

# 基于深度学习的巢湖富营养化预测

龚肇青<sup>1,2</sup>, 牛宏<sup>1</sup>, 林晶晶<sup>2,3\*</sup>, 黄小龙<sup>2</sup>, 杨超<sup>2</sup>, 余明星<sup>2</sup>

(1 中南民族大学 资源与环境学院, 湖北 武汉 430074; 2 生态环境部 长江流域生态环境监督管理局生态环境监测与科学研究中心, 湖北 武汉 430010; 3 流域水生态智能监测预警与保护湖北省重点实验室, 湖北 武汉 430010)

**摘要** 巢湖作为中国重要的淡水湖泊, 长期面临严重的水质污染和富营养化问题. 基于2014—2021年水质数据, 采用CNN、LSTM及CNN-LSTM三种模型预测巢湖的营养状态指数并进行对比, 探讨巢湖水质时空演变规律. 结果表明: CNN-LSTM混合模型预测性能最优( $R^2=0.8054$ ), 适用于复杂水环境的水质预测; 巢湖富营养化程度呈“西北湖区>中部湖区>东部湖区”的空间分异格局, 西北湖区仍处于中度富营养状态, 其部分地区有水质改善趋势; 2022—2030年预测期内湖区TLI值整体低于历史水平, 表明流域治理措施可能具有一定积极效果. 研究结果揭示了巢湖富营养化的时空分异规律, 对优化流域治理措施、提升精准高效的防控策略具有重要指导意义.

**关键词** 巢湖; 富营养化; 水质预测; 深度学习

**中图分类号** X524; TP183 **文献标志码** A **文章编号** 1672-4321(2026)06-0744-10

**doi**: 10.20056/j.cnki.ZNMDZK.20260603

## Prediction of eutrophication in Chaohu Lake based on deep learning

Gong Zhaoqing<sup>1,2</sup>, Niu Hong<sup>1</sup>, Lin Jingjing<sup>2,3\*</sup>, Huang Xiaolong<sup>2</sup>, Yang Chao<sup>2</sup>, Yu Mingxing<sup>2</sup>

(1 College of Resources and Environmental Science, South-Central Minzu University, Wuhan 430074, Hubei China;

2 Yangtze River Basin Ecological Environment Monitoring and Scientific Research Center, Ecological Environment Ministry, Wuhan 430010, Hubei China; 3 Hubei Key Laboratory of Intelligent Monitoring and Early Warning of Aquatic Ecology for Watershed Protection, Wuhan 430010, Hubei China)

**Abstract** As a pivotal freshwater lake in China, Chaohu Lake has persistently grappled with severe water pollution and eutrophication issues. Drawing upon water quality data spanning from 2014 to 2021, three models—CNN, LSTM, and CNN-LSTM—were utilized to forecast and contrast the trophic level index (TLI) of Chaohu Lake, with the objective of unraveling its spatiotemporal evolutionary dynamics. The findings reveal that the CNN-LSTM hybrid model exhibits the highest predictive efficacy ( $R^2=0.8054$ ), making it apt for water quality prediction in intricate aquatic settings. The eutrophication condition of Chaohu Lake manifests a spatial differentiation trend characterized by "the northwestern lake region>the central lake region>the eastern lake region." The northwestern area continues to be in a state of moderate eutrophication, although certain zones exhibit signs of improvement. Over the prediction period from 2022 to 2030, the TLI values across the lake are predominantly lower than historical benchmarks, implying that the prevailing watershed management initiatives have yielded tangible positive outcomes. These discoveries clarify the spatiotemporal differentiation patterns of eutrophication in Chaohu Lake, offering crucial insights for optimizing watershed management protocols and enhancing precise and effective prevention and control strategies.

**Keywords** Chaohu Lake; eutrophication; water quality prediction; deep learning

湖泊是地球陆地表面水体的关键组成部分, 对 全球范围内的水循环、水量调节以及物质和能量平

收稿日期 2026-03-29

\* 通信作者 林晶晶(1990-), 女, 高级工程师, 博士, 研究方向: 水环境保护与管理, E-mail: jjlin90@163.com

基金项目 国家重点研发计划资助项目(2023YFC3205600)

衡起着重要作用.同时,湖泊作为地球上重要的淡水来源之一,为人类生产与生活提供了水资源保障<sup>[1]</sup>.然而,人类活动和气候变化威胁着全球湖泊的水质<sup>[2]</sup>,为维护湖泊的生态、社会和经济功能,及时了解和预测湖泊水质对于科学管理水资源至关重要<sup>[3]</sup>.

传统湖库水质评价方法主要包括物理化学法、生物法和综合评价法<sup>[4]</sup>.物理化学法涉及测定水样中的各种物理化学参数,如pH、溶解氧和养分含量,以评估水的基本性质和污染水平.生物法主要是观察和分析水中生物群落的结构和活动,从而评价水环境质量<sup>[5]</sup>.虽然物理化学法和生物法能准确反映水质参数的内容<sup>[6]</sup>,但不能反映整体水质.综合评价法通过计算水质指数(water quality index, WQI),全面利用水质参数信息,并通过标准化的规范值传达水质的整体状况<sup>[7]</sup>.这些方法在处理非线性、非稳定的时间序列数据时,难以捕捉其中复杂的模式,从而无法提取有效信息.

随着人工智能的发展,机器学习(machine learning, ML)和深度学习(deep learning, DL)在水质时间序列预测中被广泛应用.机器学习模型可以捕捉数据中的非线性关系,包括支持向量机(support vector machine, SVM)、随机森林(random forest, RF)、极端梯度增强(extreme gradient boosting, XGBoost)等<sup>[8]</sup>.Jiang等<sup>[9]</sup>首次研究了SVM、最小二乘支持向量回归(least squares support vector regression, LSSVR)和广义回归神经网络(generalized regression neural network, GRNN)模型在预测环丙沙星(ciprofloxacin, CFX)方面的有效性,并基于传统易检测的水体物理化学指标,实现了郊区水面环境中抗生素浓度的粗略预测.张强英等<sup>[10]</sup>总结了机器学习在水质分类、水质参数预测、污染物扩散模拟与溯源分析、时间序列预测、实时监测与突发事件响应、气候变化对水质影响等方面的应用,阐述了ML技术在水质时间序列预测中取得的重要进展,但仍存在缺乏跨区域适用性的局限.水质数据具有时间动态性、变异性和非平稳性特征,浅层机器学习模型可以用来拟合非线性数据,但仍无法在水质时间序列中学习深度相关性,导致预测精度不理想<sup>[11]</sup>.相比之下,深度学习能够克服普通神经网络的局限,学习数据中隐含的时间和空间相关性,利用多层神经网络捕捉复杂水质数据中的高层特征和模式<sup>[12]</sup>.

