综述与专论

DOI：10.11949/0438-1157.

数据驱动的氟喹诺酮类抗生素降解性能预测

王力辰[1]，王琪[1]，李欣[1]，刘晓红[1]，唐次来[2]

（[1]延安大学化学与化工学院，陕西 延安 716000；[2]赣南师范大学地理与环境工程学院，江西 赣州 341000）

摘要： 氟喹诺酮类抗生素（Fluoroquinolones，FQs）在水环境中的残留物对公共健康构成潜在风险。活化过氧单硫酸盐催化降解FQs的过程受多重非线性条件制约，传统动力学模型难以准确外推复杂反应下的降解效能。突破传统实验试错方法局限，构建涵盖极端梯度提升、随机森林、梯度提升及类别提升的多模型预测框架，旨在从有限数据中挖掘关键变量与降解率间的映射关系。结果显示梯度提升模型在测试集上取得最佳表现（R^2=0.906，RMSE=0.067）。借助SHAP（SHapley Additive exPlanations）分析、部分依赖图以及三维特征交互图对结果解读，定量解释了催化剂投加量、初始pH及反应时间对降解率的主导贡献，揭示了变量间存在的协同或拮抗作用，显著降低实验试错成本，为活化催化降解体系提供高可信度的预测工具。

关键词： 机器学习；氟喹诺酮类抗生素；算法；SHAP；催化剂

中图分类号： TP181　　**文献标志码：** A

文章编号： 0438-1157（XXXX）XX-0001-11

Data-driven prediction of fluoroquinolone antibiotics degradation performance

WANG Lichen[1], WANG Qi[1], LI Xin[1], LIU Xiaohong[1], TANG Cilai[2]

([1] *College of Chemistry and Chemical Engineering, Yan'an University, Yan'an, 716000, Shaanxi, China*; [2] *College of Geography and Environmental Engineering, Gannan Normal University, Ganzhou, 341000, Jiangxi, China*)

Abstract: Residues of fluoroquinolone antibiotics (Fluoroquinolones, FQs) in aquatic environments have raised considerable public health concerns due to their persistent nature and widespread occurrence. The degradation of FQs via activated peroxymonosulfate-activated systems is governed by multiple nonlinear factors, rendering conventional kinetic models inadequate for accurately predicting degradation efficiency under complex reaction conditions. To address this limitation, the present study establishes a multi-model prediction framework that integrates extreme gradient boosting, random forest, gradient boosting, and categorical boosting, aiming to elucidate the mapping between key variables and degradation rates from a limited dataset. Among the evaluated models, the gradient boosting model demonstrated superior predictive performance on the test set, achieving R^2 = 0.906 and RMSE = 0.067. The feature importance ranking revealed that contact time, catalyst dosage, and initial pH were the three most influential contributors. Furthermore, SHAP (SHapley Additive exPlanations) analysis, in conjunction with partial dependence plots and three-dimensional feature interaction maps, was employed to interpret the model outputs. The results quantitatively delineated the dominant roles of catalyst dosage, initial pH, and reaction time in determining degradation efficiency, while also uncovering synergistic and antagonistic interactions among variables.

收稿日期： 2026-08-09　　**修回日期：** 2026-09-20

通信作者： 刘晓红（1972—），女，博士，教授，xiaohliu29@126.com；唐次来（1981—），男，博士，教授，bolong8111@163.com

第一作者： 王力辰（2001—），男，硕士研究生，2681951875@qq.com；王琪（2002—），女，硕士研究生，2864972729@qq.com.

基金项目： 陕西省提升公众科学素质计划项目（2020PSL(Y)047）；延安市科技项目（2025-SFGG-127）

引用本文： 王力辰, 王琪, 李欣, 刘晓红, 唐次来. 数据驱动的氟喹诺酮类抗生素降解性能预测[J]. 化工学报, XXXX, XX(XX): 1-11

Citation: WANG Lichen, WANG Qi, LI Xin, LIU Xiaohong, TANG Cilai. Data-driven prediction of fluoroquinolone antibiotics degradation performance[J]. CIESC Journal, XXXX, XX(XX): 1-11

This data-driven approach substantially reduces the cost of conventional trial-and-error experimentation and offers a robust predictive tool for the design and optimization of persulfate-based catalytic degradation systems. The proposed framework is readily transferable to other advanced oxidation processes or emerging pollutant systems, and provides a viable pathway toward data-driven decision-making in wastewater treatment process control.

Keywords: machine learning; fluoroquinolone antibiotics; algorithm; SHAP; catalyst

引 言

氟喹诺酮类抗生素(FQs)在水环境中的残留风险日益凸显[1],因其化学稳定性强、生物降解性差的特点,对传统水处理工艺构成持续挑战。基于过氧单硫酸盐(PMS)的高级氧化技术因能产生高氧化电位的硫酸根自由基、羟基自由基及非自由基活性物种,可高效破坏抗生素分子结构,且具备pH适应范围广泛、反应速率快及氧化能力强等优势[2],其技术核心在于高效催化材料的设计与适配。然而,PMS催化体系对FQs的降解效率受催化剂理化性质、反应条件及污染物分子特征等多因素协同调控[3],交互作用与构效关系复杂,传统实验方法在描述此类非线性全局优化方面存在明显局限。因此构筑数据驱动的预测模型,以系统刻画多因素与降解效率间的复杂映射关系。

机器学习(Machine Learning, ML)作为一种无需先验机理假设的数学建模方法,通过构建输入-输出的统计映射关系,可在复杂体系中实现高效预测与特征识别,已广泛应用于环境催化领域的性能预测与参数选择[4]。本研究使用四种学习模型预测与对比抗生素的降解去除率,并利用SHAP分析进一步揭示关键驱动因素。常见的模型有随机森林(RF)、梯度提升(GB)、极端梯度提升(XGBoost)、类别提升(CatBoost)等。其中RF由Leo Breiman[5]正式提出,其核心是多次从原始数据中有放回地抽样,得到多个不同的训练子集并在每个子集上独立训练一棵决策树,节点分裂时只从全部特征中随机选取一部分作为候选分裂特征。随机森林在样本量不大、特征关系不十分复杂的任务中优先考虑。GB属于串行集成方法,最早由Jerome H. Friedman[6]系统论述,其逐步修正残差初始时可以用常数作为预测起点,后续每一步都新添一棵浅层决策树,用这棵树去逼近当前模型预测值的负梯度方向,每棵新树学习前面所有树还未捕捉到的部分信号。算法通过乘以一个较小的学习率来控制每棵树的贡献,使模型慢慢逼近真实关系。XGBoost由Tianqi Chen等开发,在工程实现和算法设计上做出创新,改进了经典的梯度提升框架[7]。首先目标函数中加入了正则化项,惩罚叶子节点个数和叶子权重的L2范数,直接约束模型的复杂程度。算法支持列抽样,加快计算又降低了过拟合风险,且会自动处理部分缺失值,学习出一个最优的缺失分裂方向。XGBoost在大量数据竞赛中被证明性能卓越,但缺点是需要事先将类别特征转换为数值型,调参空间也相对复杂。CatBoost由Dorogush等开源,不使用传统独热编码而采用有序目标统计方法,在计算某一行样本的编码时会排除该样本自身的历史数据,避免了因数据泄漏导致的预测偏移[8]。此思想延伸到其整体训练流程中,形成了有序提升的策略,模型默认带有L2正则化。这四种模型都扎根于决策树,实践中CatBoost对默认超参数表现稳健,不需大量搜索参数就能输出高质量结果,适合处理大规模结构化数据;随机森林依靠降低方差来提升精度;梯度提升通过不断修正偏差把强拟合能力带入模型;XGBoost和CatBoost则分别是梯度提升在效率和易用性上的延伸。

