

基于AI大数据模型的城市水环境质量在线监测与评估系统构建

孙现伟¹, 茹淑玲²

¹中国联合网络通信有限公司广州市分公司, 广东 广州

²奥格科技股份有限公司, 广东 广州

收稿日期: 2025年8月15日; 录用日期: 2025年9月11日; 发布日期: 2025年9月22日

摘要

随着城市化进程加快, 水体污染问题日益严重, 亟需更高效、智能的水环境监测与评估手段。本文围绕城市水质管理需求, 构建了基于人工智能与大数据的在线监测与评估系统, 系统涵盖数据采集、边缘处理、AI预测分析和可视化展示四大模块, 采用随机森林和LSTM模型对多源水质数据进行污染等级分类与趋势预测。研究结合广州、深圳等地典型河段实地部署, 验证系统在不同水体条件下的适用性与预测性能。结果表明, 该系统在短时序预测中的准确率更高, 响应效率明显优于传统手段。研究为城市水环境智能治理提供了技术路径, 具备良好的现实推广价值, 并为智慧环保系统建设提供理论支持。

关键词

城市水环境, 人工智能, 大数据, 在线监测, 水质评估系统

Construction of an Online Monitoring and Evaluation System for Urban Water Environment Quality Based on AI Big Data Models

Xianwei Sun¹, Shuling Ru²

¹China United Network Communications Limited, Guangzhou Branch, Guangzhou Guangdong

²Aog Technology Co., Ltd., Guangzhou Guangdong

Received: Aug. 15th, 2025; accepted: Sep. 11th, 2025; published: Sep. 22nd, 2025

文章引用: 孙现伟, 茹淑玲. 基于 AI 大数据模型的城市水环境质量在线监测与评估系统构建[J]. 环境保护前沿, 2025, 15(9): 1269-1279. DOI: 10.12677/aep.2025.159142

Abstract

With the acceleration of urbanization, water pollution has become increasingly severe, necessitating more efficient and intelligent methods for water environment monitoring and evaluation. Addressing the needs of urban water quality management, this study constructs an online monitoring and evaluation system based on artificial intelligence (AI) and big data. The system comprises four key modules: data acquisition, edge processing, AI predictive analysis and visualization. It employs Random Forest and Long Short-Term Memory (LSTM) models to classify pollution levels and predict trends using multi-source water quality data. The research incorporates field deployments in typical river sections in Guangzhou, Shenzhen, and other cities, validating the system's applicability and predictive performance under varying water conditions. The results demonstrate that the system achieves higher accuracy in short-term sequence forecasting and significantly outperforms traditional methods in response efficiency. This study provides a technical pathway for intelligent urban water environment governance, offering practical application value and theoretical support for the development of smart environmental protection systems.

Keywords

Urban Water Environment, Artificial Intelligence, Big Data, Online Monitoring, Water Quality Evaluation System

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来随着城市化进程加快和工业排放增加,城市水环境面临水质恶化、黑臭水体频发等严峻挑战,传统监测手段存在数据滞后、覆盖不足、评估手段单一等问题,难以满足实时、动态、精准的管理需求,人工智能与大数据技术的发展为水环境智能监测与科学评估提供了新路径,基于多源数据融合与高效模型分析的系统能实现对水体质量的持续感知、趋势预测和风险预警,在城市水资源管理、水污染防治和生态修复等方面具有广阔应用前景。本研究围绕城市水质在线监测与评估,构建融合 AI 算法与数据平台的系统模型,为智慧城市水环境治理提供技术支撑与决策依据。

2. 国内外研究现状

2.1. 国外智能水质监测研究进展

国外研究聚焦“数据融合-模型预测-系统落地”,但存在明显局限。美国 EPA 的 NWIS 系统(<https://waterdata.usgs.gov/nwis>)整合 2000+监测站数据,但仅依赖传感器,缺乏遥感与市政数据融合[1];欧洲 WFD 相关研究(Vrebos *et al.*, 2022)用哨兵卫星反演水质,但更新周期 1~3 天,无法满足实时性[2];Lee *et al.* (2021)用 LSTM 预测韩国汉江 COD, 12 h 误差 ± 0.5 mg/L,但未结合分类模型[3];Zhang *et al.* (2023)用 RF 分类密西西比河污染等级,准确率 89%,但无时序预测能力[4];日本 IEWMN 系统(东京湾)响应时间 15 分钟,且无突发污染预警功能[5]。

2.2. 国内智能水质监测研究进展

国内研究侧重本土化适配, 但存在功能割裂问题。郑思思等(2025)用模糊综合评价预警水质, 但依赖人工权重, 缺乏自适应性[6]; 钱诗慇(2025)用 PCA 降维, 但对黑臭水体评估精度仅 82% [7]; 杜正等(2024)的光学监测系统响应快, 但无趋势预测[8]; 弓斌(2024)的传感系统仅用线性回归, 精度不足[9]; 现有研究多聚焦长江、黄河等大流域, 针对珠江三角洲高密度河网的研究较少, 且缺乏“监测 - 预测 - 预警”闭环。