深度学习已被成功应用于诸多科学领域,包括计算机视觉、图像分类、语音识别、时间序列与时空

预测及自然语言处理等<sup>[13]</sup>.Wang等<sup>[14]</sup>总结了深度学习在医学图像分割中的技术分支,并基于深度神经网络提高了医学图像分析的准确性.戴蔓等<sup>[15]</sup>通过构建多元线性回归(multiple linear regression, MLR)、SVM、决策树(decision tree, DT)、RF等4种传统机器学习模型与循环神经网络(recurrent neural network, RNN)、长短期记忆网络(long short-term memory, LSTM)、门控循环单元(gated recurrent unit, GRU)的复合深度学习模型,验证了不同深度学习模型的预测性能和泛化能力均优于传统机器学习模型.许景璇等<sup>[16]</sup>基于LSTM模型预测漓江阳朔水文站的日径流量,为桂林的水资源开发与保护、防汛抗旱提供参考.

LSTM是一种流行的深度学习模型,属于RNN,特别适合分析时间序列数据<sup>[17]</sup>.Elsheikh等<sup>[18]</sup>在研究中证明了LSTM在时间序列预测中的卓越性能,LSTM实现了显著的预测准确率( $R^2=0.9976$ ),表现优于传统自回归积分滑动平均模型模型(autoregressive integrated moving average, ARIMA) ( $R^2=0.0017$ ),这一显著的性能差距证明LSTM通过其门控记忆单元捕获复杂时间依赖关系的卓越能力<sup>[19]</sup>.卷积神经网络(convolutional neural network, CNN)适合处理多维图像进行分类和回归任务,可将高维图像作为输入,提取图像数据中的复杂特征,从而极大提升神经网络的解释和预测能力<sup>[20]</sup>.但CNN主要关注数据的空间特征,对于水质数据中的时间序列信息挖掘不够充分,如果单独使用,可能无法很好地捕捉水质随时间的动态变化规律.LSTM网络因其独特的门结构,能很好地处理水质数据的时间序列特性,捕捉长期的时间依赖关系,适应水质数据中的复杂变化,有效解决梯度消失问题.因此,采用融合两种模型的预测方法,既能利用CNN强大的特征提取能力捕捉水质数据的空间特征,又能借助LSTM处理时间序列的优势,对水质的动态变化进行准确建模,从而提高预测的准确性.

巢湖作为中国第五大淡水湖,长期受到工业、农业和生活污染的多重压力,水体富营养化问题尤为严重<sup>[21]</sup>.尽管近年来实施了一系列治理措施,水质有所改善,但其演变趋势仍具有高度不确定性和空间异质性.因此,开展巢湖水质的预测研究,对于评估治理效果、预警生态风险具有重要意义.

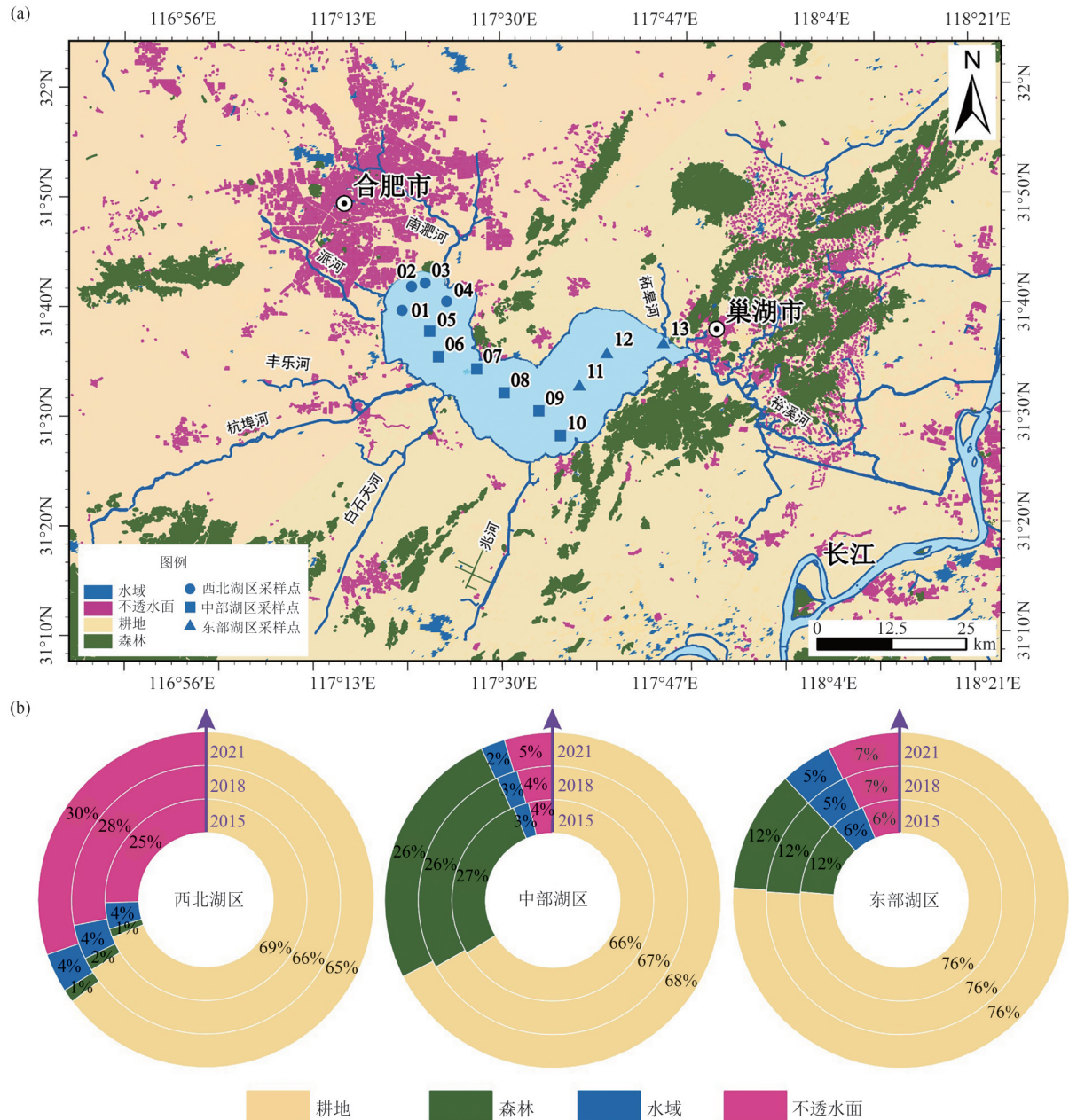
本文基于巢湖2014—2021年的水质数据,包括pH、总磷、总氮、氨氮、高锰酸盐指数、五日生化需氧量、化学需氧量、溶解氧、叶绿素 $a$ 和透明度等,构建

并比较了 CNN、LSTM、CNN-LSTM 三种模型对巢湖水质的预测性能,筛选出最优模型,预测巢湖富营养状态指数,揭示巢湖各湖区富营养化状态演变规律.