本研究以FQs为目标污染物,通过整合PMS催化降解实验的公开数据,构建涵盖催化剂性质、反应条件与抗生素分子特征的数据集。在此基础上,系统评估上述四种模型对降解去除率的预测性能,筛选最优算法,并结合SHAP(SHapley Additive exPlanations)方法[9]对关键驱动因素进行可解释性分析,揭示影响降解效率的主导因素及其非线性作用机制,为高级氧化工艺的智能优化提供理论依据与方法支撑。

1 材料与方法

1.1 数据收集与预处理

本研究收集了来自Science Direct数据库公开文献中的抗生素降解实验数据共549组数据点[10-38],包括恩诺沙星(Enrofloxacin)、环丙沙星(Ciprofloxacin)、氧氟沙星(Ofloxacin)、左氧氟沙星

(Levofloxacin)、诺氟沙星(Norfloxacin)、莫西沙星(Moxifloxacin)和洛美沙星(Lomefloxacin)。这些抗生素作为新兴有机污染物,在环境中频繁检出且难以被常规工艺去除,因此成为高级氧化过程研究的焦点。数据集所涉及的实验均基于PMS活化的催化氧化体系,记录了三类关键输入变量:(1)催化剂理化性质,包括比表面积(S_{BET}, $m^2 \cdot g^{-1}$)、总孔容(Total PV, $cm^3 \cdot g^{-1}$)和平均孔径(D_p, nm),这些参数直接影响活性位点的暴露程度和传质效率;(2)反应条件,包括抗生素初始浓度(Initial concentration, $mg \cdot L^{-1}$)、PMS投加量(The dosage of PMS, $g \cdot L^{-1}$)、催化剂投加量(The dosage of catalyst, $g \cdot L^{-1}$)、反应温度(Reaction temperature, ℃)、接触时间(Contact time, min)以及初始pH,这些因素共同决定了自由基生成速率和底物竞争反应路径;(3)抗生素分子属性,包括第一、第二和第三解离常数(pK_{a1}、pK_{a2}、pK_{a3})、辛醇-水分配系数对数(log*P*)、分子量(MW, $g \cdot mol^{-1}$)、拓扑极性表面积(TPSA, $Å^2$)以及偶极矩(Debye)。这些分子描述符反映了抗生素的电离行为、疏水性和极性,对催化剂表面的吸附亲和力及电子转移过程具有重要影响。模型输出目标为抗生素去除率(*Removal*),原始值介于0~1之间,其原始计算方式见式(1),代表PMS/催化剂体系中对目标污染物的降解率。各输入特征及目标变量的分布形态见图1。

$$Removal(\%) = \frac{C_0 - C_t}{C_0} \times 100\% \#(1)$$

其中C_0为反应开始前抗生素的初始浓度,C_t为反应终点污染物的残余浓度,所有文献数据均采用该标准定义。

原始数据在建模前经过系统的预处理流程,以消除量纲差异、处理缺失值并实现类别特征的数值化。首先根据特征的数据类型进行自动区分:非数值列,如污染物名称被识别为类别特征,其余列为数值特征。对于数值特征,采用中位数策略填充缺失值,该方法对异常值具有鲁棒性,能有效避免均值填充带来的偏差。随后应用基于Z-score的标准化(StandardScaler)将各特征转换为标准正态分布,变换公式见式(2)。

$$z = \frac{x - \mu}{\sigma} \#(2)$$

其中z为标准化后的特征值,x为原始特征值,μ和σ分别为该特征在训练集上的均值和标准差。此线性缩放操作确保了不同量纲,如反应时间与比表面积等特征在模型训练中具有同等权重,避免了数值范围差异对树模型分裂节点选择的潜在影响。对于类别特征,先以众数填补缺失项,再采用独热编码(OneHotEncoder)将其转换为二进制数值向量,从而使非数值信息被集成学习模型有效利用。所有预处理步骤均通过ColumnTransformer并行应用于相应特征列,最终生成完全数值化的特征矩阵,以供后续回归建模使用。

1.2 机器学习模型选择

抗生素降解过程受多因素非线性耦合影响,传统动力学模型难以精确描述其复杂响应关系,本研究选取了四种先进的集成学习算法——XGBoost、GB、RF以及CatBoost来构建预测模型。集成学习通过组合多个弱学习器即决策树,来降低过拟合风险并提升泛化能力,尤其适用于小样本数据集。本文将数据按9:1单次随机划分为训练集和测试集,四个模型共享同一划分,以保证公平。预设各模型的超参数,并在最终训练中固定使用:GradientBoosting:基础学习器数量为1800,最大树深度为5,节点分裂所需最小样本数为2,叶节点分裂最小样本数为1,每次迭代随机抽取90%的样本(subsample=0.90)以增强随机性;XGBoost:树数量为2000,最大深度为5,最小叶子节点权重为1,行采样率和列采样率均为0.95,并引入L2正则化参数(λ = 0.5)以抑制过拟合;CatBoost:迭代次数为1800,最大深度为6,叶节点L2正则化系数为3.0,该模型对类别特征具有原生高效处理能力;RandomForest:树数量为700,每次分裂考虑的最大特征数为1即仅使用平方根特征数,节点分裂最小样本数为2,叶节点最少样本数为1。

所有模型均采用未经特征筛选的原始特征集作为输入,以完整保留各因素对降解FQs的潜在贡献。后续将通过SHAP特征重要性分析、部分依赖图(Partial Dependence Plot, PDP)与三维特征交互分析进一步阐释关键变量对降解效率的驱动机制,并采用R^2、均方根误差(Root Mean Square Error, RMSE)综合评价模型性能。该建模框架旨在为PMS/催化剂体系的高效设计提供数据驱动的快速预测工具,减少繁琐实验试错成本。