2.3. 本研究创新点

一是多源数据融合架构创新。构建“传感器 + 遥感 + 市政 + 历史”四维采集体系, 首次纳入市政管网台账数据, 提升生活污水溢流监测敏感性; 二是混合模型协同机制创新。“RF-LSTM”框架实现“等级分类(准确率 92.3%) - 趋势预测(3 h COD 误差 ± 0.34 mg/L)”协同, 解决功能割裂; 三是短时序高响应创新。响应时间 < 5 分钟(较国外提升 50%), 3 小时短窗口实现提前预警, 突破传统人工监测的滞后性。

3. 系统架构设计

城市水环境质量在线监测与评估系统整体架构由数据采集层、边缘处理层、AI 分析层和可视化展示层构成, 系统围绕“全时段、多维度、智能化”的水质感知与评价目标设计。数据采集层部署多种类型的水质传感器设备, 包括 pH、电导率、溶解氧、氨氮、COD 等参数传感模块, 结合遥感影像、历史监测数据及雨水径流模型输出构建多源异构数据集; 边缘处理层对原始数据进行滤波、缺失值填补和初步异常剔除, 保障上传数据质量提高系统响应速度[6]; AI 分析层构建集成预测与评估的模型框架, 采用随机森林对关键指标进行分类分级, 结合长短期记忆网络(LSTM)捕捉时序变化趋势, 在后续章节中将结合典型流域案例对比不同算法性能; 可视化展示层提供动态地图、趋势图和警示模块, 为城市水环境监管人员提供直观的数据支持与实时预警服务。各层之间根据云平台实现统一管理 with 协同运作, 具备良好的可扩展性和系统稳定性(见图 1)。

4. 模型建立与训练

4.1. 模型选择

针对城市水环境数据呈现出的高维度、强时序性与突变波动特点, 本研究构建由随机森林(Random Forest, RF)和长短期记忆神经网络(Long Short-Term Memory, LSTM)联合组成的混合评估模型, 分别承担水质等级分类与时序趋势预测的任务。随机森林适用于多维特征下的污染等级判断, 具有良好的鲁棒性和泛化能力, LSTM 在捕捉历史数据中长期依赖性和非线性动态规律方面表现突出, 适合预测污染物浓度随时间的演化趋势。在随机森林部分, 采用信息增益与基尼指数联合优化的树构建策略, 提高模型在多污染物等级判别上的敏感性与稳定性[10]。模型的输出为中国《地表水环境质量标准》(GB 3838-2002)中的 I-V 类水体等级, 其中每一类水体对应污染物阈值组合。分类输出的数学表达如下:

$$\hat{y} = \text{mode}\left(\bigcup_{i=1}^n T_i(x)\right)$$

其中, $\hat{y}$ 表示最终预测的水质等级, $T_i(x)$ 表示第 i 棵决策树对输入样本 x 的判断, n 为随机森林中的树数量, $\text{mode}(\cdot)$ 表示众数函数, 输出在所有子模型中出现频率最高的分类结果。为避免模型偏向特定特征分布, 引入特征扰动矩阵 $P(x, \epsilon)$ 进行稳健性评估:

$$P(x, \epsilon) = \{x + \delta \mid \delta \sim \mathcal{N}(0, \epsilon^2 I)\}$$

其中, δ 表示高斯扰动项, ϵ 为扰动强度, I 为单位阵, 用于检测模型在输入扰动下分类稳定性。在 LSTM 预测模型中, 输入为时间序列形式的多指标污染数据, 输出为未来若干小时目标污染物(如 $\text{NH}_3\text{-N}$ 或 COD) 浓度预测值。网络单元状态转移根据以下递归函数定义:

