



机器学习赋能的水中全氟/多氟烷基物质去除技术

白明霞, 宋潮潮, 张圆正*, 牛军峰

资源环境系统优化教育部重点实验室, 华北电力大学环境科学与工程学院, 北京 102206

*通讯作者, E-mail: zhangyz@ncepu.edu.cn

收稿日期: 2025-07-03; 接受日期: 2025-09-23; 网络版发表日期: 2025-10-17

国家自然科学基金(编号: 52300084)、国家重点研发计划政府间国际科技创新合作项目(编号: 2022YFE0135700)和中央高校基本科研业务费专项资金(编号: 2024MS063)资助项目

摘要 全氟和多氟烷基物质(PFAS)因其极高的化学稳定性和环境持久性, 在全球水体中广泛存在, 且对生态系统与人类健康构成潜在威胁. 传统去除技术依赖大量实验与参数调优, 面临效率低、成本高与适应性差等瓶颈. 近年来, 基于人工智能的机器学习(ML)方法在环境科学领域快速发展, 展现出提升去除效率和优化处理过程的重要潜力. 本文综述了ML在吸附、膜分离、电化学及光催化等主流PFAS去除技术中的应用进展, 聚焦其在去除性能预测、工艺参数优化、材料筛选与合成、机制解析等方面的赋能作用. 在此基础上, 进一步剖析了当前ML方法在数据质量、模型泛化、环境适应性与毒性预测等方面面临的挑战. 最后提出通过高质量数据集、多模态模型和智能化部署等路径, 推动构建高效低碳的PFAS治理体系, 为未来研究与工程实践提供参考.

关键词 机器学习, 全氟/多氟烷基物质, 污染物, 水处理技术, 降解机制, 模型泛化

1 引言

全氟和多氟烷基物质(per- and polyfluoroalkyl substances, PFAS)是一类新型持久性有机污染物, 因其优异的化学稳定性(C-F键键能为 $531.5 \text{ kJ mol}^{-1}$)及显著的疏水疏油性^[1], 在工业领域被广泛应用于表面活性剂、不黏图层、食品包装、水成膜泡沫及半导体制造^[2,3]. 自20世纪40年代首次合成以来, PFAS已在全球多种环境介质中被广泛检出, 特别是在水体中, 包括地表水、地下水、饮用水、自来水及工业废水^[4-8]等. 此外, 空气、沉积物及生物体内也存在PFAS^[9-13]. PFAS可通过多种途径进入环境, 其来源主要分为点源和非点源两类. 点源主要包括氟化工厂、各类工业设施、

消防训练场、污水处理厂及垃圾填埋场等; 非点源则涵盖大气传输、前体化合物的生物和化学分解、地表径流、降水以及消费品分解等途径^[14-16]. 研究表明, 长期暴露于PFAS环境中可能导致人体肝功能损害、免疫系统紊乱、生殖异常及发育障碍等问题^[4,13,17-19]. 鉴于PFAS的广泛分布和潜在危害, 多国制定了相关法规以加强监管. 2024年, 美国环境保护署(U.S. Environmental Protection Agency, EPA)颁布了针对饮用水新规, 要求全氟辛酸(PFOA)、全氟辛烷磺酸(PFOS)、全氟壬酸(PFNA)以及全氟己烷磺酸(PFHxS)的限值分别为4、4、10和10 ng L^{-1} ; 我国于2022年发布的《生活饮用水卫生标准》(GB 5749-2022)规定PFOA和PFOS的最大允许浓度分别为80和40 ng L^{-1} .

引用格式: Bai M, Song C, Zhang Y, Niu J. Machine learning-enabled technologies for the removal of per- and polyfluoroalkyl substances from water. *Sci Sin Chim*, 2025, 55: 3141-3154, doi: 10.1360/SSC-2025-0195

近年来, 针对PFAS在环境中表现出的持久性和毒性问题, 相关治理技术持续取得进展。目前常用的去除技术包括吸附、膜分离、电化学与光催化等, 虽在特定条件下具备较好效果, 但普遍依赖复杂实验, 效率受限, 成本较高。吸附法对长链PFAS具有良好的去除效果, 但易受共存污染物干扰且再生困难^[20-22]。膜分离虽能高效截留PFAS, 但存在膜污染、膜浓缩物处理和高运行成本等问题^[21,23-25]。电化学和光催化技术虽然具有降解潜力, 但常面临能耗高、催化效率低和工程化难度大等挑战^[21,26-29]。这些瓶颈凸显出在PFAS治理中亟需突破传统依赖经验与试错实验的路径, 发展更高效、智能化的技术手段。

机器学习(machine learning, ML)作为人工智能(artificial intelligence, AI)的重要分支, 能够使计算机系统通过数据学习并做出预测与决策任务, 从而揭示变量间的复杂关系并预测目标结果。传统PFAS去除技术通常依赖大量实验与参数优化以确定最佳的处理条件, 过程耗时、昂贵且效率有限^[30]。ML为克服这一挑战提供了新的路径。近年来, ML技术在环境科学与工程中的应用日益广泛, 已成为提升污染物治理效率与优化处理过程的重要工具^[31]。通过数据驱动的方法, ML模型可协助开发并优化多种PFAS去除技术。研究表明, ML不仅能准确预测PFAS去除效果^[32], 还可用于工艺参数的智能优化^[33], 从而显著提高整体处理效率。

鉴于ML在揭示复杂环境问题中的潜力, 本文系统回顾了其基本原理与典型工作流程, 重点聚焦其在PFAS污染治理中的最新应用进展。综述内容涵盖ML在吸附、膜分离、电化学与光催化等主流PFAS去除技术中的关键角色, 尤其在去除效率预测、工艺参数优化、材料筛选与合成、机制解析与路径模拟等方面的赋能作用。进一步剖析了当前ML技术在实际应用中面临的数据质量、模型泛化、环境复杂性及毒性评估等挑战, 最后提出未来研究亟待突破的关键方向, 以期PFAS污染智能治理提供理论支撑与技术路径。

2 机器学习在环境中的应用

ML是一种利用计算机从历史数据中训练模型的方法, 通过发现数据中的潜在规律, 能够有效执行分类、回归、聚类等多种任务, 从而对未知数据进行预测与分析^[34]。其主要特点是利用计算机程序对数据进

行训练, 以更好地发现研究对象的发展规律; 同时, 采取交叉验证等措施, 确保训练后的数据能够对未知数据进行分析且不发生拟合问题^[35]。根据输入数据是否包含标签信息, ML可分为监督学习、无监督学习和半监督学习等类型。其中, 监督学习依赖带标签的数据集, 通过学习输入特征与已知输出之间的映射关系, 用于回归预测或分类判断; 无监督学习则在无标签条件下, 通过分析数据间的内在结构实现聚类、降维等任务, 适用于污染物来源解析等情形; 半监督学习则结合有限标签与大量无标签数据, 在标签稀缺场景下展现出较高实用性^[36]。常见ML算法及其优缺点如表1所示。典型ML流程包括数据收集、预处理、数据划分、模型训练、结果评估、模型解释与适用性域界定等步骤(图1)。环境科学与工程领域面临庞大且复杂数据集, ML正成为该领域研究的重要工具, 为复杂环境问题的分析与解决提供新的思路与方法, 有力推动该领域的技术进步与理论发展。

ML技术已广泛应用于环境分析检测的各个环节^[37], 在环境质量评估方面, 一些研究利用机器学习预测土壤质量指标^[38]; 在污染物浓度预测中, 卷积神经网络被用于预测CO₂、PM_{2.5}、CO等污染物浓度^[39]; 在空气与水体监测方面, 前馈神经网络已被用于预测印度河上游的溶解氧含量及总溶解固体水平^[40](图2a)。在污染物环境行为研究中, ML可模拟污染物扩散趋势, 揭示其迁移与转化路径, 进而为环境风险评估与管理决策提供科学依据(图2b)^[41,42]。在水处理领域, ML方法如人工神经网络(ANN)、支持向量机(SVM)、随机森林(RF)等, 已被应用于水质监测与预警、系统故障诊断及过程智能控制, 显著提升污染物识别的准确性与处理系统的运行效率(图2c)^[43,44]。此外, ML在新材料与化学物质的开发^[45]方面发挥重要作用, 涵盖材料设计、性能预测、实验条件优化与自动化筛选等环节^[46,47], 有助于加速研发周期并提高实验效率(图2d)。

尽管ML在PFAS污染源识别^[48,49]、环境行为预测与高通量筛选^[50]等方向已取得广泛进展, 但其在污染物去除技术中的应用仍相对滞后。与源解析或风险评估相比, PFAS去除涉及多尺度传质过程、材料性能异质性以及复杂水体干扰等多重变量, 显著提升了模型构建的挑战性。ML具备强大的非线性建模与变量交互识别能力, 为破解传统试错法的效率瓶颈提供了新

