

机器学习在城市河道水体浊度预测中的应用研究进展

李梦珂, 汤焯

(温州大学 生命与环境科学学院, 浙江 温州 325035)

摘要:【目的】系统梳理主流机器学习技术在河道浊度预测中的应用现状,旨在明确不同技术的适用场景与实践效能,为后续技术优化与应用拓展提供理论支撑。【方法】基于综合方法论视角,详细阐述支持向量回归、集成学习、神经网络及深度学习等模型在河道浊度预测中的应用进展,重点探讨各模型在技术适配性与实践效能方面的差异。【结果】①集成学习是河道浊度预测的优选;支持向量回归在小样本建模中表现出更强的鲁棒性;神经网络可灵活适配多因子耦合预测;时序深度学习模型适用于主干河道的动态特征提取与长时序监测。②将深度学习技术与可控实验、物理机理模型相结合,可实现多源数据扩展,缓解样本稀缺问题。③应用 Shapley 加法解释方法对模型决策逻辑进行归因分析,有助于破解“黑箱”,提升模型实用性与可信度。【结论】现有算法虽兼具计算效率与预测精度优势,但其应用仍面临数据稀缺与模型可解释性不足的双重制约。在实际模型选取中,需综合考量水体污染特征、数据可用性、运维成本等多重因素。未来研究应深化机理模型与机器学习的融合,结合可控实验构建突发环境事件浊度预测体系,以进一步释放机器学习潜力,从而为可持续水资源管理及水生态保护提供更为可靠的智能决策支持。

关键词:浊度预测;机器学习;数据扩展;深度学习;可解释性评价

中图分类号:X832;X522;TP181

文献标识码:A

浊度是水体中悬浮颗粒物对光的散射与吸收作用导致透明度降低的量化指标^[1],主要受地形地貌、气候气象、水文植被、土壤地质等自然因素,以及工农业生产、城市生活污水排放及水土流失等人为因素的影响^[2],是城市河道水环境监测重要指标。随着社会经济发展和城市扩张,土地开发及沿岸居民生活等活动加剧了悬浮物及其他污染物质向河流的输入。污染负荷超过水体自净能力,导致许多城市河道频繁出现浑浊现象,原本清澈的水体呈现灰黄或暗绿色调,甚至伴随悬浮物漂浮与异味,河道水质受到严重影响^[3-5]。高度浑浊的河流通过多重机制破坏水生态系统。悬浮颗粒物增多不仅会显著降低水体透光率,直接抑制浮游植物的光合作用,削弱初级生产力^[6],还会堵塞滤食性生物的纤毛运动与滤食结构^[7],干扰鱼类视觉与侧线感知,降低其摄食效率^[8-9]。此外,颗粒物吸附重金属、农药等污染物后加速沉降,引发底质理化性质改变^[10],破坏螺类、虾蟹等底栖生物的生存环境,严重影响

城市水环境的观感和生态健康。因此,迫切需要对河道浊度进行管控,维护健康的河流生态系统。

当前许多河道的浊度管理仍不完善,传统的水质监测及评估更加关注化学需氧量、氨氮等化学指标,对浊度以常规监测为主,仅做预警监督,治理措施滞后。同时,现有水质监测系统对暴雨、台风、工业事故等突发事件缺乏高频动态监测能力,难以精准捕捉事件对浊度的瞬时影响,应急响应时效性不足,河道浊度风险防控与应急处置能力亟待提升。此外,极端气候及突发事件可能加剧水质波动^[11-12],超过监测范围,使得可用数据稀缺。由此导致传统的过程驱动或统计模型在建模时无法充分考虑水文气候、流域特征、土地利用特征、地质地貌和社会经济活动对浊度变化的影响,难以满足现代水质管理中浊度预测的实际需求。

为应对这些问题,机器学习(Machine Learning, ML)等人工智能技术正在成为研究热点。神经网络、随机森林、深度学习等机器学习

收稿日期/Received: 2025-10-10

修回日期/Revised: 2026-3-19

基金项目: 国家重点研发计划(2023YFE0101700-02);温州市环境局项目(KH2205014)。

第一作者: 李梦珂(1999—),女,硕士研究生,从事水污染控制方面的研究。E-mail: limengke@stu.wzu.edu.cn。

通信作者: 汤焯(1990—),女,助理研究员,博士,从事流域生态治理方面的研究。E-mail: tangye@wzu.edu.cn。

习算法因其出色的非线性信息处理能力和自适应学习特性在水质预测建模中展现出优势^[13],为河道浊度预报与管控提供了新思路。从数据处理角度看,足量的可用数据是训练机器学习模型的基础^[14]。机器学习技术能够有效整合多源异构的河流水质监测数据,通过特征工程和模式识别挖掘变量间的潜在关联规律^[15]。在机理研究方面,机器学习与数值模型的融合为水质建模提供了新范式。将机器学习算法嵌入水动力-水质耦合模型,这种混合建模方法既能保持物理过程的机理描述,又能通过数据驱动的方式修正模型偏差^[16]。同时,基于特征重要性分析和可解释性人工智能(XAI)技术识别关键驱动因子,应用 Shapley 加法解释 (SHapley Additive exPlanations, SHAP) 等解释方法,不仅能够优化传统水质模型参数化方案,更能深入揭示影响河道浊度变化的主导因素及其相互作用机制^[17],显著提升水质预测的精度和可靠性。

目前文献模型中的输出指标一般为综合性的水质指数,对单一物理水质输出指标的关注较少。已有的浊度相关文献中对湖泊浊度的研究明显多于河流浊度。河流形态相较于湖泊更加弯曲、多变,且受到的污染源影响更加复杂^[18]。针对河道的浊度预测模型在处理数据稀缺性、模型泛化性和可解释性等方面仍面临巨大挑战。

本文旨在通过综述主流机器学习方法在河道浊度预测中的研究进展,探讨其在实际应用中的优势与局限性,提出多源扩展可用数据的解决方法,对河道浊度预测提出未来研究重点,强调了利用机器学习为可持续水资源管理和保护重要水生态系统的强化解决方案的潜力。

1 机器学习主流方法在河流浊度预测

中的应用

机器学习模型具体分类情况如图 1 所示^[19]。机器学习在河流水质预测中的应用主要涉及监督学习^[20],但在一定条件下也可通过无监督学习实现^[21]。浊度预测作为一种回归任务,目前主流的方法是应用监督学习,将 pH、溶解氧(Dissolved Oxygen, DO)等水质参数与温度、降雨以及土地利用等环境因素的历史数据用于模型训练。经过训练的模型可以根据新的环境数据预测河流浊度。基于功能相似性归类情况,下文对主流的支持向量回归、集成学习、神经网络和深度学习技术在河流浊度预测中的应用现状进行介绍。

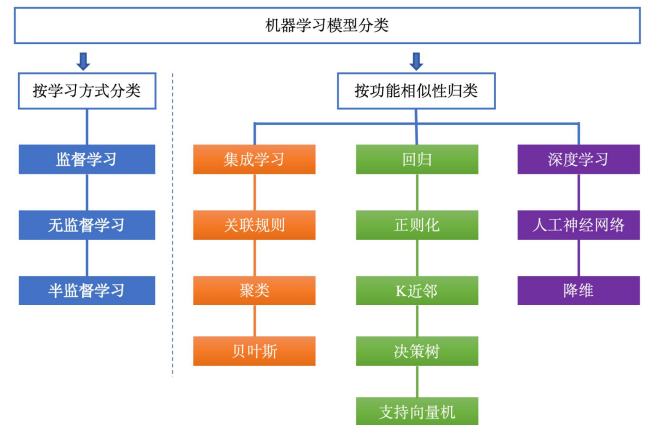


图 1 机器学习模型分类^[19]