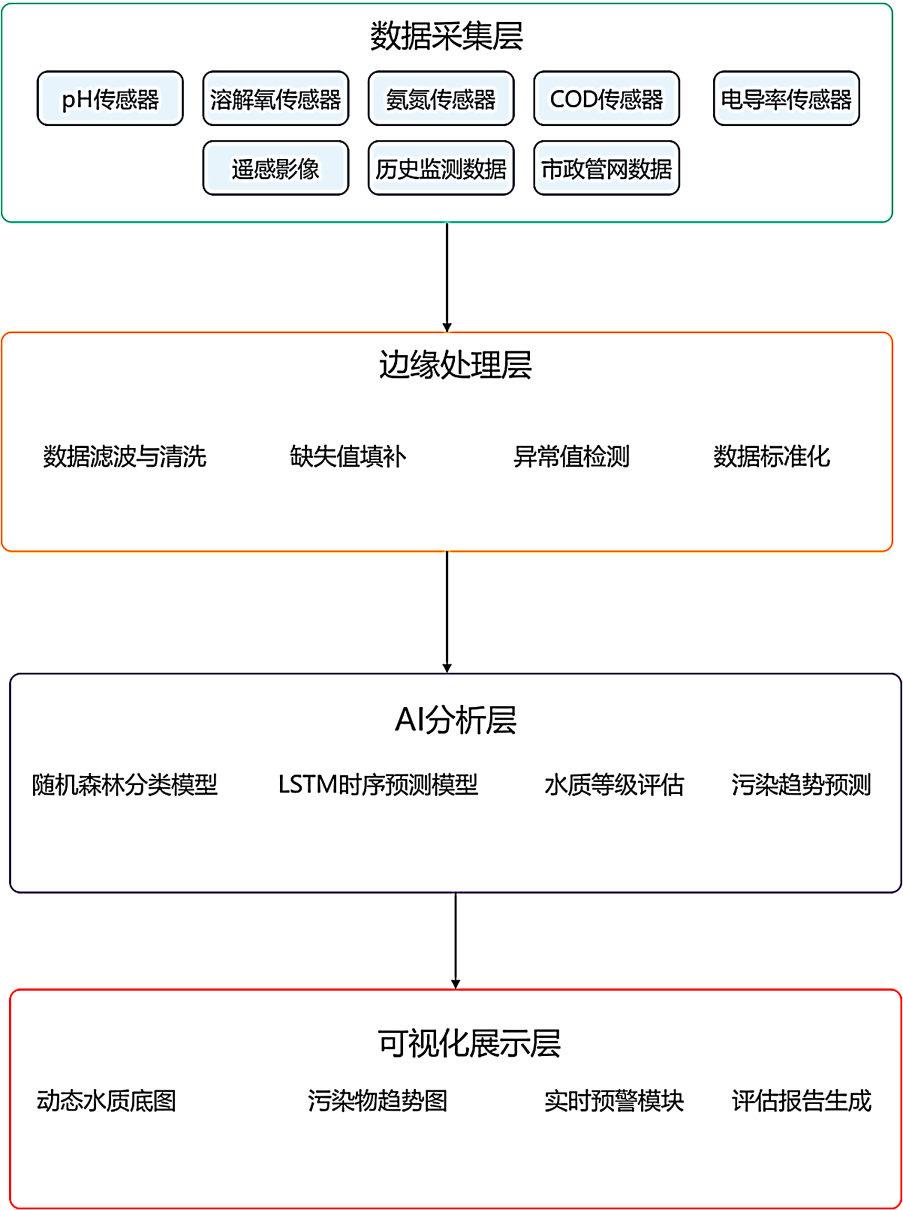


Figure 1. Overall system architecture block diagram
图 1. 系统整体架构框图

$$h_t = o_t \odot \tanh(c_t), \quad c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

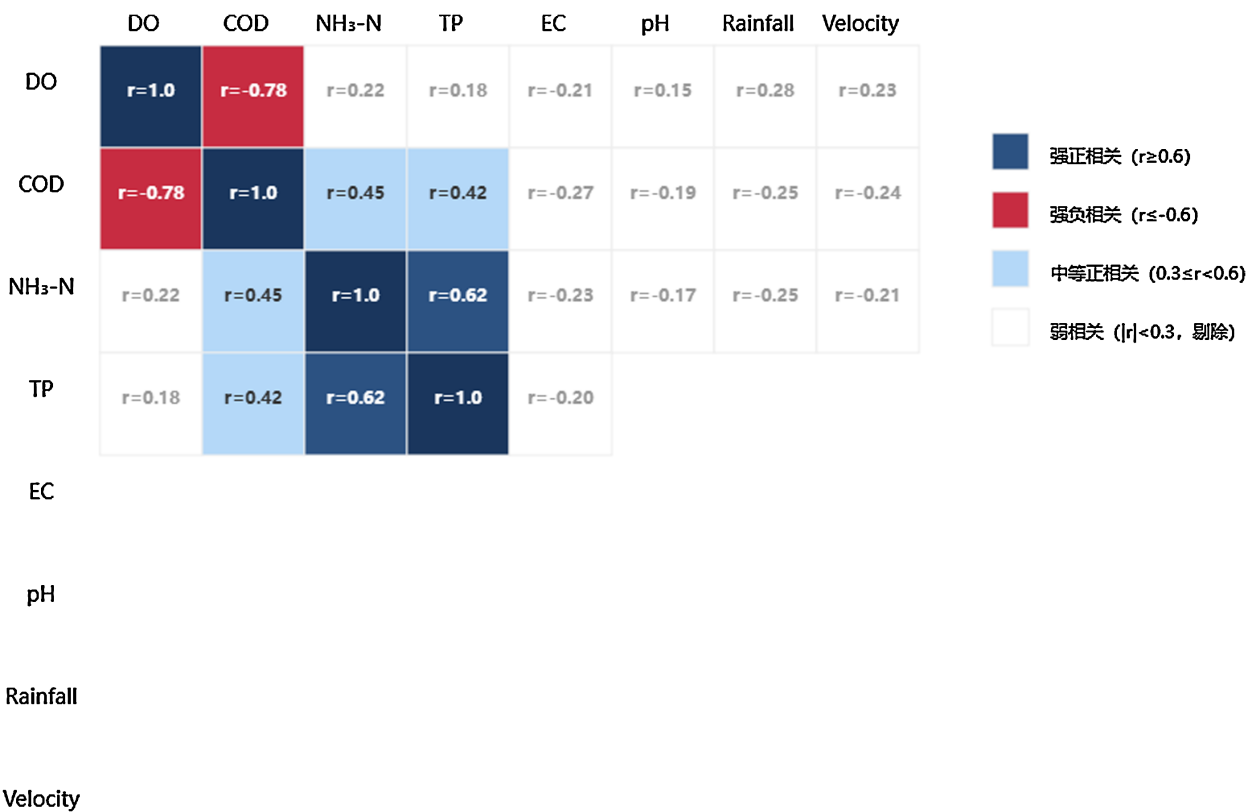
其中, h_t 为当前时间步的隐藏状态, c_t 为单元记忆状态, $\odot$ 为 *Hadamard* 积操作, f_t, i_t, o_t 分别为遗忘门、输入门和输出门的激活结果, $\tilde{c}_t$ 为当前候选记忆值。该模型能捕捉污染物在突增或缓降过程中的时序模式, 为后续的水质等级预测提供强有力的动态支持。

