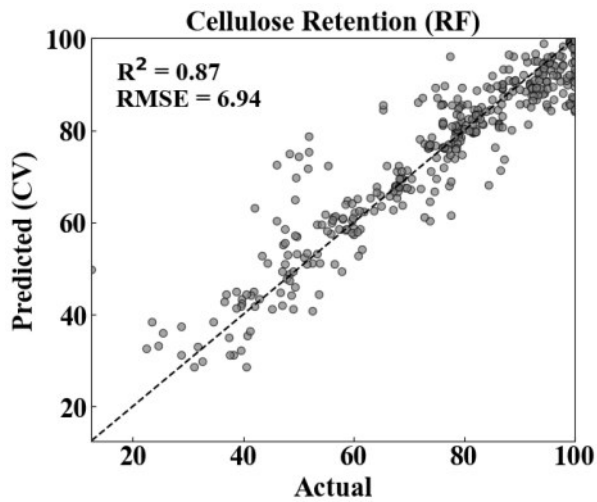
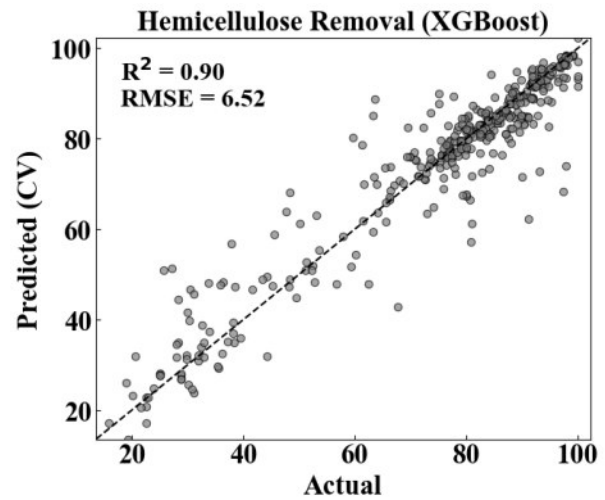


图3 皮尔森相关系数热力图

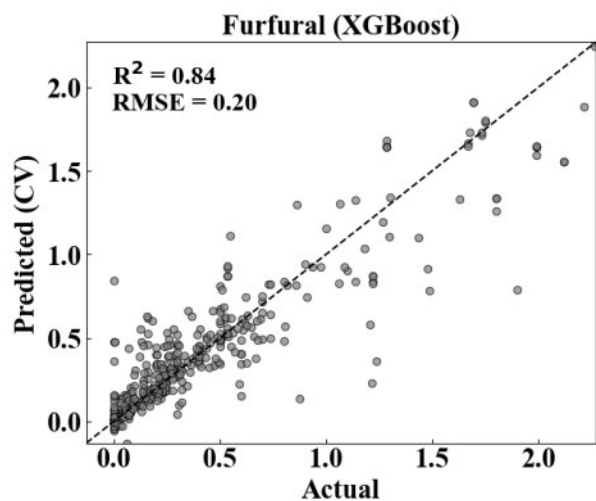
Fig.3 Heat map of Pearson's correlation coefficient