Fig.1 Classification of machine learning models^[19]

1.1 支持向量回归

支持向量机(Support Vector Machine, SVM)在机器学习领域中,常被用于二分类、多类分类及回归预测研究。支持向量回归(Support Vector Regression, SVR)是 SVM 在回归任务上的扩展,通过核函数和损失函数将数据映射到高维空间,在预定义公差边距(ϵ)内约束误差,提高了对离群值的稳健性和跨数据集的推广性^[22],在水生态系统的应用中表现出优越性能。马来西亚冷岳河流域因人口增长污染加剧,MAMAT N 等^[23]将 SVR 与 K 折交叉验证结合应用于该河流水质指数预测,通过评价测试核函数、K 折数和参数组合多维度优化模型,得到的最佳模型相关系数 $R = 0.9998$ 。SVR 模型不仅能够处理原位水质数据,也能结合遥感数据进行浊度预测,进一步提高数据利用率。WATTELEZ G 等^[24]基于大量原位浊度值以及适度分辨率成像光谱辐射计(Moderate Resolution Imaging Spectroradiometer, MODIS)的一致卫星反射值,建立了 SVR 浊度预测模型,显著提高了模型从 MODIS 图像中检索原位浊度范围的能力,证明了 SVR 与多源数据结合的技术可用性,但对大数据集和参数选择敏感,对内存和训练时间要求较高,易受异常值影响。

1.2 集成学习方法

河流浊度数据获取难度高,使得可用数据量有限且时空分布一致性不足,机器学习模型训练困难。随机森林(Random Forest, RF)和极端梯度提升(Extreme Gradient Boosting, XGBoost)等集成学习模型可以将多个弱机器学习模型相结合,通过集成的方法获得比单个学习模型更好的泛化能力和鲁棒性,在河道水质预测领域有着广泛应用。SANTOS V O 等人^[25]使用来自美国密西西比河上 12 个监测站的水质数据,结合卫星遥感数据集对浊度进行建模,对比了 XGBoost、RF、SVR、k 近

邻 (K-Nearest Neighbors, KNN) 等算法, 发现 XGBoost 对浊度的预测精度最高, 决定系数 R^2 为 0.757。该研究结果相较其他同类文献更具优势, 可为遥感数据驱动的浊度建模提供替代方案。ZOU T L 等^[26]在此基础上, 基于 MODIS 数据构建 RF 和 XGBoost 模型对洞里萨湖的浊度进行反演, 并通过趋势分析和相关性分析, 揭示浑浊度的时空特征和驱动机制, 有效增强了模型的泛化性和可解释性。河流流域特征与水质之间存在复杂非线性关系, 对

预测模型提出更高的技术要求。ZARE M S 等^[27]引入了一种新的集成方法用于解决不同深度的水质差异问题, 技术路线如图 2 所示。通过收集阿曼 Wadi Dayqah 大坝水库的表层水数据 (盐度、密度、温度), 利用集成学习模型结合聚类与贝叶斯优化加权的策略, 预测 1~35m 深度的 DO、叶绿素 a 和浊度。该方法较传统模型误差降低 4%~16%, 对数据预处理要求低, 适用于动态捕捉浊度变化, 展现了集成算法与优化特征工程相结合的潜力。

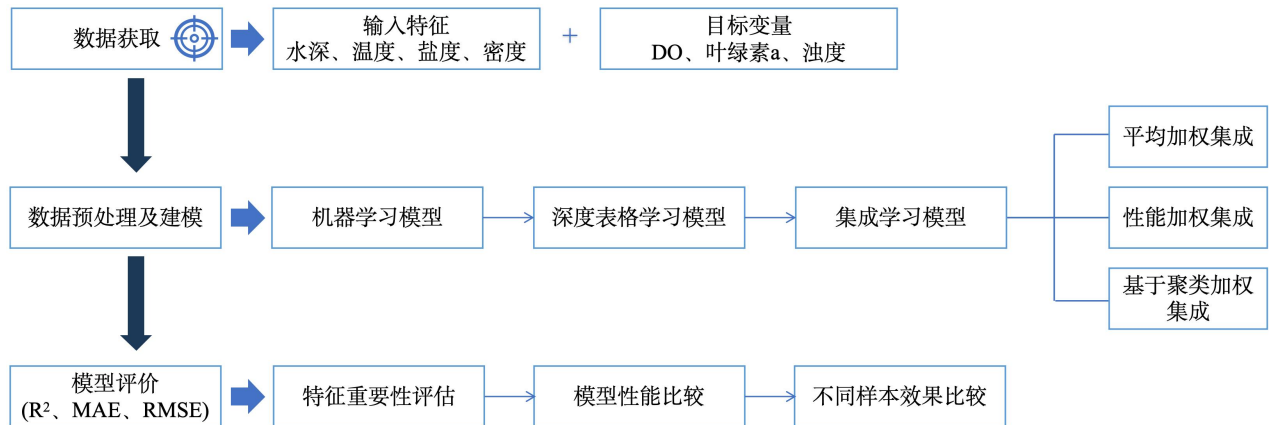


图 2 基于集成学习的浊度预测模型技术路线

Fig.2 The technical route of turbidity prediction model based on ensemble learning

1.3 神经网络

神经网络 (Artificial Neural Network, ANN) 是一类仿生物神经系统设计的机器学习模型, 具备强非线性逼近能力, 浊度预测精度较高, 还可有效缓解 SVR 等传统回归算法的过拟合问题^[28]。NAJWA MOHD RIZAL N 等^[28]应用了 SVM、ANN 和回归模型来预测马来西亚冷岳河的总悬浮固体和总固体, 结果发现 SVM 和高斯过程回归模型表现出过拟合的特征 ($R^2 = 0.490$), ANN 模型表现最佳 ($R^2 = 0.999$)。通过多样化的数据优化程序、训练算法和选择合适的输入技术可以进一步提高 ANN 模型的有效性。MULIA I E 等^[29]将 ANN 与遗传算法 (Genetic Algorithm, GA) 相结合, 开发了精度更好的 ANN-GA 联用模型, 可用于预测未来 14 天的浊度。也有研究人员引入协同变量提供更多样本信息以实现 ANN 模型改良。IGLESIAS C 等^[30]在研究中引入了由两个原始输入参数的乘积生成的协同变量 (溶解氧*温度) 后模型精度得到提升 ($R^2 = 0.8$)。

反向传播神经网络 (Back Propagation Neural Network, BPNN)、径向基神经网络 (Radial Basis Function Neural Network, RBFNN) 是更先进的 ANN 模型, 已被逐渐应用于河道水质预测中。TOMPERI J 等^[31]在 MULIA I E 的基础上构建了以河流水位和水温为输入变量的神经网络线性回归

(Multiple Linear Regression, MLR) 模型, 可以准确估计训练期间的水浊度 ($R = 0.77$)。针对传统 BPNN 收敛速度慢、易陷入局部最优等问题, 刘洁等^[32]建立了基于遗传-神经网络 (Improved Genetic Algorithm-Back Propagation Neural Network, IGA-BPNN) 的河流水质预测模型 ($R^2 = 0.876$), 并对美国波托马克河流的浊度进行预测, 新模型相较于传统 BPNN 预测性能更好。GUO Q Z 等^[33]利用高分二号遥感影像和实测悬浮物浓度数据, 结合偏最小二乘法和粒子群优化算法构建了 BPNN 优化模型, 实现了海河流域大范围水质参数反演。综合来看, ANN 能对流域复杂关系进行建模, 适应大量数据特征, 但模型计算成本高, 网络结构选择困难, 同样面临数据需求、计算资源消耗和模型解释性等挑战。