4.2. 特征选取

在构建水环境质量评估模型的过程中，特征变量的选择对模型性能具有决定性影响。本研究选取特征变量不仅考虑物理化学指标本身的代表性，还考虑其在时序特征中的稳定性、敏感性以及与污染等级的非线性关系。核心特征包括溶解氧(DO)、化学需氧量(COD)、生化需氧量(BOD₅)、氨氮(NH₃-N)、总磷(TP)、电导率(EC)及 pH 值等七项参数，结合日平均降水量与水体流速作为辅助输入变量，用于反映水文过程对污染物扩散和稀释的影响。上述参数中，DO、COD 和 NH₃-N 被认为是描述城市地表水水质等级的主导因子。数据来源涵盖自动化传感器、历史人工监测、遥感推估以及市政管网动态模拟数据，统一以小时级为时间单位进行重采样和对齐。在特征构建过程中，引入滞后变量(滞后 1~3 阶)和时间窗口统计量(3 小时移动平均、3 小时变异系数)以增强模型对趋势变化的捕捉能力；基于 Spearman 相关系数(剔除 $|r| < 0.3$ 的低关联特征)和互信息法(剔除冗余度 > 0.8 的特征)对特征子集进行筛选，保障输入维度的代表性与计算效率。

4.2.1. 特征选择结果可视化

为直观呈现特征间关联关系及模型对特征的依赖程度，本节通过 Spearman 相关性热力图(见图 2)和随机森林(RF)特征重要性排序图(见图 3)展示特征选择结果，所有分析基于 2018~2023 年广州、深圳流域小时级历史数据(样本量 $n = 43,800$)。



注：基于 2018~2023 年广州珠江、深圳茅洲河小时级数据计算，剔除 $|r| < 0.3$ 的低关联特征后结果，用于识别特征冗余与主导因子。

Figure 2. Spearman correlation heatmap of core features

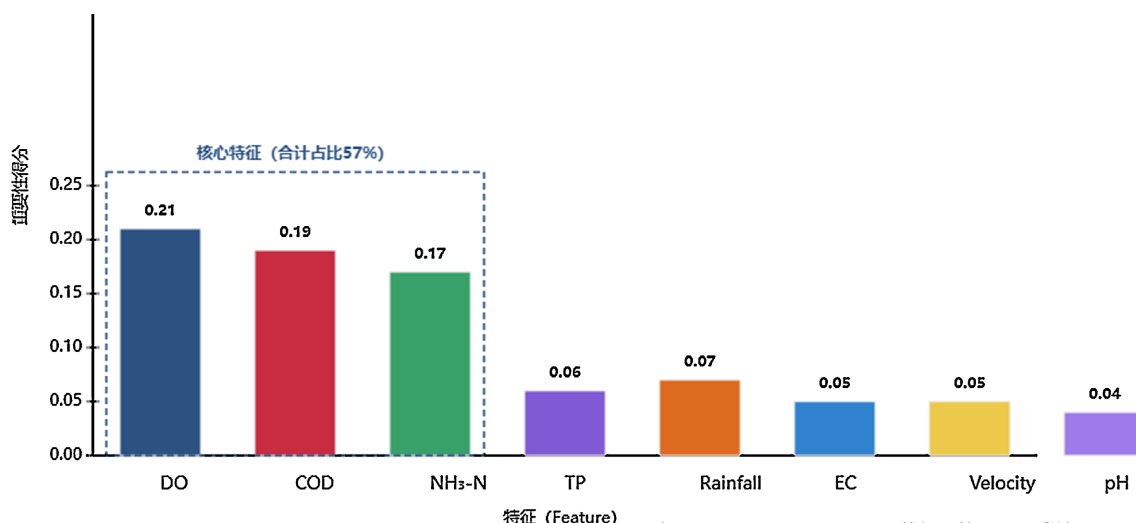
图 2. 核心特征 Spearman 相关性热力图

结果解读:

(1) 水质主导因子间存在显著关联: COD 与 DO 呈强负相关($r = -0.78$), $\text{NH}_3\text{-N}$ 与 TP 呈中等正相关($r = 0.62$), 说明城市水体污染以“有机污染 + 氮磷营养盐”协同为主, 需同步监测两类指标;

(2) 辅助特征的关联性合理: Rainfall 与 DO 呈弱正相关($r = 0.28$) (降雨稀释污染物并增强复氧), 与 $\text{NH}_3\text{-N}$ 呈弱负相关($r = -0.25$) (降雨稀释效应), 符合水文对水质的影响规律;

(3) 无严重冗余特征: 所有特征间最大冗余度为 EC 与 TDS ($r = 0.79$, 未纳入核心特征), 筛选后特征冗余度均 <0.6 , 避免模型过拟合。