1 研究区域概况

巢湖位于安徽省(117°29'E—117°84'E, 31°42'N—31°71'N),地处长江和淮河之间,为安徽省最大湖泊、中国五大淡水湖之一<sup>[22]</sup>.湖泊整体呈 V

形,受地形影响南北沿宽窄相近的东西方向伸展[图 1(a)].湖泊面积 1.34 万 km<sup>2</sup>,平均水深 2.7 m,属大型亚热带浅水湖泊.流域年降水量 1100 mm,年流入量 48 亿 m<sup>3</sup>,年流出量 34 亿 m<sup>3</sup>,年均气温 15~16 °C<sup>[23]</sup>.在空间分布上,巢湖流域被划分为西北湖区、中部湖区和东部湖区三个区域[图 1(a)].流域土地利用类型主要由耕地、不透水面和森林组成,其中耕地占总面积 60% 以上[图 1(b)];不透水面主要集中在西北湖区排水区域内[图 1(a)],且由于城市扩张而持续增加[图 1(b)].



(a)巢湖流域位置及采样点;(b)巢湖流域土地利用情况

图1 2014—2021年巢湖流域概况图

Fig. 1 Chaohu Lake Basin overview map from 2014 to 2021

## 2 研究方法

### 2.1 数据处理

依据资料的连续性和完整性,本文选取了巢湖2014—2021年13个采样点的pH、TP、TN、 $\text{NH}_4^+\text{-H}$ 、 $\text{COD}_{\text{Mn}}$ 、 $\text{BOD}_5$ 、COD、DO、Chl-*a*和SD等数据,所用数据均由安徽省水文局提供.采样点分布如图1(a)所示,采样点01—04位于西北湖区,05—10位于中部湖区,11—13位于东部湖区.

数据异常值会增加序列的不确定性,导致模型预测结果失真,因此采用 $3\sigma$ 原则对数据进行识别和检测<sup>[24]</sup>.为确保数据的完整性和可靠性,剔除异常值并对数据中的缺失值进行插值填充.此外,为消除变量之间的量纲影响、将不同数量级的数据转换为统一尺度的统计量,采用Min-Max标准化方法进行归一化处理,归一化后的无量纲值落在 $[0,1]$ 区间内.

将选取的共8年的水质数据分为训练集和验证集:训练集为2014—2018年水质数据,验证集为2019—2021年数据.在保证训练数据充足的同时,验证集能覆盖多个季节性周期且能反映近期水质变化趋势,避免模型仅拟合早期模式.

### 2.2 Pearson相关分析

皮尔逊相关性分析是一种用于评估不同参数之间线性关系的统计方法,可减少多源数据的维度,筛选重要因素.通过计算相关系数,可以确定哪些参数之间存在显著的相关性,识别巢湖不同湖区的污染源差异.其相关系数由估计样本的协方差和标准差组成<sup>[25]</sup>,计算公式如式(1):

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

式中: $n$ 是参数指标数量; $X_i$ 是第*i*种参数数据; $Y$ 是该站点对应的TLI值; $\bar{X}$ 和 $\bar{Y}$ 分别是参数与TLI的均值; $r$ 为皮尔逊相关系数, $-1 < r < 1$ (当 $0.6 \leq |r| \leq 1$ ,为强相关;当 $0.4 \leq |r| < 0.6$ ,为中等相关;当 $0.2 \leq |r| < 0.4$ ,为弱相关;当 $0 < |r| < 0.2$ ,为极弱相关).

### 2.3 营养状态指数

营养状态指数(trophic level index, TLI)是评估湖泊和水库富营养化状态的常用方法,也是大多数水质评估中应用最广泛的方法之一<sup>[26-27]</sup>.TLI方法通过将单变量分析的简便性与多元分析的准确性相结合,基于Chl-*a*、TP、TN、SD和 $\text{COD}_{\text{Mn}}$ 五项指标进行加权计算,以反映湖泊的营养级水平.计算公式如式(2)-(8)<sup>[27]</sup>:

$$TLI(\Sigma) = \sum_{j=1}^5 W_j \times TLI(j), \quad (2)$$

$$TLI(\text{Chl} - a) = 10[2.5 + 1.086 \ln C(\text{Chl} - a)], \quad (3)$$

$$TLI(\text{TP}) = 10[9.436 + 1.624 \ln C(\text{TP})], \quad (4)$$

$$TLI(\text{TN}) = 10[5.453 + 1.694 \ln C(\text{TN})], \quad (5)$$

$$TLI(\text{SD}) = 10[5.118 - 1.94 \ln C(\text{SD})], \quad (6)$$

$$TLI(\text{COD}_{\text{Mn}}) = 10[0.109 + 2.661 \ln C(\text{COD}_{\text{Mn}})], \quad (7)$$

$$W_j = r_{(j)}^2 / \sum_{j=1}^5 r_{(j)}^2, \quad (8)$$

式中:TLI( $\Sigma$ )称为综合营养状态指数,是5个参数TLI的加权和,为无量纲值; $W_j$ 为第*j*种参数营养状态指数的相关权重;TLI(*j*)为第*j*种参数的营养状态指数,为无量纲; $C(\text{Chl}-a)$ 、 $C(\text{TP})$ 、 $C(\text{TN})$ 、 $C(\text{SD})$ 、 $C(\text{COD}_{\text{Mn}})$ 分别为叶绿素*a*、总磷、总氮、透明度、高锰酸盐指数的浓度,单位依次为 $\text{mg} \cdot \text{m}^{-3}$ 、 $\text{mg} \cdot \text{L}^{-1}$ 、 $\text{mg} \cdot \text{L}^{-1}$ 、 $\text{m}$ 、 $\text{mg} \cdot \text{L}^{-1}$ ;  $r_{(j)}$ 为参数与叶绿素*a*之间的相关系数,对于中国的主要湖泊,经验常数 $r(\text{Chl}-a)$ 、 $r(\text{TP})$ 、 $r(\text{TN})$ 、 $r(\text{SD})$ 、 $r(\text{COD}_{\text{Mn}})$ 依次取值为1、0.84、0.82、-0.83、0.83.