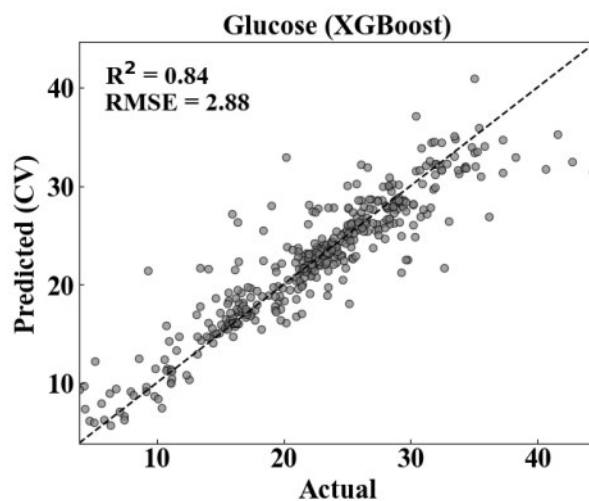
(a) 纤维素保留率最佳预测模型



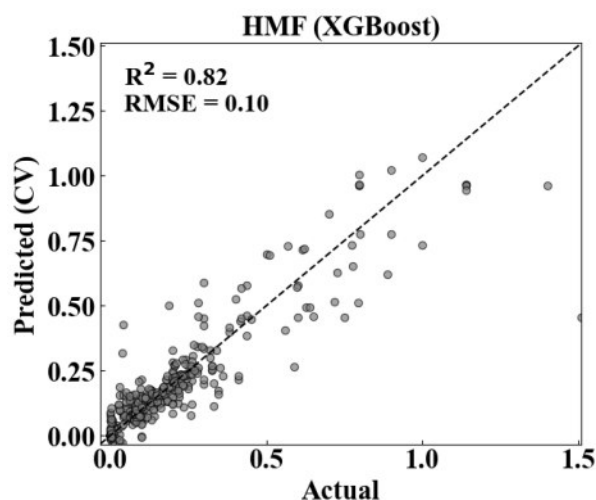
(b) 半纤维素去除率最佳预测模型



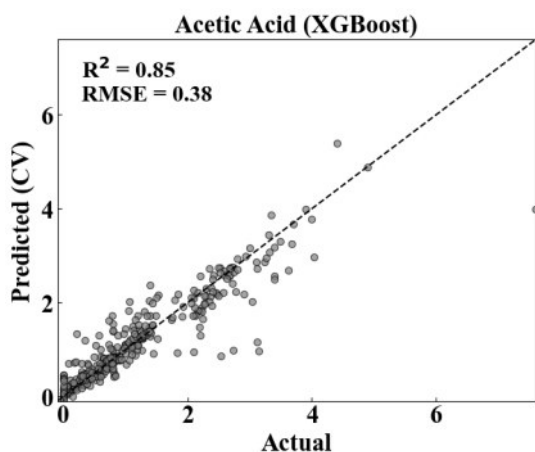
(c) 糠醛最佳预测模型



(f) 葡萄糖产量最佳预测模型



(d) HMF最佳预测模型



(e) 乙酸最佳预测模型

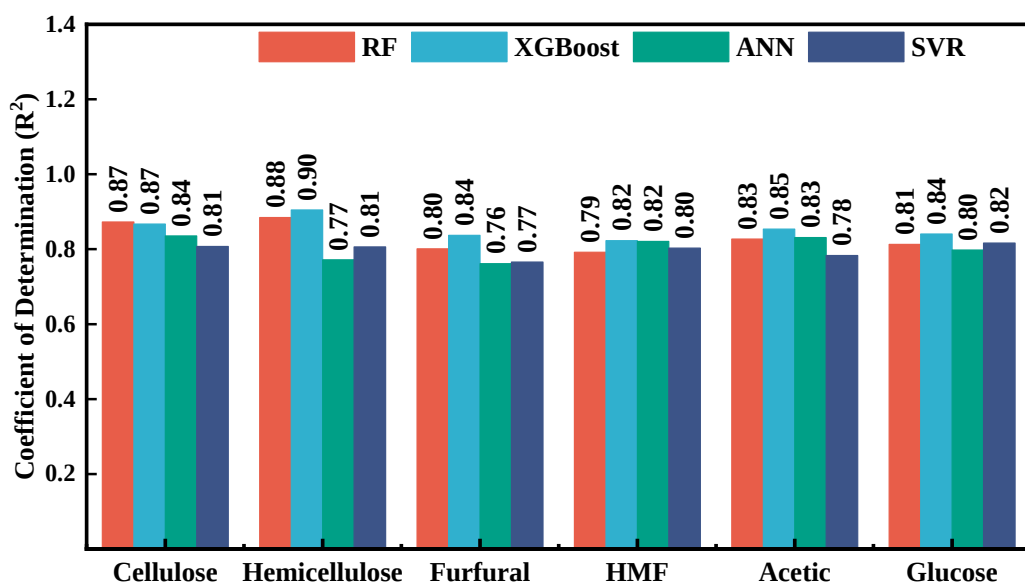
图5 各特征最佳预测模型拟合图

Fig.5 Fitting plot of the optimal prediction model for each feature

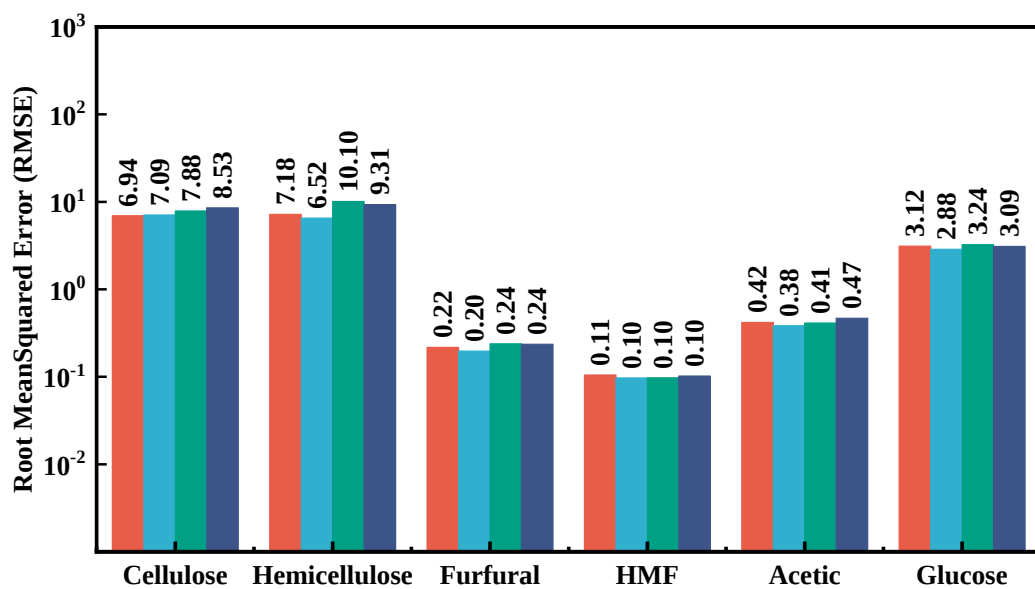
2.3 SHAP特征重要性分析

为了深入解析机器学习模型对蒸汽爆破流程中关键输出变量的预测机制,对最优模型采用SHAP方法进行了可解释性分析。SHAP值可量化输出特征对于目标变量预测结果的相对贡献与影响方向,有效评估不同输出特征对预测结果的影响程度。

图6展示了不同目标特征重要性排序及SHAP分布,揭示了生物质木质纤维素组成特性与预处理条件对预处理后固相组分转化及副产物生成的差异化影响规律。图6(a,b)中不同生物质半纤维素含量影响固体残渣纤维素保留率,蒸汽爆破温度、pKa与酸性催化剂在蒸汽爆破预处理中的添加量则影响最终固体残渣的半纤维素去除率。模型在预测过程中,部分特征的高、低取值同时分布于正、负SHAP值区域,表明这些特征对模型输出的影响方向并不固定,可能存在非线性响应,并受到原料组成与蒸汽爆破工艺参数的共同调节^[8]。浸出液糠醛与浸出液HMF主要由半纤维素和纤维素汽爆过程中水解得到的戊糖与己糖进一步脱水形成^[47]。图6(c)蒸汽爆破温度和生物质半纤维素含量对浸出液糠醛预测贡献最高,图6(d)中HMF则主要受蒸汽爆破温度、酸浓度影响,上述结果说明,模型能够一定程度识别半纤维素含量、高温处理与浸出液糠醛形成过程之间的关联;同时,较高汽爆温度对应正SHAP值,说明预处理强度提升会促进HMF的形



(a) 预测模型决定系数



(b) 预测模型误差

图4 不同特征预测精度对比

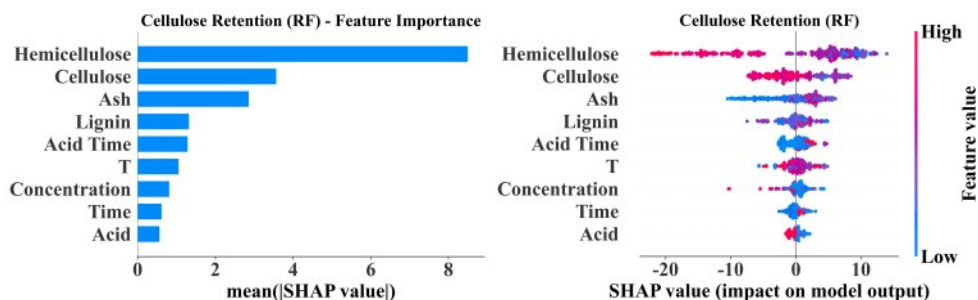
Fig.4 Comparison of prediction accuracy based on different features

成^[48]。HMF 特征贡献较为分散,说明其生成可能受到多种因素共同控制,这可能是其预测性能略低于糠醛的原因之一。图 6(e)中酸浓度与原料半纤维素含量对于乙酸产量具有较高的平均绝对 SHAP 值,且表现为正向贡献,可能与木聚糖侧链乙酰基

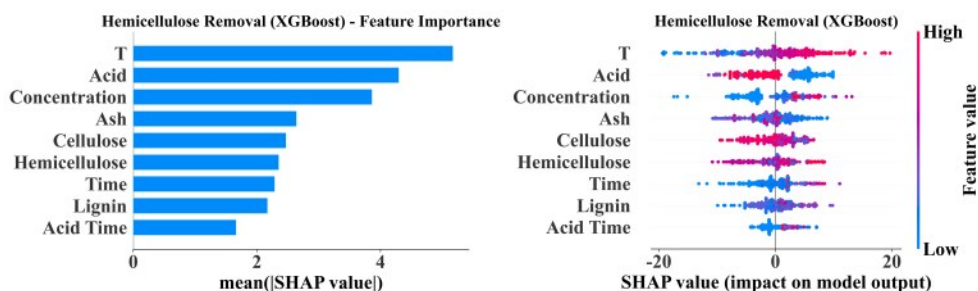
在预处理过程中脱乙酰反应相关^[8]。较高酸浓度可能促进木聚糖脱乙酰,进而增加浸出液中的乙酸含量。木质素虽为乙酸相关输入中 SHAP 特征重要性贡献排序的最高位,但红、蓝 SHAP 值存在明显重叠,说明其对乙酸预测的影响方向较为复杂。木质