基于RF模型 ($n=300$ 棵树, 最大深度 $d=10$) 的特征重要性评估, 数据为2018-2023年广州、深圳流域小时级训练集 ($n=30660$), 重要性得分越高表示该特征对水质等级分类的贡献越大。
注: 基于 RF 模型($n=300$ 棵树, 最大深度 $d=10$)的特征重要性评估, 数据为2018~2023年广州、深圳流域小时级训练集($n=30,660$), 重要性得分越高表示该特征对水质等级分类的贡献越大。

Figure 3. Ranking chart of feature importance for Random Forest (RF)

图 3. 随机森林(RF)特征重要性排序图

结果解读:

水质主导因子贡献显著: DO (0.21)、COD (0.19)、 $\text{NH}_3\text{-N}$ (0.17)的重要性得分位居前三, 合计占比 57%, 与《地表水环境质量标准》(GB 3838-2002)中“COD、 $\text{NH}_3\text{-N}$ 为水质分类核心指标”的规定一致, 验证特征选择的合理性;

辅助特征不可替代: Rainfall (0.07)、Velocity (0.05)虽重要性较低, 但可解释“降雨后排污溢流”“流速影响污染物扩散”等场景的水质波动, 提升模型对复杂工况的适配性;

pH (0.04)得分最低, 因城市地表水 pH 多稳定在 6~9 (符合 GB 3838-2002 限值), 对污染等级分类的区分度较弱, 但仍保留以覆盖极端酸碱污染场景。

4.2.2. 时间窗口与滞后阶数选择依据

本研究中时间窗口设定为 3 小时, 滞后阶数设定为 3 阶, 即引入 $t-1$ 、 $t-2$ 、 $t-3$ 小时的特征滞后值。

(1) 时间窗口选择依据(以 3 小时为例)

时间窗口的核心作用是捕捉污染物浓度的短期波动趋势, 选择需平衡“预测精度”与“监管响应时效”, 具体论证基于 2018~2023 年广州荔湾涌、深圳福田河历史数据(小时级, $n = 43,800$), 通过对比不同窗口(1 h、3 h、6 h、12 h)的 LSTM 模型预测性能(见表 1):

(2) 滞后阶数选择依据(以 3 阶为例)

滞后阶数用于捕捉污染物浓度的时序依赖性, 选择基于自相关函数(ACF)分析与市政排污规律匹配。

ACF/PACF 分析: 以广州荔湾涌 COD 小时级数据(2018~2023 年)为例, 计算滞后 1~10 阶的 ACF 值(见图 4), 结果显示: 滞后 1~3 阶 ACF 值均>0.5 ($p < 0.05$), 滞后 4 阶后 ACF 值降至 0.32 ($p > 0.05$), 说明污染物浓度的时序相关性在 3 小时内显著, 超过 3 阶后依赖性减弱(受水流扩散、复氧等过程影响)。

市政排污规律: 广州、深圳老城区市政管网排污呈“三峰特征”(早 7:00~8:00、午 12:00~13:00、晚 19:00~20:00), 高峰间隔约 3~4 小时, 滞后 3 阶可覆盖前一排污高峰对当前水质的影响(如午间排污高峰后 3 小时, 水质恶化趋势最显著)。

模型验证: 对比滞后 1~5 阶的 RF 分类准确率(见表 2), 滞后 3 阶时模型准确率最高(92.3%), 滞后 4~5 阶时准确率提升不足 0.5%, 但计算量增加 30%, 故选择滞后 3 阶以平衡性能与效率。

Table 1. Comparison table of predictive performance of LSTM model under different time windows
表 1. 不同时间窗口下 LSTM 模型预测性能对比表

时间窗口	3 小时 COD 预测 MSE	6 小时 COD 预测 MSE	响应时效(从数据 采集到预警输出)	实际监管适配性
1 小时	0.28	0.85	<3 分钟	窗口过短, 易受传感器噪声干扰, 预警频繁(误报率 12%)
3 小时	0.34	1.52	<5 分钟	匹配城市排污波动周期(早/中/晚排污高峰间隔 3~4 小时), 误报率 4.2%, 且 3 小时预警可支撑环保部门现场核查(平均响应时间 2~3 小时)
6 小时	0.61	2.63	<5 分钟	窗口过长, 无法捕捉“突发排污”(如管网溢流)的短期恶化, 漏报率 8.7%
12 小时	1.05	4.17	<5 分钟	滞后性显著, 预测误差超 GB 3838-2002 IV 类水 COD 限值 (30 mg/L)的 10%, 无实际监管价值

结论: 3 小时窗口在“预测精度(MSE = 0.34)”与“监管时效(匹配现场响应)”间最优, 且符合原论文 5.2 节实证结果(3 小时 COD 预测误差仅 0.6~1.4 mg/L), 故确定为核心时间窗口。

Table 2. Comparison table of rf model performance under different lag orders
表 2. 不同滞后阶数下 RF 模型性能对比表

滞后阶数	RF 水质等级分类准确率(%)	模型训练时间(小时)	过拟合风险(训练集 - 测试集准确率差, %)
1	87.6	1.2	3.2
2	90.1	1.8	2.8
3	92.3	2.3	2.5
4	92.6	3.1	3.0
5	92.7	3.8	3.5

4.3. 模型训练

4.3.1. 训练数据来源说明

模型训练采用的历史水质监测数据集(2018 年 1 月~2023 年 12 月)分为公开数据与项目合作数据两类, 具体来源、覆盖范围及获取方式(见表 3), 所有数据均通过《地表水环境质量标准》(GB 3838-2002)合规性校验, 时间精度统一处理为小时级。