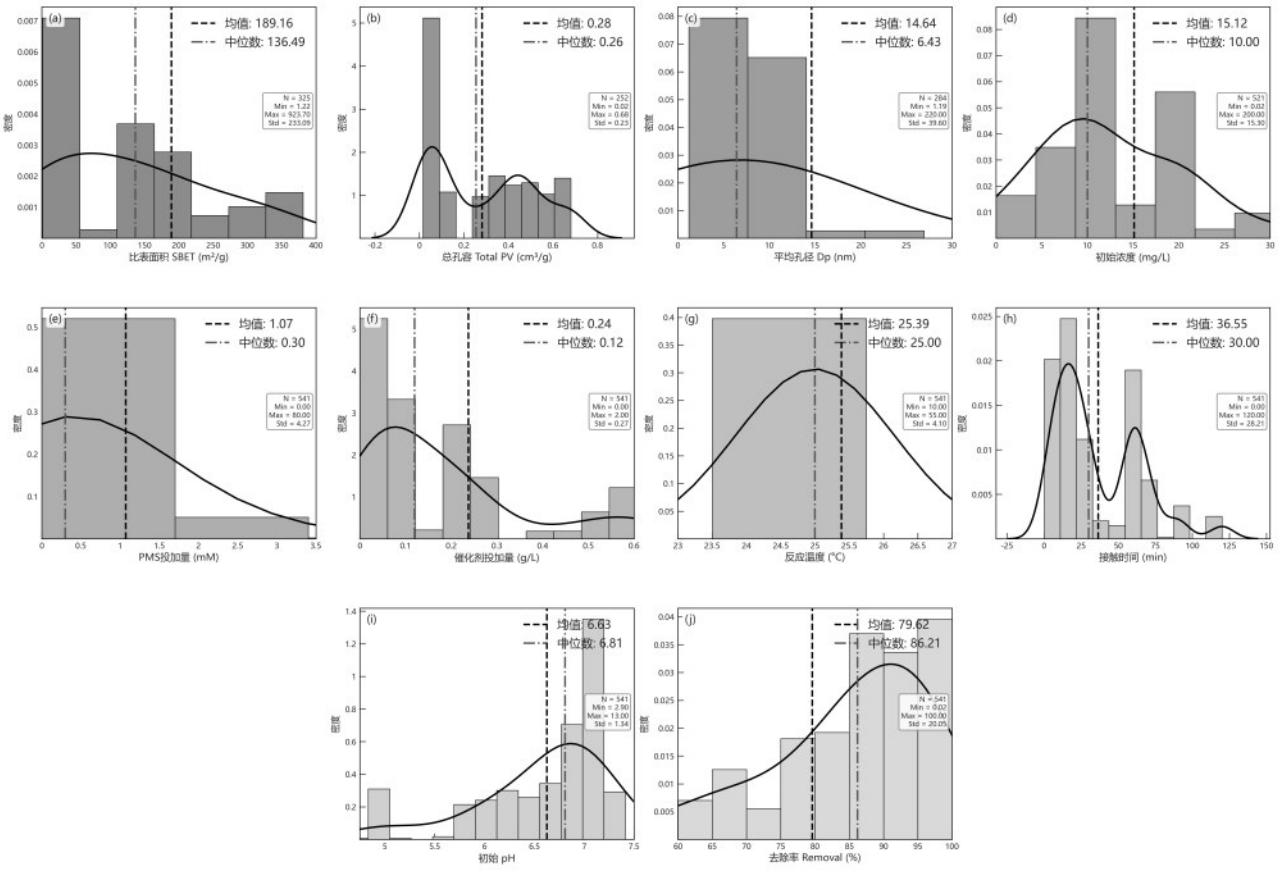


图1　各输入特征及目标变量的分布形态

Fig.1　The distribution patterns of input features and target variable

2　结果与讨论

2.1　模型性能对比

为系统量化各机器学习模型对PMS/催化剂体系降解抗生素的预测可靠性，本研究选取决定系数(R^2)和均方根误差(RMSE)作为核心评价指标，其计算公式分别如下，见式(3)和式(4)。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \#(3)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \#(4)$$

其中，y_i 表示第 i 个样本的真实观测值，$\hat{y}_i$ 为模型对第 i 个样本的预测值，$\bar{y}$ 为所有真实值的均值，n 为样本总量。R^2 取值范围为[0l1]，其值越接近1，表明模型对目标变量方差的解释能力越强；而RMSE则反映预测误差的幅值，数值越低意味着模型预测精度越高。

本研究的输出目标变量设定为抗生素的降解去除率，用*Removal*表示，以定量反映PMS/催化降解体系在特定反应条件下对目标污染物的氧化削减能力。基于上述指标，本研究对RF、XGBoost、CatBoost及GB四种集成模型在测试集上的预测性能进行了系统比较。结果表明，四种模型在预测抗生素催化氧化降解效能方面均展现出良好的适用性，各模型训练集结果符合预期，分别为GB(R^2=0.918，RMSE = 0.058)、XGBoost(R^2=0.895，RMSE = 0.060)、CatBoost(R^2=0.869，RMSE=0.072)与RF(R^2=0.869，RMSE=0.070)。其中GB模型取得了最优预测效果，其测试集 R^2 达到0.906，RMSE低至0.067；紧随其后的XGBoost(R^2=0.884，RMSE = 0.073)、CatBoost(R^2=0.836，RMSE=0.087)与RF(R^2=0.813，RMSE=0.093)亦表现出较好的拟合优度。GB的$|\Delta R^2|$=0.014，|ΔRMSE|=0.009，说明模型未过拟合；RF和CatBoost差距略大，但仍在合理范围，在有限数据规模的复杂催化体系中尤为重要，直接关系模型对多维反应条件的非线性外推可靠性。为进一步直观评估模型预测值与实际值之间的一致性，图2绘制了四种模型在测试集上的预测–实测散点图。图中红色虚线为拟合回归线，各数据点均匀且紧密

地聚集于理想对角线($y = x$)附近,未表现出明显的系统偏差。该直观视觉结果充分验证了集成学习模型在捕获多因素协同与拮抗交互作用及精准预估抗生素氧化削减效率方面的稳健性。

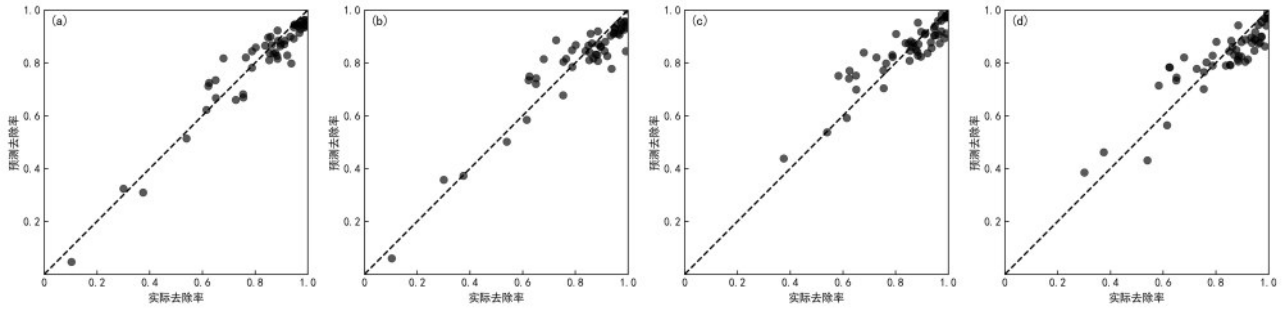


图2 四种模型的预测-实测散点图:(a)XGBoost;(b)CatBoost;(c)GB;(d)RF

Fig.2 Scatter plots of predictions vs. actual measurements for four models: (a) XGBoost; (b) CatBoost; (c) GB; (d) RF

2.2 模型解释与可视化

2.2.1 各特征平均绝对SHAP值排序图 基于GB模型得到各特征平均绝对SHAP值的排序图,见图3。结果显示接触时间的平均绝对SHAP值最高,表明其对预测去除FQs的贡献最大,是影响降解效率最关键的参数;催化剂投加量次之;随后依次为初始pH、总孔容、初始浓度、PMS投加量、比表面积和平均孔径;反应温度的平均绝对SHAP值最小,对模型输出的影响最弱。pK_{a1}、pK_{a2}、pK_{a3}、log$P$、MW、TPSA以及Debye这7个分子描述符的平均|SHAP|值较低,未进入重要性排序前列,其在同一污染物内为恒定值,实际等价于污染物身份标识,信息被实验条件变量所覆盖。此外,本文模型并未将催化剂类型作为独立的类别变量,而是通过可量化的催化剂理化性质与催化剂投加量来表征催化剂差异。SHAP分析表明,这些参数与接触时间、初始pH等反应条件共同主导了去除率的预测,说明模型有效捕捉到催化剂对降解性能的影响。该排序结果与高级氧化工艺的实际机理一致:催化剂投加量决定活性位点数量与自由基生成速率[16];接触时间提供了充分的反应窗口[18];初始pH则通过影响自由基形态和抗生素解离状态调控降解效率[24]。