TLI指数取值范围从0~100,按等级对湖泊和水库的营养状态进行分类<sup>[28]</sup>,见表1.

表1 营养状态分类  
Tab. 1 Classification of trophic status

| 营养状态  | 取值范围      |
|-------|-----------|
| 贫营养   | [0, 30]   |
| 中营养   | (30, 50]  |
| 轻度富营养 | (50, 60]  |
| 中度富营养 | (60, 70]  |
| 重度富营养 | (70, 100] |

### 2.4 模型架构与评估

#### 2.4.1 LSTM模型

LSTM是一种改进的递归神经网络(recurrent neural network, RNN),利用门控机制调节信息流入和流出记忆单元,从而解决梯度消失问题.LSTM引入了记忆单元及遗忘门、输入门、输出门等门控结构,通过选择性更新机制,利用sigmoid激活函数和逐元素乘法控制信息流动<sup>[29]</sup>.LSTM的网络呈链式结构(图2).

记忆单元与门控机制的工作过程主要包括以下三个步骤:

遗忘门( $f_t$ )先对前一时刻的输出 $h_{t-1}$ 和当前时刻*t*下的输入值 $x_t$ 进行处理,控制上一单元信息的选择性遗忘[式(9)].

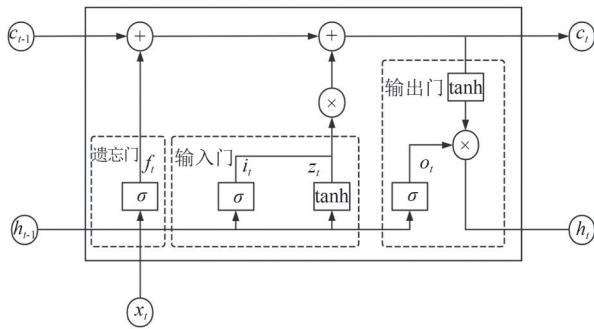


图2 LSTM网络结构图

Fig. 2 LSTM network structure diagram

输入门( $i_t$ )决定当前时刻哪些信息应该被存储至记忆单元. 首先使用 sigmoid 激活函数处理数据, 控制信息的写入[式(10)], 再通过 tanh 函数指示被允许存储信息的重要程度[式(11)].

输出门( $o_t$ )决定当前记忆单元  $c_t$  输出到  $h_t$  的信息, 即下一时刻的隐藏状态. 首先使用 sigmoid 激活函数筛选达到输出标准的记忆单元[式(12)], 再使用 tanh 函数计算各记忆单元的权重, 两者的输出相乘得到新时刻的隐藏状态[式(13)]. 新的记忆单元和隐藏单元信息会被带入下一时刻循环处理. 所需计算公式如式(9)-(14)所示<sup>[30]</sup>:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \quad (9)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \quad (10)$$

$$g_t = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g), \quad (11)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \quad (12)$$

$$H_t = o_t \tanh(c_t), \quad (13)$$

$$\sigma = \frac{1}{1 + e^x}, \quad (14)$$

式中:  $f_t, i_t, g_t, o_t, h_t$  分别表示遗忘门、输入门、中间输出、输出门和记忆单元的状态;  $W_{xf}, W_{hf}, W_{xi}, W_{hi}, W_{xg}, W_{hg}, W_{xo}, W_{ho}$  分别是对应门的矩阵权重乘以输入  $x_t$  和输出  $h_{t-1}$ ;  $b_f, b_i, b_g, b_o$  是对应门的偏置项.

#### 2.4.2 CNN 模型

CNN 是一种在图像和时序数据处理中非常有效的深度学习模型, 主要用于处理具有空间相关性的数据<sup>[31]</sup>. CNN 由卷积层、池化层、激活层和全连接层组成, 结构如图 3 所示. CNN 的模型架构中, 卷积层通过滑动窗口机制实现局部特征捕获, 池化层执行特征降维以抑制过拟合, 而全连接层则完成全局特征分类或回归任务<sup>[32]</sup>.

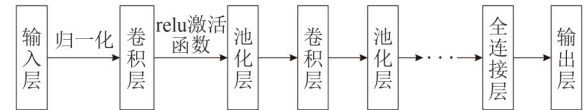


图3 CNN网络结构图

Fig. 3 CNN network structure diagram

CNN 的卷积操作通过以下公式描述<sup>[33]</sup>:

$$Y = X * W + b, \quad (15)$$

式中:  $X$  是输入特征数据;  $W$  是卷积核, 也叫滤波器;  $*$  是卷积操作;  $b$  为偏置项;  $Y$  是输出特征数据.

#### 2.4.3 CNN-LSTM 模型

CNN-LSTM 是一种深度学习算法, 它将 CNN 和 LSTM 相结合(图 4). 通过 CNN 能够提取时间序列数据中的空间特征, 而 LSTM 则能够捕捉时间序列数据中的长期依赖关系<sup>[34]</sup>.