素并非乙酸的直接来源,而是作为反映生物质类型、半纤维素结构及乙酰化程度的关联特征参与模型预测^[49]。图6(f)展示了酶解时间和蒸汽爆破温度是影响最终酶解葡萄糖产量的主要操作特征,两个输入变量具备较高的正SHAP值,表明延长酶解时间并提高汽爆温度会促进最终酶解后葡萄糖转化。生物质木质素含量在相同区间对应的SHAP值存在

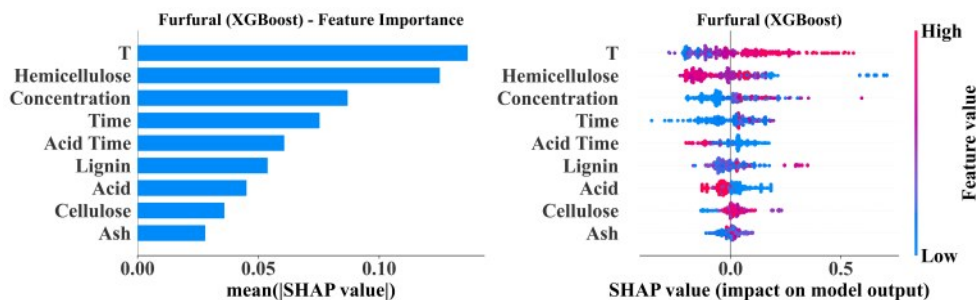
差异,表明木质素含量对酶解效果的影响具有复杂的非线性特征。这表明酶解过程中葡萄糖的转化并不仅取决于生物质纤维素含量,还与生物质中木质纤维素交联形成的抗降解屏障密切相关。蒸汽爆破能够促进半纤维素水解及木质素-碳水化合物复合结构的破坏,提高纤维素的酶可及性,从而改善后续酶解效果^[47]。



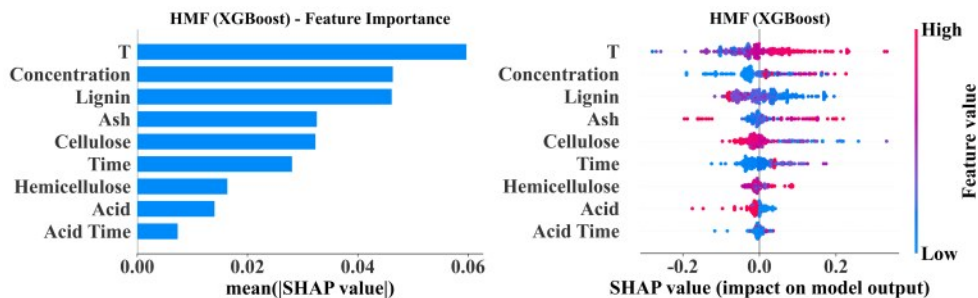
(a) 纤维素保留率特征重要性



(b) 半纤维素去除率特征重要性



(c) 糠醛特征重要性



(d) HMF特征重要性

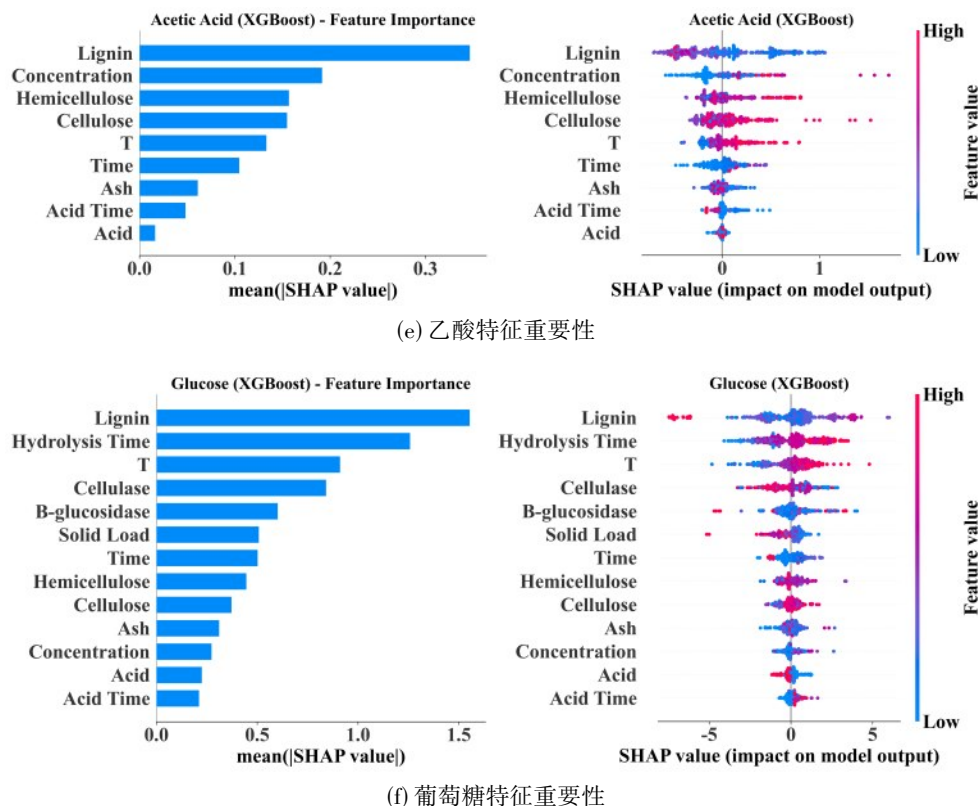


图6 特征重要性分析

Fig.6 Feature importance analysis

2.4 偏依赖图分析

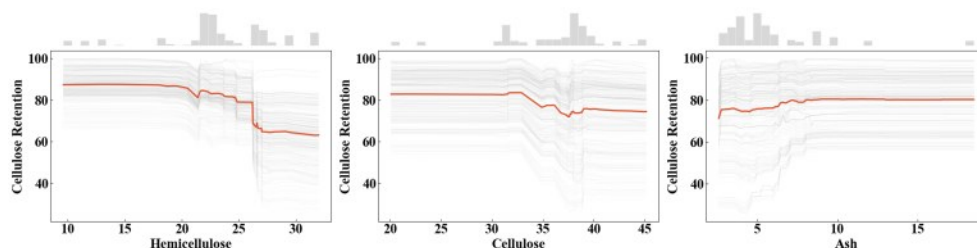
偏依赖图(Partial Dependence Plot, PDP)分析在考虑了模型中所有其他特征的影响后,特定特征的变化对预测结果的平均影响是线性、非线性等关系,可以直观的观察该特征对模型预测结果的贡献程度。

图7(a)中生物质半纤维素含量呈阶梯式下降,在 9.60–18.00 wt% 含量区间保留率为 87.19–87.52 %;当含量增至 21.40 wt%时保留率降至 81.15 %;含量达 27.00 wt%后进一步降至 65.12 %,此后缓降至 63.08 %。该阶梯式下降特征与蒸汽爆破过程中半纤维素的选择性溶出规律一致,半纤维素在蒸汽爆破中优先被水解去除,其脱除程度与预处理强度因子($\log R_0$)密切相关^[50]。生物质纤维素含量的最低点在 37.50 wt%,此前(20.10–32.90 wt%)纤维素保留率稳定于 82.52–83.52 %,在 37.50 wt%处达到最低(保留率 71.99 %),此后稳定在大约 75.00 %。半纤维素和未定型区纤维素优先被去除后,结晶纤维素的抗降解性使其在固相中相对富集^[50, 51]。图7(b)蒸汽爆破温度转折在 169.80 °C,此前(140–160 °C)半纤维素去除率从 61.19 %缓慢升