表 1 机器算法及其优缺点
Table 1 Machine learning algorithms: advantages and disadvantages

机器学习算法	优点	缺点
线性回归	结构简洁、计算高效, 结果可解释性强, 小样本表现良好	非线性拟合能力弱, 对异常值敏感, 依赖特征独立性假设
逻辑回归	可输出概率预测, 适用性广; 能处理非线性关系和交互效应, 结构简单	需大样本以保证稳定性, 对高度相关特征敏感; 且性能依赖特征选择和特征工程
决策树	非参数模型, 可处理非线性和特征交互, 适应多种数据类型, 易理解与可视化	易过拟合, 高维数据表现差, 对微小数据变动敏感, 多类别特征偏向
支持向量机	准确率高, 有效避免过拟合; 高维数据表现良好	训练慢; 性能对参数(如惩罚系数和核函数类型)、噪声和异常值敏感
朴素贝叶斯	算法简单, 训练快速, 无需手动设置参数; 小数据集和高维数据表现良好	依赖特征独立假设, 对数值型特征处理有限, 预测概率可能不准确
K近邻	简单易实现, 适合分类和回归, 对非线性边界适应良好	计算量大, 性能依赖邻居数 k , 对噪声和高维数据敏感
随机森林	计算快, 可扩展性强, 抗噪声与过拟合, 能评估特征重要性	树多时运行慢, 可解释性低于单棵决策树, 极端不平衡数据表现一般
梯度提升机	预测精度高, 可处理非线性和复杂数据, 对缺失值和不平衡数据有鲁棒性	训练慢, 参数多, 易过拟合, 可解释性低
K-Means聚类	算法结构简单, 计算快, 处理大规模数据表现良好, 对球状簇聚类效果尤佳	需预设簇数 K , 初值敏感, 可易陷入局部最优; 对噪声和异常值敏感, 仅适用于凸形簇, 复杂簇表现不佳
主成分分析	可实现降维和去相关, 具有去噪能力、高维数据可视化; 线性算法使计算高效	仅能捕捉线性关系, 降维后主成分难以解释, 对异常值敏感
人工神经网络	非线性建模能力强, 适合复杂模式, 应用范围广, 鲁棒性强	训练慢, 参数多且敏感, 需大数据和计算资源, 可解释性差, 易过拟合
卷积神经网络	高效提取局部与层级特征; 参数量少, 易于优化; 可利用预训练模型迁移	长程依赖建模能力有限; 缺乏旋转/尺度等几何不变性; 难以处理非规则结构数据
图神经网络	直接处理非欧几里得结构; 支持节点、边及图级表示学习; 分子建模、社交网络等表现优异	深层网络易过平滑; 大规模图训练成本高; 对图结构噪声敏感
变换器	有效建模长距离依赖; 高度并行, 适合大规模训练; 框架通用, 适用于文本、图像、语音及多模态任务	注意力计算对长序列/高分辨率任务昂贵; 数据与算力需求高; 缺乏局部归纳偏置
物理信息神经网络	数据依赖性低; 保证物理一致性; 可用于正/逆问题与参数辨识	多目标优化难以平衡; 高维或高频解收敛困难; 工程规模应用受限
生成模型	可生成高维时空数据; 具备小尺度与不确定性建模能力, 能捕捉复杂非线性与极端事件分布	高度依赖大规模高质量数据, 物理一致性与可解释性不足, 长时序预测易积累误差

机遇. 因此, 本文聚焦ML在PFAS去除领域的应用潜力, 系统梳理其在主要处理技术中的典型实践与关键进展, 为实现智能高效的污染控制体系提供理论支撑.

3 ML赋能的PFAS去除技术

3.1 吸附技术

吸附技术因高效截留性能与良好工程适用性, 已广泛用于水体PFAS污染治理, 主要途径包括碳基材料物理吸附与离子交换树脂富集. 如图3a所示, PFAS在固-液界面的吸附机制主要涉及静电作用、氢键络合、疏水作用及配位交换等过程^[51]. 其中, 长链PFAS主要由疏水作用和配位交换主导; 相比之下, 短链

PFAS则主要依赖静电吸引和氢键作用. 吸附剂的孔隙结构、比表面积及表面官能团等理化性质决定了其对不同碳氟链长PFAS的亲合力. 同时, 水体基质参数(如天然有机物浓度、离子强度)及工艺参数(如水力停留时间、pH值)亦通过竞争吸附与扩散传质过程影响去除效率^[52]. 然而, 吸附剂、污染物与环境介质之间存在多尺度、非线性耦合关系, 传统经验驱动的试错策略难以实现多目标参数协同优化. 近年来, ML被引入吸附研究, 助力揭示材料特性与吸附性能的潜在关联, 有效支撑材料精准筛选与复杂工况下参数优化.

为破解传统试错法在确定PFAS高效去除所需树脂特性与操作条件效率低下的难题, 研究者构建了贝叶斯优化与ML耦合框架(图3b)^[53], 系统评估树脂特

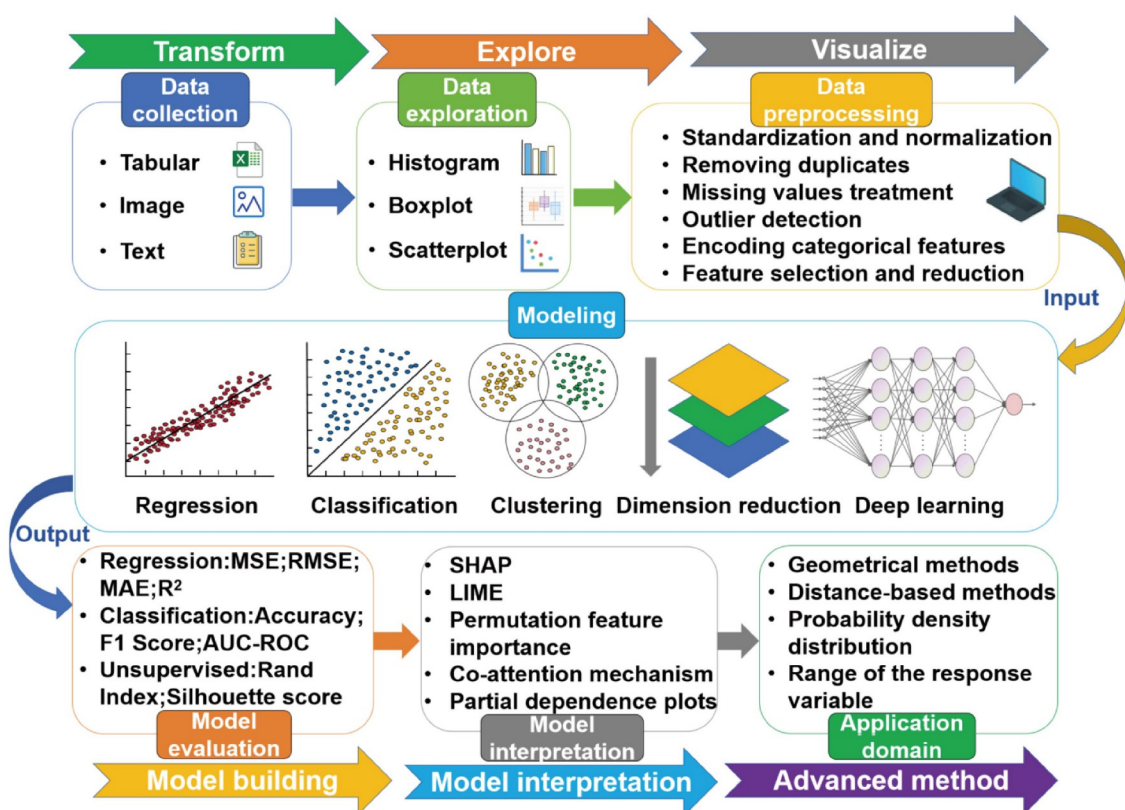


图1 (网络版彩图)机器学习工作流程. MSE: 均方误差; RMSE: 均方根误差; MAE: 平均绝对误差; R^2 : 决定系数; AUC-ROC: 曲线下面积-受试者工作特征曲线; SHAP: shapley additive explanations; LIME: local interpretable model-agnostic explanations

Figure 1 (Color online) Illustrates the machine learning workflow. MSE stands for mean squared error; RMSE stands for root mean square error; MAE stands for mean absolute error; R^2 stands for the coefficient of determination; AUC-ROC stands for the area under the receiver operating characteristic curve; SHAP stands for shapley additive explanations; LIME stands for local interpretable model-agnostic explanations.

性、水质参数与操作条件对去除率(RE)的影响. 该模型显著提升了预测准确性, 在实验条件下对长链(PFOA、PFOS)与短链(PFBS、GenX) PFAS的RE预测偏差仅为1%~10%. 基于模型识别的最优参数组合, 在天然湖泊水体中开展验证实验, 结果表明, 筛选出的聚苯乙烯树脂D201 (PFBS:RE=88.32%)、聚丙烯酸树脂IRA958 (GenX:RE=86.76%; PFOA:RE=98.77%; PFOS:RE=98.66%)显著优于WA10 (聚丙烯酸树脂)、A111S、IRA400及IRA910等聚苯乙烯树脂, 验证了模型筛选有效性.

为进一步提升模型对材料吸附行为的预测精度, Ersan团队^[54]提出了机器学习-线性溶剂化能关系(ML-LSER)模型. 该模型融合了吸附剂特性、水化学参数与操作条件, 显著提升了活性炭在复杂水体中对PFAS吸附行为的拟合精度. 相比常规LSER模型($R^2 < 0.1$),

ML-LSER模型 R^2 值提升至0.13~0.80, 进一步结合主成分回归(PCR)后可达0.65~0.99, 为复杂体系中吸附性能的量化预测和材料智能筛选提供了有效工具.