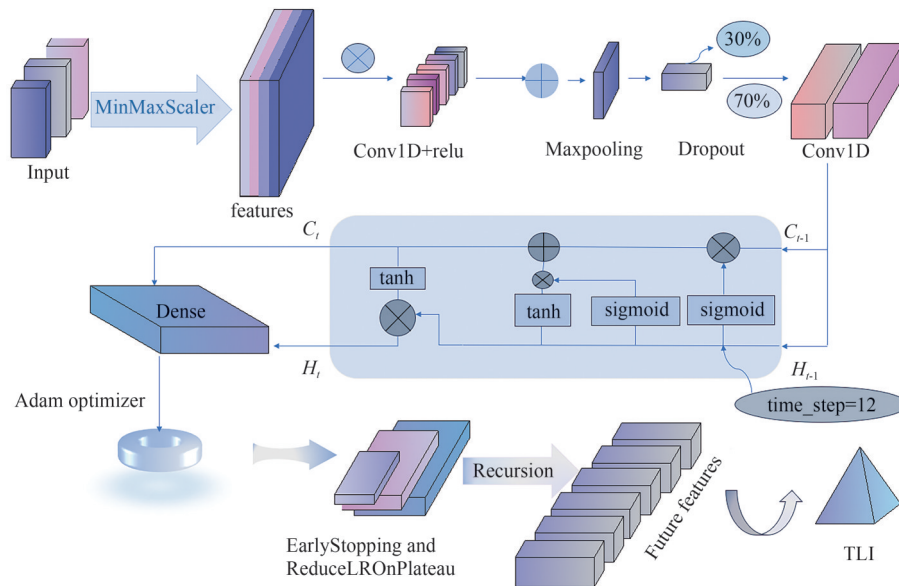


图4 CNN-LSTM网络结构图

Fig. 4 CNN-LSTM network structure diagram

在CNN模块中,一维卷积核通过时间窗口滑动,自动提取不同水质参数之间的局部空间特征.CNN模块在西北湖区优先提取TN与TP的波动模式,以及BOD<sub>5</sub>与COD的协同变化特征;中部湖区侧重Chl-*a*与TP、COD<sub>mn</sub>的组合,呈现季节性藻类增殖特征;东部湖区的特征响应则以Chl-*a*与BOD<sub>5</sub>为主,体现藻类生物量与有机分解过程的动态平衡.LSTM通过门控机制选择性地记忆或遗忘历史信息,捕捉的时间依赖关系包括:上月水质对下月的影响,冬春季节营养物质累积对夏季藻类爆发的影响,以及年际尺度上污染物质的累积与消减变化.

在CNN-LSTM模型的预处理阶段,CNN模块对选取的特征变量采用Min-Max标准化以消除不同水质参数的量纲差异,以12个月为一个时间步长.构建特征预测模型参数:层数为2,第一层卷积核64个,第二层卷积核128个,LSTM神经元分别为128个和64个,Adam学习率0.001.之后对预测的特征数据逆标准化,得到原始量纲数据.TLI指数预测与特征预测模型相同,通过相同工作机理得出.

#### 2.4.4 模型评估

为了比较三种模型的预测精度,采用平均绝对误差(MAE)、平均相对误差(MRE)和确定性系数( $R^2$ )衡量模型预测值和实际值的偏差程度.绝对误差和相对误差越接近0,预测误差越小; $R^2$ 越接近1,说明预测结果与实测值的拟合度越高.计算公式如式(16)-(18):

$$MAE = \frac{1}{N} |y_{ci} - y_{oi}|, \quad (16)$$

$$MRE = \frac{|y_{ci} - y_{oi}|}{y_{oi}}, \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{ci} - y_{oi})^2}{\sum_{i=1}^n (y_{oi} - \bar{y}_o)^2}, \quad (18)$$

式中: $y_{ci}$ 是第*i*个样本的预测值; $y_{oi}$ 是第*i*个样本的实际值; $\bar{y}_o$ 是实际值样本的均值; $n$ 是一种样本总数.

### 3 水质现状

本文采用克里金法对13个采样点的TLI值进行空间插值,以生成巢湖全湖区的富营养化状态空间分布图.在本文中,13个采样点呈“西北密集、东部稀疏”的分布特征:西北湖区布设4个采样点,中部湖区6个,东部湖区仅3个.克里金法能够有效利用空间自相关权重,合理预测东部湖区等采样稀疏区

域的TLI值.在采样点分布不均匀的情况下,相较于反距离权重法等确定性插值方法,其插值结果具有更高的统计无偏性和最小估计方差.

图5显示了2014—2021年巢湖富营养状态指数(TLI)的空间分布以及不同湖区、不同时期的TLI指数.从空间分布来看,历年的TLI指数大致为西北湖区>中部湖区>东部湖区.西北湖区年度TLI范围为56.07~69.34,富营养状态介于轻度富营养至中度富营养之间;中部湖区各年TLI在53.33~59.54之间,富营养状态为中度富营养;东部湖区各年TLI在51.69~57.72之间,富营养状态为中度富营养.从年际变化来看,西北湖区TLI在2014—2017年变化不显著,从2018年开始至2020年逐年下降,由中度富营养向轻度富营养乃至贫营养状态转变.中、东部湖区TLI值从2014年到2020年先升高后降低,峰值出现在2018年,富营养状态经历了从贫营养到轻度富营养再到贫营养的转变.2021年,三个湖区的TLI均有不同程度的再次升高,富营养水平均提升一个等级.

从空间分异机制来看,尽管三个湖区呈现稳定的营养状态梯度,但其内部动态响应机制却存在显著差异.西北湖区作为全湖最高营养负荷区,其TLI值长期高于60,在2014—2017年间虽波动不大,但始终维持在中度富营养化水平,反映出该区域是污染物输入的关键节点.西北湖区临近合肥市以及周边的城镇建成区,不透水面比例系数大,产生的工业和生活废水通过南淝河、派河等入湖河流集中输入,形成了较高浓度的污染模式[图1(a)].相比之下,中、东部湖区的汇水范围大部分为耕地[图1(b)],污染输入以农业化肥施用产生的氮磷流失为主,来自工业污染的部分较少.此外,巢湖的水文流向整体自西向东流,西北湖区的污染物会随水流向东迁移扩散,在这个过程中伴随着稀释、沉降、生物转化等自净作用,使得污染物浓度沿流向逐渐降低.

### 4 水质预测结果

#### 4.1 水质指标相关性分析

采用皮尔逊相关系数法表示变量之间的相关性(图6),根据计算结果,各项水质指标与TLI指数之间的相关性如下:TP、TN、COD<sub>mn</sub>、BOD<sub>5</sub>、COD、Chl-*a*与TLI呈较强正相关关系,相关系数依次为0.67、0.54、0.69、0.6、0.66及0.49.因此,选取这些指标作为特征变量.