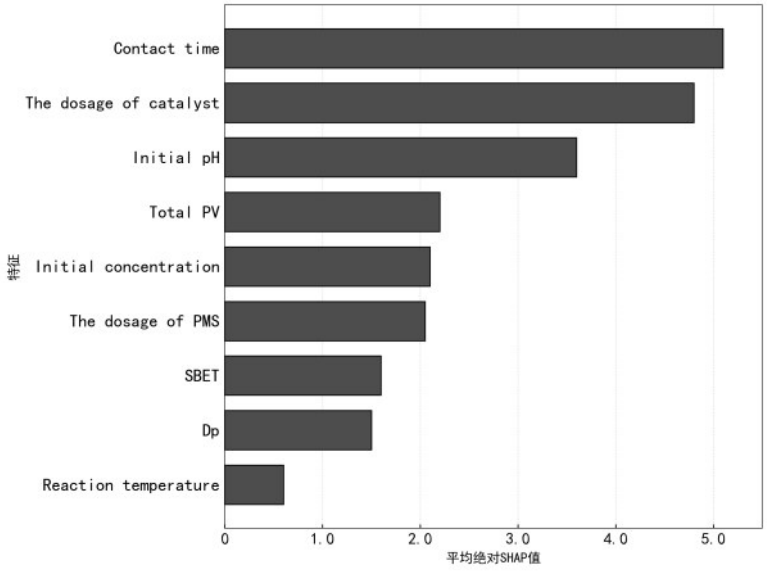


图3 平均绝对SHAP特征重要性排序

Fig.3 Average absolute SHAP feature importance ranking

2.2.2 皮尔逊相关系数分析 对各输入特征进行皮尔逊相关系数(Pearson correlation coefficient, r)分析,以初步判断特征之间及特征与目标变量之间的线性相关程度,皮尔逊相关系数衡量两个变量之间的线性相关强度[39],其计算公式见式(5)。r =

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}\#(5)$$

式中：X_i 和 Y_i 分别为两个变量的第 i 个观测值；$\bar{X}$ 和 $\bar{Y}$ 分别为两个变量的均值；n 为样本数量。r 的取值范围为 [− 1,1]，正值表示正相关，负值表示负相关，绝对值越接近 1 表明线性相关越强，越接近0表明线性相关越弱。结果见图4，比表面积与总孔容呈正相关（r=0.46），表明高比表面积材料通常伴随较大的孔容；总孔容与初始浓度同样呈正相关（r=0.49）；比表面积与平均孔径呈负相关（r=−0.44），反映出多孔材料中比表面积与孔径的权衡关系；其余特征间的相关系数绝对值大多低于0.3，如 S_{BET} 与反应温度相关系数接近0，D_P 与PMS投加量相关系数为−0.15等。所有特征间的相关系数绝对值均未超过0.8，不存在高度共线性问题，因此无需因多重共线性而剔除特征。

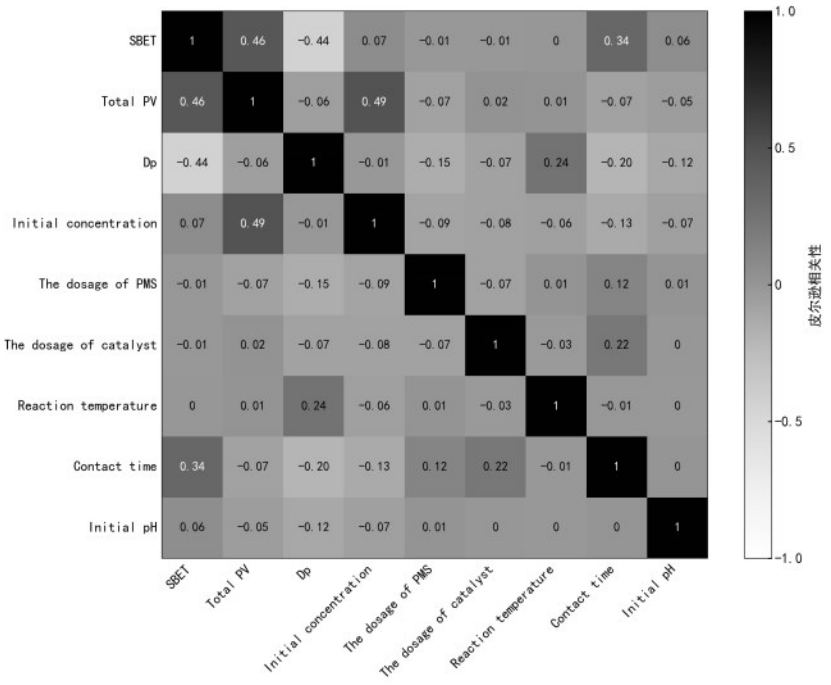


图4　皮尔逊相关系数矩阵

Fig.4　Pearson correlation coefficient matrix

2.3　SHAP总结与部分依赖分析

2.3.1　SHAP总结　采用SHAP方法对最优模型进行全局解释。图5展示了各特征对模型预测输出的SHAP总结图，其中每个点代表一个样本，横坐标为该特征对预测值的SHAP贡献，正值表示正向驱动，负值表示抑制作用，纵坐标按特征重要性从高到低排列，颜色表示该特征值的高低，红色为高值，蓝色为低值。从图5可以看出，接触时间的SHAP值分布范围最广，高值多集中在正半轴，表明延长反应时间显著提升去除率，与自由基充分氧化的机理一致[16]；催化剂投加量的SHAP值同样分布较宽，高投加量对应正贡献，符合增加活性位点数量的预期，低投加量则抑制反应[40]；初始pH SHAP值呈现分散特征，表明存在最适pH区间，偏离该区间会降低降解效率；PMS投加量SHAP值分布较为分散且正负混杂，说明其对去除率的影响受其他特征调节，存在非线性交互作用；初始浓度SHAP值以负贡献为主，说明高浓度抗生素会消耗有限自由基，起抑制作用；总孔容、比表面积、平均孔径和反应温度的SHAP值分布相对集中，其中总孔容的高值更倾向于正向贡献，而反应温度的影响最弱。

2.3.2　部分依赖分析　部分依赖图直观展示单个特征对模型预测去除率的边际效应[41-42]，揭示可能存在的非线性关系。图6绘制了9个输入特征的部分依赖图，横坐标为特征原始值，纵坐标为SHAP值，正值表示正向驱动去除率，负值则表示抑制作用。SHAP值基于标准化特征计算，但依赖图横坐标统一采用原始特征值展示，便于机理解读。图6(a)、(c)和(g)显示，比表面积的SHAP值随着其增大缓慢上升，大比表面积提供更多活性位点；平均