在高选择性吸附材料开发方面, 研究者构建了高通量ML与真实溶剂似导体筛选(COSMO-RS)平台(图3c)^[55], 用于从1000余种离子液体中筛选对PFOA具有高亲和力的分子印迹聚合物(MIP), 最终筛选出的(Vim)C₃(L-Pro)@MIP结合分子印迹技术, 构建了具有多重氢键位点的纳米腔结构材料. 实验结果表明, 该材料在室温下对PFOA的最大吸附量达568.18 mg/g, 去除率超过99%; 其吸附过程符合Langmuir等温模型, 表明主导机制为单分子层化学吸附. Coutinho等^[57]以丝瓜海绵为原料, 经一步热化学处理制备出比表面积达1127 m²/g的多孔碳基吸附剂TCL1, 并结合响应面法的中心复合可旋转设计(RSM-CCRD)与ANN对

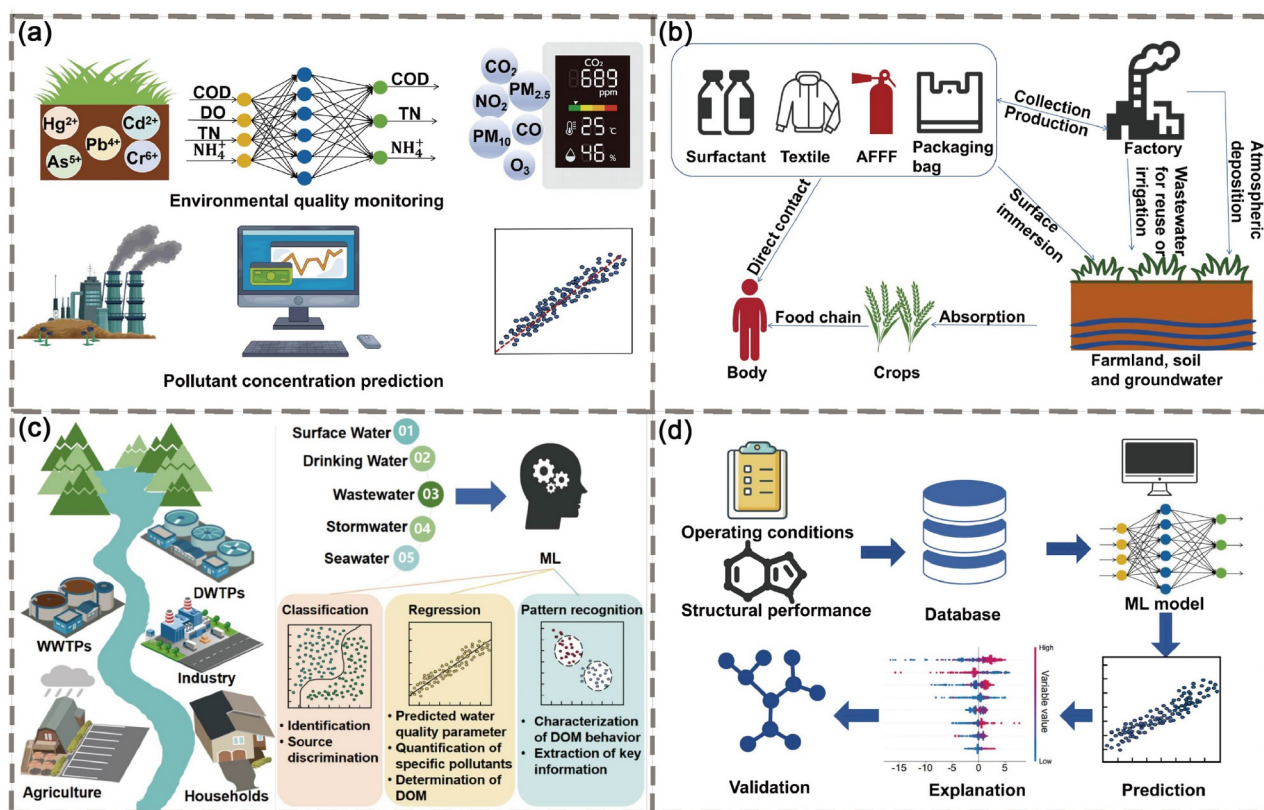


图2 (网络版彩图)机器学习在环境中的应用。(a) 环境分析检测过程; (b) 污染物环境行为研究; (c) 水处理过程, DWTPs: 饮用水厂; WWTPs: 污水处理厂; (d) 新材料和化学物质的开发

Figure 2 (Color online) Applications of machine learning in the environment. (a) Environmental analysis and detection process; (b) environmental behavior of pollutants; (c) water treatment processes, DWTPs: drinking water plants; WWTPs: sewage treatment plants; (d) development of new materials and chemicals.

PFOA的吸附过程进行优化建模。实验结果表明, 该吸附剂在22℃、pH 3.4条件下对PFOA的去除率达到 $95.3\% \pm 1.24\%$, Langmuir模型最大吸附容量达 $2.18 \times 10^{-3} \mu\text{g/g}$ 。ANN模型在复杂非线性条件下的预测精度($R^2=0.9796$)高于RSM-CCRD ($R^2=0.9035$), 凸显了人工智能方法在实验优化中的优势。Jiang等^[58]则采用RF、极限梯度提升(XGBoost)、分类提升(CatBoost)等集成机器学习方法, 结合实验数据与特征重要性分析, 系统解析了多孔碳材料的物化参数对PFOA吸附性能的非线性影响。结果显示, 比表面积($1200\sim 1600 \text{ m}^2/\text{g}$)、总孔隙体积($1.0\sim 1.4 \text{ cm}^3/\text{g}$)以及碳/氧含量为关键控制因素, 偏离最优区间会导致吸附性能下降。进一步通过实验合成的多孔碳材料(如C-900)验证了模型预测, 同时借助密度泛函理论(DFT)计算揭示了静电作用、疏水作用与表面官能团(如-COOH)结合的协同吸附机理。

针对工艺参数的优化, 研究人员采用贝叶斯优化

构建不同操作条件的PFAS去除性能预测模型(图3d), 并通过100次训练/测试集的随机交叉验证以降低过拟合风险。特征重要性分析显示, pH值(得分59.94)为影响PFAS去除效率的首要因素, 其次为吸附剂粒径(41.94)和初始浓度(34.41)^[56]。Shapley加性解释(SHAP)分析进一步指出, 低pH值通过增强静电吸引显著提升吸附能力, 而小粒径吸附剂由于比表面积较大, 可有效缩短吸附平衡时间。实验数据显示, IRA67树脂在优化条件下(pH 3, $t=48 \text{ h}$)吸附容量可达 3067 mg g^{-1} , 显著优于金属基氧化物吸附剂(仅约为 0.007 mg g^{-1})。上述研究为吸附材料筛选与工艺优化提供了智能决策支持。

3.2 膜分离技术

膜分离技术(如反渗透RO和纳滤NF)因操作简便、分离效率高及低药剂需求, 成为PFAS去除的重要

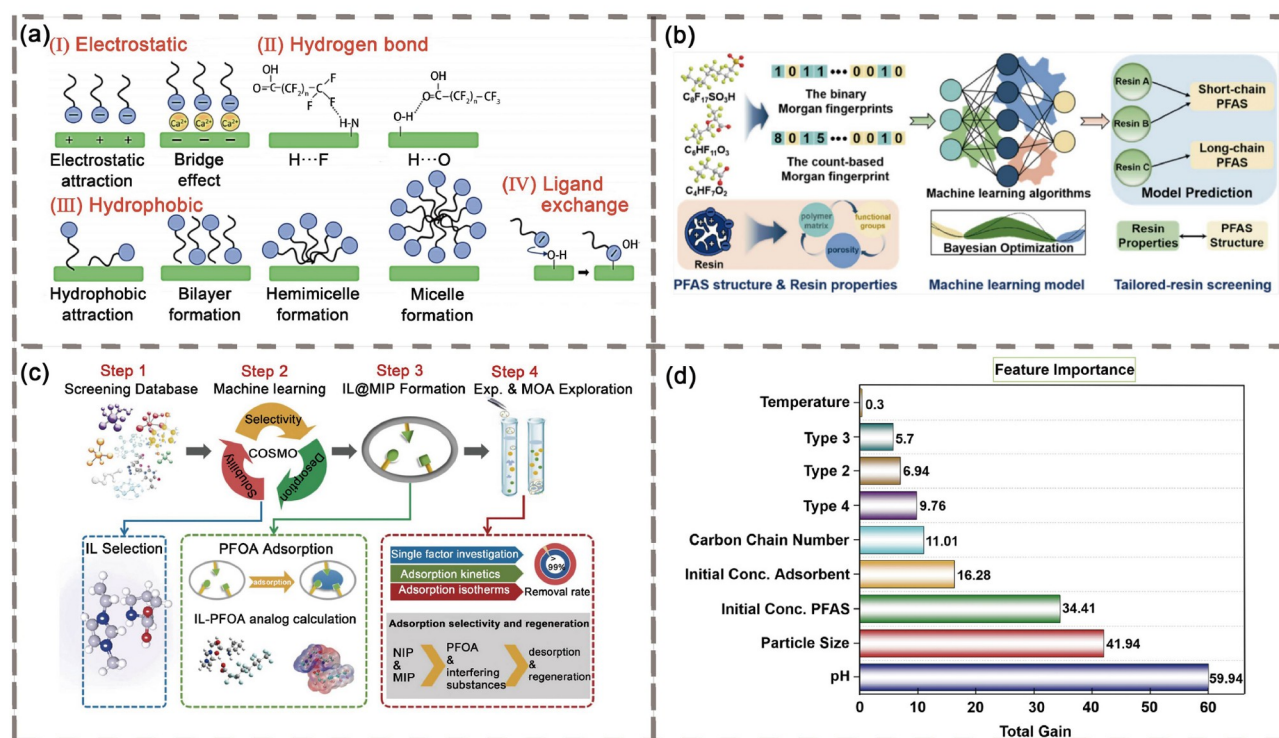


图3 (网络版彩图) ML赋能的PFAS吸附材料筛选与机制解析。(a) PFAS在吸附剂表面的作用机制示意图^[51];(b) 贝叶斯优化与ML协同树脂材料筛选^[53];(c) 高通量ML与COSMO-RS联合筛选分子印迹聚合物材料^[55];(d) XGBoost模型识别PFAS去除的关键特征变量, Conc: Concentration. 数据来源: 参考文献^[56]

