

DOI:10.19853/j.zgjsps.1000-4602.2026.11.003

基于机器学习模型优化电凝聚法去除地下水中氟化物

王陈波¹, 王越¹, 余华荣¹, 瞿芳术¹, 运嘉宁²

(1. 广州大学 土木与交通工程学院, 广东 广州 510006; 2. 广东生活环境无害化处理中心有限公司, 广东 广州 510000)

摘要: 本研究提出一种基于机器学习的电凝聚工艺参数优化框架,用于高效去除地下水中的氟化物。通过整合文献中收集的204个试验数据点,构建了涵盖反应条件与处理结果的多维数据集,并系统比较了6种典型机器学习算法的预测性能。结果表明,XGBoost模型在剩余氟浓度和能耗的双重预测任务中表现最佳($R^2_{\text{test}}=0.76\sim0.89$, $\text{MSE}_{\text{test}}\leq 0.0047$),优于KNN、SVR、RF、LightGBM和MLP。基于SHAP的特征重要性分析揭示,初始氟浓度、电解时间和电流密度是决定除氟效果的关键因子,而电解时间、电流密度与阳极面积则主导能耗水平。进一步将训练良好的XGBoost模型嵌入快速精英多目标遗传算法(NSGA-II),实现了“去除率最大化-能耗最小化”的多目标优化,获得了多组帕累托前沿解。验证结果显示,模型预测值与实测值的相对误差在10%~25%之间。该框架有效展示了数据驱动逆向设计的潜力,为电化学水处理工艺的智能化设计提供了新思路,提升了除氟工艺的效率。

关键词: 地下水; 电凝聚; 除氟; 机器学习; 参数优化

中图分类号: TU991 **文献标识码:** A **文章编号:** 1000-4602(2026)11-0016-10

Machine Learning-based Optimization of Electrocoagulation for Fluoride Removal from Groundwater

WANG Chenbo¹, WANG Yue¹, YU Huarong¹, QU Fangshu¹, YUN Jianing²

(1. School of Civil and Transportation Engineering, Guangzhou University, Guangzhou 510006, China; 2. Guangdong Living Environment Harmless Treatment Center Co. Ltd., Guangzhou 510000, China)

Abstract: This study presented a machine learning-based framework for optimizing the electrocoagulation (EC) process parameters to efficiently remove fluoride in groundwater. By integrating 204 experimental data points collected from the literature, a multidimensional dataset encompassing reaction conditions and treatment results was constructed. The performance of six typical machine learning algorithms was systematically compared. The results showed that the XGBoost model performed the best in the dual prediction task of residual fluoride concentration and energy consumption ($R^2_{\text{test}}=0.76\sim0.89$, $\text{MSE}_{\text{test}}\leq 0.0047$), outperforming KNN, SVR, RF, LightGBM, and MLP. Feature importance analysis based on SHAP revealed that initial fluoride concentration, electrolysis time, and current density were the key factors determining fluoride removal efficiency, while electrolysis time, current density, and anode area dominated energy consumption levels. Furthermore, the well-trained XGBoost model was integrated

基金项目: 国家重点研发计划项目(2023YFC3207903)

通信作者: 余华荣 E-mail: huarongyu@gzhu.edu.cn

into a non-dominated sorting genetic algorithm (NSGA-II) for multi-objective optimization of removal rate maximization-energy consumption minimization, yielding multiple Pareto front solutions. Experimental validation showed that the relative error between model predictions and actual measurements ranged from 10% to 25%. This framework effectively demonstrates the potential of data-driven reverse design, providing new insights for the intelligent design of electrochemical water treatment processes, improving fluoride removal efficiency.

Keywords: groundwater; electrocoagulation; fluoride removal; machine learning; parameter optimization

电凝聚(EC)工艺因其去除污染物高效、无需额外化学药剂、设备简单且易于系统集成等优势,近年来受到了广泛关注。已有研究表明,EC工艺能有效去除包括氟化物、砷、磷酸盐以及多种有机和无机污染物在内的各种污染物,并在实际水处理中展现出良好的适用性^[1-3]。为了实现EC工艺的高效、经济运行,合理设计电解反应器并优化运行参数至关重要。然而,EC的除氟效果与能耗水平受电流密度、电解时间、pH、电极间距和电导率等条件的耦合影响,如何在多变量条件下快速获得最优操作参数,仍然是推广应用中的一大挑战。传统研究方法通常依赖单因子试验、正交试验或中试等手段来确定最佳运行条件,不仅耗时费力,还难以全面揭示参数间的非线性关系,限制了EC技术在复杂水处理场景中的应用。因此,亟须开发新的方法来突破试验优化的局限,实现除氟工艺的高效与智能化。

近年来,机器学习(ML)作为一种数据驱动的建筑模型工具,在环境科学与工程领域展现出巨大潜力。与传统依赖机理方程的模型不同,ML能够从历史数据中自动学习复杂的非线性关系,在污染物去除、膜污染诊断以及工艺能耗评估等方面表现出了出色的预测能力^[4-5]。同时,将ML与特征解释方法(如SHAP)相结合,能够提供高精度预测并揭示关键操作参数对反应结果的作用机制,从而提升模型的透明性与可解释性。因此,通过引入ML与多目标优化,解决了传统试错方法的局限性,为EC除氟工艺的高效设计与智能化发展提供了新思路。

1 材料和方法

1.1 数据集的清洗

从已发表的文章中收集了EC的反应条件(即反应器类型、电极类型及其个数与面积、氟初始浓度、电流密度、停留时间、电极间距、反应器体积和

pH),共204个试验数据点^[6-30],采用反应条件为输入数据集,EC出水氟浓度和能耗为输出数据集。为减少单次随机划分带来的偶然性,采用5折交叉验证对模型进行训练与验证。将全部样本随机划分为5个子集,每次选取其中4个子集作为训练集用于模型训练,剩余1个子集作为测试集用于评估模型的泛化性能。该过程循环5次,使每个子集均作为测试集使用一次。最终,以5折交叉验证所得测试集均值作为模型预测的性能指标。