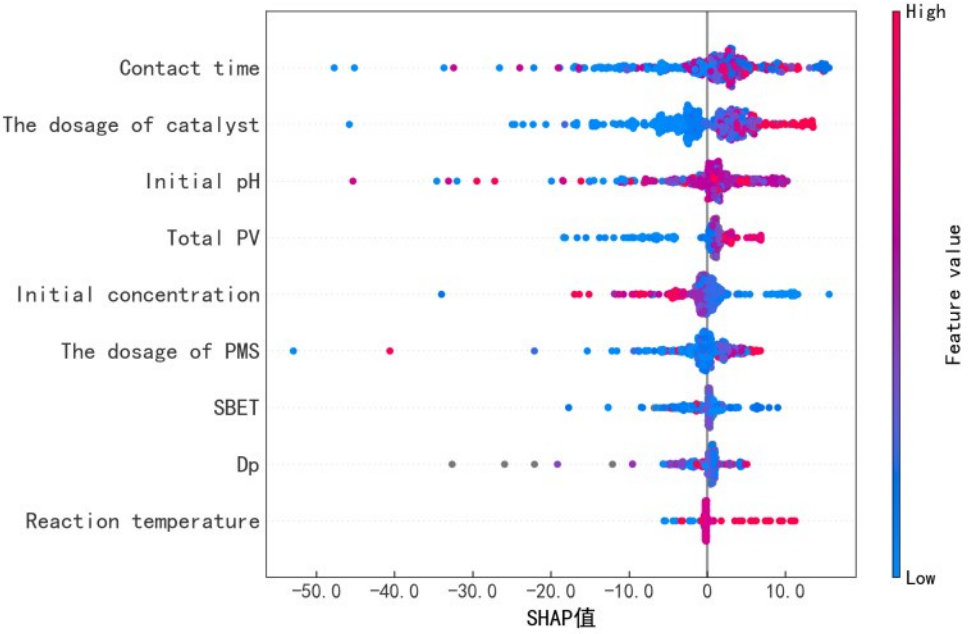


图5 SHAP值摘要图

Fig.5 SHAP value summary chart

孔径在较小孔径时SHAP值偏低,孔径增大时SHAP值提升并趋于稳定;反应温度的SHAP值分布较分散,整体影响最弱,温度升高可加速反应但也会促进自由基猝灭[43],建议控制在30~45 ℃。由图6(b)知,总孔容SHAP值随总孔容增大单调上升,总孔容小于0.2 $cm^3 \cdot g^{-1}$时SHAP值偏低,孔容过小限制反应物扩散;当孔容大于0.25 $cm^3 \cdot g^{-1}$时,正向贡献显著增强,较大的孔容可为自由基和抗生素分子提供更充足的反应空间,延长有效接触时间,但需注意孔容过大可能伴随比表面积下降,应与孔径协同优化。由图6(d)得,初始浓度SHAP值随初始抗生素浓度增加而持续下降,整体为负或接近零,表明高浓度底物会显著抑制去除率。当初始浓度大于20 $mg \cdot L^{-1}$时,SHAP负值幅度加大,说明体系中自由基总量固定,高浓度抗生素竞争消耗自由基,导致单位质量去除率下降,建议对高浓度废水进行预处理稀释或同步适量增加催化剂/PMS投加量。从图6(e)看出,PMS投加量SHAP值呈非单调变化,投加量增至0.18 $g \cdot L^{-1}$时达到峰值正贡献,由于1.0 $g \cdot L^{-1}$时正向贡献最明显,推荐最佳PMS投加量为1.0 $g \cdot L^{-1}$。图6(f)显示,催化剂投加量SHAP值随投加量增加呈先快速上升后趋于平缓的趋势。低投加量时SHAP值为负,自由基生成不足;当投加量增至0.3~0.8 $g \cdot L^{-1}$时,SHAP值转为正并达到较高水平,表明活性位点充足;继续增加投加量增益有限,部分样本出现轻微下降,可能因催化剂团聚或光遮蔽,推荐最佳投加量范围为0.3~0.8 $g \cdot L^{-1}$。图6(h)接触时间的SHAP值随反应时间增加上升,由负转正后持续增加,说明去除率随时间增加而显著提高,10 min以内SHAP值多为负,15 min后正向贡献明显,60min后增益放缓,建议反应时间不低于10 min,实际工艺可根据能耗需求可选择15~60 min。图6(i)初始pH的SHAP值在pH 5~8区间内为正且较高,pH过低或过高时转为负值。弱酸性至中性条件最有利于PMS活化产生硫酸根自由基和羟基自由基;强酸或强碱环境会改变自由基形态或加速无效分解[44],推荐初始pH控制在5~8。综上所述,SHAP部分依赖图验证了催化剂投加量、接触时间、初始pH等关键因素对FQs降解去除率的驱动作用,并量化了各特征的最优操作区间,为PMS/催化剂体系的工艺参数优化提供了明确的数据驱动依据。

2.4 三维特征交互分析

基于SHAP重要性排序构建了反应时间、催化剂投加量、初始pH值和PMS投加量的四组三维交互分析,揭示了强非线性协同与拮抗机制,见图7。图7(a)中反应时间与催化剂投加量的交互关系明显,低剂量催化剂即0.5~1.1 $g \cdot L^{-1}$并配合10~30 min反应时间,去除率可达80%以上,但投加量过高会引发团聚导致边际效益递减;图7(b)中反应时间与