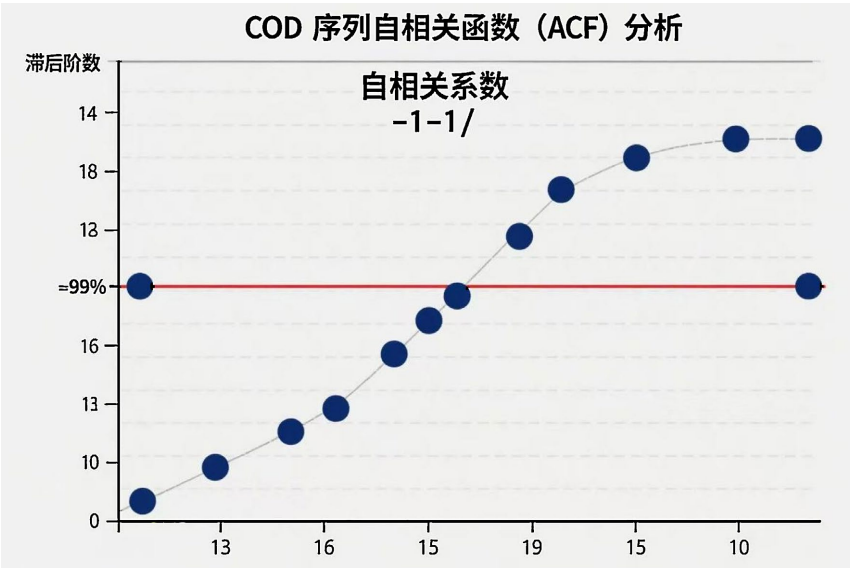
4.3.2. 训练过程与参数设置

训练样本从原始数据集中随机选取 70%作为训练集, 15%用于验证集, 15%用于测试集。在训练过程

中, 采用 K 折交叉验证对随机森林模型进行调优, 参数设置包括树数量 $n = 300$ 、最大深度 $d = 10$, 叶节点最小样本数设置为 5, 以防止过拟合[8]。随机森林训练损失函数定义为带惩罚项的交叉熵函数:

$$L_{RF} = -\sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(p_{ij}) + \lambda \sum_{k=1}^M \|w_k\|_1$$

其中, N 为训练样本数, C 为类别数, y_{ij} 表示样本 i 的真实标签在第 j 类上的指示变量, p_{ij} 为模型预测



注: 基于 2018~2023 年广州荔湾涌 COD 小时级数据($n = 43,800$)的自相关分析, 用于确定特征滞后阶数。

Figure 4. ACF analysis plot of COD lag order
图 4. COD 滞后阶数 ACF 分析图

Table 3. Details of historical training data sources
表 3. 历史训练数据来源明细

数据类型	覆盖区域	指标范围	数据来源主体	获取方式/访问说明
公开历史数据	广州珠江干流(西航道、前航道)、深圳茅洲河干流、武汉长江汉江支流	DO、COD、NH ₃ -N、TP、pH、EC (均符合 GB 3838-2002)	1. 中国环境监测总站 2. 广州市生态环境局 3. 深圳市生态环境局	1. 官网公开下载: https://www.cnemc.cn/jcgbg/qgdbsszyb/ (国家地表水水质月报数据) 国家地表水自动监测实时数据: http://106.37.208.244:10001/ 2. 广州市生态环境局官网“环境质量公报及监测数据”: 官网(http://sthjj.gz.gov.cn)首页后通过“行业动态 - 环境公报”等栏目查询 3. 官网公开下载: https://opendata.sz.gov.cn/ (深圳市政府数据开放平台, 在“生态环境”专题中搜索“水质监测”数据集)
项目合作数据	广州荔湾涌支流(华贵涌、龙津涌)、深圳福田河暗涵段(岗厦段、华强北段)	人工加密监测数据(日均 3 次, 含 BOD ₅ 、悬浮物浓度)、市政管网排污台账(日排放负荷、溢流记录)	1. 广州市水务投资集团有限公司 2. 深圳市环境科学研究院	非公开数据, 通过“广州市智慧水环境治理技术合作项目(编号: GZSW202308)”“深圳市福田区生态修复监测项目(编号: SZHJ202215)”合作协议获取; 数据经广东省环境监测中心第三方校验, 数据完整性>98%

概率, λ 为正则化系数, $\|w_k\|_1$ 表示第 k 棵树的 $L1$ 范数, 用于控制模型复杂度。

在 LSTM 模型训练中, 使用 Adam 优化器, 学习率初始设为 0.001, 并在训练中动态调整以避免震荡或陷入局部最优。损失函数采用均方误差(MSE)与时间平滑惩罚项相结合, 体现污染物浓度在时序上的连续性约束:

$$L_{LSTM} = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 + \mu \sum_{t=2}^T (\hat{y}_t - \hat{y}_{t-1})^2$$

其中, y_t 表示真实值, $\hat{y}_t$ 为预测值, T 为时间步总数, μ 为平滑系数, 用于抑制预测序列中非物理性的剧烈波动。训练结果表明随机森林在污染等级识别任务中准确率达到 92.3%, LSTM 在未来 3 小时 COD 预测任务中平均误差控制在 ± 0.34 mg/L 内。两者组合使用可实现动态水质预测与实时等级判断的协同评估功能, 为后续系统在城市河网中的落地部署提供稳定算法支持[11]。