在机器学习建模前,数据预处理至关重要。首先,使用Spearman相关性去除高相关性的特征,避免冗余信息。接着,采用四分位距法(IQR)识别并去除异常值。为了处理缺失值,使用公式法插补能耗缺失值,条件估算插补反应器体积缺失值,其他特征使用K近邻插补(KNN)。对于分类特征,采用one-hot编码转换为数字形式。最后,计算z-score对输入特征和输出目标进行归一化和标准化,以优化建模效果。

1.2 ML模型的开发与评估

为了建立EC过程反应条件与出水剩余氟浓度和能耗的关系模型,采用KNN、多层感知机(MLP)、随机森林(RF)、支持向量回归(SVR)、梯度提升决策树模型(LightGBM)和极限梯度提升(XGBoost),通过Python编程从scikit-learn软件包中获得。数据集划分采用5折交叉验证。使用optuna进行模型的超参数调整。在超参数调整过程中,执行了5折交叉验证方法,以避免过度拟合。为了定量评估模型性能,计算了决定系数(R^2)和均方误差(MSE),以评估模型预测的准确性。

1.3 特征分析

ML模型是“黑箱”模型,为了解释EC过程中输入特征对目标输出的影响,将SHapley加性解释

(SHAP)方法与ML相结合进行特征分析。SHAP方法可以克服“黑箱”模型解释能力的不足,因为它能够计算特征对模型输出的边际贡献,将全局预测分解为局部特征,从而有助于理解这些输入特征如何影响目标变量。

1.4 参数优化与试验验证

为了研究ML模型在实际应用中的有效性并确定电凝聚过程的最佳反应条件,本研究采用基于模型的试验测试方法。为此,使用了广泛应用的优化算法,即快速精英多目标遗传算法(NSGA-II)。NSGA-II通过精英策略防止个体流失,将父代和子代所有个体混合后进行非支配排序。

将基于出水氟浓度和能耗最小化的ML模型嵌入到NSGA-II中,给定污染物的初始浓度和目标约束后,基于ML模型的多目标优化算法可搜索到最佳反应参数,并得到帕累托前沿。为了验证优化结果,选取帕累托前沿的8个解进行EC试验验证。

在本研究中,用于优化的适应度函数、目标函数及约束条件分别如式(1)~(3)所示。

$$f(\text{inputs}) = [F_{\text{ML},1}, F_{\text{ML},2}] = F_{\text{ML}}(\text{inputs}) \quad (1)$$

$$\min[F_{\text{ML},1}, F_{\text{ML},2}] = \min[F_{\text{ML}}(\text{inputs})] \quad (2)$$

$$\text{LB}_i \leq \text{input}_i \leq \text{UB}_i, \quad i \in (1, 2, \dots, n) \quad (3)$$

式中: F_{ML} 为用于关联输入和输出的ML模型; $F_{\text{ML},1}$ 和 $F_{\text{ML},2}$ 为ML模型的两个目标输出。这些输入特征中的7个是可控的,包括电流密度、阳极面积、电极间距、反应器体积、pH、电解时间、电导率。 LB_i 和 UB_i 分别为第*i*个输入参数的下限和上限; n 为输入参数的数量。

2 结果与讨论

2.1 数据集的描述性分析

为了初步了解各输入变量之间的相互关系,并避免冗余特征给模型训练带来无差异的信息,采用Spearman秩相关系数(ρ)对各变量进行了成对相关性分析。参考已有统计学及环境科学研究中对Spearman系数大小的经验划分,将 $|\rho| \geq 0.7$ 视为强相关, $0.3 \leq |\rho| < 0.7$ 视为一般相关, $|\rho| < 0.3$ 视为弱相关。除氟各因素之间的Spearman相关性热图如图1所示。可见,电解时间(min)与极距(cm)之间存在较强的正相关性(Spearman系数为0.74),表明这两个变量可能在氟的去除过程中共同作用,具有较高的相关性。因此,在后续模型开发中,对这两个特征

进行降维或合并可能会提高模型的效率和性能。其余输入特征之间的相关性较弱,大多数变量的Spearman系数介于-0.5~0.6之间,表明它们之间相对独立。这为特征选择提供了依据,建议在开发预测模型时保留这些特征,同时避免冗余特征的干扰。此外,电流密度与电解时间之间存在轻微的负相关(Spearman系数为-0.43),暗示这两个变量在电解反应过程中可能存在一定的互斥效应。

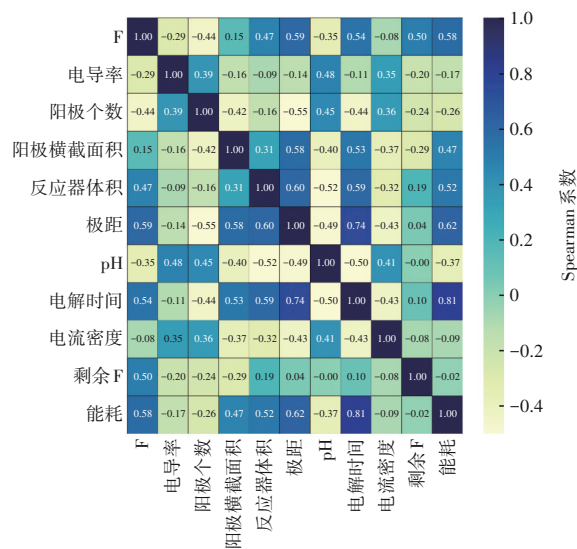


图1 Spearman相关性热图

Fig.1 Spearman correlation heatmap

为了深入理解原始数据的整体结构,进一步使用Python进行了描述性分析。将目标变量按降序排列后,划分为100个等频分区得到的样本分布见图2,每条曲线表示输入变量在各区间的归一化均值变化趋势。分析结果表明,不同特征对剩余氟浓度和能耗的影响存在显著差异。如图2(a)所示,前30个分位区间内,剩余氟浓度较高且能耗较大,表明除氟效果较差,需优化反应条件。随着分位区间增大,氟浓度逐渐下降,能耗减少,除氟效率提高。如图2(b)所示,电导率与氟浓度和能耗呈正相关,高电导率对应较高的氟浓度和能耗,影响反应效率。图2(c)和(d)显示,阳极个数和横截面积的增加在某些区间并未显著提高除氟效率,反而可能导致能耗增加,特别是在高氟浓度区域,需进一步优化阳极选择。图2(e)和(f)显示,反应器体积和极距增大可能导致较低的除氟效率和较高的能耗,需进一步优化反应器设计。图3(g)~(i)表明,pH、电解时间和电流密度对除氟效率和能耗有显著影响。