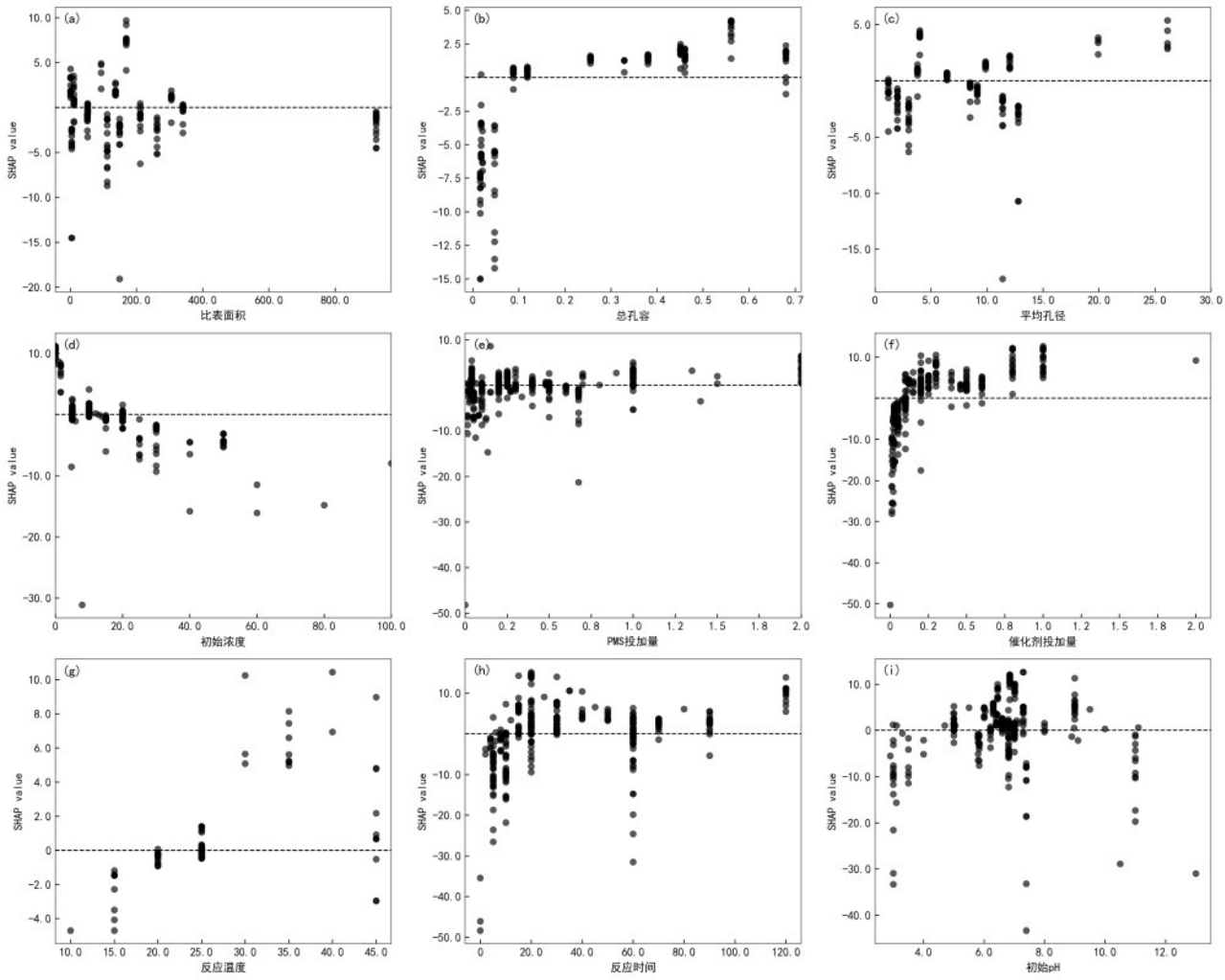


图6　各特征的部分依赖图:(a)比表面积;(b)总孔容;(c)平均孔径;(d)初始浓度;(e)PMS投加量;(f)催化剂投加量;(g)反应温度;(h)反应时间;(i)初始pH

Fig.6　Partial dependency graphs of features: (a) S_{BET} ; (b) Total pore volume; (c) Average aperture; (d) Initial concentration; (e) PMS dosage; (f) Catalyst dosage; (g) Reaction temperature; (h) Reaction time; (i) Initial pH

初始pH的协同效应则集中于中性区间(pH 6~8、10~30 min),强酸下的表面质子化排斥及PMS非目标消耗、强碱下的自由基猝灭均可能使效率显著衰减;图7(c)中反应时间与PMS投加量的交互凸显了氧化剂用量阈值,适量的PMS为自由基链式反应提供充足物质基础,但过量投加会诱发自由基猝灭及活性位点竞争,从而导致去除率降低;图7(d)中催化剂投加量与初始pH在空间上交互作用明显,中性pH(6~8)与低投加量的精准匹配可协同突破90%。综合四组高亮区域,最优工艺窗口高度收敛于pH 5~8、反应时间10~30 min、适量的催化剂与PMS投加量区间,充分体现了数据驱动可视化在突破单因素优化局限、精准锁定高维空间协同窗口及规避极值陷阱中的直观价值。

2.5　模型对比与优越性分析

本研究的模型优势在于小样本情况下,在对抗生素降解的预测中GB模型的测试集R^2达到0.906。Li等[45]采用了600条实验记录,系统比较了12种机器学习算法预测羟基胺改性过渡金属氧化物/PMS体系中的抗生素降解效率,其中CatBoost取得了最佳性能(测试集R^2=0.811),模型结构更为复杂;Bi等[46]提出的CatBoost双任务模型预测抗生素降解动力学(测试集R^2=0.807),在识别活性物种生成方面具有较高精度。在模型解释方面,本研究采用SHAP分析方法识别出催化剂投加量、接触时间和初始pH为影响去除率的三个最关键的驱动因素,这与Li等研究中识别出PMS投加量、pH、初始污染物浓度和催化剂负载量为主要驱动因子的结论类

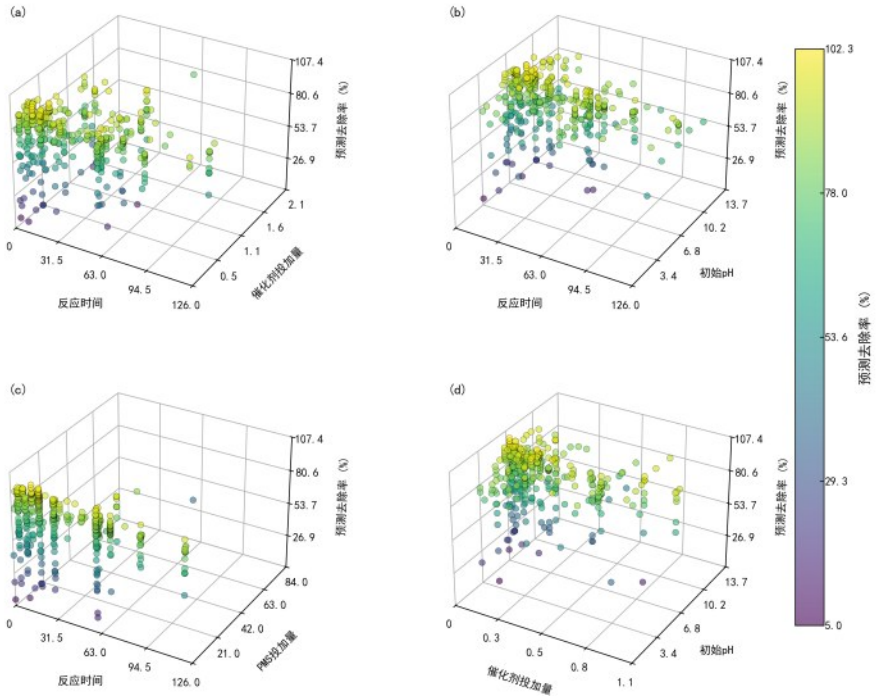


图7 三维特征联合作用图:(a)反应时间与催化剂投加量;(b)反应时间与初始pH;(c)反应时间与PMS投加量;(d)催化剂投加量与初始pH

Fig.7 Three-Dimensional Feature Interaction Plots: (a) Reaction Time versus Catalyst Dosage; (b) Reaction Time versus Initial pH; (c) Reaction Time versus PMS dosage; (d) Catalyst Dosage versus Initial pH

似,验证了SHAP分析在揭示高级氧化工艺中关键参数方面的可靠性。在数据质量差异方面,本研究在数据预处理阶段进行了严格的文献筛选,纳入明确报告了催化剂理化性质、反应条件及抗生素分子属性的文献,相对完整的特征覆盖使得模型能够更充分地学习多因素间的非线性映射关系。其次,在任务复杂度差异方面,Li等的研究涵盖了改性金属氧化物/PMS体系,催化剂类型多样但体系复杂度较高;本研究聚焦于PMS/催化剂体系对FQs的降解,反应机理相对统一,降低了模型需要学习的体系差异,使得有限样本中的映射关系更易于被捕捉。在模型算法的适配方面,本研究采用的GB模型无需依赖复杂的超参数调优策略即可获得良好的精度预测,与其他基于树模型的集成方法相比,GB模型在549个样本的小规模数据集上仍能有效捕捉特征间的非线性交互作用。本研究在有限的实验数据条件下实现了较为优秀的预测精度,且模型结构清晰,为抗生素废水处理工艺的参数优化提供了一种高效可靠的学习方案。