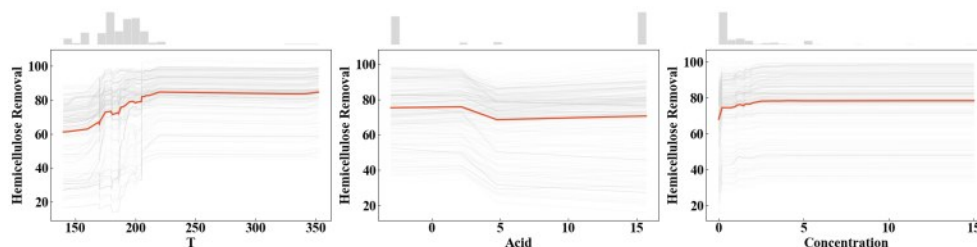
至 62.87 %,此后持续上升至 82.00 % (205.20 °C),此后半纤维素去除率变化平缓(82–85 %)。温度升高,半纤维素乙酰基断裂释放的乙酸使体系 PH 降低,自催化效应增强,促进半纤维素溶出^[52]。强酸(硫酸, pK_a : -3)去除率范围在 75–76 %,到弱酸(乙酸, pK_a : 4.76)降至 68.58 %,表明强酸对不同生物质半纤维素去除效果优于弱酸。硫酸等强酸在水中完全解离,可提供相对于弱酸更高的氢离子浓度,催化效率更高^[37]。图7(c)和图7(d)两者对蒸汽爆破温度的相应规律相似,转折均在 170 °C附近,此前预测值较为平缓(糠醛 0.23 wt%、HMF 0.11 wt%),170 °C后缓慢上升,糠醛在 195 °C、HMF 在 197.85 °C再次上升,之后趋于平稳。糠醛整体量约为 HMF 的 2 倍,对温度的敏感性更高。两者对酸浓度的响应则差异明显,糠醛在酸浓度为 0.25 wt%处上升至 0.46 wt%,又在酸浓度为 2.10 wt%处提升至 0.58 wt%,在酸浓度为 2.70 wt%达到最高(0.73 wt%)后回落稳定约 0.61 wt%,HMF 则对浓度变化不敏感,全程从 0.12 wt%缓慢上升至 0.29 wt%,无明显转折点。正交试验研究表明,糠醛的生成虽以温度为主导因素,但酸浓度对其产率仍有影响,而 HMF 的生成主

要受温度控制,对酸浓度变化的响应不显著^[53]。图7(e)酸浓度转折点在2.70 wt%,此前(0–2.55 wt%)乙酸含量缓慢从0.91 wt%增至1.40 wt%,酸浓度为2.70 wt%时提升至1.77 wt%,当酸浓度达到3.00 wt%时乙酸含量最高(2.21 wt%)。图7(f)展示了酶解时间和蒸汽爆破温度对酶解葡萄糖产量的影响,结果表明酶解时间的转折点在70.55 h,此前缓慢从20.55 g/L增至21.68 g/L,超过70.55 h后两次增加,

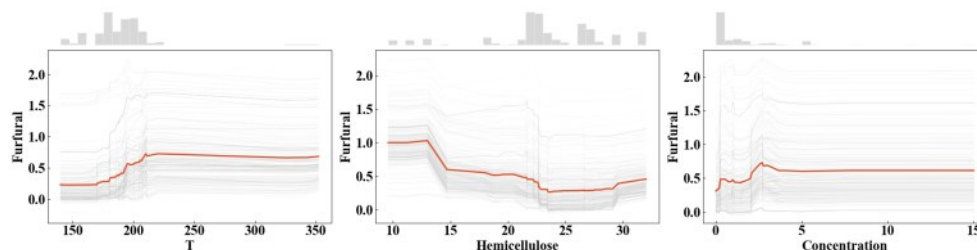
至80 h后稳定在24.41 g/L。长时间的酶解使酶逐渐渗透至纤维素内部,但随着位点趋于饱和,水解速率下降并最终进入平台期^[54]。温度的转折在169.80 °C,此前预测值约20.35 g/L,温度增加至200 °C达到23.54 g/L,随后缓慢增加。蒸汽爆破温度升高增强了半纤维素溶出和纤维结构破坏,增大了纤维素对纤维素酶的可及性,从而提高酶解葡萄糖效率^[50]。