较低的pH有助于氟去除,但伴随较高能耗;增加电解时间能提高除氟效果,但能耗也增加;增加电流密度能显著提升除氟效率,但同样会导致能耗增

加。各特征与目标变量之间存在显著的非线性关系,传统线性模型难以捕捉其复杂规律,ML回归方法更适合用于后续的EC过程多目标预测^[31-32]。

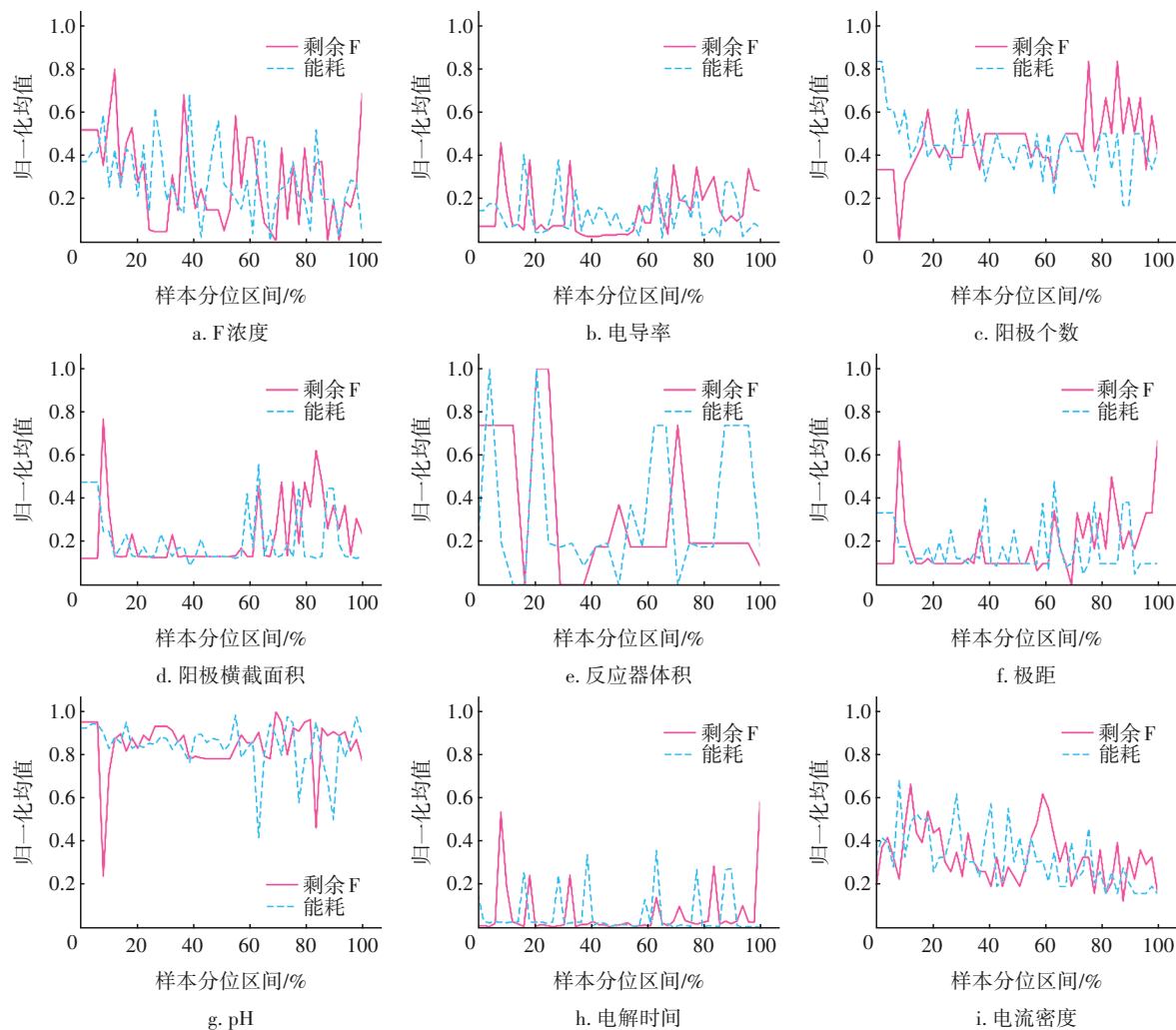


图2 特征数据分布

Fig.2 Feature data distribution map

2.2 模型的开发与验证

在数据清洗后,采用ML算法开发了用于预测EC出水氟离子浓度和处理能耗的模型。由于预处理后的数据已包含明确的标签信息,并假设训练集与测试集在统计特性上具有一致性,采用监督学习方法是合理的。ML方法通常围绕“预测精度”和“模型可解释性”两个核心目标展开,但提升预测精度往往伴随着模型可解释性的下降^[32]。因此,本研究优先选择预测性能较强的判别模型而非生成模型。与生成模型基于联合分布 $P(x, y)$ 进行建模不同,判别模型直接学习决策边界或条件概率 $P(y|x)$,从而实现更有效的类别边界划分。这种方法在面

对未知样本时通常能够提供更高的预测准确度,适合本研究中对预测精度要求较高的任务。

本研究选用了6种预测精度较高的判别模型,即KNN、SVR、XGBoost、RF、LightGBM (LGBM) 和MLP,其性能对比见表1。对于剩余氟浓度,除KNN外,其他5种模型的决定系数均较高,训练集在0.9以上,测试集在0.7以上。所有模型的均方误差(MSE)无论在训练集还是测试集上均小于0.01。尽管所有模型在训练集上的表现优于测试集,但较高的 R^2 和较低的MSE仍然证明了模型预测的准确性。在其他相关研究中, R^2 值在0.6~1.0之间被认为预测值与实际值之间具有较强的相关性^[31]。