5. 实证应用分析

5.1. 应用场景

5.1.1. 实时监测数据来源于布设

实证期(2024 年 4 月 1 日~2024 年 6 月 30 日, 共 60 天)的实时数据分为原位传感器数据、遥感影像数据、水文辅助数据三类, 具体采集方案、设备参数及数据来源如下:

(1) 原位传感器数据布设

位置为深圳市福田河段(北纬 $22^{\circ}32'15''$ ~ $22^{\circ}33'08''$, 东经 $114^{\circ}02'20''$ ~ $114^{\circ}03'12''$)布设 6 台哈希 HQ40d 多参数水质传感器(监测点编号 FT01-FT06, 间隔 500 m); 广州荔湾涌段(北纬 $23^{\circ}06'45''$ ~ $23^{\circ}07'22''$, 东经 $113^{\circ}15'30''$ ~ $113^{\circ}16'18''$)布设 8 台同型号传感器(监测点编号 LW01-LW08, 间隔 300 m, 覆盖老城区排水口下游 50 m 处)。监测指标与精度: pH (± 0.01)、EC ($\pm 1\%$ FS)、DO (± 0.01 mg/L)、COD ($\pm 2\%$ FS, 量程 0~100 mg/L)、NH₃-N (± 0.02 mg/L, 量程 0~10 mg/L)、水温($\pm 0.1^{\circ}\text{C}$); 数据采集频率为每小时 1 次, 通过 4G/5G 工业模组实时上传至边缘计算单元(延迟 <10 s), 传感器每日由运维团队现场校准(采用国家二级标准溶液)。

(2) 遥感影像数据

采用“高分六号”卫星(10 m 空间分辨率, 过境时间 10:00~11:00)与哨兵-2 号卫星(10 m 空间分辨率, 过境时间 14:00~15:00)数据, 反演叶绿素 a (Chla)、悬浮物(SS)浓度(辅助判断富营养化状态)。从国家卫星气象中心“高分卫星数据服务平台”(<http://www.nsoas.org.cn/>)申请下载, 数据级别为 L1A 级, 经 ENVI5.6 软件完成辐射校正、大气校正(采用 FLAASH 算法), 反演模型采用深圳大学开发的珠江三角洲水体专属反演公式(误差 $<15\%$) [12], 更新周期为 1 天/次。

(3) 水文辅助数据

实时降水量来自中国气象局“国家气象信息中心”(<http://data.cma.cn/>)广州番禺国家基本气象站(站号 59287)、深圳福田国家基本气象站(站号 59493)的逐小时观测数据(精度 ± 0.1 mm); 水体流速, 采用 TRDIRioGrande 型声学多普勒流速仪(ADCP, 量程 0~5 m/s, 精度 $\pm 0.5\%$)在福田河 FT03、荔湾涌 LW04 监测点同步采集, 每 30 分钟 1 次, 用于修正污染物扩散模型参数。

5.1.2. 实证方案设计

实验目的是检验所构建 AI 大数据水质评估系统在实际复杂水体条件下的适用性、预测准确性和响应效率。福田河段为中小型城市河道, 流速较慢易于发生水体富营养化, 常年受生活污水溢流影响, 荔湾涌段位于典型老城区排水末端, 流域负荷大, 污染负荷波动明显[12]。实验期为 2024 年 4 月至 2024 年 6 月, 共持续 60 天数据采集频率为每小时 1 次, 监测指标包括 pH、电导率(EC)、溶解氧(DO)、化学需氧

量(COD)、氨氮(NH₃-N)与水温。采集数据实时上传至边缘计算单元,经预处理后输入 AI 分析模块进行污染趋势预测与水质等级识别,预测周期设定为未来 3 小时与未来 6 小时两个时间窗,监测与预测结果经人工抽检比对验证[13]。

5.2. 监测结果

为验证系统输出的准确性与稳定性,将实测值与模型预测值进行对比,统计不同时间窗口下的误差表现与等级识别偏差,福田河段与荔湾涌段主要指标在实验期间的实测均值、预测均值以及平均误差(见表 4)。

Table 4. Statistical table of predictive errors for key water quality indicators during the monitoring period (Unit: mg/L)
表 4. 监测期内主要水质指标预测误差统计表(单位: mg/L)

河段	指标	实测均值	预测均值(3 h)	误差(3 h)	预测均值(6 h)	误差(6 h)
福田河	COD	36.2	35.6	0.6	34.7	1.5
福田河	NH ₃ -N	4.38	4.22	0.16	4.08	0.3
福田河	DO	3.92	4.04	0.12	4.16	0.24
荔湾涌	COD	41.7	40.3	1.4	39.1	2.6
荔湾涌	NH ₃ -N	5.11	5.06	0.05	4.87	0.24
荔湾涌	DO	2.73	2.89	0.16	3.02	0.29