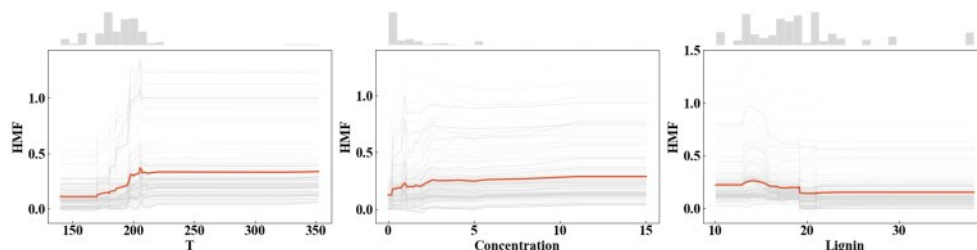
(a) 纤维素保留率单特征偏依赖图



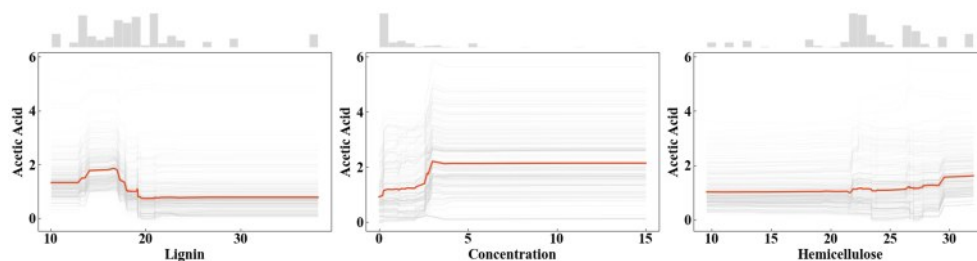
(b) 半纤维素去除率单特征偏依赖图



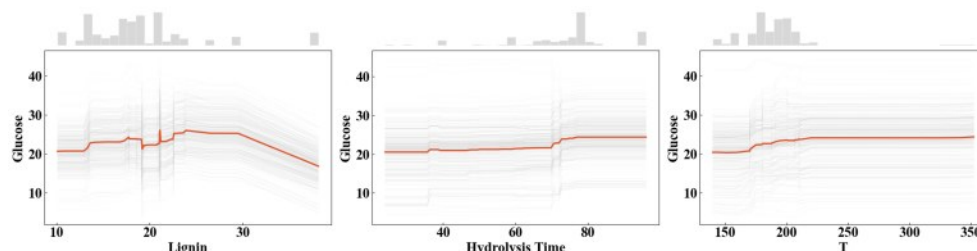
(c) 糠醛单特征偏依赖图



(d) HMF单特征偏依赖图



(e) 乙酸单特征偏依赖图



(f) 葡萄糖单特征偏依赖图

图7 单特征偏依赖图

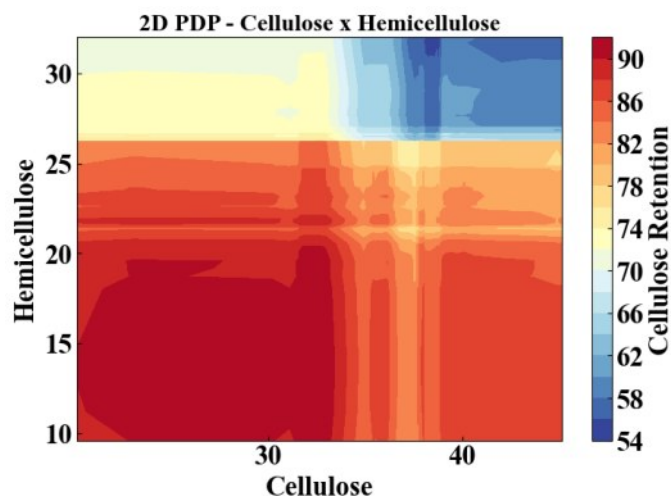
Fig.7 single feature dependence diagram

为解析关键特征对模型输出的非线性影响,采用二维偏依赖图(2D PDP)表示不同特征组合与目标变量之间的响应关系,结果如图8所示。图8(a)展示了生物质纤维素和半纤维素含量对预处理后固体残渣纤维素保留率的联合影响。当生物质半纤维素含量低于18.00 wt%、纤维素含量约为31.80 wt%时,纤维素保留率达到最高预测值91.73%;随着生物质半纤维素含量升高,纤维素保留率总体下降,表明原料组成影响纤维素选择性保留。研究表明不同物种在相同预处理条件下半纤维素去除率和纤维素保留率差异显著^[50, 52]。Auxenfans等人对多种原料的比较研究表明,草本生物质中半纤维素含量较高时,在预处理中更易被溶出,但过高的半纤维素初始含量也意味着更多的乙酰基释放和更强的自催化效应,导致固相纤维素部分热水解损失^[55]。图8(b)揭示了蒸汽爆破温度与pKa对半纤维素去除率的协同调控。低温条件下,不同酸之间的预测差异较大;随着温度升高,该差异逐渐缩小,表明提高温度可部分补偿弱酸催化能力的不足。在本文数据范围内,高温与较低pKa(强酸)的组合对应较高的半纤维素去除率。Rocha等人中试规模上对比了蒸汽爆破(190℃,自催化)与稀硫酸(120℃)预处理甘蔗渣的效果:半纤维素溶出率达89%、纤维素转化率90%,稀硫酸预处理(93%溶出、79%转化),从实验上证实了高温可补偿酸催化能力不足^[56]。图8(c)呈现了糠醛预测值随蒸汽爆破温度和

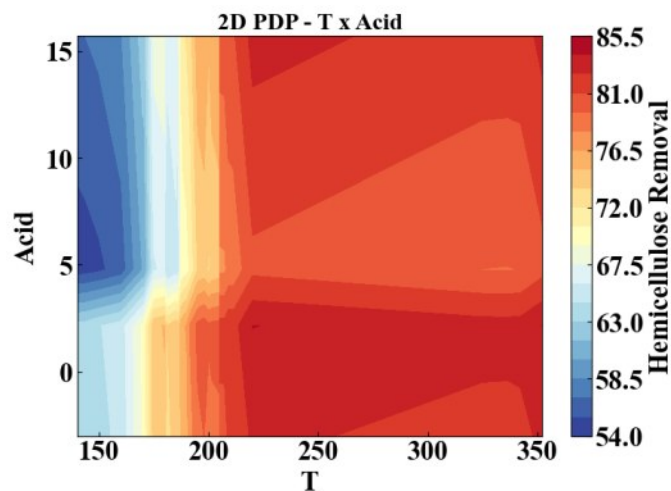
生物质半纤维素含量呈非线性变化。当半纤维素含量为9.60–13.00 wt%时,糠醛预测值较高。当爆破温度为220℃、生物质半纤维素含量为13.00 wt%时,糠醛达到最高预测值1.15 wt%。该趋势可能源于半纤维素水解、戊糖脱水及糠醛进一步降解之间的竞争^[57]。HMF主要受汽爆温度控制,酸浓度的影响相对较弱(图8 d)。温度低于170℃时,HMF预测值维持在0.08–0.11 wt%;超过该温度后迅速升高并趋于稳定。当酸浓度为0.95–1.00 wt%、温度为205.20℃时,HMF达到最高预测值0.50 wt%,表明高温促进己糖脱水,而酸浓度达到一定水平后边际效应减弱。HMF对温度的依赖性和对酸浓度的弱响应性有其动力学基础:Asghari和Yoshida在亚临界水中对果糖脱水的研究表明,温度是决定HMF产率的最显著因素,酸浓度的进一步增加并不能选择性提高HMF生成路径的优势^[58]。酸浓度与生物质半纤维素含量共同影响乙酸生成(图8 e)。乙酸预测值随生物质半纤维素含量升高而增加,但在酸浓度达到一定水平后趋于稳定。这可能是由于较高半纤维素含量提供了更多乙酰基,而酸催化促进其水解脱除。木质素含量与酶解时间对固体残渣酶解葡萄糖得率呈阶段性联合响应(图8 f)。酶解时间低于70 h时,固体残渣酶解葡萄糖产量预测值变化较小;超过70 h后明显升高,并在木质素含量约26.50 wt%、酶解时间77.82 h时达到最高值28.06 g/L。木质素含量过高时固体残渣酶解葡萄糖预测产

量下降,可能与其结构屏蔽及对纤维素酶的非生产性吸附有关。Rahikainen等人系统研究了木质素对纤维素酶的抑制机制,证实木质素通过疏水相互作用和静电作用非生产性吸附纤维素酶,尤其是含碳水化合物结合模块(carbohydrate binding module,

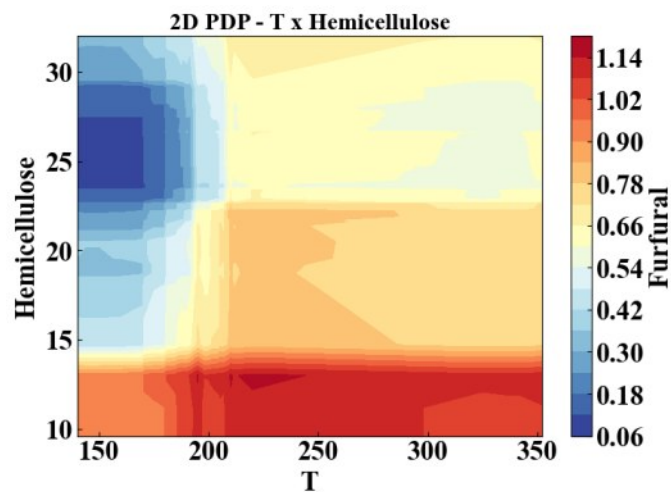
CBM)的纤维二糖水解酶(Cellobiohydrolase, CBH)和内切葡聚糖酶(Endoglucanase, EG)最易受此影响,导致酶解效率下降^[59]。同时过高的木质素含量构成纤维素周围致密物理屏障,阻碍酶分子向底物扩散^[60]。