表 1 不同优化模型的性能比较

Tab.1 Comparison of performance of different optimization models

项目			XGBoost	RF	LightGBM	SVR	MLP	KNN
F 浓度	R^2	训练集	0.949 0	0.976 0	0.978 0	0.934 0	0.909 0	0.721 0
		测试集	0.762 0	0.763 0	0.846 0	0.785 0	0.708 0	0.579 0
	MSE	训练集	0.001 0	0.000 5	0.000 4	0.001 0	0.001 4	0.005 5
		测试集	0.004 7	0.004 7	0.002 6	0.003 1	0.004 4	0.008 4
能耗	R^2	训练集	0.989 0	0.993 0	0.969 0	0.851 0	0.954 0	0.657 0
		测试集	0.885 0	0.879 0	0.771 0	0.769 0	0.754 0	0.494 0
	MSE	训练集	0.000 2	0.000 1	0.000 6	0.002 6	0.000 8	0.005 8
		测试集	0.001 9	0.002 0	0.004 5	0.004 0	0.004 3	0.008 6

以污染物浓度和反应条件作为输入特征时,基于 6 种不同算法(5 折交叉验证)训练的模型性能见图 3。由图 3(a)可知,SVR 和 MLP 的 R^2 值范围较其他模型更广,提示可能存在过拟合现象,这通常是由模型复杂性与小数据规模不匹配所致。为了减轻过拟合问题,需使用更多的数据来训练深度学习模型。对于能耗预测,表 1 显示 XGBoost、RF 表现优异,测试集 R^2 值均大于 0.8,而 LightGBM、SVR 和 MLP 的表现稍逊,测试集 R^2 值介于 0.7~0.8 之间。KNN 模型的测试集 R^2 值范围变化较大,进一步验证了其在该任务中的不适用性,如图 3(b)所示。综合考虑预测精度、模型稳定性和计算效率,XGBoost 被认为是本研究中最优的模型,适用于剩余氟浓度和能耗的同时预测。XGBoost 作为一种梯度提升框架,结合了加权分位数分裂策略和多线程并行处理机制,不仅在训练效率上表现优异,还具有良好的正则化能力,能够有效避免过拟合问题,特别是在小规模高维数据集上的应用优势尤为明显。尽管 RF 在能耗预测上和 LightGBM 在 F 浓度预测上表现良好,综合两者则 XGBoost 的性能仍略优,这可能是由于 RF 在节点上使用随机特征选择,且不包含复杂的分割算法,尤其对于特征关系复杂的数据集,而 LightGBM 虽然是一个高效的梯度提升模型,但其逐叶树生长策略在小数据集或缺乏正则化时可能导致过拟合,较 XGBoost 更易不稳定。KNN 模型未能有效预测这两项任务,表现最差,因此不建议在类似任务中使用。因此,鉴于电流密度、电解时间、极距、电导率等多参数之间存在显著的非线性耦合关系,本研究最终选用 XGBoost 模型进行 EC 除氟预测。XGBoost 通过集成多棵回归树并在目标函数中加入正则化项,可有效刻画复杂的非线性映射与特征交互,并抑制样本量有限条件下的过拟合,从

而在除氟效率与能耗预测中表现出较高的精度与稳定性,成为小数据集建模的最优选择,为后续更复杂预测系统的构建提供了可靠的技术支持。

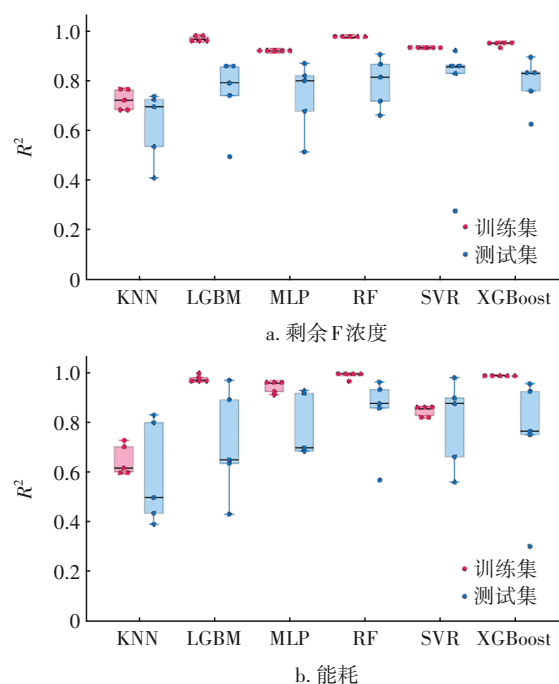


图 3 基于 6 种不同算法(5 折交叉验证)训练的模型性能
Fig.3 Performance of models trained using six different algorithms (5-fold cross-validation)

2.3 特征重要性分析

为定量评估各操作参数对除氟效率和能耗预测的贡献,采用 SHAP 方法解释 XGBoost 模型。SHAP 通过比较特征加入或剔除时模型输出的变化,计算出样本级 SHAP 值,并以各特征在全部样本上的平均绝对 SHAP 值作为全局重要性指标,确定影响剩余氟浓度(剩余 F)和能耗的主要参数。尽管排列重要性方法能够有效识别对模型预测性能影响较大的关键特征,但其本质是一种全局层面的“黑箱”评估手段,仅能衡量特征的重要程度,无法

揭示特征取值变化与预测结果之间的因果关联,也无法区分特征对目标变量的调控方向。换句话说,置换重要性可以衡量特征的重要程度,但无法判断特征对预测结果的具体作用是促进还是抑制。相比之下,SHAP方法能够针对每个特征的具体取值,提供局部层面的重要性分解,从而实现对ML模型预测过程的机制性解释。SHAP不仅有助于量化单个特征的贡献大小,还能进一步揭示其对目标变量的正向或负向影响。