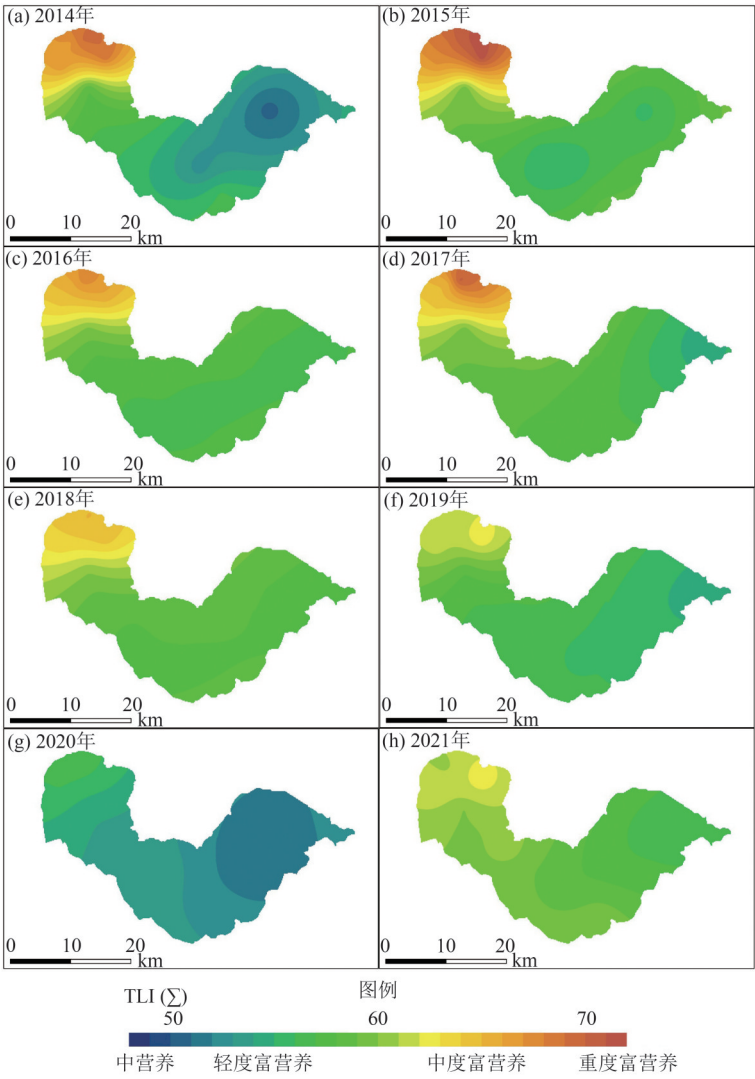


图5 基于TLI的巢湖富营养化年度分析

Fig. 5 Annual analysis of eutrophication in Chaohu Lake based on TLI

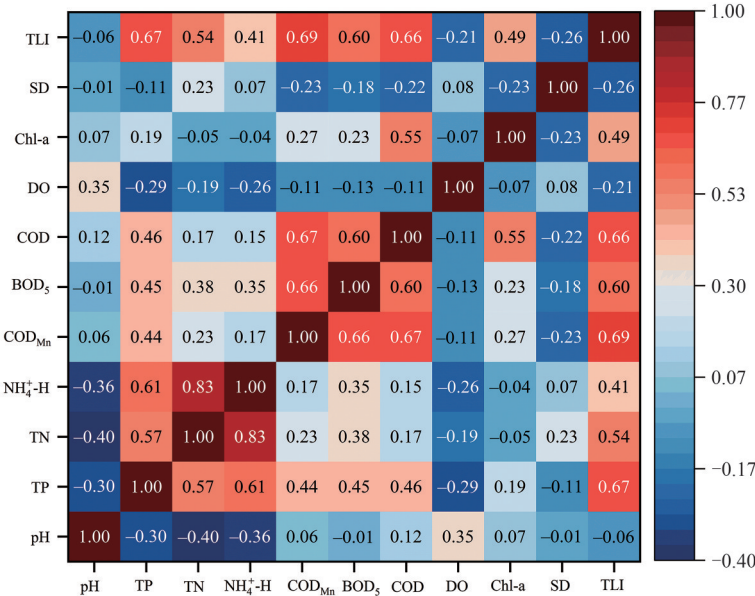


图6 Pearson相关性热图

Fig. 6 Pearson correlation heatmap

不同指标与TLI的相关性存在差异,反映出各指标在富营养化进程中的影响不同.TP、COD<sub>Mn</sub>的相关系数最高,表明磷营养物质和有机物是驱动巢湖富营养化状态变化的主导因素.其中,TP的浓度水平直接调控藻类的生物量<sup>[35]</sup>;COD<sub>Mn</sub>表征水体中的有机物总量,反映了外源有机污染物的输入以及藻类代谢产生的有机质累积对水体富营养化的影响.COD和BOD<sub>5</sub>同样呈现较强的正相关,二者均为有机污染程度的综合指标,进一步支撑了有机污染负荷与富营养化之间的耦合关系.TN属于中等强度正相关,尽管氮也是藻类生长的重要营养元素,但在巢湖这类浅水湖泊中,非固氮藻类对氮的响应可能受到磷限制的调节,因而其与TLI的相关性略低于TP.Chl-*a*作为浮游植物生物量的直接度量指标,其相关系数虽相对较低,但也属于中等相关,可能与该指标的季节性波动有关.

4.2 误差分析

根据计算巢湖各采样点(01-13)基于CNN、LSTM与CNN-LSTM混合模型的预测误差数据,不同模型的预测性能存在明显空间异质性.总体而言,CNN-LSTM混合模型在多数站点上表现出更优的预测精度与稳定性(表2).

图7展示了不同模型在各个站点验证集的预测误差.在绝对误差方面,CNN模型在不同站点的误

表2 各模型预测误差表  
Tab. 2 Model prediction error

| 模型          | 平均绝对误差 | 平均相对误差 | 确定性系数  |
|-------------|--------|--------|--------|
| CNN 模型      | 0.9436 | 1.6811 | 0.7564 |
| LSTM 模型     | 0.7965 | 1.4190 | 0.7765 |
| CNN-LSTM 模型 | 0.7069 | 1.2618 | 0.8054 |

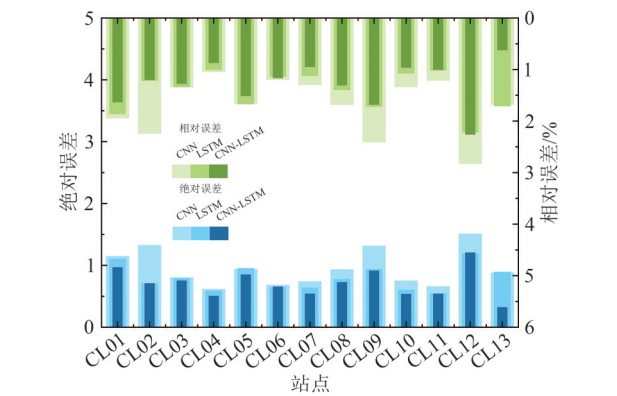


图7 各模型验证集误差  
Fig. 7 Validation errors for each model

差最高,表明其较难捕捉复杂的时空变化;LSTM模型的绝对误差较CNN有所降低;CNN-LSTM模型在大多数站点的绝对误差低于或接近前两者.在相对误差方面,CNN-LSTM模型仍较两种单一模型的误差更小,稳定性更高.