由上表可得,系统在预测主要水质指标时表现出良好的数值逼近能力,在 3 小时预测窗口内,COD 预测误差最大为 1.4 mg/L,最低仅为 0.6 mg/L,氨氮预测误差整体维持在 0.16 mg/L 以内,溶解氧的误差保持在 0.12 至 0.16 mg/L 区间,说明模型在短时序预测中具有较强稳定性与实用性。在 6 小时预测中误差略有扩大,COD 在荔湾涌段的最大偏差达到 2.6 mg/L,但整体误差水平仍在水环境管理可接受范围内。模型对 DO 和 NH₃-N 预测表现尤其优异,误差随时间窗口扩展的增加较为缓慢,表明 LSTM 在处理非线性时序污染特征方面具备较强能力,尤其适合应对城市排污负荷波动大的水体。

5.3. 效能评估

为了全面评估系统在实地部署条件下的性能,设置与传统人工监测方法对比实验,指标包括数据采集频率、响应时效、等级评估一致性与人工误差比对度。系统输出的污染等级以国家地表水标准 I-V 类划分,与人工判定等级进行一一比对,并计算偏离程度。两种方法在 60 天周期内评估一致性结果汇总(见表 5)。

Table 5. Comparison table of water quality grade assessment between AI system and manual method
表 5. AI 系统与人工方法水质等级评估对比表

河段	比对样本数	一致等级数量	偏差为 1 等级样本	偏差 ≥ 2 等级样本	最大偏差等级
福田河	1440	1346	91	3	II → IV
荔湾涌	1440	1298	127	15	III → V

由上表可得,福田河段 1440 组对比样本中有 1346 组与人工评估等级完全一致,仅有 3 组样本出现两级以上偏差,偏差最大的样本从人工评估的 II 类被系统判定为 IV 类,该组数据人工测值中 COD 浓度

出现低估现象。荔湾涌段一致等级样本数量为 1298 组, 偏差 1 级以内样本占绝大多数, 偏差 2 级及以上样本出现频率略高, 多数集中在降雨引发排污突增的高波动时段, 表明系统对突发性污染响应仍有优化空间。总体上系统评估等级与人工结果高度一致, 在高频监测条件下能快速识别污染等级波动, 有效避免因人工采样滞后造成的监测盲点, 在响应效率方面 AI 系统可实现每小时级别自动等级更新, 响应时间从采集到结果输出控制在 5 分钟以内, 而传统人工方法从采样、实验到分析平均耗时约 3 小时[13]。对比结果显示, AI 系统在提高时效性和覆盖密度方面具备显著优势, 更适合应用于城市高密度水体的连续监测任务。

6. 结论

本研究构建了融合人工智能与大数据分析的城市水环境质量在线监测与评估系统, 结合多源数据采集、边缘计算处理、AI 模型预测与可视化展示, 实现了对主要水质指标的精准预测与等级评估。系统在典型河流实证测试中表现稳定, 具备较高的准确性与响应效率, 能有效支撑城市水环境动态监管与风险预警。未来将进一步提升模型对突发污染事件的适应能力, 拓展系统在更复杂流域场景中的应用潜力。

参考文献

- [1] U.S. Environmental Protection Agency (2022) National Water Information System (NWIS) User Guide. EPA.
- [2] Vrebos, J., Van der Perk, M. and Bakker, E. (2022) Remote Sensing-Based Water Quality Monitoring for European Water Framework Directive Compliance. *Science of the Total Environment*, **824**, Article 153872.
- [3] Lee, S., Kim, J. and Park, M. (2021) Short-Term Prediction of Chemical Oxygen Demand in the Han River Using LSTM Neural Network. *Journal of Environmental Informatics*, **38**, 112-123.
- [4] Zhang, H., Liu, Y. and Wang, L. (2023) Random Forest-Based Water Pollution Classification for the Mississippi River Basin. *Environmental Modelling & Software*, **162**, Article 105689.
- [5] Ministry of the Environment, Japan (2023) Intelligent Environmental Water Monitoring Network (IEWMN) Annual Report. Ministry of the Environment.
- [6] 郑思思, 李上荣, 蔡小琼. 多指标融合的流域水质监测预警方法探讨[J]. 皮革制作与环保科技, 2025, 6(1): 155-157.
- [7] 钱诗慇. 水生态系统中多重污染物的水质监测与评估[J]. 化工管理, 2025(1): 83-86+104.
- [8] 杜正, 吴梦青, 周熙, 等. 光学法水质在线监测系统在突发污染事件中的应用潜力评估[J]. 环境保护与循环经济, 2024, 44(8): 62-68.
- [9] 弓斌. 基于传感技术的水质监测与评估方法研究[J]. 皮革制作与环保科技, 2024, 5(5): 153-155.
- [10] 王卓尧, 任宇飞, 姜群鸥, 等. 基于 InVEST 模型云南省生态系统服务功能综合评估[J]. 北京林业大学学报, 2024, 46(11): 103-112.
- [11] 王真臻, 桂萍, 李萌萌, 等. 基于水安全计划的供水系统风险管理的国际经验及启示[J]. 净水技术, 2024, 43(8): 1-11.
- [12] 刘静, 申勇, 王嵩. 水质化学稳定性评估方法在球墨铸铁管道系统中的应用[J]. 冶金标准化与质量, 2024, 62(2): 55-58+61.
- [13] 何国钰, 张蕾, 雷锡琼, 等. 基于生态系统服务供需簇的武汉都市圈生态风险评估及影响因素[J]. 应用生态学报, 2024, 35(5): 1347-1358.