3 结论

1)本研究构建基于集成学习多模型的预测框架,用于预估PMS/催化剂体系对氟喹诺酮类抗生素(FQs)的降解去除率。通过统一数据预处理与随机划分,系统比较了梯度提升(GB)、极端梯度提升(XGBoost)、类别提升(CatBoost)和随机森林(RF)四种集成学习算法的预测性能。结果表明,GB模型在测试集上取得了最优预测精度(R^2=0.906,RMSE=0.067),展现出突出的抗过拟合能力。该模型在处理有限样本、高维特征及非线性耦合关系等方面展现出优越性,表明梯度提升算法适用于环境催化领域的复杂回归任务,为后续工艺参数快速筛选与响应预测提供了可靠的数据驱动工具。

2)SHAP全局特征重要性分析揭示,催化剂投加量、接触时间和初始pH是影响降解去除率的前三位调控因子,该排序与PMS活化氧化机理高度一致。SHAP图的量级分析与方向性进一步表明,各特征对输出预测的影响并非恒定,而是随其取值区间发生非线性变化,凸显了单一线性假设在描述该类复杂催化体系时的局限性。因此借助部分依赖图(PDP)与三维特征交互图共同对上述关键变量的

有效操作窗口进行了量化界定：催化剂投加量宜维持在0.3~0.8 g·L^{-1}，初始pH建议控制在6~8，接触时间不低于10 min。上述基于模型推演解析所得出的定量边界，为后续实验与工艺设计提供了明确的数据支撑。

3）本研究建立的机器学习预测框架，能够在无需预设动力学模型的条件下，从有限实验样本中高效提取输入特征与去除率之间的隐性物化规律，显著降低了传统参数优化过程中对大量实验试错的依赖。该策略为抗生素废水处理工艺的精准调控提供了低成本、高效率的决策支持。相较于纯实验研究，数据驱动范式在加速材料筛选与反应条件优化方面具有显著潜力，契合绿色化学与智能水处理的发展方向。

参考文献

[1] Jiang S Y, Zhou Y R, Xu W, et al. Machine learning-driven optimization and application of bimetallic catalysts in peroxymonosulfate activation for degradation of fluoroquinolone antibiotics[J]. Chemical Engineering Journal, 2024, **486**: 150297.[LinkOut]

[2] Jiang C L, Ji Y F, Shi Y Y, et al. Sulfate radical-based oxidation of fluoroquinolone antibiotics: Kinetics, mechanisms and effects of natural water matrices[J]. Water Research, 2016, **106**: 507-517.[LinkOut]

[3] Adly A, Galal M M, Matta M E. Catalytic degradation of norfloxacin using persulfate activation by Ni-Fe layered double hydroxide catalyst supported on activated carbon[J]. Scientific Reports, 2025, **15**: 5132.[LinkOut]

[4] Dihan M R, Alam M A, Akter S, et al. Statistical modeling and application of machine learning for antibiotic degradation using UV/persulfate-peroxide based advanced oxidation process[J]. Heliyon, 2025, **11**(13): e43664.[LinkOut]

[5] Breiman L. Random forests[J]. Machine Learning, 2001, **45**(1): 5-32.[LinkOut]

[6] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of Statistics, 2001, **29**(5): 1189-1232.[LinkOut]

[7] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. ACM, 2016: 785-794.[LinkOut]

[8] Dorogush A V, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support[PP/OL]. arXiv: 1810.11363 [cs. LG].[LinkOut]

[9] Zhang S T, Wang H H, Wu T T, et al. Regulating surface band structure of anisotropic $SrTiO_3$ via ultrathin TiO_2 layer to promote spatial charge separation[J]. Chemical Engineering Journal, 2025, **507**: 160059.[LinkOut]

[10] Zheng K W, Xiao L. Fe loading 3D micro-meso-porous carbon sphere derived from natural cellulose of sawdust activating peroxymonosulfate for degradation of enrofloxacin[J]. International Journal of Biological Macromolecules, 2024, **259**: 129366.[LinkOut]

[11] Feng D, Li X C, Cao S H, et al. Efficient degradation of Enrofloxacin with novel magnetic $MnFe_2O_4/NiS_2$ composite as an activator of peroxymonosulfate[J]. Chemical Engineering Journal, 2024, **500**: 156942.[LinkOut]

[12] Cao S H, Yin Y Q, Zheng S G, et al. Peroxymonosulfate activation by Ni-Co nanoparticles encapsulated in nitrogen-doped carbon derived from Co-doped Ni metal-organic frameworks for efficient enrofloxacin degradation[J]. Chemical Engineering Journal, 2025, **503**: 158281.[LinkOut]

[13] He C, Xia W, Zhou C Y, et al. Rational design to manganese and oxygen Co-doped polymeric carbon nitride for efficient nonradical activation of peroxymonosulfate and the mechanism insight[J]. Chemical Engineering Journal, 2022, **430**: 132751.[LinkOut]

[14] Tang Y G, Wei Z S, Wang C, et al. One-step synthesis of magnetic catalysts containing Mn_3O_4-Fe_3O_4 from manganese slag for degradation of enrofloxacin by activation of peroxymonosulfate [J]. Chemical Engineering Journal, 2024, **499**: 156505.[LinkOut]

[15] Li Y, Wang Z, Zou Z Y, et al. Mn-Co/γ-Al_2O_3 coupled with peroxymonosulfate as efficient catalytic system for degradation of norfloxacin[J]. Separation and Purification Technology, 2022, **302**: 122132.[LinkOut]

[16] Wang X R, Wei J, Zhang H, et al. $CoFe_2O_4$@BC as a heterogeneous catalyst to sustainably activate peroxymonosulfate for boosted degradation of enrofloxacin: Properties, efficiency and mechanism[J]. Separation and Purification Technology, 2024, **345**: 127349.[LinkOut]

[17] He B, Song L X, Zhao Z X, et al. $CuFe_2O_4$/CuO magnetic nano-composite activates PMS to remove ciprofloxacin: Ecotoxicity and DFT calculation[J]. Chemical Engineering Journal, 2022, **446**: 137183.[LinkOut]

[18] Liu Z F, He M, Tang L, et al. Dual redox cycles of Mn(Ⅱ)/Mn(III) and Mn(III)/Mn(IV) on porous Mn/N Co-doped biochar surfaces for promoting peroxymonosulfate activation and ciprofloxacin degradation[J]. Journal of Colloid and Interface Science, 2023, **634**: 255-267.[LinkOut]

[19] Zhang H Q, Lee Smith R, Shen F, et al. ZnFe layered double hydroxide wrapped metal-organic framework derived cobalt-nitrogen-doped carbon material as peroxymonosulfate activator for ciprofloxacin degradation[J]. Chemical Engineering Journal, 2024, **496**: 154191.[LinkOut]

[20] Huang Y, Nengzi L C, Zhang X Y, et al. Catalytic degradation of ciprofloxacin by magnetic $CuS/Fe_2O_3/Mn_2O_3$ nanocomposite activated peroxymonosulfate: Influence factors, degradation pathways and reaction mechanism[J]. Chemical Engineering Journal, 2020, **388**: 124274.[LinkOut]

[21] Yang Z Y, Li X, Huang Y Z, et al. Facile synthesis of cobalt-iron layered double hydroxides nanosheets for direct activation of peroxymonosulfate (PMS) during degradation of fluoroquinolones antibiotics[J]. Journal of Cleaner Production, 2021, **310**: 127584.[LinkOut]