在剩余氟浓度预测中,初始氟浓度、电解时间、电流密度、电导率、极距、阳极横截面积、pH、阳极个数、阳极Al、反应器体积、阳极Fe、阳极Fe+Al、阴极Al、阴极Fe、连续运行、间歇运行的平均绝对SHAP值分别为0.719、0.522、0.354、0.333、0.321、0.242、0.217、0.204、0.125、0.110、0.066、0.036、0.032、0.024、0.020、0.018(R^2 降序排列)。

平均绝对SHAP值作为特征重要性的度量,表明对于剩余F浓度预测任务,初始浓度被认为是最重要的输入特征。较高的特征值(红色)对应正的SHAP值,而较低的特征值(蓝色)则对应负的SHAP值(见图4),这表明初始污染物浓度与剩余F浓度之间存在较强的正相关性。其次,电解时间和电流密度是第二和第三重要的输入特征,说明较长的处理时间和较高的电流密度有助于降低剩余F浓度,从而体现了电化学反应增强传质过程的正效应。

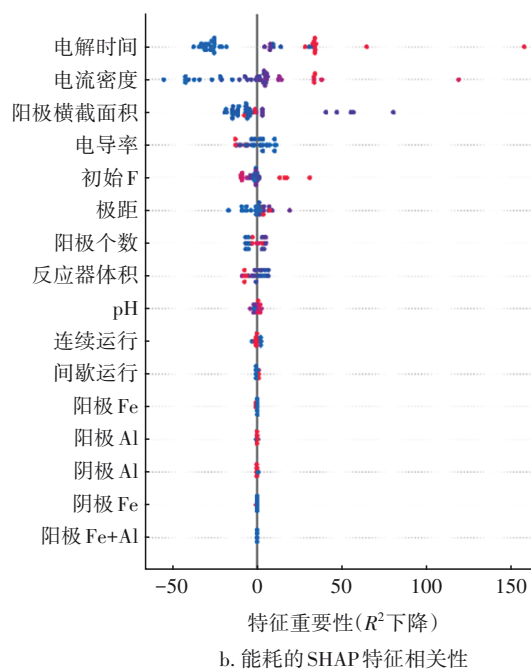
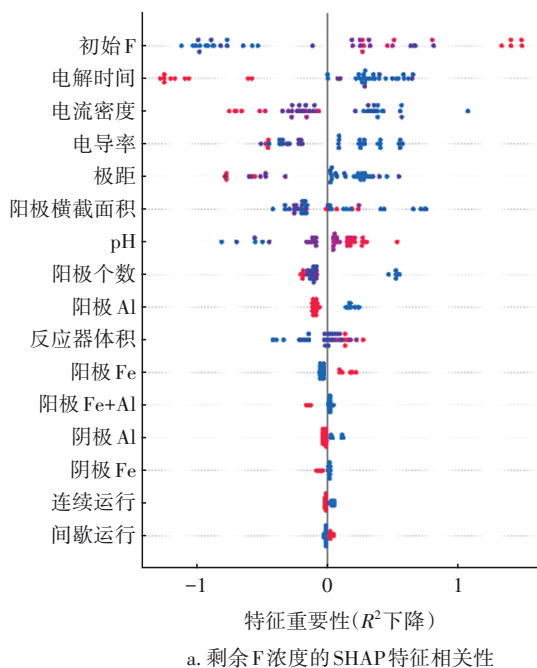


图4 XGBoost模型的SHAP特征重要性与相关性分析

Fig.4 SHAP-based feature importance and dependence analysis for trained XGBoost model

相比之下,在能耗预测任务中,电解时间、电流密度、阳极横截面积、电导率、初始氟浓度、极距、阳极个数、反应器体积、pH、连续运行、间歇运行、阳极Fe、阳极Al、阴极Al、阴极Fe、阳极Fe+Al的平均绝对SHAP值分别为29.955、20.763、14.164、5.834、4.795、4.594、4.369、3.377、1.318、0.761、0.498、0.277、0.165、0.159、0.056、0.029(R^2 降序排列),即电解时间、电流密度和阳极横截面积是最主要的影响因素,其平均绝对SHAP值均远高于其他变量,这也符合能耗与处理时间、电流量成正比的物理意义。图4(b)进一步揭示了这些特征的局部贡献方向。较高的电解时间和电流密度会导致较高的能耗;而在固定电流和反应器体积条件下,阳极面积的增大会降低单位面积的电流密度,从而减小极化电压和能量损失,降低系统的总能耗。这一趋势与原始数据集特征选择中的结果一致。原始数据集的特征选择中的排列重要性排序是一种评估特征对模型预测贡献度的方法,其核心思想是:通过随机打乱某个特征值,观察模型性能下降程度,性能下降越多说明该特征越重要。上述分析结果与原始数据集特征选择中的排列重要性排序具有一致性。

为了进一步解释和验证特征重要性的可靠性,选取剩余F浓度和能耗作为计算SHAP值的实例,

结果见图5。在剩余F浓度预测中,初始氟浓度是最具影响力的特征,对模型输出具有显著的负向贡献($\text{SHAP}=-0.78$),表明较高的进水氟浓度有助于提升去除效果,从而降低出水浓度。电解时间和电极间距均表现出正向作用(SHAP 为0.23和0.16),说明延长反应时间和适当的电极间距能够进一步增强除氟效果。此外,阴极Al和其余10个特征在SHAP分析中呈现出较小的贡献幅度,表明在当前数据分布和建模框架下,其对剩余F浓度预测的边际影响相对有限。然而,这些特征仍可能通过非线性关系或特征交互,对模型输出产生间接调节作用。因此,从整体趋势上仍对预测结果起到补充作用。