1.4 深度学习模型

深度学习 (Deep Learning, DL) 因其具有更强大的特征学习和拟合能力, 已成为预测和评估河流水质、解决复杂非线性问题的有力工具。常用的深度学习算法包括长短期记忆神经网络 (Long Short-Term Memory, LSTM)、卷积神经网络 (Convolutional Neural Network, CNN)、循环神经网络 (Recurrent Neural Network, RNN) 等。LSTM 模型能够从历史数据中学习规律并对未来的水质状况做出预测, 善于处理具有时序特征的水质数据,

具备快速收敛、稳定性和高精度的优势^[34]。CNN是深度学习在计算机视觉领域的基石,通过卷积层、池化层和全连接层的层级结构,自动提取数据的局部特征与全局模式,在浊度分类工作中发挥出色。LOPEZ-BETANCUR D等^[35]使用液体样品的图像数据集训练 CNN,通过不同颜色照明条件下的数据集拟合 MLR 模型,估算液体样品的总悬浮固体及浊度值,展示了 CNN 通过图像数据进行浊度反演的潜力,技术路线如图 3 所示。门控循环单元(Gated Recurrent Unit, GRU)是 RNN 的优化变体,能捕捉数据中的长期依赖关系。MEI P 等^[36]提出了一种 CNN-GRU-Attention (CGA) 模型对轻工农业污染原水的浊度进行预测,模型中的 CNN 层可以

提取水质数据的短期特征,GRU 层可以提取长期特征,注意力层可以根据神经元之间的相关性调整不同神经元的权重,增加它们的相互联系。ALIZAMIR M 等^[37]在预测河水中的每日荧光溶解有机物 (FDOM) 浓度时发现,与独立机器学习架构相比,混合 LSTM-CNN 模型表现出更高的预测精度 ($R^2 = 0.979$)。深度学习在处理高维数据时,其多层结构可以有效地提取水质数据中的关键特征,具有优异的稳定性与可靠性,在河流浊度预测的迁移应用中展现出良好前景。但该类模型的训练依赖大规模高质量标记数据,若数据集规模有限或覆盖范围较窄,易导致模型泛化能力不足。

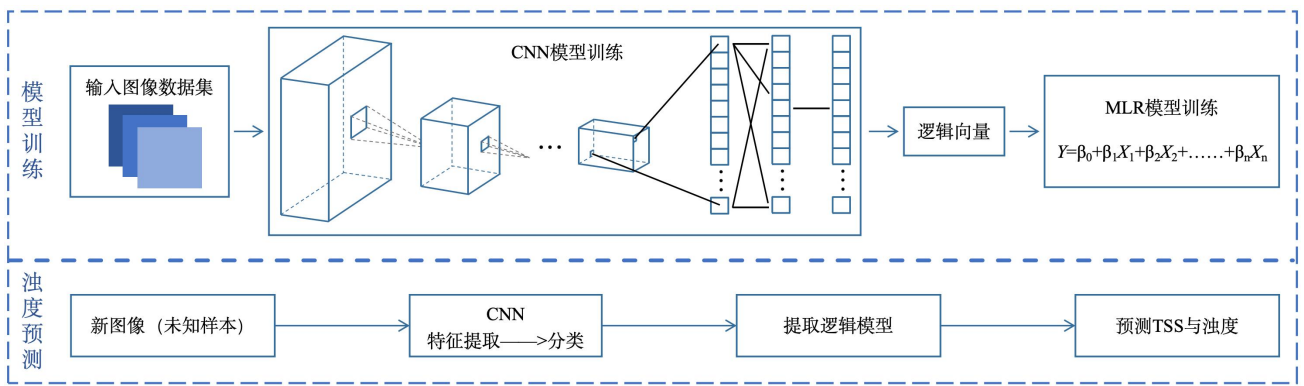


图 3 基于 CNN-MLR 的浊度预测模型技术路线

Fig.3 The technical route of turbidity prediction model based on CNN-MLR

1.5 主流模型对比总结

表 1 总结了机器学习模型在河流浊度预测中的典型应用案例,通过对比各模型的 R^2 、均方根误差 (Root Mean Square Error, RMSE) 及平均绝对误差 (Mean Absolute Error, MAE) 等指标,综合评估模型精度与实用性。XGBoost、RF 在多数河道场景中均能实现高精度与运行稳定性的平衡,且对数据规模和硬件资源要求适中,是河道浊度预测的优选方案。SVR 能在数据稀缺条件下保持可靠泛化能力,适用于小样本河流建模。LSTM 时序模型能够利用历史数据对缺失信息进行技术填充,扩大模型适用范围,尤其适用于主干河道时序监测。

ANN、BPNN 等性价比突出,可在无需高端硬件的情况下适配多因子耦合预测,且模型精度较高。

城市河道水质波动大,易受雨污混流、面源污染等外部干扰,污染物浓度偏高,尤其是特别事件情况下的浊度等参数数据不足。表 2 进一步梳理了 SVR、集成学习方法、ANN 和 DL 四种模型的优势、局限性及适用场景。各类模型在数据需求、算力要求和预测精度方面各具特点,因此,机器学习模型在城市河道水体浊度预测的实际应用中,需具体分析水体污染特征、数据可用性结合硬件设备购置、软件工具授权和云服务费用等计算成本进行综合考量,通过多算法对比以确定最优建模方案。

表 1 机器学习河流浊度预测模型的典型应用案例

Tab.1 Summary of mainstream machine learning models for river turbidity prediction

研究区域	输入指标类型	模型类型	模型应用重点	最佳模型性能评价	文献	时间
法国新喀里多尼亚泻湖	2 个水质参数和遥感数据	SVR	结合 MODIS 图像预测原位浊度	$R^2 = 0.553$ RMSE = 0.22 NTU MAE = 0.136 NTU	[24]	2017
美国密西西比河	遥感数据	XGBoost、RF、GMDH、SVR、KNN、LASSO	结合 Sentinel-2 遥感光谱数据预测建模	$R^2 = 0.757$ RMSE = 27.53 FNU MAE = 12.89 FNU	[25]	2025

柬埔寨洞里萨湖	遥感数据	RF 、XGBoost	结合 MODIS 图像预测建模; 基于 RF 填充缺失数据	$R^2 = 0.78$ RMSE = 30.68 NTU	[26]	2024
美国东南部大西洋地区流域	气候数据和流域特征数据	RF 、 BRT	引入土壤、地形等因素作为流域特征进行预测	RMSE = 2.22 NTU MAE = 1.51 NTU	[27]	2022
西班牙纳隆河流域	5 个水质参数和 3 个协同变量	ANN	引入协同变量(两个水质变量的乘积)进行预测	$R^2 = 0.80$ $R = 0.89$	[30]	2013
芬兰南部流域	2 个水质参数	MLR 、PSLR、 NN	仅用 2 个变量得到高精度预测模型	$R = 0.80$ RMSE = 7.74 NTU	[31]	2025
美国波托马克河	5 个水质参数	BPNN 、 IGA-BPNN	模型预测区间可进行异常识别和预警	$R^2 = 0.876$ RMSE = 0.026 NTU MAE = 0.102 NTU	[32]	2020
青草沙水库	2 个水质参数和 风场数据	LSTM	引入预测风场数据改善实际浊度预测	RMSE < 0.05 NTU	[34]	2020
受轻工农业污染原水	8 个水质参数	CNN-LSTM-Attention 、 LSTM 、 GRU 、 CNN-GRU-Attention	建立改进的 TOPSIS 模型综合评估水源质量	$R^2 = 0.979$ RMSE = 0.029 NTU	[36]	2022