[22] Luo J M, Bo S F, Qin Y N, et al. Transforming goat manure into surface-loaded cobalt/biochar as PMS activator for highly efficient ciprofloxacin degradation[J]. Chemical Engineering Journal, 2020, **395**: 125063.[LinkOut]

[23] Le V R, Nguyen T B, Huang C P, et al. Enhanced

peroxymonosulfate activation for rapid degradation of ciprofloxacin using a novel ternary ZIF-67@MnOOH@GO composite[J]. Separation and Purification Technology, 2024, **345**: 127360.[LinkOut]

[24] Liu Z Y, Tian N, Zhang Z Y, et al. Key role of La in La@nZVI/rGO for the efficient and steady PMS activation towards ofloxacin degradation via a non-radical process[J]. Chemical Engineering Journal, 2023, **476**: 146635.[LinkOut]

[25] Zhao Q Y, Wu X, Fu W X, et al. Boosting peroxymonosulfate activation by $MnFe_2O_4/C_3N_5$ S-scheme heterojunction for ofloxacin removal under visible-light irradiation: Performance, mechanism, and degradation pathway[J]. Chemical Engineering Journal, 2025, **522**: 168132.[LinkOut]

[26] He Y X, Qian J, Wang P F, et al. Acceleration of levofloxacin degradation by combination of multiple free radicals via MoS_2 anchored in manganese ferrite doped perovskite activated PMS under visible light[J]. Chemical Engineering Journal, 2022, **431**: 133933.[LinkOut]

[27] Lu P, Bao J C, Ma Y X, et al. pH-dependent oxidation pathway by jarosite residue as a peroxymonosulfate activator for fast levofloxacin removal[J]. Separation and Purification Technology, 2026, **380**: 135169.[LinkOut]

[28] Liu Y, Guo H G, Zhang Y L, et al. Heterogeneous activation of peroxymonosulfate by sillenite $Bi_{25}FeO_{40}$: Singlet oxygen generation and degradation for aquatic levofloxacin[J]. Chemical Engineering Journal, 2018, **343**: 128-137.[LinkOut]

[29] Xu M H, Yang J J, Wang Y, et al. Novel urchin-like Co_5Mn-LDH hierarchical nanoarrays: Formation mechanism and its performance in PMS activation and norfloxacin degradation[J]. Separation and Purification Technology, 2022, **300**: 121822.[LinkOut]

[30] Liu B M, Song W B, Wu H X, et al. Degradation of norfloxacin with peroxymonosulfate activated by nanoconfinement Co_3O_4@CNT nanocomposite[J]. Chemical Engineering Journal, 2020, **398**: 125498.[LinkOut]

[31] Zhang X Y, Yang Z Q, Cui X D, et al. Cobalt/calcium bimetallic oxides based on bio-waste eggshells for the efficient degradation of norfloxacin by peroxymonosulfate activation[J]. Journal of Colloid and Interface Science, 2022, **621**: 1-11.[LinkOut]

[32] Lv H R, Yu W C, Li Y Y, et al. Cyclodextrin-supported sulfide zero-valent iron as PMS activator for simultaneous removing norfloxacin and ARGs in reclaimed water: Activation and controlled release of active components[J]. Chemical Engineering Journal, 2025, **505**: 159656.[LinkOut]

[33] Gu D X, Liu Y, Zhu H Y, et al. Cobalt-doped iron-based metal - organic gels as efficient visible-light-driven peroxymonosulfate activators for the efficient degradation of norfloxacin with simultaneous antibacterial activity[J]. Chemical Engineering Journal, 2024, **486**: 149892.[LinkOut]

[34] Zhang J, Xing Z, Wang X, et al. Degradation of norfloxacin by highly magnetic $CoFe_2O_4$ loaded 0D-CQDS activated PMS: Mechanistic analysis and environmental impacts[J]. Separation and Purification Technology, 2025, **362**: 131989.[LinkOut]

[35] Li X Y, Fu X X, Wu J D, et al. MOF derived bimetallic-nanoreactor as peroxymonosulfate activator for highly efficient levofloxacin degradation: Performance and mechanism[J]. Chemical Engineering Journal, 2025, **515**: 163948.[LinkOut]

[36] Gao C F, Sun Y, Yu S S, et al. Type-II heterojunction Se-g-C_3N_4/ZnO coupled $MnCo_2O_4$ photocatalyst for enhanced levofloxacin hydrochloride degradation activated by peroxymonosulfate[J]. Chemical Engineering Journal, 2024, **500**: 156944.[LinkOut]

[37] Xu A L, Wu D H, Zhang R, et al. Bio-synthesis of Co-doped FeMnOx and its efficient activation of peroxymonosulfate for the degradation of moxifloxacin[J]. Chemical Engineering Journal, 2022, **435**: 134695.[LinkOut]

[38] Liu L L, Mi H S, Zhang M, et al. Efficient moxifloxacin degradation by $CoFe_2O_4$ magnetic nanoparticles activated peroxymonosulfate: Kinetics, pathways and mechanisms[J]. Chemical Engineering Journal, 2021, **407**: 127201.[LinkOut]

[39] Li N, Ye J Y, Dai H X, et al. A critical review on correlating active sites, oxidative species and degradation routes with persulfate-based antibiotics oxidation[J]. Water Research, 2023, **235**: 119926.[LinkOut]

[40] Hu H W, Chen D C, Liang Y H, et al. Understanding the active sites and associated reaction pathways of metal-free carbocatalysts in persulfate activation and pollutant degradation [J]. Environmental Science: Nano, 2024, **11**(4): 1368-1393.[LinkOut]

[41] Molnar C, Freiesleben T, König G, et al. Relating the Partial Dependence Plot and Permutation Feature Importance to the Data Generating Process[C]//Explainable Artificial Intelligence. Cham: Springer Nature Switzerland, 2023: 456-479.[LinkOut]

[42] Scholbeck C A, Casalicchio G, Molnar C, et al. Marginal effects for non-linear prediction functions[J]. Data Mining and Knowledge Discovery, 2024, **38**(5): 2997-3042.[LinkOut]

[43] Tang Z C, Zhou X Q, Du M Y, et al. Crystal plane regulation promotes the oriented conversion of radicals in heterogeneous persulfate catalyzed oxidation process[J]. Small, 2024, **20**(30): 2312173.[LinkOut]

[44] 李斌, 郝帅, 张秋月, 等. 不同酸碱体系中 Pt/SiO_2 催化肼分解产生自由基[J]. 核化学与放射化学, 2024, **46**(1): 54-59.
Li B, Hao S, Zhang Q Y, et al. Free radical generation from hydrazine decomposition catalyzed by Pt/SiO_2 in different acid-base systems[J]. Journal of Nuclear and Radiochemistry, 2024, **46**(1): 54-59.[维普]

[45] Li Z X, Li J W, Zeng A, et al. Machine learning-driven kinetic optimization of hydroxylamine-modified transition metal oxide/peroxymonosulfate system for antibiotic degradation[J]. Water, 2026, **18**(8): 945.[LinkOut]

[46] Bi H, Zhou Z R, Liu Q, et al. Machine learning decoding of electronic thresholds in prediction of oriented generation of solar-induced oxidative reactive species and antibiotics degradation[J]. Journal of Hazardous Materials, 2025, **499**: 140062.[LinkOut]