Figure 3 (Color online) ML-empowered screening and mechanism analysis of PFAS adsorption materials. (a) Schematic diagram of the mechanism of PFAS on the surface of adsorbents ^[51]; (b) Bayesian optimization and ML collaborative resin material screening ^[53]; (c) high-throughput ML and COSMO-RS joint screening of molecularly imprinted polymer materials ^[55]; (d) XGBoost model identifies key feature variables for PFAS removal. The data source: reference ^[56].

手段. 研究表明, RO与NF膜对PFAS的去除率普遍可达90%~99%, 其分离机制主要包括尺寸排阻、电荷排斥及疏水作用(图4a)^[59]. 其中, 尺寸排阻与疏水作用对长链PFAS截留更具优势, 而电荷排斥为短链PFAS的主导去除机制. 相较RO膜, NF膜兼具高通量与优良截留性能而受到日益关注. PFAS的去除效能受膜类型、分子性质(如链长、官能团、浓度)、操作条件(如pH、压力)及水基质成分等多因素共同影响.

为提升膜分离去除PFAS性能的预测准确性, 研究人员逐步引入ML方法. Hosseinzadeh等^[60]构建了以RF、梯度提升机(GBM)和自适应提升算法(AdaBoost)为核心的预测框架, 综合考虑膜类型、温度、初始浓度、pH、施加压力及多价态阳离子浓度等变量对NF膜去除PFOS的影响. 其中, GBM模型预测性能最优(测试集 $R^2=0.975$, $MSE=2.450$, $MAE=0.861$). 通过排列变量重要性(PVI)深入分析, RF模型识别出施加压力、初

始PFOS浓度和膜类型为主导因素; 而pH值和阳离子价态影响较小. Singh等^[61]进一步比较了多元线性回归(MLR)、最小绝对收缩和选择算子(Lasso)回归、岭(Ridge)回归、RF和ANN五种模型在PFOS去除预测中的表现, 结果表明ANN模型预测能力最强($R^2=0.89$). 变量重要性分析揭示, 膜类型与温度的交互作用对PFOS去除效率影响最大, PFOS浓度与pH之间的协同效应也不可忽视. 在较高pH条件下, 膜表面负电荷密度增加, 使带负电的PFOS分子受到更强的静电排斥作用, 从而提高NF膜对PFOS的去除效率. 在低至中等浓度范围内, PFOS去除率随进水浓度升高而有所改善; 然而, 当浓度接近或超过其临界胶束浓度时, 膜表面可能出现吸附饱和或污染加剧现象, 导致截留性能下降^[21]. 因此, pH与进水浓度的协同作用对PFOS截留表现出双重效应: 适宜的pH与浓度条件能够提升去除效率, 而过高浓度则可能削弱甚至抵消pH带来的正向作

用(图4b).

近年来, 针对PFOS在水体中去除难度大、传统工艺效率低和NF工艺优化依赖大量实验的问题, 研究者

逐渐将深度学习方法引入膜过程建模与优化. Xiao等^[33]提出了一维梯度提升卷积神经网络(1D-GBCNN)(图4c), 在326条实验数据的基础上实现了对PFOS去除

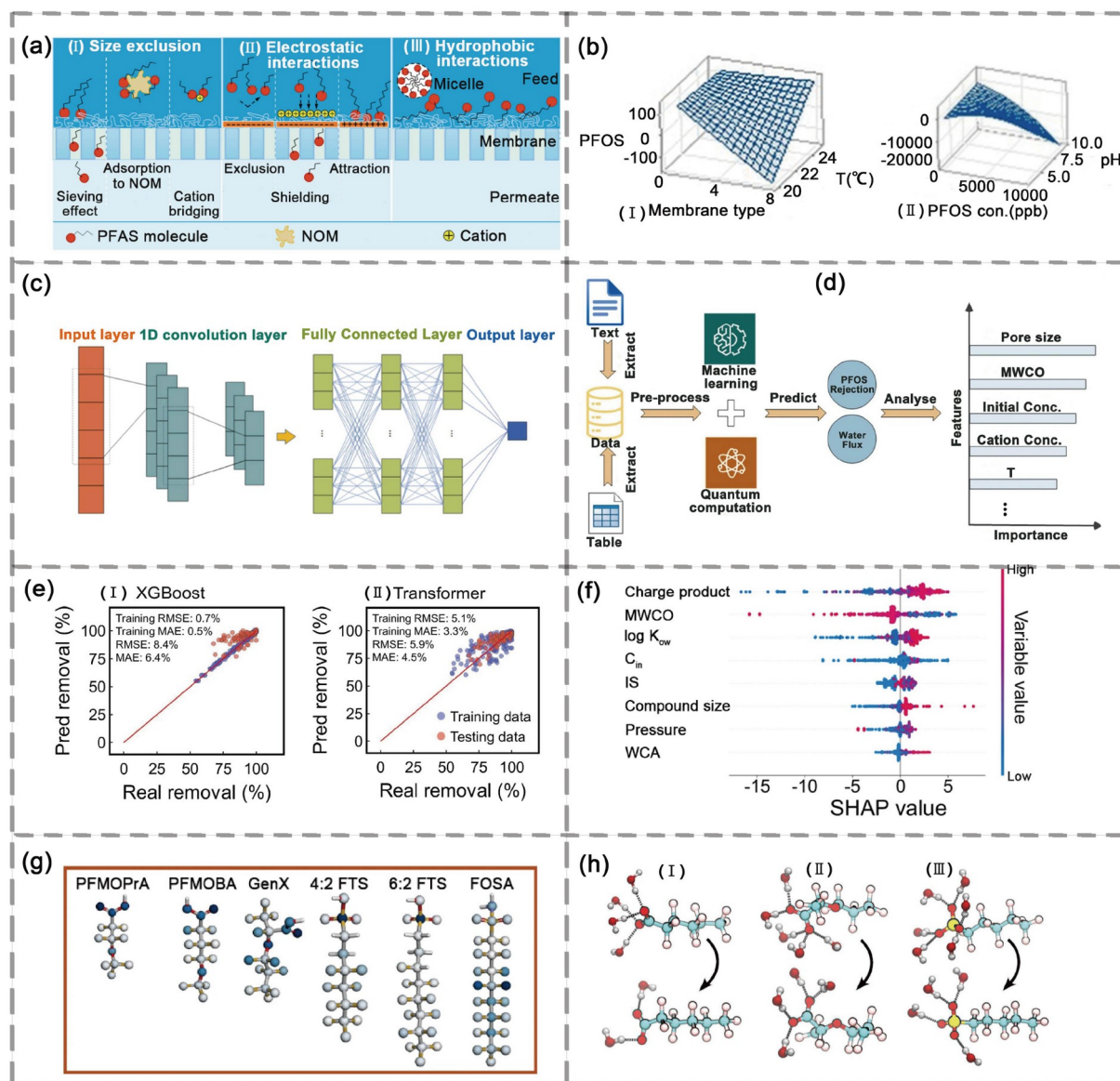


图 4 (网络版彩图) ML赋能的膜分离PFAS去除关键因子识别与机制解析. (a) NF/RO膜去除PFAS机制示意图^[59]; (b) 膜类型与温度、初始浓度与pH交互作用对PFOS去除性能影响^[61], T : temperature; (c) 一维卷积模型结构^[33]; (d) 量子增强梯度提升神经网络去除PFOS及关键影响因素识别^[62]; (e) XGBoost与Transformer模型对PFAS去除率的预测性^[63]; (f) SHAP分析识别出的PFAS去除关键影响因子^[63]; (g) 注意力机制揭示PFAS分子中关键基团对去除率预测的贡献^[63]; (h) 分子动力学模拟PFHxA (I)、GenX (II)和PFPeS (III)在水相(上)与聚酰胺膜层(下)中的氢键行为比较^[63]

Figure 4 (Color online) ML-empowered membrane separation for PFAS removal: key factor identification and mechanism analysis. (a) Schematic of the PFAS removal mechanism by NF/RO membranes ^[59]; (b) the impact of membrane type, temperature, initial concentration, and pH interaction on PFOS removal performance ^[61]; (c) diagram of the 1D convolutional model structure ^[33]; (d) quantum-enhanced gradient boosting neural networks for PFOS removal and identification of key influencing factors ^[62]; (e) predictive performance of XGBoost and Transformer models for PFAS removal rates ^[63]; (f) key factors identified by SHAP analysis for PFAS removal ^[63]; (g) attention mechanism reveals the contribution of key groups in PFAS molecules to removal rate prediction ^[63]; (h) molecular dynamics simulation comparing the hydrogen bond behavior of PFHxA (I), GenX (II), and PFPeS (III) in water phase (top) and polyamide membrane layer (bottom) ^[63].