注:字体加粗模型为对比模型中性能最佳模型。表中, GMDH 为多项式前馈神经网络(Group Method of Data Handling), LASSO 为最小绝对收缩与选择算子(Least Absolute Shrinkage and Selection Operator), BRT 为提升回归树(Boosted Regression Trees), PSLR 为偏最小二乘回归(Partial Least Squares Regression), NN 为神经网络(Neural Networks)。

表 2 机器学习河流浊度预测模型对比

Tab.2 Comparison of mainstream machine learning models for river turbidity prediction

模型类型	优势	局限性	适用场景
支持向量回归	小样本泛化能力强; 能有效处理高维、非线性数据	核函数选择依赖经验; 对数据预处理敏感	小样本、数据维度适中的河道浊度单项指标预测
集成学习方法	多模型融合精度高, 鲁棒性强; 适配多因子交互与多源数据	训练耗时; 对硬件要求高; 可解释性差	大样本、多因子耦合的河道水质综合预测; 复杂河网水质空间分布估算
人工神经网络	可拟合复杂非线性关系; 结构灵活, 适配多变量预测	需大量数据; 可解释性差; 易陷入局部最优	中大规模样本、水质因子交互复杂的河道综合水质预测
深度学习模型	可自动提取河道高维特征; 大规模数据下性能优异; 结合迁移学习可以跨区域预测	数据、算力要求极高; 可解释性差	大规模、高频河道时序浊度预测; 结合卫星遥感的流域级河道水质空间分布及长期趋势预测

2 机器学习多源扩展可用数据

水污染数据识别与收集是水质认知及机器学习建模的基础^[38]。受采样成本高、河道环境复杂制约, 目前河道水质数据存在两大瓶颈: 时空一致性差, 数据波动大、缺失多、分布散; 对突发污染事件的响应性监测覆盖不足。机器学习与已有数据获取技术的深度融合, 正推动水质数据从“可用”向“好用”跃升。下文将围绕多源数据扩展方法, 重点探讨如何结合机器学习特性, 破解水质数据的双重瓶颈。

2.1 结合可控实验扩展数据

当前城市河道的固定监测站点或定期采样等

数据通常离散, 特别是缺乏突发污染事件和洪涝、台风等极端事件的数据。对此, 可控实验已成为弥补实际监测数据不足的重要手段。利用模拟人工河道装置, 研究者可主动生成洪水、污染泄漏等极端水文事件情况下的高质量标注数据。机器学习模型通过学习标注数据中的知识, 掌握不同工况下的水质响应规律, 如降雨径流增加、排污量新增、闸坝调控等, 量化不同因素对水质变化的影响贡献, 帮助理解污染物迁移转化机理。YANG J 等^[39]搭建砂箱实验装置模拟狄利克雷边界条件下非饱和流中污染物的输运, 并以河道为例建立了污染物迁移对监测水质影响的模型, 计算值与实验数据吻合较好。可控实验设计应以补充缺口数据为核心, 遵循

“参数对标实际结合变量可控”原则, 优先模拟实际河道可用数据有限的场景, 提升实验数据的补充价值。室内外模拟实验借助强化学习可以动态调整参数^[40], 通过测试降解速率、扩散系数等关键参数的敏感性, 指导实测数据的关键采集方向, 落脚风险管控。JAEGER A 等^[41]利用河流模拟引水槽研究了沉积物中有机微污染物的转化规律。尽管模拟水槽微生物群落相较实际河道存在部分偏差, 但测试不同的实验条件时, 水槽实验与埃尔佩河实地实验相比更加有效。JAVERNICK L 等^[42]利用辫状河水槽实验中高精度的时空数据集, 校准并评估二维模型 Delft3D 预测推移质输移的能力, 结果表明该模型能达到较高的淹没精度 (77%~84%), 验证了实验室数据的迁移有效性。可控实验补充数据应以实测数据为基准修正与实际环境的偏差, 并通过相似比换算等方法消除尺度效应, 实现模拟与实际数据的融合。

2.2 结合机理模型扩展数据

基于过程引导的数值模型也被广泛应用, 这类模型利用水质迁移转化的“内在机理”, 如污染物扩散、降解、生物循环, 输入污染源、水文参数等已知的边界条件, 模拟生成水质数据。流体动力学预测的研究进步给数据扩展提出了新的研究思路^[43-44]。WEISSCHER S A H 等^[45]用数值模拟的水动力模型对河流和河口的规模实验进行补充, 得到质量更好的连续时空数据。FEI K 等^[46]建立水文-水动力耦合风暴潮数值模型, 结合 XGBoost 成功计算每个驱动因子对水位的相对贡献。许多研究已经通过实验模型耦合流体力学进行水质预测, 证明流体动力学模型和机器学习的组合可以提高分析可靠性和计算效率^[47]。KIM J Y 等^[48]聚焦叶绿素 a 的三维扩展, 提出“数据增强-数值模拟-机器学习”三阶段整合框架, 解决水质高光谱图像数据 (Hyperspectral Data, HSD) 垂直分布缺失问题。研究者利用神经网络模型对有限的野外水质数据进行扩充生成高频水质数据, 然后通过环境流体动力学模型生成三维数据, 最后借助机器学习模型将 HSD 从二维表层扩展为三维数据。结果显示, 高斯过程回归模型表现最优, R^2 达 0.99。该研究验证了集成机器学习模型与数值模型进行数据扩展的有效性, 为将来的水质变化分析提供了参考。在实际应用中, 这种联用方法需要重点关注模型的物理合理性、参数化方法、数据输入质量以及与实际系统的匹配性, 平衡物理机理的完备性与计算可行性。

表 3 机器学习数据扩展方法对比

Tab.3 Comparison of machine learning methods for data imputation

扩展方法	适用场景	适用条件	局限性
------	------	------	-----

2.3 深度学习填充数据

将数据从监测良好的地点迁移到未设监测站的地点进行预测是长期以来的挑战, 深度学习在对未监测流域的预测方面显示出了希望。DL 模型可以将监测地点的水质数据与气象数据、遥感数据或流域特征等空间特征结合在一起构建模型, 然后推广到未设监测站的河流, 实现数据迁移和填充。ZHI W 等^[49]整合了 580 条美国河流的水文气象、生物地球化学数据及流域属性, 以其中 480 条河流数据为基础构建了大规模 LSTM 模型, 并使用剩下的 100 条河流数据进行性能验证, 结果显示模型在东北部流域的纳什效率系数 (Nash-Sutcliffe efficiency coefficient, NSE) 达到 0.65 以上 (NSE > 0.5 时模型可信), 展示了利用深度学习模型进行数据填充的前景。此外, 数据观测训练与基于过程模型的训练, 对于深度学习模型的迁移应用具有良好效果。CHEN Z 等^[50]通过 LSTM 与迁移学习 (TrAdaBoost) 相结合的新算法, 捕捉时间序列长期依赖性, 利用源域完整数据集的知识填补目标域大规模连续缺失值。并以钱塘江流域 DO 浓度数据为案例, 验证了 TrAdaBoost-LSTM 较传统方法能够提升 15%~25% 的填补精度, 为水质数据驱动分析提供科学支撑。ZHENG Y 等^[51]提出一种基于表示学习的深度学习模型, 通过源河流域尺度时空预训练并结合站点气象数据引导微调, 解决数据稀缺及异质性条件下的跨流域水质预测问题。模型平均 NSE 可达 0.80 (NSE > 0.75 时模型性能优秀), 证明基于表示学习开发的模型在跨流域水质预测方面具有重大潜力, 能够通过融合算法“整合碎片化信息”, 为水质预测数据扩展提供高效解决方案。