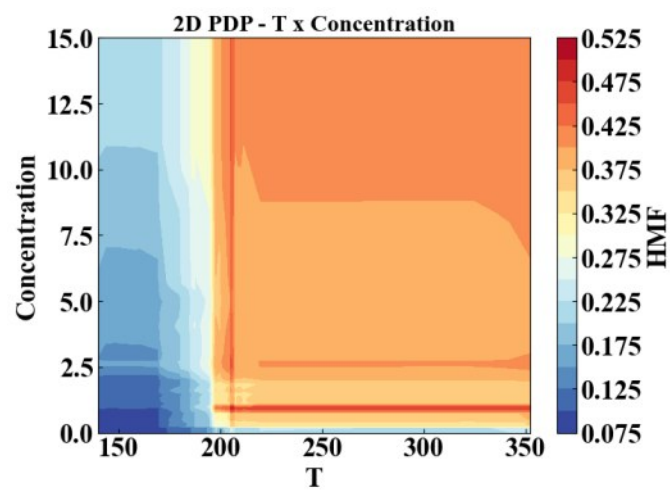
(a) 纤维素双特征偏依赖图



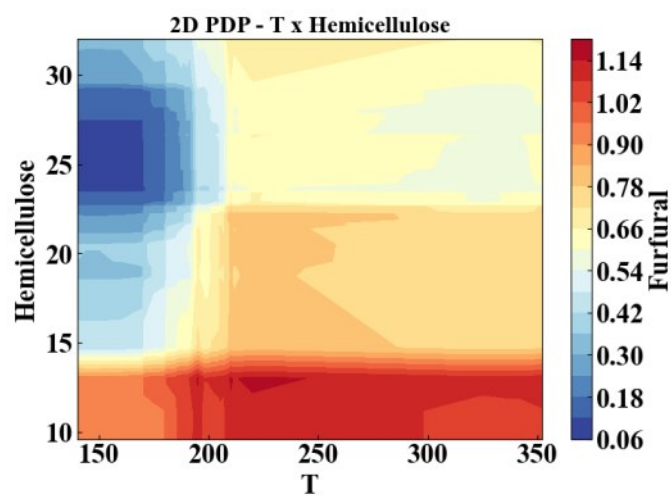
(b) 半纤维素双特征偏依赖图



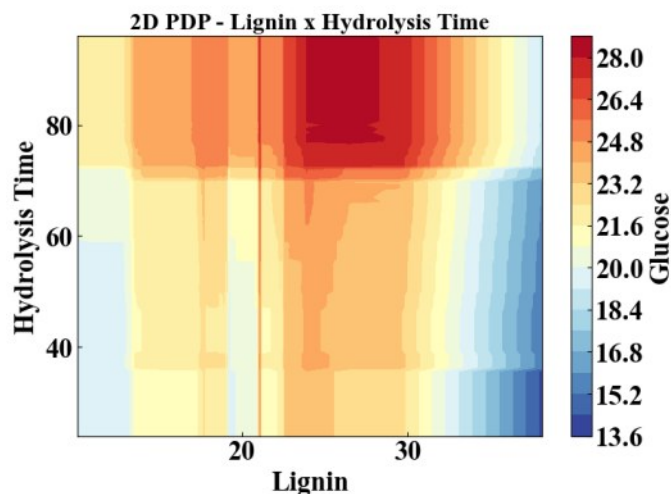
(c) 糠醛双特征偏依赖图



(d) HMF 双特征偏依赖图



(e) 乙酸双特征偏依赖图



(f) 葡萄糖双特征偏依赖图

图8 双重特征偏依赖图

Fig.8 double feature dependence diagram

2.5 页面可视化

虽然现有研究在预测精度方面已取得显著成果,但仍普遍面临复现成本高昂及操作流程冗长等挑战。为克服上述局限性,本文首次开发了以蒸汽爆破为背景、基于Web端的交互平台,旨在通过封装已训练的模型,提升算法的可落地性与用户访问的便捷性。通过运行StrawMLPredictor.exe文件可以直接通过浏览器打开网页。图9展示了交互页面主要操作流程,步骤一:选择预测目标产量,包括纤维素保留率、半纤维素去除率、糠醛、HMF、乙酸、酶解后葡萄糖六个变量(图9(a));步骤二:以纤维素保留率为例,输入原料组成、预处理条件共9个操作参数,原料组成以纤维素30.20%、半纤维素22.30%、木质素15.30%^[32];预处理条件:蒸汽爆破温度170℃、时间5 min;酸性催化剂以pKa=-3、酸浓度5%为例(图9(b))。输出结果纤维素保留率90.49%(图9(c))。与原文结果(91.30%)对比仅相差0.89%^[61]。

3 结论

本研究构建了基于贝叶斯优化的机器学习模型,旨在为生物质蒸汽爆破工程化应用提供智能化解决方法。该模型通过文献数据与实验数据构建数据集,基于RF、XGBoost、ANN和SVM四种机器学习模型进行蒸汽爆破6种产物产量预测,建立的最优模型预测 R^2 均超过0.80,展现出较好的预测性能。

为实现产物的定向调控,采用SHAP分析了模型输入特征的重要性。结果表明:生物质含量是影响固体残渣纤维素保留率的重要因素;半纤维素去除率则与汽爆温度和pKa有关;浸出液糠醛和浸出液HMF则受汽爆温度和酸浓度影响最大;乙酸受pKa和生物质影响较大,木质素含量、酶解时间和汽爆温度是影响固体残渣酶解葡萄糖的重要因素。

通过PDP进一步分析,以固体残渣纤维素保留率最大为目标产物,最佳生物质条件为低半纤维素含量(<18 wt%);以固体残渣半纤维素去除率最大为目标产物,最佳工艺条件是高温(>170℃)强酸;以浸出液糠醛产量最小为目标,最佳生物质条件是低半纤维素含量;以浸出液HMF最小为目标产物,最佳工艺条件是低浓度酸和高温;以浸出液乙酸产量最小为目标,最佳生物质条件是低半纤维素含量;以固体残渣酶解葡萄糖产量最大为目标产物,最佳工艺条件参数为大于70 h的酶解时间。本研究还建立了基于Web的使用界面,便于模型在实践中应用,从而展现了酸处理蒸汽爆破机器学习在相关工程应用上的价值,更好的促进实际应用。受限于生物质种类、酸浓度、酸种类等样本数据量,后续仍需不断丰富数据库,进一步完善和验证模型,为未来开展生物质蒸汽爆破工程化应用提供新方法。