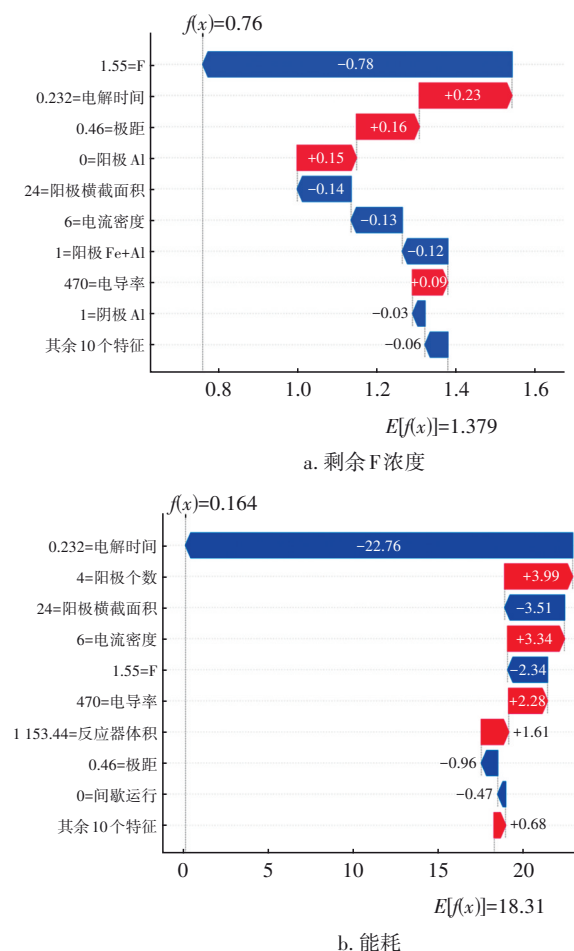


图5 XGBoost模型的SHAP值计算实例

Fig.5 Example of SHAP value calculation for XGBoost model

在能耗预测中,电解时间在所有特征中表现出最大的SHAP绝对值($|\text{SHAP}| = 22.76$),是影响能耗预测结果的主导因素。在所选代表性样本中,电解时间对应的SHAP值为负,说明在该工况下其对预

测能耗产生了负向贡献。然而,从整体样本分布来看,电解时间与能耗之间呈现出显著的正相关趋势,是决定系统能效的关键因素。反应器体积(1.61)和电流密度(3.34)也表现出正向贡献,说明在高电流密度和大反应体积条件下,能耗明显增加。相比之下,电极间距(-0.96)和电极横截面积(-3.51)则表现出负向贡献,意味着在特定条件下合理优化电极结构有助于降低能耗。电极种类(Fe-Al、Al)的作用幅度较小,但其正负差异提示了不同材料在能耗调控上的差异。

2.4 基于ML的EC工艺参数优化及验证

采用NSGA-II算法对EC除氟工艺进行双目标优化(最小化出水剩余氟浓度与单位体积能耗)。优化问题包含8个连续决策变量:电导率(100~2500 $\mu\text{S}/\text{cm}$)、阳极个数(1~8个)、阳极面积(5~200 cm^2)、反应器体积(100~6000 mL)、极距(0.1~2.5 cm)、pH(0.1~12)、电解时间(0.1~150 min)和电流密度(0.1~20 mA/cm^2),以及2个目标函数,均为最小化问题。NSGA-II算法采用实数编码,种群规模设为100,最大进化代数数为500,交叉算子为模拟二进制交叉(SBX),交叉概率 $P_c=0.9$,变异算子为多项式变异(PM),变异概率 $P_m=0.125$,并启用重复个体剔除以维持种群多样性。

在优化过程中,假设污染物浓度和电极材料保持不变,将以下反应条件作为可调变量进行优化:电流密度、阳极面积、电极间距、反应器体积、pH、电解时间和电导率。随后,将已训练完成的XGBoost模型嵌入NSGA-II的初始化参数的适应度函数中,模型输入为优化变量,输出则作为适应度函数的返回值,以指导NSGA-II算法评估每组操作条件的优劣,从而实现剩余F浓度和能耗的最优平衡。优化过程中,NSGA-II算法生成了多个优选解,这些解构成了优化后的帕累托前沿。表2展示了在初始氟浓度为5 mg/L条件下,以Fe为阳极、铝为阴极时,通过NSGA-II算法优化得到的8组反应条件及其性能数据。这些组合涵盖了不同电流密度、阳极面积、pH、电解时间等变量,既能满足处理要求,又能最大化能效、降低系统能耗,为实际应用中的不同需求提供多样化方案。通过这些优化结果,在满足剩余氟浓度要求的基础上,能够有效平衡能效与能耗,为电解反应器的优化设计与操作提供科学依据。

表2 基于模型的逆过程设计方法的EC处理优化参数

Tab.2 EC optimization parameters of model-based reverse process design method

电导率/ ($\mu\text{S}/\text{cm}$)	阳极个数/ 个	阳极面积/ cm^2	反应器体 积/ mL	极距/ cm	pH	电解时间/ min	电流密度/ (mA/cm^2)	剩余 F/ (mg/L)	能耗/ ($\text{kW}\cdot\text{h}/\text{m}^3$)
1 146	7	31	4 703	0.75	5.90	71	0.12	0.617	0.01
659	6	31	5 574	0.51	8.00	67	0.76	0.548	0.10
654	7	31	4 770	0.75	7.45	70	6.46	0.280	0.14
654	7	30	5 669	0.79	7.89	74	6.46	0.217	0.23
764	7	30	5 988	0.75	7.54	70	6.39	0.179	0.75
724	7	30	5 686	0.68	7.86	65	6.39	0.117	1.36
912	8	33	5 668	0.51	7.67	66	6.92	0.082	3.31
982	7	31	5 194	0.56	7.97	70	6.82	0.059	5.40