2.4 数据扩展方法对比总结

围绕机器学习数据扩展方法的适用场景、适用条件和局限性进行总结分析, 由表 3 可知, 可控实验需完成模拟场景靶向筛选与跨场景数据迁移验证, 机理模型构建依赖足量水文基础数据支撑, 而深度学习数据填充技术则对硬件计算资源提出较高要求, 三类扩展方法均需通过实测数据校验以保障数据可靠性。未来河道浊度预测数据获取需强化跨学科融合, 联合水文水利、环境遥感、地理信息科学等领域技术优势, 构建“遥感-水文-实测-模拟扩展”多维立体数据体系。同时推动地方环保、水利、气象等部门的数据互通共享, 统一数据标准格式, 建立规范化、可复用的河道水质数据库, 为模型训练提供全面、高质量的数据支撑。

ML 结合可控实验	补充暴雨或突发污染等极端工况的标注数据	具备可控实验环境与参数调控能力; 实验设计贴合实际水体特性	实验成本高、周期长; 模拟数据易存实验室偏差
ML 结合机理模型	河道水动力-水质耦合预测; 获得物理意义明确的扩展数据	具备地形、水文等基础数据; ML 与机理模型输出格式适配	基础数据精度要求高; 复杂流域参数难定; 大尺度计算成本高
深度学习	补全时空监测断点数据; 相似流域数据迁移复用	数据具备时空关联性; 具备部分高质量实测数据或相似场景迁移数据源	对数据关联性要求高; 可解释性弱; 依赖硬件算力

3 应用 SHAP 增强模型可解释性

浊度作为水环境的核心指标, 其预测结果直接服务于防洪减灾、水质治理、生态修复等决策场景。模型可解释性评价是连接“模型精度”与“实际应用价值”的关键桥梁。传统黑箱模型虽能提供高精度预测, 但它们缺乏对模型结构的明确呈现和对输入变量影响的解释, 导致决策者对模型结果的信任度不足。针对以上问题, 人们提出了几种解释技术, 如局部代理模型、反事实解释、个体条件期望图和 Shapley 加法解释, 通过揭示模型的决策逻辑、关键驱动因素及潜在局限性, 可显著提升浊度预测的实用性和可信度。上述技术中, SHAP 因为不受模型架构限制、解释范围广且可视化直观, 而成为机器学习模型可解释性评价的重要方法。SHAP 的可解释性评价一般分为全局可解释性评价和局部可解释性评价。

3.1 SHAP 全局可解释性评价

SHAP 全局可解释性评价关注模型对浊度变

化的整体决策逻辑, 旨在验证模型是否符合水文规律等专业知识, 并识别影响浊度变化的关键因素。全局可解释性评价主要包括以下两个维度:

1) 特征重要性分析, 识别主导因素。SHAP 能够量化每个特征对预测的边际贡献^[52], 揭示哪些因素(流速、降雨)长期主导浊度变化, 适合非线性模型。HAN J W 等^[53]在预测汉江下游流域的叶绿素 a 时, 使用了 SHAP 对开发的 XGBoost 模型 ($R^2 = 0.949$) 进行特征重要性分析和解释, 指出 pH 值、浊度对叶绿素 a 浓度的积极贡献 (SHAP 值分别为 3.593 和 1.720), 并验证了 XAI 在河流水质预测方面的效用。同时, SHAP 法可辅助研究者进行全局特征重要性排序, 确定模型重点输入参数。ALIZAMIR M 等^[37]根据 SHAP 结果, 确定溶解氧、pH 是各模型中影响河水 FDOM 预测效果最重要的参数 (SHAP 值分别为 0.23 和 0.19), 构建了基于 LSTM-CNN 的稳健且可解释的深度学习模型 (如图 4 所示)。

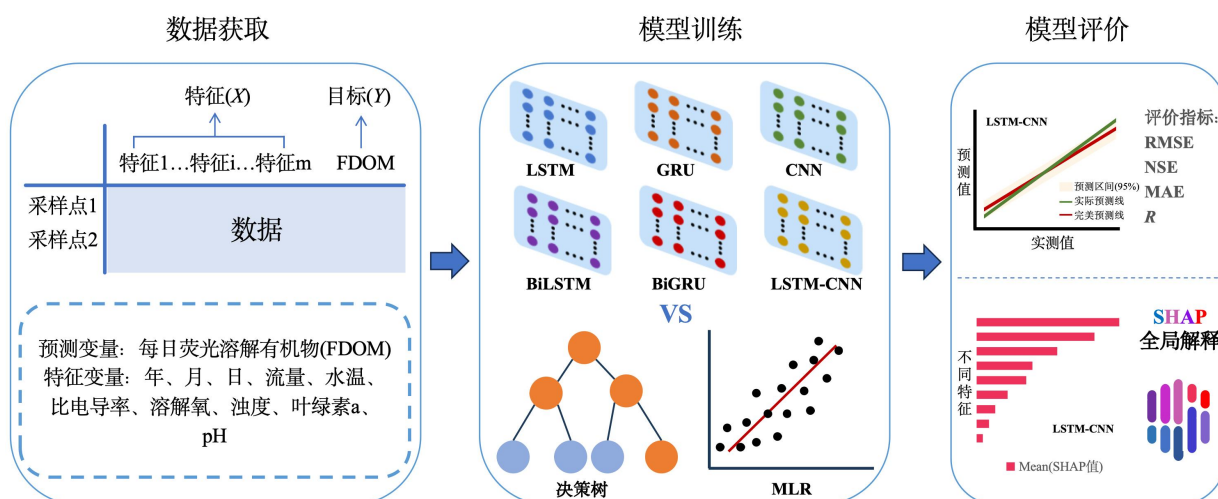


图 4 基于 LSTM-CNN 模型结合 SHAP 可解释性评价预测河水 FDOM 技术路线

Fig.4 The technical route of turbidity prediction model based on LSTM-CNN incorporating SHAP analysis

2) 特征贡献模式分析, 验证因果关系。正 SHAP 值表示该特征对预测结果起正向促进作用, 而负值则代表其具有负向影响。数值的绝对值越大, 表明该特征对预测的影响越显著^[20]。MAKUMBURA R K 等^[17]将机器学习模型与 XAI

技术相结合加强水质评估和预测, 结果显示化学需氧量和生化需氧量是决定水质的最有影响力的因素 (SHAP 值分别为 6.92 和 4.45)。部分依赖图能够结合 SHAP 值与特征取值, 对某一特征在不同取值范围内对模型预测的平均影响进行可视化, 验证模

型是否符合领域知识。ALNAHIT A O 等^[54]应用 SHAP 解释树模型对溪流浊度的预测情况,部分依赖图中的阈值显示,当土壤水力电导率增加时,径流减少使溪流的浊度水平降低,与领域认知一致,评价结果指出研究需要确定关键流域特征及其关键阈值才能有效应用模型。

3.2 SHAP 局部可解释性评价

SHAP 局部可解释性评价能够用于追溯单次预测的具体驱动因素。旨在回答“某时刻或某地点的高浊度是否由当日暴雨或排污口排放引起”,为应急响应提供精准依据。SHAP 值为每个预测实例生成局部解释,通过量化每个特征对输出结果的影响方向与程度,帮助理解模型的决策机制^[24]。局部可解释性评价主要包括以下两个内容:

1) 单样本 SHAP 值分解, 归因具体特征。城市河道环境因素和人为影响之间存在复杂的相互作用,对于洪水、面源污染等突发事件的浊度预测管理理解亟待加强。GAO Z Y 等^[55]提出了一种可解释的注意力门控循环单元模型(平均 $R^2 = 0.907$),该模型整合了半干旱大黑河流域的水质、气象和水文数据,SHAP 分析结果表明气象因子在高污染条件下影响有限,发生极端降水事件会暂时提高污染物浓度。局部可解释方法可用于解释特定高浊度突发事件的预测过程,对超标和极端情况的分析进一步突出了最有效的监管干预措施及特定时间。

2) 局部一致性评估, 确保解释可靠。通过对比 SHAP 解释结果与现场监测数据、遥感数据等,评估局部解释的可靠性,识别虚假关联,增强模型可

信度和实用性。MERABET K 等^[56]通过 SHAP 局部解释力图,对比了美国伊利诺伊河不同单样本中各个水质输入指标对浊度的影响方向变化和特征重要性差异,证明了 XGBoost 模型的鲁棒性较强 ($R > 0.9, NSE > 0.8$)。SOLEYMANI HASANI S 等^[57]使用将站点特定模型与 SHAP 集成的空间建模方法,识别和量化影响浊度的环境因素,发现进水和水位是湖口浊度的主要驱动因素,而风速和气温影响湖中部浊度。这项研究促进了对湖泊水质动态的理解,对相对动态的河流浊度的预测具有一定的指示意义。但在实际应用中,SHAP 也存在一定挑战:深度学习等复杂模型的 SHAP 评价可能会占用大量计算资源^[58];数据中的噪声、缺失值或异常值容易导致 SHAP 计算不准确,需要提前进行严格预处理,避免数据偏差问题带来的影响^[59]。

3.3 SHAP 可解释性评价方法对比总结

表 4 对全局可解释性评价与局部可解释性评价的适用场景、核心条件、局限性及可视化应用进行深入讨论。全局可解释性评价适用于大样本场景下的特征筛选,可明确降雨、流速、污染源等主导因素的优先级,但对高维数据与异常值较为敏感,且受多特征耦合解读难度高、特征独立性假设的制约。局部可解释性评价是浊度异常峰值溯源与环保执法单例审计的关键工具,但其解释结果不可迁移,存在一致性缺乏统一量化阈值、水文异质性易引发误判等问题。不同评价方向需结合实操工具与具体场景需求灵活选用,方能实现模型解释科学性与实用性的有机平衡。

表 4 SHAP 可解释性评价方法对比
Tab.4 Comparison of SHAP explainability evaluation methods

SHAP 评价方法	适用场景	核心条件	局限性	可视化应用
全局可解释性评价	特征重要性分析 模型特征筛选; 主导因子识别; 决策逻辑审查	样本量不少于 100 组,覆盖全特征梯度; Pearson 相关系数小于 0.8	受异常值影响大; 特征超过 20 个时计算耗时翻倍	SHAP 摘要图、特征重要性条形图
局部可解释性评价	特征贡献模式分析 特征交互效应解析; 预测结果与水文机理验证	样本覆盖特征关键阈值	多特征耦合时解读难度大; 交互量化受特征独立性假设影响	SHAP 依赖图、交互作用热力图
局部可解释性评价	单样本 SHAP 值分解 浊度异常峰值溯源; 个体预测信任度验证	单样本特征数据无缺失; 基线值采用全局浊度平均预测值; 需标注样本对应的水文场景	无法迁移至同类样本; 小样本场景基线偏差大	SHAP 瀑布图、力导向图
局部可解释性评价	局部一致性评估 模型微调效果校验; 同一流域不同河段预测逻辑比对	聚类构建相似样本集群, 集群样本量 ≥ 30 组; 以 SHAP 值分布标准差作为一致性度量	一致性无统一量化阈值; 河段水文异质性大易导致误判; 集群数量 > 5 组时计算复杂度剧增	SHAP 值分布箱线图、样本相似度热力图

4 结语与展望

本文综述了机器学习算法在城市河道水体浊度预测模型中的应用进展,详细论述了机器学习结合可控实验、机理模型多源扩展可用数据的方法,介绍了 SHAP 可解释性评价以提高模型透明度,破解“黑盒”问题。鉴于河流水环境的复杂性及机器学习方法将持续面临数据稀缺和过程理解不完整的挑战,当前研究正在向多源技术联用方向发展,后期还需在以下方面开展进一步研究:

1) 结合可控实验构建突发环境事件浊度预测体系。扩大深度学习与可控实验数值模拟联用,利用可控实验模拟获得精准的“因果”标注数据,弥补特定情景观测数据的不足,收集隐含新知识的补充数据集。应用深度学习挖掘遥感、传感器网络等多源数据的内在规律,预测极端环境下的浊度。这种模式能够对未来进行预测,并将假设的场景投射到未来的需求中,而不仅仅是事后分析。但需要对新的机器学习模型进行严格的泛化性检验,包括在无监测数据河流和超前预报中进行时空外推检验,以及与其他已建立的方法或模型进行基准比较测试。这些测试能够证明新模型阐释复杂过程和投影到新场景的能力,支持决策者预测和应对水质挑战。

2) 增强机理模型与机器学习联用。城市河流受

近岸人类活动影响极大,河道沉积物和浊度的变化机制复杂,未来浊度预测的核心方向是融合物理机制与机器学习,深入可解释性挖掘驱动机理。发展物理信息神经网络将描述物理过程的偏微分方程作为损失函数的一部分,直接嵌入神经网络的训练过程中,得到的预测模型既可正向预测浊度,又能从浊度数据中反演扩散系数或污染源强度。开发水文机理模型与机器学习耦合的新模型。实现路径包括将传统模型模拟的流场、扩散场等物理量作为机器学习模型输入特征,利用残差网络修正机理模型输出的系统误差等,增强模型输入的物理信息量,提升预测精度。

3) 发展模型轻量化及应用场景自适应。模型应在确保预测精度满足水质监测、应急预警需求的同时,避免资源冗余或性能损耗,平衡模型复杂性与计算资源限制。针对日常浊度监测场景开发轻量级 GRU 模型,通过剪枝移除冗余连接。突发高浊度事件场景则启用复杂模型并结合早停机制避免无效迭代。此外,通过针对性分配硬件与流程,建立动态调整机制。设定例如 GPU 显存占用 80%、推理时间 500 ms 等阈值,结合场景触发模型切换,如终端传感器用轻量化模型,后台分析用复杂模型,最终实现河流浊度预测中模型性能与资源消耗的动态适配。