4.3 预测结果

图8采用相同的克里金插值参数设置,展示了

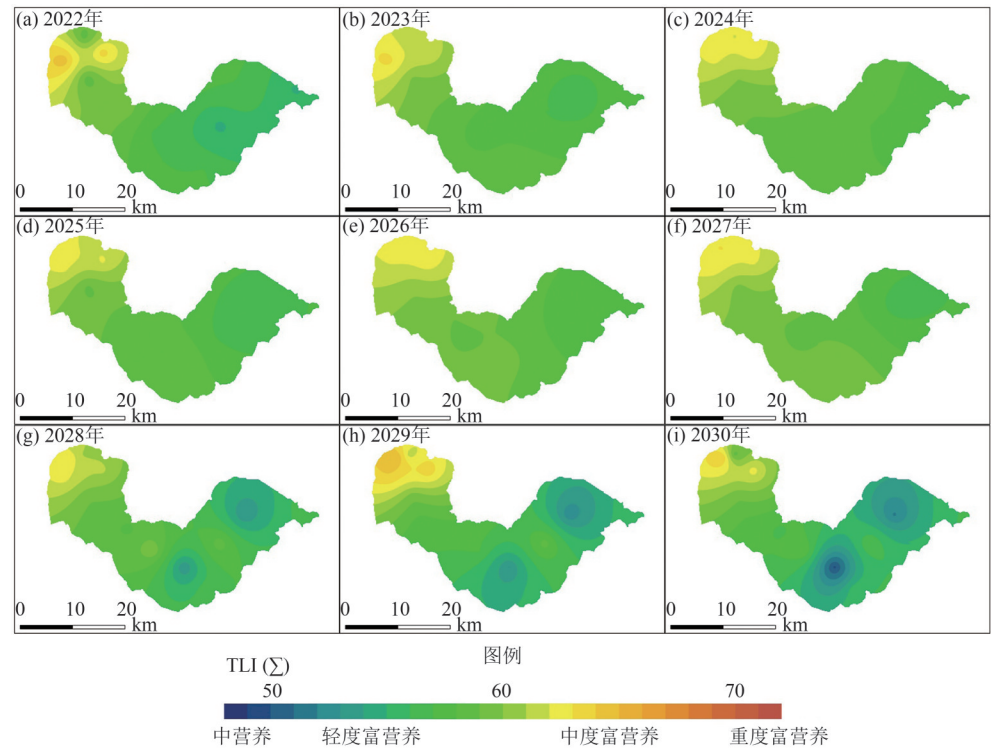


图8 2022—2030年巢湖富营养化预测分析  
Fig. 8 Eutrophication prediction analysis of Chaohu Lake from 2022 to 2030

基于CNN-LSTM模型预测的2022—2030年巢湖富营养状态的结果,直观呈现了巢湖富营养化水平的时空演变规律.从空间格局看,巢湖整体呈现“西高东低”的梯度分布:西北湖区TLI值长期处于60~70区间,属于中度富营养化;中、东部湖区TLI值集中于50~60,整体处于轻度富营养化水平.不同年份间的空间梯度保持稳定,反映出污染物输入、水文条件等对富营养化状况的持续影响.

时间维度上,富营养化过程表现出阶段性变化特征.2022—2024年间,巢湖整体水质富营养状态相对稳定,各湖区的富营养等级波动较小;2025—2027年,西北湖区中度富营养化范围略有扩张,TLI较高值区向中部湖区迁移,而东部湖区仍保持相对稳定;2028—2030年,预测结果显示巢湖富营养化程度有所减缓,中、东部湖区基本维持轻度富营养化状态,西北部湖区仍是中度富营养化,但其TLI值先增大后减小,反映出富营养化压力在达到阶段性峰值后有所缓解.

整体上,预测年份的水质状态较2014—2021年有所好转,污染物浓度相对降低,TLI值减小.位于西北湖区边缘、靠近中部湖区过渡带的采样点04及其周边区域,以及受入湖污染影响相对较轻的采样点01所在子区域,在预测期内TLI值呈现下降趋势,逐渐从中度富营养化向轻度富营养化转变.相较之下,位于主要入湖河口附近的采样点02和03区域改善幅度较小,仍长期维持中度富营养化状态.这一内部差异表明,西北湖区的富营养化改善趋势并非空间均一,而是优先出现在远离主要污染输入源、水动力交换条件相对较好的局部区域.

## 5 结论

本文以巢湖2014—2021年的多维度水质数据为基础,通过数据预处理、多模型构建及对比分析,系统探讨了深度学习模型在水质预测中的应用潜力以及巢湖的水质演变特征,主要结论如下:

(1)通过对2014—2021年水质时空演变规律的分析,巢湖不同湖区的TLI呈现西北湖区>中部湖区>东部湖区的空间格局.西北湖区因受城市工业、生活污染以及水文流向的影响,长期处于中度富营养水平,而中、东部湖区主要受农业面源污染影响,富营养化程度相对较轻但改善缓慢.

(2)CNN-LSTM模型在预测精度与稳定性上优

于单一模型,其误差最低,确定性系数最高( $R^2=0.8054$ ),表明该模型能够有效耦合水质时空演变特征,更适用于巢湖这类受多重因素驱动、具有空间异质性湖泊的水质预测.

(3)基于CNN-LSTM模型的预测显示,2022—2030年巢湖整体水质状态呈现先稳后缓的趋势.西北湖区仍处于中度富营养化状态,其部分地区有向轻度富营养化转变的趋势;中、东部湖区基本维持轻度富营养化状态,与历史时期相比呈改善趋势,表明流域治理措施可能发挥了积极作用.