在上述 8 组试验中,EC 出水 F 浓度分别为 0.636、0.516、0.324、0.238、0.212、0.147、0.110、0.077 mg/L。可见,模型预测值与试验值呈现出相似的变化趋势,尤其是在前几组样本(样本 1~3)中,预测值与实际值之间的差异较小。然而,随着剩余 F 浓度的降低,预测值与实际值之间的差距逐渐加大,尤其是在后续组别,预测值的偏差显著增大。模型的相对误差分别为 2.99%、6.20%、13.58%、8.82%、15.57%、20.41%、25.45%、23.38%。这些相对误差揭示了不同试验条件下,模型的预测精度和误差的波动性。值得注意的是,随着剩余 F 浓度的降低,模型的预测误差逐渐增大。这一现象可能与以下几个因素有关:首先,当电解反应接近理想终点时,剩余 F 浓度较低,反应速率和电解时间等因素的微小变化会对浓度的预测结果产生较大的影响,从而放大了预测误差。在这种情况下,模型对微小变化的敏感性较高,因此产生了较大的相对误差。其次,低浓度样本的波动性较大,尤其是在极端情况下,试验数据可能出现较大的噪声。由于模型的拟合能力主要依赖于训练数据的分布,当训练数据中的低浓度样本较少时,模型在处理这些特殊情况时的预测能力会有所不足。从这些相对误差的变化趋势来看,尽管大多数试验样本的预测结果与实际值较为接近,但在某些试验组(尤其是低浓度组)中,预测值与实际值之间的偏差较大。这表明,尽管 XGBoost 模型在大部分情况下具有较高的预测精度,但在处理低浓度样本时预测准确性有所下降。为进一步提升模型在低浓度样本上的预测能力,未来的研究可以通过优化模型参数、增加低浓度样本的训练数据以及应用数据增强技术,来提高模型的鲁棒性和适应性。嵌合了 XGBoost 模型的

NSGA-II 优化算法在本研究中表现出色,在大多数试验条件下能够较好地拟合实际数据,预测精度较高。然而,针对低浓度样本的优化仍然是未来研究的重点。通过进一步优化和数据增强,模型有望在不同试验场景中展现出更高的预测准确性,特别是在低浓度区间的表现将得到显著改善。由于数据来源涵盖多种反应器结构与运行工况,模型在该参数空间内具有一定泛化能力,可为类似水质条件下的 EC 除氟工艺优化提供参考。然而,现有文献普遍对共存离子、有机物等水质背景信息报道有限,本模型尚未考虑这些因素;当目标水体的水质组成或运行参数明显超出本文数据范围时,直接应用该模型可能会产生较大误差。因而,更为合理的推广方式是在目标水体条件下补充一定数量的 EC 试验数据,对模型进行再训练或迁移学习,或在相同的 XGBoost-NSGA-II 框架下重建面向特定水体的预测与多目标优化模型。

3 结论

- ① 不同输入特征在不同分位区间内对剩余氟浓度和能耗的影响呈现复杂的非线性关系。优化电解反应条件(例如电流密度、阳极面积、电解时间等)能够有效提高除氟效率并降低能耗,从而提升整体处理效果。由于大部分输入特征与目标变量之间的关系无法通过传统线性模型准确捕捉,ML 方法在应对该复杂非线性关系方面展现出了更大的优势。
- ② 基于 ML 的 XGBoost 回归模型能够有效捕捉输入特征与目标变量之间的非线性关系,特别是在剩余氟浓度和能耗的双重预测中,展现了较高的预测精度和稳定性(R^2 值为 0.76~0.89, $\text{MSE}\leq 0.0047$)。其强大的拟合能力可准确反映输入特征

与多目标之间的复杂关联,适用于电凝聚(EC)工艺的多目标优化。

③ 嵌入XGBoost模型的NSGA-II优化算法能够生成多个优选解,在不同操作条件下有效平衡剩余氟浓度与能耗的优化需求。尽管优化结果的相对误差为10%~25%,且在低浓度下波动较大,模型仍能为EC工艺的优化提供多样化的操作方案,并为实验室与实际应用中的参数优化提供科学依据。

④ 该模型目前适用于实验室工况,尚不适用于实际工业应用。为实现工艺放大的智能设计,仍需获取更大尺度下的试验数据。然而,目前此类数据资源稀缺,且低浓度样本的优化仍是未来研究的重点。未来的工作可以通过进一步优化模型参数、增加低浓度样本的训练数据和应用数据增强技术,以提高模型的鲁棒性和适应性,从而提升低浓度区间的预测精度和整体模型的稳定性。

参考文献:

- [1] AL-RAAD A A, HANAFIAH M M. Removal of inorganic pollutants using electrocoagulation technology: a review of emerging applications and mechanisms[J]. Journal of Environmental Management, 2021, 300: 113696.
- [2] BOINPALLY S, KOLLA A, KAINTHOLA J, et al. A state-of-the-art review of the electrocoagulation technology for wastewater treatment[J]. Water Cycle, 2023, 4: 26-36.
- [3] SANDOVAL M A, FUENTES R, THIAM A, et al. Arsenic and fluoride removal by electrocoagulation process: a general review[J]. Science of the Total Environment, 2021, 753: 142108.
- [4] CHO K T, COTTON A, SHIBATA T. A framework for optimal parameter selection in electrocoagulation wastewater treatment using integrated physics-based and machine learning models[J]. Sustainability, 2025, 17 (10): 4604.
- [5] NIU C, LI X, DAI R, et al. Artificial intelligence-incorporated membrane fouling prediction for membrane-based processes in the past 20 years: a critical review[J]. Water Research, 2022, 216: 118299.
- [6] CASTAÑEDA L F, COREÑO O, CARREÑO G, et al. Electrocoagulation with Fe-Al hybrid electrodes for the removal of arsenic, fluoride, and silica from natural groundwater[J]. Chemical Engineering and Processing, 2023, 190: 109434.
- [7] CASTAÑEDA L F, COREÑO O, NAVA J L, et al. Removal of fluoride and hydrated silica from underground water by electrocoagulation in a flow channel reactor[J]. Chemosphere, 2020, 244: 125417.
- [8] CASTAÑEDA L F, COREÑO O, NAVA J L. Simultaneous removal of arsenic, fluoride, and hydrated silica from deep well water by electrocoagulation using hybrid Al-Fe electrodes[J]. Process Safety and Environmental Protection, 2022, 166: 290-298.
- [9] VALENTÍN-REYES J, TREJO D B, COREÑO O, et al. Abatement of hydrated silica, arsenic, and coexisting ions from groundwater by electrocoagulation using iron electrodes[J]. Chemosphere, 2022, 297: 134144.
- [10] DAS D, NANDI B K. Removal of co-existing Fe(II), As(V) and fluoride ions from groundwater by electrocoagulation[J]. Groundwater Sustainable Development, 2022, 17: 100752.
- [11] GOREN A Y, KOBAYASHI M. Arsenic removal from groundwater using an aerated electrocoagulation reactor with 3D Al electrodes in the presence of anions[J]. Chemosphere, 2021, 263: 128253.
- [12] VALENTÍN-REYES J, COREÑO O, NAVA J L. Concurrent elimination of arsenic and hydrated silica from natural groundwater by electrocoagulation using iron electrodes[J]. Chemical Engineering Research and Design, 2022, 184: 103-112.
- [13] NYANGI M J, CHEBUDE Y, KILULYA K F, et al. Effects of coexisting ions on simultaneous removal of fluoride and arsenic from water by hybrid Al-Fe electrocoagulation[J]. International Journal of Environmental Science and Technology, 2022, 19(7): 6667-6680.
- [14] ROSALES M, COREÑO O, NAVA J L. Removal of hydrated silica, fluoride and arsenic from groundwater by electrocoagulation using a continuous reactor with a twelve-cell stack[J]. Chemosphere, 2018, 211: 149-155.
- [15] THAKUR L S, MONDAL P. Simultaneous arsenic and fluoride removal from synthetic and real groundwater by electrocoagulation process: parametric and cost evaluation[J]. Journal of Environmental Management, 2017, 190: 102-112.
- [16] GUZMÁN A, NAVA J L, COREÑO O, et al. Arsenic and fluoride removal from groundwater by electrocoagulation using a continuous filter-press reactor[J]. Chemosphere, 2016, 144: 2113-2120.
- [17] SANDOVAL M A, FUENTES R, NAVA J L, et al.