率的准确预测($R^2=0.902$, $RMSE=2.379$), 并利用SHAP方法揭示了膜孔径、分子量截留(MWCO)、温度、水通量及离子浓度等为NF膜分离PFOS的关键影响因素, 开发了基于1D-GBCNN的在线预测平台, 为膜选择与工艺优化提供了便捷工具. 在此基础上, 进一步发展出结合量子神经网络与梯度提升框架的量子增强梯度提升神经网络(QGBNN)^[62], 在小样本条件下表现更为优异, 预测精度提升至 $R^2=0.931$ 、 $RMSE=2.018$, 并同样确认了膜物理特性与运行条件在PFOS去除中的主导作用(图4d). 两项研究表明, 基于数据驱动的建模方法能够有效降低实验成本并提高预测与优化效率, 其中QGBNN在捕捉复杂非线性机制和提升小样本预测性能方面展现出更强潜力.

除基于实验数据的过程建模外, 部分研究融合分子模拟手段, 深入解析PFAS在膜过程中的迁移机制. Jeong等^[63]基于457条文献数据和77条实验数据, 构建了XGBoost和多模态变换器(transformer)模型的高精度预测体系, 对PFAS去除率的预测表现优异($RMSE$ 为5.9%、 MAE 为4.5%) (图4e). SHAP分析结果表明, 电荷积、MWCO、疏水性参数($\log K_{ow}$)和初始PFAS浓度是关键变量, 证实静电排斥和尺寸排阻主导了聚酰胺NF与RO膜的分离(图4f). 此外, 注意力机制识别出PFAS分子中电子密度较高的基团(如羧酸根、磺酸根及醚基)对模型预测贡献显著, 强调了分子结构对迁移行为的关键作用(图4g). 为验证上述推断, 研究进一步展开分子动力学模拟, 证实PFAS的跨膜过程主要受排斥势能($55\sim65\text{ kJ mol}^{-1}$)控制, 而非膜表面的吸附能. 同时, PFAS分子的脱水合过程(以氢键损失数量计: $PFHxA < GenX < PFPeS$)及偶极相互作用在其迁移行为中发挥重要作用(图4h). 该研究深化了对PFAS在膜分离体系中迁移规律的机制性理解, 并为定向开发新型膜材料和高效PFAS去除技术提供理论依据与技术支撑.

3.3 电化学技术

电化学技术主要包括还原与氧化两类过程, 其基本反应原理如图5a所示, 电还原通过电子转移断裂C-F键, 使F原子被 H^+ 取代生成C-H键, 最终得到部分或完全氢化产物^[64]. 电氧化过程则依赖阳极在通电条件下产生羟基自由基($\cdot OH$)等强氧化性物种, 对PFAS分子进行逐步攻击, 降解生成短链中间体, 最终矿化为 CO_2 、 H_2O 和F等无机产物^[21,27,65]. 相较还原过程, 电

氧化在PFAS治理中表现出更高的降解效率与更广泛的适应性, 但其普遍面临能耗高、工艺参数复杂等工程瓶颈.

为突破上述限制, Alnaimat等^[66]引入ML方法对电氧化过程进行建模与优化, 通过识别关键参数与反应条件, 有效提升了处理效率. 研究系统评估了五种模型的预测性能, 包括K近邻(KNN)、决策树(DT)、RF、梯度提升决策树(GBDT)和深度学习(DL). 结果表明, RF模型在 $RMSE$ (7.7)、 MAE (5.6)、平均绝对百分比误差MAPE (20%)与相关系数($CC=0.965$)多个方面均表现最优预测精度与稳定性(图5b). 进一步结合基尼(Gini)不纯度与平均精度下降(MDA)两种变量重要性分析(图5c, d), 识别出电解时间、电流密度及阳极材料为影响PFOA去除效率的关键因素, 其中电解时间在各方法中权重最高. 此外, 电解质的浓度和种类亦显著影响去除效率, 而水体基质的作用相对较小. 该研究建立的预测与优化体系不仅实现了电氧化降解机制的定量解析, 也为开发高效、低能耗的PFOA治理工艺优化提供了数据支撑.

3.4 光催化技术

光催化因其高效与环境友好性, 被认为是PFAS去除的有前景策略, 其基本原理为光催化剂在光照激发下生成强氧化($\cdot OH$ 、空穴)或还原(水合电子)性活性物种, 通过攻击PFAS分子末端羧酸或磺酸基团, 诱导脱羧、脱硫和逐步脱氟反应, 最终生成短链降解产物(图6a)^[67]. 然而, 受限于现有催化剂活性、选择性和反应条件控制, 光催化降解PFAS的整体效率仍有较大提升空间. 因此, 利用ML方法开展污染物行为预测、降解路径模拟以及催化剂性能优化, 成为提升光催化应用潜力的重要发展方向.

针对部分PFAS降解路径尚不明确的问题, Zhong等^[68]结合DFT与ML方法, 系统评估了48种PFHxS前体物的键解离焓(BDE)及其降解潜力. 研究识别出两类主要断键路径, 分别生成 $C_6F_{13}SO_2^-$ 和 $C_6F_{13}^-$ 基团. 基于偏最小二乘回归(PLS)、偏最小二乘判别分析(PLS-DA)与SVM算法构建了回归和分类模型, 并结合定量构效关系(QSAR)方法探索前体结构与BDE之间的关联性. 结果显示, SVM回归模型在两类断键路径预测中的 R^2 值分别为0.82与0.89, SVM分类模型在降解潜力识别中的准确率达80%~100%. 特征贡献分析表明,

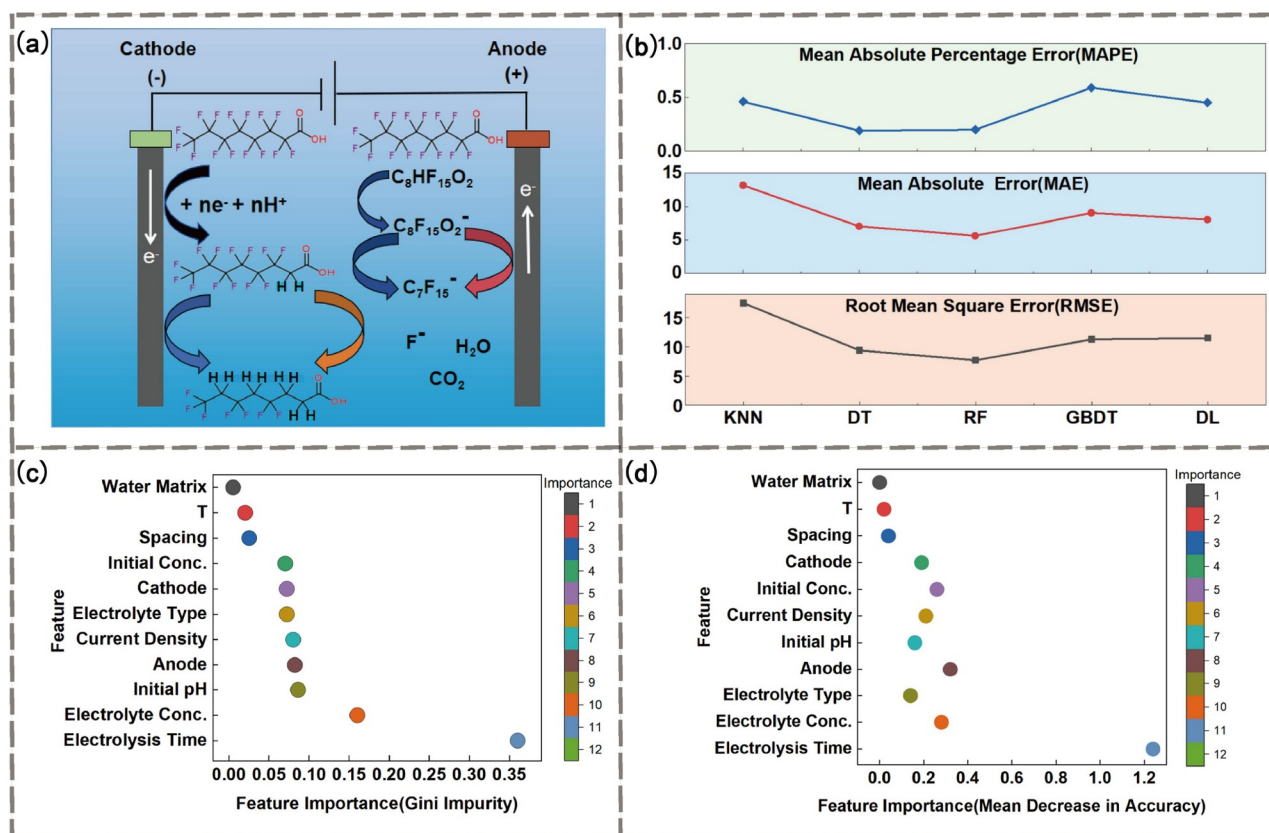


图5 (网络版彩图) ML赋能的电化学技术去除关键因子识别。(a) PFAS电还原与电氧化去除机制示意图; (b) 多监督算法对PFOA去除率的预测性能; (c) 基于Gini不纯度的RF模型特征重要性分析; (d) 基于MDA的RF模型特征重要性分析。数据来源: 参考文献[66]

Figure 5 (Color online) ML-empowered electrochemical technology for key factor identification. (a) Schematic of the removal mechanisms for PFAS through electroreduction and electrooxidation; (b) prediction performance of PFOA removal rates by multi-supervised algorithms; (c) feature importance analysis of the RF model based on Gini impurity; (d) feature importance analysis of the RF model based on MDA. The data source: reference [66].