## 参 考 文 献

- [1] Pi X, Luo Q, Feng L, et al. Mapping global lake dynamics reveals the emerging roles of small lakes [J]. *Nature Communications*, 2022, 13: 5777.
- [2] Huang A, Liu X, Dong F, et al. A holistic assessment of water quality in the lake and rivers of Lake Chaohu Basin, China [J]. *Processes*, 2025, 13(10): 3125.
- [3] Zheng Y, Zhang Q, Zhang X, et al. A spatial-temporal trend-aware neural network model for accurate water quality prediction in river [J]. *Water Research*, 2025, 287: 124389.
- [4] Zhao Z, Fan B, Zhou Y, et al. An effective data-driven water quality modeling and water quality risk assessment method [J]. *Engineering Applications of Artificial Intelligence*, 2024, 138: 109457.
- [5] 刘东征, 张耀琨, 赵德霜, 等. 湖库水质评价方法研究进展 [J]. *环境生态学*, 2025, 7(9): 7-15, 24.
- [6] 郭金萍. 小流域水环境评价及影响因素研究 [J]. *生态与资源*, 2025(10): 34-36.
- [7] Wu J, Cheng S P, He L Y, et al. Assessing water quality in the Pearl River for the last decade based on clustering: Characteristic, evolution and policy implications [J]. *Water Research*, 2023, 244: 120492.
- [8] Ding F, Hao S, Jiang M, et al. An improved graph neural network integrating indicator attention and spatio-temporal correlation for dissolved oxygen prediction [J]. *Ecological Informatics*, 2025, 87: 103126.
- [9] Jiang S, Cheng X, Shi B, et al. Optimal selection of machine learning algorithms for ciprofloxacin prediction based on conventional water quality indicators [J]. *Ecotoxicology and Environmental Safety*, 2025, 289: 117628.
- [10] 张强英, 陈家乐, 曹阳, 等. 机器学习模型在水质评价与预测中的应用研究进展简析 [J]. *生态与资源*, 2025(6): 1-4.
- [11] Luo W, Huang L, Shu J, et al. Predicting water quality

- in municipal water management systems using a hybrid deep learning model [J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108420.
- [12] Mallick J, Alqadhi S, Alsubih M, et al. Automated interpretation of deep learning-based water quality assessment system for enhanced environmental management decisions[J]. *Applied Water Science*, 2025, 15(5): 110.
- [13] Barzegar R, Aalami M T, Adamowski J. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model[J]. *Stochastic Environmental Research and Risk Assessment*, 2020, 34(2): 415-433.
- [14] Wang R, Lei T, Cui R, et al. Medical image segmentation using deep learning: A survey[J]. *IET Image Processing*, 2022, 16(5): 1243-1267.
- [15] 戴蔓, 胡丹心, 李名升, 等. 基于深度学习的地下水水质预测方法研究——以大湾区(广州)地下水多层监测基地氨氮为例[J]. *中国环境监测*, 2025, 41(1): 191-200.
- [16] 许景璇, 李勇军, 罗运辉, 等. 基于LSTM的漓江阳朔水文站日径流量预测研究[J]. *广西水利水电*, 2026(1): 56-60.
- [17] Gao Z, Chen J, Wang G, et al. A novel multivariate time series prediction of crucial water quality parameters with Long Short-Term Memory (LSTM) networks [J]. *Journal of Contaminant Hydrology*, 2023, 259: 104262.
- [18] Elsheikh A H, Katekar V P, Muskens O L, et al. Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate [J]. *Process Safety and Environmental Protection*, 2021, 148: 273-282.
- [19] Nong X, He Y, Chen L, et al. Machine learning-based evolution of water quality prediction model: An integrated robust framework for comparative application on periodic return and jitter data[J]. *Environmental Pollution*, 2025, 369: 125834.
- [20] Pyo J, Park L J, Pachepsky Y, et al. Using convolutional neural network for predicting cyanobacteria concentrations in river water[J]. *Water Research*, 2020, 186: 116349.
- [21] 卜松, 周戎星, 陈倩倩. 集对分析法在巢湖水质富营养化时空变化中的应用[J]. *黑龙江工业学院学报(综合版)*, 2018, 18(8): 153-156.
- [22] 刘阳, 张劲松, 王欢, 等. “十三五”期间巢湖水质变化及影响因素[J]. *中国环境监测*, 2025, 41(3): 70-82.
- [23] 严圣希, 姚亦鹏, 李典鹏, 等. 巢湖湖区水质状态对浮游植物多样性时空格局的影响[J]. *湿地科学与管理*, 2024, 20(4): 50-54.
- [24] Peng W, Tao D, Ma Q, et al. Clipped seismic record recovery analysis based on the cubic spline interpolation algorithm [J]. *Journal of Seismology*, 2024, 28(3): 843-857.
- [25] Su X, Shang S, Xu Z, et al. Assessment of dependent performance shaping factors in SPAR-H based on Pearson correlation coefficient [J]. *Computer Modeling in Engineering & Sciences*, 2024, 138(2): 1813-1826.
- [26] Wang J, Fu Z, Qiao H, et al. Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China[J]. *Science of the Total Environment*, 2019, 650: 1392-1402.
- [27] Woelmer W, Gregersen R, Özkundakci D. Lake trophic state indicator shows non-stationary relationships with environmental drivers [J]. *Ecological Indicators*, 2025, 180: 114347.
- [28] Zeng Y, Wang Z, Zhao Q, et al. Assessment of the effects and contributions of natural and human factors on the nutrient status of typical lakes and reservoirs in the Yangtze River Basin[J]. *Water*, 2025, 17(4): 559.
- [29] 龚畅, 马悦, 程硕. 基于LSTM-Adam算法的某水库洪水模拟分析[J]. *水利科技与经济*, 2026, 32(2): 5-10.
- [30] 刘桦楠. 基于LSTM和BP神经网络的湖泊水质变化趋势分析与预测研究[D]. 武汉: 华中科技大学, 2021.
- [31] 高小平, 朱宏博, 赵凯, 等. 基于卷积神经网络-长短期记忆人工神经网络-自注意力机制的岩性识别方法[J]. *科学技术与工程*, 2026, 26(2): 544-552.
- [32] 刘永胜, 沈军宏, 李达, 等. 基于CNN-Transformer-ARG的双护盾TBM掘进速度预测模型[J]. *河海大学学报(自然科学版)*, 2026, 54(1): 112-118, 176.
- [33] 卓小龙, 邵继. 基于贝叶斯优化CNN-BiLSTM模型的渭河流域水质预测方法研究[J]. *水利技术监督*, 2025(11): 248-253.
- [34] 杨璐霞, 王梦冉, 林兴亮, 等. 基于CNN-LSTM的工业出水水质预测模型[J]. *净水技术*, 2024, 43(5): 143-149.
- [35] 孙云瀚, 余俊爽, 王松波, 等. 武汉北湖沉积物磷含量和分布特征[J]. *中南民族大学学报(自然科学版)*, 2026, 45(1): 24-31.

(责编&校对 刘钊)