(a) 操作步骤一



(b) 操作步骤二



(c) 模型预测结果

图9 页面可视化

Fig.9 Interface Visualization

参考文献

- Giummarella N, Pu Y Q, Ragauskas A J, et al. A critical review on the analysis of lignin carbohydrate bonds[J]. Green Chemistry, 2019, **21**(7): 1573–1595.
- Tu W C, Hallett J P. Recent advances in the pretreatment of lignocellulosic biomass[J]. Current Opinion in Green and Sustainable Chemistry, 2019, **20**: 11–17.
- 何晓琴, 李苇舟, 李富华, 等. 蒸汽爆破预处理在农产品加工副产物综合利用中的应用[J]. 食品与发酵工业, 2019, **45**(8): 252–257.
- He X Q, Li W Z, Li F H, et al. Application of steam-explosion pretreatment in utilizing agricultural by-products[J]. Food and Fermentation Industries, 2019, **45**(8): 252–257.
- Guerrero A B, Ballesteros I, Ballesteros M. Optimal conditions of acid-catalysed steam explosion pretreatment of banana lignocellulosic biomass for fermentable sugar production[J]. Journal of Chemical Technology & Biotechnology, 2017, **92**(9): 2351–2359.
- Kumar V, Yadav S K, Kumar J, et al. A critical review on current strategies and trends employed for removal of inhibitors and toxic materials generated during biomass pretreatment[J]. Bioresource Technology, 2020, **299**: 122633.
- Xie T T, Fan M S. Machine learning models for predicting enzymatic hydrolysis yields of lignocellulosic biomass after various pretreatments[J]. Industrial Crops and Products, 2025, **235**: 121644.
- Zhang Y, Zhao T, Xiao L, et al. Machine learning prediction of glucose production from lignocellulosic biomass through various pretreatment methods[J]. Biomass and Bioenergy, 2026, **207**: 108811.
- Huang X Y, Zhang X, He Y, et al. Leveraging machine learning for acid catalyzed steam explosion pretreatment: Towards supporting fermentation by the trade-off between glucose and inhibitors[J]. Journal of Cleaner Production, 2024, **448**: 141530.
- Zhang Y, Feng Y J, Ren Z H, et al. Tree-based machine learning model for visualizing complex relationships between biochar properties and anaerobic digestion[J]. Bioresource Technology, 2023, **374**: 128746.
- Liu M X, Tao J Y, Mu L, et al. Machine learning-driven predictions of biochar yield and NPK composition: insights into biomass pyrolysis with data augmentation and model interpretability[J]. Carbon Research, 2025, **4**(1): 62.
- Huang X, Bai X P, Yang Y F, et al. Machine learning-based prediction and optimization of heavy metal adsorption performance of biochar[J]. Forests, 2026, **17**(3): 326.
- Sun Y G, Ma Y L, Wang L Q, et al. Physicochemical properties of corn stalk after treatment using steam explosion coupled with acid or alkali[J]. Carbohydrate Polymers, 2015, **117**: 486–493.
- Wang L, Fan X G, Tang P, et al. Xylitol fermentation using hemicellulose hydrolysate prepared by acid pre-impregnated steam explosion of corncob[J]. Journal of Chemical Technology & Biotechnology, 2013, **88**(11): 2067–2074.
- 朱均均, 杨金龙, 张玲玲, 等. 稀酸蒸汽爆破玉米秸秆预水解液的脱毒及乙醇发酵[J]. 林产化学与工业, 2014, **34**(05): 60–66.
- Zhu J J, Yang J L, Zhang L L, et al. Detoxification and Ethanol Fermentation of the Acid-catalysed Steam-exploded Corn Stover Pre-hydrolyzate[J]. Chemistry and Industry of Forest Products, 2014, **34**(5), 60–66.
- Bondesson P M, Galbe M, Zacchi G. Ethanol and biogas production after steam pretreatment of corn stover with or without the addition of sulphuric acid[J]. Biotechnology for Biofuels, 2013, **6**(1): 11.
- Fan F M, Liu H X, Xia C X, et al. Study of the pre-treatment technology of corn stover by two-step steam explosion[J]. Environment, Development and Sustainability, 2025.
- Zimbardi F, Viola E, Nanna F, et al. Acid impregnation and steam explosion of corn stover in batch processes[J]. Industrial Crops and Products, 2007, **26**(2): 195–206.
- 苏昊. 蒸汽爆破秸秆半纤维素湿热转化机制的研究[D]. 天津: 天津科技大学, 2020.
- Su H. Study on hydrothermal conversion mechanism of corn stalk hemicellulose by steam explosion[D]. Tianjin: Tianjin University