磺酰胺基团与苯基结构是影响BDE的关键因素。实验验证进一步支持理论分析, 前体物N-MeFHxSA的降解率为51.90%, 显著高于FHxSA (27.20%), 证实了较大BDE值对应较低的键解离能力。该研究从理论与实验双重维度挑战了《斯德哥尔摩公约》中关于PFHxS前体物的界定, 指出含 $C_6F_{13}SO_2^-$ 基团的化合物不一定转化为PFHxS (图6b), 并建议修订前体物分类标准, 以更科学地评估环境风险。

为探索ML在PFOA光催化降解效率预测中的应用潜力, Navidpour等^[69]对7种监督学习算法(包括MLR、岭回归、SVM、多层感知器、AdaBoost、RF和GBM)进行了系统比较。基于涵盖18类催化剂和9个关键反应变量的大型数据集, 通过网格搜索优化超参数, 并采用PVI方法分析变量重要性。结果表明, GBM

模型表现最优($R^2=0.878$, $MSE=106.660$, $MAE=6.009$), RF次之, AdaBoost表现最差($R^2=0.574$, $MSE=388.369$, $MAE=16.480$) (图6c)。PVI分析揭示光照时间、催化剂类型、催化剂投加量和溶液pH值是影响PFOA降解效率的主要变量(图6d)。该研究充分展示了ML在光催化参数优化与效率预测中的强大潜力, 为高效PFAS光催化降解反应条件筛选与催化剂设计提供了有力技术支持。

4 关键挑战与局限

尽管近年来ML在PFAS污染控制领域的应用已在上述处理技术中初显成效, 但仍存在诸多挑战, 限制了其模型的有效性和准确性。表2总结了ML模型在PFAS

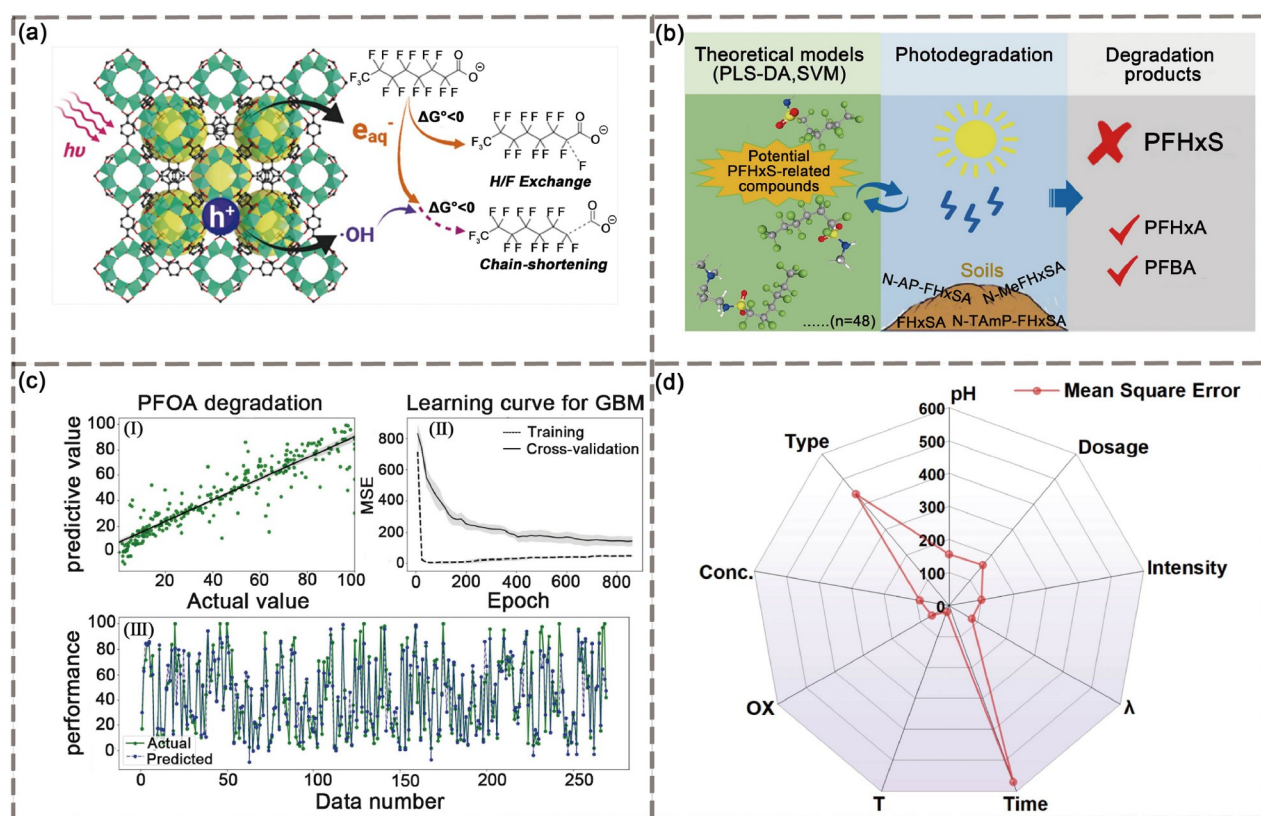


图6 (网络版彩图) ML赋能的光催化技术在PFAS去除中的应用. (a) PFAS光催化去除机制示意图^[67]; (b) PFHxS相关化合物的降解产物^[68]; (c) GBM模型在PFOA降解预测中性能评估: 相关系数(I), 学习曲线(II)与预测强度(III)^[69]; (d) GBM模型识别的光催化反应关键变量

Figure 6 (Color online) The application of ML-empowered photocatalytic technology in PFAS removal. (a) Schematic diagram of the photocatalytic mechanism for PFAS removal ^[67]; (b) degradation products of PFHxS-related compounds ^[68]; (c) performance evaluation of the GBM model in predicting PFOA degradation: correlation coefficient (I), learning curve (II), and prediction strength (III) ^[69]; (d) key variables of the photocatalytic reaction identified by the GBM models.

去除技术中的性能对比, 具体表现在以下4个方面.

(1) 数据质量与覆盖范围有限: 当前研究普遍缺乏高质量、系统化的实验数据, 尤其在新兴PFAS化合物与新型处理技术方面表现突出. 由于PFAS种类繁多、结构差异显著, 不同处理路径间的数据积累严重不均. 目前, ML赋能的研究主要集中于吸附与膜分离技术, 在电化学氧化、光催化等降解技术的应用尚处于探索阶段, 其他处理技术如超声、等离子体、热解等几乎未被涉及, 导致训练数据覆盖面窄, 制约了模型的拓展性与应用广度.

(2) 模型泛化能力与适应性不足: 多数ML模型在静态训练集条件下表现出良好预测性能, 但在面对新型PFAS种类或复杂波动水体环境条件时, 常因泛化能力不足而性能下降. 此外, 当前主流模型调参与验证

过程仍依赖人工经验与繁琐的交叉验证, 难以满足工程实际中对模型部署的高效性与实时性要求, 特别是在处理高通量或异质数据场景中, 其效率和适应能力亟待提升. 值得注意的是, 这种泛化能力的不足不仅源于模型自身的算法与参数极限, 也与其所依赖的实验条件偏离实际环境密切相关.

(3) 实验条件与真实水体环境脱节: 目前研究多基于理想化的实验室条件构建模型, 忽视了实际水环境中存在的大量共存污染物, 如天然有机物、无机离子及其他有机污染物等. 这些共存物质会竞争吸附位点、堵塞膜孔或参与副反应, 导致PFAS处理效率在实际应用中显著下降. 因此, 亟需构建更具代表性和可转化性的实验体系, 并将复杂背景因素纳入模型特征变量, 以提升预测能力与实际工程指导价值.