- Simultaneous removal of fluoride and arsenic from groundwater by electrocoagulation using a filter-press flow reactor with a three-cell stack [J]. *Separation and Purification Technology*, 2019, 208: 208–216.
- [18] LÓPEZ-GUZMÁN M, ALARCÓN-HERRERA M T, IRIGOYEN-CAMPUZANO J R, et al. Simultaneous removal of fluoride and arsenic from well water by electrocoagulation[J]. *Science of the Total Environment*, 2019, 678: 181–187.
- [19] THAKUR L S, GOYAL H, MONDAL P. Simultaneous removal of arsenic and fluoride from synthetic solution through continuous electrocoagulation: operating cost and sludge utilization [J]. *Journal of Environmental Chemical Engineering*, 2019, 7(1): 102829.
- [20] SILVA F J, GRAÇA S N, RIBEIRO M A, et al. Electrocoagulation process for the removal of co-existent fluoride, arsenic and iron from contaminated drinking water[J]. *Separation and Purification Technology*, 2018, 197: 237–243.
- [21] HALPEGAMA J U, HEENKENDA K Y, WU Z, et al. Concurrent removal of hardness and fluoride in water by monopolar electrocoagulation [J]. *Journal of Environmental Chemical Engineering*, 2021, 9 (5) : 106105.
- [22] MENA V F, BETANCOR-ABREU A, GONZÁLEZ S, et al. Fluoride removal from natural volcanic underground water by an electrocoagulation process: parametric and cost evaluations [J]. *Journal of Environmental Management*, 2019, 246: 472–483.
- [23] ZHAO X, ZHANG B, LIU H, et al. Simultaneous removal of arsenite and fluoride via an integrated electro-oxidation and electrocoagulation process[J]. *Chemosphere*, 2011, 83(5): 726–729.
- [24] SANDOVAL M A, FUENTES R, NAVA J L, et al. Fluoride removal from drinking water by electrocoagulation in a continuous filter press reactor coupled to a flocculator and clarifier[J]. *Separation and Purification Technology*, 2014, 134: 163–170.
- [25] LOPEZ R L, CORENO O, NAVA J L. Abatement of hydrated silica and simultaneous removal of coexisting ions from deep well water by electrocoagulation using an up-flow reactor [J]. *Journal of Water Process Engineering*, 2019, 32: 100923.
- [26] ROBLEDO-PERALTA A, LÓPEZ-GUZMÁN M, MORALES-AMAYA C G, et al. Arsenic and fluoride in groundwater, prevalence and alternative removal approach[J]. *Processes*, 2021, 9(7): 1191.
- [27] THAKUR S L, MONDAL P. Techno-economic evaluation of simultaneous arsenic and fluoride removal from synthetic groundwater by electrocoagulation process: optimization through response surface methodology [J]. *Desalination and Water Treatment*, 2016, 57(59): 28847–28863.
- [28] GARCÍA I, CASTAÑEDA L F, NAVA J L, et al. Continuous electrocoagulation–flocculation–sedimentation process to remove arsenic, fluoride, and hydrated silica from drinking water [J]. *Journal of Water Process Engineering*, 2025, 69: 106571.
- [29] DANIEL M, NINA C S, VEIT M M. Arsenic removal from highly contaminated groundwater by iron electrocoagulation: investigation of process parameters and iron dosage calculation [J]. *Water*, 2021, 13 (5): 687.
- [30] CASTAÑEDA L F, GARCÍA I, NAVA J L, et al. Concurrent arsenic, fluoride, and hydrated silica removal from deep well water by electrocoagulation: comparison of sacrificial anodes (Al, Fe, and Al-Fe) [J]. *Journal of Environmental Management*, 2024, 365: 121597.
- [31] WANG X, LI Y, QIAO Q, et al. Water quality prediction based on machine learning and comprehensive weighting methods [J]. *Entropy*, 2023, 25(8): 1186.
- [32] SHIRKOOHI M G, TYAGI R D, VANROLLEGHEM P A, et al. A comparison of artificial intelligence models for predicting phosphate removal efficiency from wastewater using the electrocoagulation process [J]. *Digital Chemical Engineering*, 2022, 4: 100043.

作者简介:王陈波(2002—),男,广东湛江人,硕士研究生,从事给水处理技术研究。

E-mail: 1955614559@qq.com

收稿日期: 2025-10-14

修回日期: 2025-11-18

(编辑:李德强)