- of Science & Technology, 2020.
- [19] Wang W, Ling H Z, Zhao H. Steam explosion pretreatment of corn straw on xylose recovery and xylitol production using hydrolysate without detoxification[J]. *Process Biochemistry*, 2015, **50**(10): 1623–1628.
- [20] Lizasoain J, Trulea A, Gittinger J, et al. Corn stover for biogas production: Effect of steam explosion pretreatment on the gas yields and on the biodegradation kinetics of the primary structural compounds[J]. *Bioresource Technology*, 2017, **244**: 949–956.
- [21] Zhao C, Deng L, Fang H, et al. Microbial oil production by *Mortierella isabellina* from corn stover under different pretreatments[J]. *RSC Advances*, 2017, **7**(89): 56239–56246.
- [22] Feng X T, Yao Y N, Xu N, et al. Pretreatment affects profits from xylanase during enzymatic saccharification of corn stover through changing the interaction between lignin and xylanase protein[J]. *Frontiers in Microbiology*, 2021, **12**: 754593.
- [23] Liu Z H, Chen H Z. Xylose production from corn stover biomass by steam explosion combined with enzymatic digestibility[J]. *Bioresource Technology*, 2015, **193**: 345–356.
- [24] Chen W H, Pen B L, Yu C T, et al. Pretreatment efficiency and structural characterization of rice straw by an integrated process of dilute–acid and steam explosion for bioethanol production[J]. *Bioresource Technology*, 2011, **102**(3): 2916–2924.
- [25] Semwal S, Raj T, Kumar R, et al. Process optimization and mass balance studies of pilot scale steam explosion pretreatment of rice straw for higher sugar release[J]. *Biomass and Bioenergy*, 2019, **130**: 105390.
- [26] Chen W H, Tsai C C, Lin C F, et al. Pilot–scale study on the acid–catalyzed steam explosion of rice straw using a continuous pretreatment system[J]. *Bioresource Technology*, 2013, **128**: 297–304.
- [27] Gaur R, Agrawal R, Kumar R, et al. Evaluation of recalcitrant features impacting enzymatic saccharification of diverse agricultural residues treated by steam explosion and dilute acid[J]. *RSC Advances*, 2015, **5**(75): 60754–60762.
- [28] Semwal S, Sivagurunathan P, Sattlewal A, et al. An efficient and cost–effective pretreatment of rice straw using steam explosion: a pilot scale experience[J]. *Waste and Biomass Valorization*, 2024, **15**(4): 1975–1986.
- [29] Lotfi Aski A, Borghei A, Zenouzi A, et al. Effect of steam explosion on the structural modification of rice straw for enhanced biodegradation and biogas production[J]. *BioResources*, 2019, **14**(1): 464–485.
- [30] Linde M, Jakobsson E L, Galbe M, et al. Steam pretreatment of dilute H₂SO₄–impregnated wheat straw and SSF with low yeast and enzyme loadings for bioethanol production[J]. *Biomass and Bioenergy*, 2008, **32**(4): 326–332.
- [31] Alvira P, Negro M J, Ballesteros I, et al. Steam explosion for wheat straw pretreatment for sugars production[J]. *Bioethanol*, 2016, **2**(1): 66–75.
- [32] Ballesteros I, Negro M J, Oliva J M, et al. Ethanol Production From Steam–Explosion Pretreated Wheat Straw[J]. *Twenty–Seventh Symposium on Biotechnology for Fuels and Chemicals*, 2006: 496–508.
- [33] Liu M R, Hui L F. Acetic acid catalyzed steam explosion for improving the sugar recovery of wheat straw[J]. *BioResources*, 2014, **9**(3): 4703–4709.
- [34] Cui L, Liu Z, Si C L, et al. Influence of steam explosion pretreatment on the composition and structure of wheat straw[J]. *BioResources*, 2012, **7**(3): 4202–4213.
- [35] Pal S, Joy S, Kumbhar P, et al. Pilot–scale pretreatments of sugarcane bagasse with steam explosion and mineral acid, organic acid, and mixed acids: synergies, enzymatic hydrolysis efficiencies, and structure–morphology correlations[J]. *Biomass Conversion and Biorefinery*, 2017, **7**(2): 179–189.
- [36] Pitarello A P, da Fonseca C S, Chiarello L M, et al. Ethanol production from sugarcane bagasse using phosphoric acid–catalyzed steam explosion[J]. *Journal of the Brazilian Chemical Society*, 2016, **27**(10): 1889–1898.
- [37] Fockink D H, Sánchez J H, Ramos L P. Comprehensive analysis of sugarcane bagasse steam explosion using autocatalysis and dilute acid hydrolysis (H₃PO₄ and H₂SO₄) at equivalent combined severity factors[J]. *Industrial Crops and Products*, 2018, **123**: 563–572.
- [38] do Espírito Santo M C, Cardoso E B, Guimaraes F E G, et al. Multifaceted characterization of sugarcane bagasse under different steam explosion severity conditions leading to distinct enzymatic hydrolysis yields[J]. *Industrial Crops and Products*, 2019, **139**: 111542.
- [39] Iroba K L, Tabil L G, Sokhansanj S, et al. Pretreatment and fractionation of barley straw using steam explosion at low severity factor[J]. *Biomass and Bioenergy*, 2014, **66**: 286–300.
- [40] Dai N H, Huynh K T T, Nguyen T A D, et al. Hydrothermal and steam explosion pretreatment of bambusa stenostachya bamboo[J]. *Waste and Biomass Valorization*, 2021, **12**(7): 4103–4112.
- [41] Yuan Z Y, Li G D, Wei W Q, et al. A comparison of different pre–extraction methods followed by steam pretreatment of bamboo to improve the enzymatic digestibility and ethanol production[J]. *Energy*, 2020, **196**: 117156.
- [42] Hu Y H, Wu R Y, Lin Y C, et al. A novel MissForest–based missing values imputation approach with recursive feature elimination in medical applications[J]. *BMC Medical Research Methodology*, 2024, **24**(1): 269.
- [43] de O Buanafina M M. Feruloylation in grasses: current and future perspectives[J]. *Molecular Plant*, 2009, **2**(5): 861–872.
- [44] Kootstra A M J, Beftink H H, Scott E L, et al. Comparison of dilute mineral and organic acid pretreatment for enzymatic hydrolysis of wheat straw[J]. *Biochemical Engineering Journal*, 2009, **46**(2): 126–131.
- [45] Lloyd T A, Wyman C E. Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids[J]. *Bioresource Technology*, 2005, **96**(18): 1967–1977.
- [46] Świątek K, Gaag S, Klier A, et al. Acid hydrolysis of lignocellulosic biomass: sugars and furfurals formation[J]. *Catalysts*, 2020, **10**(4): 437.
- [47] Chen H Z, Liu Z H. Steam explosion and its combinatorial pretreatment refining technology of plant biomass to bio–based products[J]. *Biotechnology Journal*, 2015, **10**(6): 866–885.
- [48] Martín C, Wu G C, Wang Z, et al. Formation of microbial inhibitors in steam–explosion pretreatment of softwood impregnated with sulfuric acid and sulfur dioxide[J]. *Bioresource Technology*, 2018, **262**: 242–250.
- [49] De Clercq D, Wen Z G, Fei F, et al. Interpretable machine

- learning for predicting biomethane production in industrial-scale anaerobic co-digestion[J]. *Science of the Total Environment*, 2020, **712**: 134574.
- [50] Jacquet N, Maniet G, Vanderghem C, et al. Application of steam explosion as pretreatment on lignocellulosic material: a review[J]. *Industrial & Engineering Chemistry Research*, 2015, **54**(10): 2593–2598.
- [51] Sun X F, Xu F, Sun R C, et al. Characteristics of degraded cellulose obtained from steam-exploded wheat straw[J]. *Carbohydrate Research*, 2005, **340**(1): 97–106.
- [52] Sarker T R, Pattnaik F, Nanda S, et al. Hydrothermal pretreatment technologies for lignocellulosic biomass: a review of steam explosion and subcritical water hydrolysis[J]. *Chemosphere*, 2021, **284**: 131372.
- [53] Fa Xian J, Yong X U, Jun Jun Z H U, et al. Degradation of Four Non-lignin Components in Corn Stover by Pretreatment of Dilute Sulfuric Acid[J]. *Chemistry and Industry of Forest Products*, 2015, **35**(2): 38–46.
- [54] Sulzenbacher D, Atzmüller D, Hawe F, et al. Optimization of steam explosion parameters for improved biotechnological use of wheat straw[J]. *Biomass Conversion and Biorefinery*, 2023, **13**(2): 1035–1046.
- [55] Auxenfans T, Crônier D, Chabbert B, et al. Understanding the structural and chemical changes of plant biomass following steam explosion pretreatment[J]. *Biotechnology for Biofuels*, 2017, **10** (1): 36.
- [56] Rocha G J M, Gonçalves A R, Nakanishi S C, et al. Pilot scale steam explosion and diluted sulfuric acid pretreatments: Comparative study aiming the sugarcane bagasse saccharification [J]. *Industrial Crops and Products*, 2015, **74**: 810–816.
- [57] Hermundsgård D H, Ghoreishi S, Tanase-Opedal M, et al. Maximizing yields of furfural and 5-hydroxymethylfurfural in side streams from steam explosion of lignocellulosic residues[J]. *Biofuels, Bioproducts and Biorefining*, 2025, **19**(2): 319–336.
- [58] Salak Asghari F, Yoshida H. Acid-catalyzed production of 5-hydroxymethyl furfural from d-fructose in subcritical water[J]. *Industrial & Engineering Chemistry Research*, 2006, **45**(7): 2163–2173.
- [59] Rahikainen J L, Martin-Sampedro R, Heikkinen H, et al. Inhibitory effect of lignin during cellulose bioconversion: The effect of lignin chemistry on non-productive enzyme adsorption[J]. *Bioresource Technology*, 2013, **133**: 270–278.
- [60] Ko J K, Um Y, Park Y C, et al. Compounds inhibiting the bioconversion of hydrothermally pretreated lignocellulose[J]. *Applied Microbiology and Biotechnology*, 2015, **99**(10): 4201–4212.
- [61] 钟禹, 李由, 吴越, 等. 低酸喷雾协同蒸汽爆破预处理对小麦秸秆结构及酶水解的影响[J]. *中国造纸学报*, 2022, **37**(4): 16–24.
- Zhong Y, Li Y, Wu Y, et al. Effects of low-acid spraying combined with steam explosion pretreatment on structure and enzyme hydrolysis of wheat straw[J]. *Transactions of China Pulp and Paper*, 2022, **37**(4): 16–24.