表 2 ML模型在PFAS去除技术中的性能对比
Table 2 Performance comparison of ML models in PFAS removal technologies

技术	模型类型	核心功能	数据量	预测精度	局限性
吸附	RF, XGBoost	预测	—	$R^2=0.863\sim0.933$	材料来源范围窄; 缺乏长期稳定性和环境适应性研究 ^[58]
	ANN	预测, 吸附行为模拟	—	$R^2=0.9796\sim0.9968$	仅针对特定吸附剂展开研究; 未考虑吸附剂长期性能及复杂水质条件下效果 ^[57,70]
	XGBoost	工艺参数优化	234	—	主要针对PFOA和PFOS的吸附去除, 数据覆盖范围不全面 ^[56]
	XGBoost, GBR, RF, CatBoost	筛选材料, 优化操作条件	1159	$R^2=0.77\sim0.94$	排除部分长链季铵基团树脂(如PFA694E、CalRes230), 研究样本选择不全面 ^[53]
	LASSO, RR, LAD-LASSO	预测	197	$R^2=0.65\sim0.99$	样本数量和种类有限; 未区分不同类别PFAS的吸附特性 ^[54]
	COSMO-RS	材料筛选与合成	1000余种	$R^2=0.8997\sim0.9305$	主要聚焦PFOA的吸附去除; 实验室条件限制实际适用性 ^[55]
膜分离	1D-GBCNN	工艺优化	326	$R^2=0.902$	未考虑废水中除PFOS之外的PFAS及溶解性有机物影响 ^[33]
	QGBNN	预测, 机制研究	334	$R^2=0.931$	模型在预测性能上相较NN略有提升, 性能优势未完全发挥 ^[62]
	RF, GBM, AdaBoost	预测, NF膜去除PFOS行为模拟	289	$R^2=0.930\sim0.975$	NF去除PFOS内在复杂机制的解释不足 ^[60]
	MLR, Lasso, Ridge, RF, ANN	预测, 工艺优化	291	$R^2=0.61\sim0.89$	复杂参数组合的交互作用研究不足 ^[61]
	XGBoost, Transformer	预测	457	RMSE=5.9%~8.4%, MAE=4.5%~6.4%	数据主要为阴离子型PFAS, 缺少阳离子、两性离子及非离子型PFAS数据 ^[63]
电化学	KNN, DT, RF, GBDT, DL	预测	—	RMSE=7.7~17.5	基于Gini不纯度和MDA的特征重要性评估存在差异, 关键因素识别准确性不足 ^[66]
光催化	AdaBoost, RF, GBM	预测	—	$R^2=0.574\sim0.878$	未评估模型在实际光催化实验中表现; 缺乏机制解析 ^[69]
	PLS, PLS-DA, SVM	预测	48	$R^2=0.70\sim0.89$	未考虑实际降解差异; 难以排除PFHxS生物转化产生不同降解产物的可能性 ^[68]

(4) 中间产物及毒性风险尚未纳入预测体系: 现有模型多聚焦于PFAS母体化合物的降解效率, 而对处理过程中可能生成的中间产物及其毒性缺乏系统考虑. 尤其在光催化、电化学、等离子体等高能处理技术中, PFAS易发生断链反应, 生成具有潜在环境持久性和毒理效应的转化产物. 目前相关毒性数据与行为机制研究仍较薄弱, 尚无法有效支撑预测模型的全面构建, 亟需将中间产物的环境风险整合进模型框架, 以提升其风险评估能力与决策支持功能.

5 总结与展望

基于ML赋能的水中PFAS去除技术在去除性能预测、工艺参数优化、材料筛选与降解行为探究等方面

展现出重要潜力. 然而, 当前研究仍面临上述多方面挑战, 在一定程度上限制了ML模型在PFAS实际去除过程中的精度与实用性, 亟需在理论支撑与工程转化方面持续突破. 未来研究可重点关注以下方向, 以推动ML赋能PFAS去除技术的高效、安全与可工程化发展.

(1) 建立系统化和标准化的实验数据平台: 通过制定涵盖新兴PFAS化合物及多样化处理技术的数据采集规范, 涵盖处理条件、浓度、温度等变量; 构建可扩展的开放数据库, 实现跨实验室的数据共享与校验; 以及利用数据增强或合成实验数据技术, 弥补稀缺PFAS种类与处理路径的数据空白.

(2) 鲁棒模型与高通量工艺开发: 构建基于迁移学习或自适应算法的模型体系, 提升在复杂水质条件下的适应性与鲁棒性, 辅助开发面向实际水体的高通量

低能耗处理工艺。

(3) 构建代表性实验体系与特征工程: 将水体中共存污染物(如天然有机物、无机离子及其他有机污染物)等复杂背景因素纳入模型特征变量, 如共存物对吸附位点的竞争、膜孔堵塞以及副反应参与等, 从而提升模型在实际水环境中的预测能力。此外, 可探索基于实测现场数据的混合建模方法, 实现实验室模型与现场实际应用的有效衔接, 增强模型的工程指导价值。

(4) 复杂反应路径建模与毒性预测: 结合图神经网络与DFT模拟技术, 揭示关键反应步骤, 强化对中间产

物的毒性预测与环境归趋评估, 提升降解路径完整性与模型风险判别能力。

(5) 数据驱动的材料设计与多机制协同构建: 依托高通量实验与模拟数据, 结合监督学习与特征选择方法, 推动兼具吸附-催化功能的一体化材料开发, 实现材料结构-性能关联的精准挖掘。

(6) 智能化控制与工程转化路径拓展: ML有望与数字孪生技术、过程模拟与智能控制系统深度融合, 推动PFAS处理系统向动态感知、自主优化与能效平衡方向演进, 支撑智能化、低碳化污染治理体系构建。

参考文献

- Liang Y, Yang L, Tang C, Yang Y, Liang S, Wang A, Xu J, Huang Q, Lin H. *Nat Commun*, 2025, 16: 1972
- Ateia M, Scherlinger M. *Science*, 2024, 385: 256–258
- Gaines LGT. *Am J Ind Med*, 2023, 66: 353–378
- Evich MG, Davis MJB, McCord JP, Acrey B, Awkerman JA, Knappe DRU, Lindstrom AB, Speth TF, Tebes-Stevens C, Strynar MJ, Wang Z, Weber EJ, Henderson WM, Washington JW. *Science*, 2022, 375: eabg9065
- Jensen CR, Genereux DP, Solomon DK, Knappe DRU, Gilmore TE. *Environ Sci Technol*, 2024, 58: 17926–17936
- Sadia M, Nollen I, Helmus R, ter Laak TL, Béen F, Praetorius A, van Wezel AP. *Environ Sci Technol*, 2023, 57: 3062–3074
- Chow SJ, Ojeda N, Jacangelo JG, Schwab KJ. *Water Res*, 2021, 201: 117292
- Ruyle BJ, Pennoyer EH, Vojta S, Becanova J, Islam M, Webster TF, Heiger-Bernays W, Lohmann R, Westerhoff P, Schaefer CE, Sunderland EM. *Proc Natl Acad Sci USA*, 2025, 122: e2417156122
- Paris-Davila T, Gaines LGT, Lucas K, Nylander-French LA. *Am J Industrial Med*, 2023, 66: 393–410
- Morales-McDevitt ME, Dunn M, Habib A, Vojta S, Becanova J, Lohmann R. *Environ Toxicol Chem*, 2022, 41: 334–342
- Zhao M, Yao Y, Dong X, Baqar M, Fang B, Chen H, Sun H. *Environ Sci Technol*, 2023, 57: 20194–20205
- Tonelli F, Masiero C, Aresi C, Torriani C, Villani S, Premoli G, Rossi A, Forlino A. *Sci Rep*, 2025, 15: 2295S
- Taibl KR, Dunlop AL, Barr DB, Li YY, Eick SM, Kannan K, Ryan PB, Schroder M, Rushing B, Fennell T, Chang CJ, Tan Y, Marsit CJ, Jones DP, Liang D. *Nat Commun*, 2023, 14: 3120
- Bolan N, Sarkar B, Yan Y, Li Q, Wijesekara H, Kannan K, Tsang DCW, Schauerte M, Bosch J, Noll H, Ok YS, Scheckel K, Kumpiene J, Gobindlal K, Kah M, Sperry J, Kirkham MB, Wang H, Tsang YF, Hou D, Rinklebe J. *J Hazard Mater*, 2021, 401: 123892
- Brase RA, Mullin EJ, Spink DC. *Int J Mol Sci*, 2021, 22: 995
- Wee SY, Aris AZ. *npj Clean Water*, 2023, 6: 57
- Wang Z, Yang W, Xu M, Li B, Chen M, Hu J, Wu P, Wu W. *Sci Rep*, 2024, 14: 25176
- Ryu S, Burchett W, Zhang S, Jia X, Modaresi SMS, Agudelo Areiza J, Rodrigues D, Zhu H, Sunderland EM, Fischer FC, Slitt AL. *Environ Sci Technol*, 2024, 58: 14641–14650
- Luo Y, Li X, Li J, Gong X, Wu T, Li X, Li Z, Zhai Y, Wei Y, Wang Y, Jiang G. *Environ Sci Technol*, 2024, 58: 13624–13635
- Abbasian Chaleshtari Z, Foudazi R. *ACS EST Water*, 2022, 2: 2258–2272
- Tshangana CS, Nhlengethwa ST, Glass S, Denison S, Kuvarega AT, Nkambule TTI, Mamba BB, Alvarez PJJ, Muleja AA. *npj Clean Water*, 2025, 8: 41
- Gomri C, Benkhaled BT, Cretin M, Semsarilar M. *Macro Chem Phys*, 2024, 225: 2400012
- Minhas S, Pandey RP, Hasan SW. *J Environ Manage*, 2025, 373: 123888
- Yadav S, Ibrar I, Al-Juboori RA, Singh L, Ganbat N, Kazwini T, Karbassiyazdi E, Samal AK, Subbiah S, Altaee A. *Chem Eng Res Des*, 2022, 182: 667–700