参 考 文 献

- [1] ZHANG W H, LI X B, HUANG Y Z, et al. Turbidity-graded adaptive underwater image enhancement[J]. *Optics & Laser Technology*, 2026, 194: 114496.
- [2] MAGER S, HORTON S. Particle size and geochemistry of suspended material and its influence on nephelometric turbidity[J]. *River Research and Applications*, 2025, 41(6): 1300-1314.
- [3] YU J F, ZHAO L, LIANG X-Z, et al. The mediatory role of water quality on the association between extreme precipitation events and infectious diarrhea in the Yangtze River Basin, China[J]. *Fundamental Research*, 2024, 4(3): 495-504.
- [4] JIANG Y J, TIAN S H, LI H Y, et al. Harnessing microbial electrosynthesis for a sustainable future[J]. *The Innovation Materials*, 2023, 1(1): 100008.
- [5] 董素涵, 刘萌硕, 蔡闻琪, 等. 磺胺甲噁唑淡水水生生物水质基准与生态风险评估[J]. *环境科学学报*, 2023, 43(5): 496-504. [DONG S H, LIU M S, CAI W Q, et al. Water quality criteria and ecological risk assessment of sulfamethoxazole freshwater aquatic organisms[J]. *Acta Scientiae Circumstantiae*, 2023, 43(5): 496-504.]
- [6] CUSSIOLI M C, SEEGER D, PRATT D R, et al. Spectral differences in the underwater light regime caused by sediment types in New Zealand estuaries: implications for seagrass photosynthesis[J]. *Geo-Marine Letters*, 2020, 40(2): 217-225.
- [7] ALTINDA A, BERDI D. Rising temperatures, changing waters: the influence of abiotic factors on freshwater zooplankton community[J]. *Desalination and Water Treatment*, 2025, 321: 100991.
- [8] LERIS I, KOEPCHEN-THOMÄ L, SMETI E, et al. Turbidity and predation risk: behavioural responses of a freshwater minnow[J]. *Animal Behaviour*, 2022, 186: 1-9.
- [9] RODRIGUES J N, ORTEGA J C G, PETSCH D K, et al. A meta-analytical review of turbidity effects on fish mobility[J]. *Reviews in Fish Biology and Fisheries*, 2023, 33(4): 1113-1127.
- [10] MIRANDA L S, WIJESIRI B, AYOKO G A, et

- al. Water-sediment interactions and mobility of heavy metals in aquatic environments[J]. *Water Research*, 2021, 202:117386.
- [11] HUANG Y P, CAI Y P, HE Y H, et al. A water quality prediction approach for the Downstream and Delta of Dongjiang River Basin under the joint effects of water intakes, pollution sources, and climate change[J]. *Journal of Hydrology*, 2024, 640:131686.
- [12] MCQUILLAN C. Planetary health research digest[J]. *The Lancet Planetary Health*, 2024, 8(12):e986.
- [13] 张鹏, 梅书浩, 石成春, 等. 基于机器学习的水库溶解氧预测模型比较研究[J]. *华北水利水电大学学报(自然科学版)*, 2025, 46(01):87-95. [ZHANG P, MEI S H, SHI C C, et al. Comparative study of dissolved oxygen prediction models for reservoirs based on machine learning[J]. *Journal of North China University of Water Resources and Electric Power(Natural Science Edition)*, 2025, 46(01):87-95.]
- [14] MOGHA G, AHLAWAT K, SINGH A P. Performance analysis of machine learning techniques on big data using apache spark[C]//International conference on recent developments in science, engineering and technology. Singapore: Springer Singapore, 2017: 17-26.
- [15] LIANG Y C, DING F Y, LIU L, et al. Monitoring water quality parameters in urban rivers using multi-source data and machine learning approach[J]. *Journal of Hydrology*, 2025, 648:132394.
- [16] SU H, ZOU R, ZHANG X L, et al. Exploring the type and strength of nonlinearity in water quality responses to nutrient loading reduction in shallow eutrophic water bodies:insights from a large number of numerical simulations[J]. *Journal of Environmental Management*, 2022, 313:115000.
- [17] MAKUMBURA R K, MAMPITIYA L, RATHNAYAKE N, et al. Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature[J]. *Results in Engineering*, 2024, 23:102831.
- [18] BESSA SANTOS R M, FARIAS DO VALLE JUNIOR R, ABREU PIRES DE MELO SILVA M M, et al. A framework model to integrate sources and pathways in the assessment of river water pollution[J]. *Environmental Pollution*, 2024, 347:123661.
- [19] NIKOO M R, AMRI A A, ETRI T, et al. A review of machine learning, remote sensing, and statistical methods for reservoir water quality assessment[J]. *Journal of Hydrology*, 2025, 659:133323.
- [20] LIU X, ZHANG F J, HOU Z Y, et al. Self-supervised learning: generative or contrastive[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1):857-876.
- [21] MOZELLI A, TAHERINEJAD N, JANTSCH A. A study on confidence: an unsupervised multiagent machine learning experiment[J]. *IEEE Design & Test*, 2022, 39(3):54-62.
- [22] BARZEGAR R, ASGHARI MOGHADDAM A, ADAMOWSKI J, et al. Comparison of machine learning models for predicting fluoride contamination in groundwater[J]. *Stochastic Environmental Research and Risk Assessment*, 2017, 31(10):2705-2718.
- [23] MAMAT N, MOHAMAD HAMZAH F, JAAFAR O. Hybrid support vector regression model and K-Fold cross validation for water quality index prediction in Langat River, Malaysia[J]. *BioRxiv*. 2021:431242.
- [24] WATTELEZ G, DUPOUY C, LEFÈVRE J, et al. Application of the support vector regression method for turbidity assessment with MODIS on a shallow coral reef lagoon (Voh-Koné-Pouembout, New Caledonia)[J]. *Water*, 2017, 9(10):737.
- [25] SANTOS V O, ROCHA P A C, THÉ J V G, et al. Evaluation of machine learning methods for forecasting turbidity in river networks using Sentinel-2 remote sensing data[J]. *Ecological Informatics*, 2025, 90:103313.
- [26] ZOU T L, YANG K, PAN M E, et al. Analysis of the temporal and spatial evolution of turbidity in Tonle Sap Lake and its influencing factors[J]. *Science of the Total Environment*, 2024, 943:173618.
- [27] ZARE M S, NIKOO M R, AL-RAWAS G, et al. Integrated ensemble learning approach for multi-depth water quality estimation in reservoir environments[J]. *Journal of Water Process Engineering*, 2024, 66:105840.
- [28] NAJWA MOHD RIZAL N, HAYDER G, MNZOOL M, et al. Comparison between regression models, support vector machine (SVM), and artificial neural network (ANN) in river water quality prediction[J]. *Processes*, 2022, 10(8):1652.
- [29] MULIA I E, TAY H, ROOPSEKHAR K, et al. Hybrid ANN-GA model for predicting turbidity and chlorophyll-a concentrations[J]. *Journal of Hydro-environment Research*, 2013, 7(4):279-299.
- [30] IGLESIAS C, MARTÍNEZ TORRES J, GARCÍA NIETO P J, et al. Turbidity prediction in a river basin by using artificial neural networks: a case study in northern