- 25 Lee T, Speth TF, Nadagouda MN. *Chem Eng J*, 2022, 431: 134023
- 26 Sharma S, Shetti NP, Basu S, Nadagouda MN, Aminabhavi TM. *Chem Eng J*, 2022, 430: 132895
- 27 Tan BTW, Abu Bakar NHH, Lee HL. *J Environ Chem Eng*, 2025, 13: 114990
- 28 Venkatesh Reddy C, Kumar R, Chakraborty P, Karmakar B, Pottipati S, Kundu A, Jeon BH. *Chem Eng J*, 2024, 492: 152272
- 29 Zhang Z, Sarkar D, Biswas JK, Datta R. *Bioresource Tech*, 2022, 344: 126223
- 30 Rahman MHU, Sikder R, Tonmoy TA, Hossain MM, Ye T, Aich N, Gadhamshetty V. *J Water Process Eng*, 2025, 70: 106941
- 31 Lei L, Pang R, Han Z, Wu D, Xie B, Su Y. *Crit Rev Environ Sci Tech*, 2023, 53: 1817–1835
- 32 Olawade DB, Ijiwade JO, Fapohunda O, Ige AO, Olajoyetan DO, Wada OZ. *Process Saf Environ Prot*, 2025, 196: 106869
- 33 Xiao J, Li H, Hu J, Wang Y, Wang Y. *J Environ Chem Eng*, 2025, 13: 117384
- 34 Jordan MI, Mitchell TM. *Science*, 2015, 349: 255–260
- 35 Hu XG, Wang ZJ, Deng P, Yu FB, Mu L, Wang S, Zhou QX. *Sci Sin Tech*, 2024, 54: 1838–1853 (in Chinese) [胡献刚, 王张佳, 邓鹏, 于福波, 穆莉, 王赛, 周启星. 中国科学: 技术科学, 2024, 54: 1838–853]
- 36 Barbierato E, Gatti A. *Electronics*, 2024, 13: 416
- 37 Srivastava S, Wang W, Zhou W, Jin M, Vikesland PJ. *Environ Sci Technol*, 2024, 58: 20830–20848
- 38 Lima HS, Oliveira GFV, Ferreira RS, Castro AG, Silva LCF, Ferreira LS, Oliveira DAS, Silva LF, Kasuya MCM, de Paula SO, Silva CC. *J Environ Manage*, 2024, 356: 120559
- 39 An F, Yang D, Sun X, Wei H, Chen F. *Environ Model Software*, 2025, 188: 106438
- 40 Shah MI, Javed MF, Alqahtani A, Aldrees A. *Process Saf Environ Prot*, 2021, 151: 324–340
- 41 Hu B, Dai Y, Zhou H, Sun Y, Yu H, Dai Y, Wang M, Ergu D, Zhou P. *J Hazard Mater*, 2024, 474: 134865
- 42 Fabregat-Palau J, Ershadi A, Finkel M, Rigol A, Vidal M, Grathwohl P. *Environ Sci Technol*, 2025, 59: 7678–7687
- 43 Bernardelli A, Marsili-Libelli S, Manzini A, Stancari S, Tardini G, Montanari D, Anceschi G, Gelli P, Venier S. *Water Sci Tech*, 2020, 81: 2391–2400
- 44 Yang W, Li H. *ACS EST Water*, 2025, 5: 511–524
- 45 Xu P, Chen H, Li M, Lu W. *Advcd Theor Sims*, 2022, 5: 2100565
- 46 Guo K, Fan X, Chen L, Zhang X, Zhou Z. *Sci Sin Chim*, 2025, 55: 1660–1673 (in Chinese) [郭柯香, 范欣宇, 陈乐添, 张旭, 周震. 中国科学: 化学, 2025, 55: 1660–1673]
- 47 Liu Y, Ke Z. *Sci Sin Chim*, 2025, 55: 539–541 (in Chinese) [刘裕邦, 柯卓锋. 中国科学: 化学, 2025, 55: 539–541]
- 48 Kibbey TCG, Jabrzemski R, O'Carroll DM. *Chemosphere*, 2020, 252: 126593
- 49 Kibbey TCG, Jabrzemski R, O'Carroll DM. *Chemosphere*, 2021, 275: 130124
- 50 Li W, Liang T, Lin Y, Wu W, Li S. *Prog Chem*, 2022, 34: 2619–2637 (in Chinese) [李炜, 梁添贵, 林元创, 吴伟雄, 李松. 化学进展, 2022, 34: 2619–2637]
- 51 Zeeshan M, Tabraiz S, Hashmi SI, Iqbal A, Dittmann D, Abbas Z, MacLeod CL, Ruhl AS. *Bioresource Tech Rep*, 2025, 29: 102077
- 52 Gagliano E, Sgroi M, Falciglia PP, Vagliasindi FGA, Roccaro P. *Water Res*, 2020, 171: 115381
- 53 Zhang J, Fu K, Zhong S, Luo J. *Environ Sci Technol*, 2025, 59: 3603–3612
- 54 Ersan G, Lukman A. *ACS EST Water*, 2025, 5: 479–487
- 55 Zhang Y, Luo Y, Tong J, Liu X, Chen Y, Xu T. *Sep Purif Tech*, 2023, 327: 124894
- 56 Karbassiyazdi E, Fattahi F, Yousefi N, Tahmassebi A, Taromi AA, Manzari JZ, Gandomi AH, Altaee A, Razmjou A. *Environ Res*, 2022, 215: 114286
- 57 Coutinho R, Hoshima HY, Marques M. *Biomass Bioenergy*, 2025, 202: 108208
- 58 Jiang Y, Wei B, Shang Y, Wu D, Xia C, Zhou J, Cao X, Kan Y. *Sep Purif Tech*, 2025, 363: 132089
- 59 Liu C, Zhao X, Faria AF, Deliz Quiñones KY, Zhang C, He Q, Ma J, Shen Y, Zhi Y. *Sep Purif Tech*, 2022, 302: 122161
- 60 Hosseinzadeh A, Zhou JL, Zyaie J, AlZainati N, Ibrar I, Altaee A. *Separation Purification Tech*, 2022, 289: 120775
- 61 Singh S, Suthar G, Bhushan Gupta A, Bezbaruah AN. *ACS EST Water*, 2025, 5: 1216–1228
- 62 Hu J, Xiao J, Yang W, Wang Y, Wang Y. *Desalination Water Treat*, 2025, 323: 101266
- 63 Jeong N, Park S, Mahajan S, Zhou J, Blotevogel J, Li Y, Tong T, Chen Y. *Nat Commun*, 2024, 15: 10918
- 64 Calvillo Solís JJ, Sandoval-Pauker C, Bai D, Yin S, Senftle TP, Villagrán D. *J Am Chem Soc*, 2024, 146: 10687–10698

- 65 Lu D, Sha S, Luo J, Huang Z, Zhang Jackie X. *J Hazard Mater*, 2020, 386: 121963
- 66 Alnaimat S, Mohsen O, Elnakar H. *J Environ Manage*, 2024, 370: 122857
- 67 Wen Y, Rentería-Gómez Á, Day GS, Smith MF, Yan TH, Ozdemir ROK, Gutierrez O, Sharma VK, Ma X, Zhou HC. *J Am Chem Soc*, 2022, 144: 11840–11850
- 68 Zhong H, Liu W, Li N, Ma D, Zhao C, Li J, Wang Y, Jiang G. *J Hazard Mater*, 2022, 436: 129240
- 69 Navidpour AH, Hosseinzadeh A, Huang Z, Li D, Zhou JL. *Catal Rev*, 2022, 66: 687–712
- 70 Nasrollahpour S, Tanhadoust A, Pulicharla R, Brar SK. *iScience*, 2024, 27: 111140

Machine learning-enabled technologies for the removal of per- and polyfluoroalkyl substances from water

Mingxia Bai, Chaochao Song, Yuanzheng Zhang^{*}, Junfeng Niu

MOE Key Laboratory of Resources and Environmental Systems Optimization, College of Environmental Science and Engineering, North China Electric Power University, Beijing 102206, China

^{*}Corresponding author (email: zhangyz@ncepu.edu.cn)

Abstract: Per- and polyfluoroalkyl substances (PFAS), characterized by their exceptional chemical stability and environmental persistence, are ubiquitous in global water systems and pose potential risks to both ecosystems and human health. Conventional removal technologies typically rely on extensive experimental trials and parameter optimization, yet they are often constrained by low efficiency, high operational costs, and limited adaptability. In recent years, machine learning (ML) approaches, driven by artificial intelligence, have rapidly advanced within the environmental science domain, demonstrating significant potential to enhance removal efficiency and optimize treatment processes. This research comprehensively summarizes the application of ML in mainstream PFAS removal technologies, including adsorption, membrane separation, electrochemical methods, and photocatalysis, emphasizing its role in predicting removal performance, optimizing process parameters, guiding material screening and synthesis, and elucidating underlying mechanisms. Furthermore, current challenges associated with ML approaches are critically discussed, encompassing data quality, model generalizability, environmental adaptability, and toxicity prediction. Finally, we propose strategies such as leveraging high-quality datasets, developing multimodal models, and implementing intelligent deployment, aimed at establishing an efficient and low-carbon PFAS management framework, thereby providing a valuable reference for future research and engineering applications.

Keywords: machine learning, per- and polyfluoroalkyl substances, contaminants, water treatment technologies, degradation mechanism, model generalization

doi: [10.1360/SSC-2025-0195](https://doi.org/10.1360/SSC-2025-0195)