- Spain[J]. *Water Resources Management*, 2014, 28(2): 319-331.
- [31] TOMPERI J, ISOKANGAS A, RUUSUNEN M. Practical data-based modelling approach for estimating river water turbidity and total organic carbon[J]. *Environmental Technology*, 2025: 1-17.
- [32] 刘洁, 祝榕婕, 姜德迅, 等. 基于遗传-神经网络的实时水质预测模型 [J]. *南水北调与水利科技 (中英文)*, 2020, 18(06):93-100. [LIU J, ZHU R J, JIANG D X, et al. Real-time water quality prediction model based on IGA-BPNN method[J]. *South-to-North Water Transfers and Water Science & Technology*, 2020, 18(6):93-100.]
- [33] GUO Q Z, WU H H, JIN H Y, et al. Remote sensing inversion of suspended matter concentration using a neural network model optimized by the partial least squares and particle swarm optimization algorithms[J]. *Sustainability*, 2022, 14(4):2221.
- [34] SONG C Y, ZHANG H P. Study on turbidity prediction method of reservoirs based on long short term memory neural network[J]. *Ecological Modelling*, 2020, 432: 109210.
- [35] LOPEZ-BETANCUR D, MORENO I, GUERRERO-MENDEZ C, et al. Convolutional neural network for measurement of suspended solids and turbidity[J]. *Applied Sciences*, 2022, 12(12): 6079.
- [36] MEI P, LI M, ZHANG Q, et al. Prediction model of drinking water source quality with potential industrial-agricultural pollution based on CNN-GRU-Attention[J]. *Journal of Hydrology*, 2022, 610:127934.
- [37] ALIZAMIR M, HEDDAM S, KIM S. A robust and explainable deep learning model based on an LSTM-CNN framework for reliable FDOM prediction in water quality monitoring: incorporating SHAP analysis for enhanced interpretability[J]. *Process Safety and Environmental Protection*, 2025, 201: 107594.
- [38] KOELMANS A A, MOHAMED NOR N H, HERMSEN E, et al. Microplastics in freshwaters and drinking water: critical review and assessment of data quality[J]. *Water Research*, 2019, 155: 410-422.
- [39] YANG J, TAO Y Z, REN W X, et al. Numerical simulation of groundwater contaminant transport in unsaturated flow[J]. *Water Supply*, 2020, 20(8): 3730-3738.
- [40] 刘宪亮, 许新勇, 陈晓楠, 等. 基于智慧水力学的长距离调水工程调度参数预测方法研究[J]. *华北水利水电大学学报 (自然科学版)*, 2024, 45(03): 18-25. [LIU X L, XU X Y, CHEN X N, et al. Research on scheduling parameter prediction method of long-distance water transfer project based on intelligent hydraulics[J]. *Journal of North China University of Water Resources and Electric Power(Natural Science Edition)*, 2024, 45(03): 18-25.]
- [41] JAEGER A, POSSELT M, SCHAPER J L, et al. Transformation of organic micropollutants along hyporheic flow in bedforms of river-simulating flumes[J]. *Scientific Reports*, 2021, 11(1): 13034.
- [42] JAVERNICK L, REDOLFI M, BERTOLDI W. Evaluation of a numerical model's ability to predict bed load transport observed in braided river experiments[J]. *Advances in Water Resources*, 2018, 115: 207-218.
- [43] PATRYNIAK K, COLLU M, CORADDU A. Rigid body dynamic response of a floating offshore wind turbine to waves: identification of the instantaneous centre of rotation through analytical and numerical analyses[J]. *Renewable Energy*, 2023, 218: 119378.
- [44] SANADA Y, KIM D-H, SADAT-HOSSEINI H, et al. Assessment of EFD and CFD capability for KRISO Container Ship added power in head and oblique waves[J]. *Ocean Engineering*, 2022, 243: 110224.
- [45] WEISSCHER S A H, BOECHAT ALBERNAZ M, LEUVEN J R F W, et al. Complementing scale experiments of rivers and estuaries with numerically modelled hydrodynamics[J]. *Earth Surface Dynamics*, 2020, 8: 955-972.
- [46] FEI K, DU H X, GAO L. The contribution of typhoon local and remote forcings to storm surge along the Makou-Dahengqin tidal reach of Pearl River Estuary[J]. *Science of the Total Environment*, 2023, 899: 165592.
- [47] XU G J, JI C J, XU Y, et al. Machine learning in coastal bridge hydrodynamics: a state-of-the-art review[J]. *Applied Ocean Research*, 2023, 134: 103511.
- [48] KIM J Y, SEO D I. Three-dimensional augmentation for hyperspectral image data of water quality: an Integrated approach using machine learning and numerical models[J]. *Water Research*, 2024, 251: 121125.
- [49] ZHI W, OUYANG W Y, SHEN C P, et al. Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers[J]. *Nature Water*, 2023, 1(3): 249-260.
- [50] CHEN Z, XU H, JIANG P, et al. A transfer Learning-Based LSTM strategy for imputing Large-Scale consecutive missing data and its application in a water quality

- prediction system[J]. Journal of Hydrology, 2021, 602:126573.
- [51] ZHENG Y, ZHANG X R, ZHOU Y C, et al. Deep representation learning enables cross-basin water quality prediction under data-scarce conditions[J]. npj Clean Water, 2025, 8(1):33.
- [52] SCHÄFER B, BECK C, RHYS H, et al. Machine learning approach towards explaining water quality dynamics in an urbanised river[J]. Scientific Reports, 2022, 12(1):12346.
- [53] HAN J W, KIM T H, LEE S C, et al. Machine learning and explainable AI for chlorophyll-a prediction in Namhan River Watershed, South Korea[J]. Ecological Indicators, 2024, 166:112361.
- [54] ALNAHIT A O, MISHRA A K, KHAN A A. Stream water quality prediction using boosted regression tree and random forest models[J]. Stochastic Environmental Research and Risk Assessment, 2022, 36(9):2661-2680.
- [55] GAO Z Y, WANG G Q, ZHU Y, et al. Prediction of water quality parameters and pollution exceedance analysis in typical rivers of semi-arid regions based on interpretable deep learning models[J]. Environmental Pollution, 2025, 383:126801.
- [56] MERABET K, DI NUNNO F, GRANATA F, et al. Predicting water quality variables using gradient boosting machine: global versus local explainability using SHapley Additive Explanations (SHAP)[J]. Earth Science Informatics, 2025, 18(3):298.
- [57] SOLEYMANI HASANI S, ARIAS M E, NGUYEN H Q, et al. Leveraging explainable machine learning for enhanced management of lake water quality[J]. Journal of Environmental Management, 2024, 370:122890.
- [58] SLATHIA A S, SHARMA A, KRISHNA P B, et al. SHERA: SHAP-Enhanced resource allocation for VM scheduling and efficient cloud computing[J]. IEEE Access, 2025, 13:92816-32.
- [59] QU Z K, XIAO R Z, YANG K, et al. Enhancing meteorological data reliability: an explainable deep learning method for anomaly detection[J]. Journal of Environmental Management, 2025, 374:124011.

Research Progress on the Application of Machine Learning in Turbidity Prediction of Urban River Water

LI Mengke, TANG Ye

(College of Life and Environmental Science, Wenzhou University, Wenzhou 325035, China)

Abstract:【Objective】 To systematically review the application status of mainstream machine learning techniques in river turbidity prediction, clarify the adaptation scenarios and practical performance of different technologies, and provide theoretical references for subsequent technical optimization and application expansion. **【Methods】** From the perspective of a comprehensive methodology, this paper elaborates on the application progress of machine learning models (including support vector regression, ensemble learning, artificial neural networks, and deep learning models) in river turbidity prediction, with a focus on discussing the differences in technical adaptability and practical efficacy among these models.

【Results】 (1) Ensemble learning is the preferred choice for river turbidity prediction; support vector regression demonstrates stronger robustness in small sample modeling; artificial neural networks can flexibly adapt to multi-factor coupling prediction; time-series deep learning models are suitable for dynamic feature extraction and long-term monitoring of main river channels. (2) Combining deep learning technology with controlled experiments and physical mechanism models can achieve multi-source data expansion and alleviate the problem of scarce samples. (3) Applying the Shapley additive explanation method to analyze the decision logic of the model helps to break the "black box" and improve the practicality and credibility of the model.

【Conclusion】 Although the existing algorithms have advantages in computational efficiency and prediction accuracy, their application still faces dual constraints of data scarcity and insufficient model interpretability. In the selection of actual models, multiple factors such as water body pollution characteristics, data availability, and operation and maintenance costs need to be comprehensively considered. Future research should deepen the integration of mechanism models and machine learning, constructing a turbidity prediction system for sudden environmental events based on controlled experiments, in order to further release the potential of machine learning and

provide more reliable intelligent decision support for sustainable water resource management and water ecological protection.

Keywords: turbidity prediction; machine learning; data expansion; deep learning; interpretability evaluation