融合卷积神经网络的混凝投药模型研究*

李泽楷，章　杰
（福州大学　物理与信息工程学院　微纳器件与太阳能电池研究所，福建　福州　350108）

摘　要：以东南某百万人口城市水厂为对象，针对传统水质监测效率低及混凝剂投加量预判困难的问题，提出基于卷积神经网络（CNN）的混凝剂预测模型。通过数据预处理提升数据质量后，采用信息增益比率筛选出关键特征，构建包含一维卷积层、池化层和全连接层的 CNN 模型，采用 ReLU 激活函数优化特征表达能力。实验显示模型预测结果的 RMSE 为 68.550，MAE 为 50.709，拟合优度达 0.926，较传统方法显著提升。

关键词：水质监测；混凝剂预测；卷积神经网络
中图分类号：TP391　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097－1788.2025.05.005
引用格式：李泽楷，章杰. 融合卷积神经网络的混凝投药模型研究［J］. 网络安全与数据治理，2025，44(5)：29－34.

Research on a coagulation dosing model incorporating convolutional neural networks*

Li Zekai，Zhang Jie
（Institute of Micro-Nano Devices and Solar Cells，College of Physics and Information Engineering，Fuzhou University，Fuzhou 350108，China）

Abstract：This study focuses on a water treatment plant in a southeastern Chinese city with a population of one million. To address the inefficiency of traditional water quality monitoring and the difficulty in pre-determining coagulant dosage，we proposed a prediction model based on convolutional neural networks（CNN）. After enhancing data quality through preprocessing，key features were selected using the information gain ratio. A CNN architecture incorporating 1D convolutional layers，pooling layers，and fully connected layers was developed，with ReLU activation functions optimizing feature representation. Experimental results demonstrated superior performance of the model，achieving a root mean square error（RMSE）of 68.550，mean absolute error（MAE）of 50.709，and goodness-of-fit of 0.926.

Key words：water quality monitoring；coagulant prediction；convolutional neural network

0　引言

随着我国城市化建设步伐的推进和城市经济的快速发展，水污染问题日益突出[1]。水处理过程，由于受很多因素影响，具有高度复杂性、不确定性以及非线性，水质净化的难度也是数倍增加。

目前，污水处理过程中存在水质监测时效性低、出水水质超标、运行能耗过高、多种净水剂投放量无法预测等诸多问题，具有实时监测和净水剂投放量预测功能的监测站点作为改善污水处理运行效果和提高运行效率的关键，已成为污水处理厂的重要选择。由于原水水质指标和进水流量之间是非线性关系，并具有一定的耦合性，同时还会受到诸如季节等因素的影响而变化，净水剂中的混凝剂投放量的控制尤为困难。混凝工艺作为水处理系统的核心环节，主要通过化学作用去除水体悬浮物[2]。其处理效能直接决定着后续工艺的稳定性，因此需要根据原水水质的周期性波动、季节性差异及动态变化进行实时调控。然而传统控制策略因水质参数的强时变特性，往往难以实现精准的药剂投加量调节，导致混凝剂投放精度难以保障。我国供水企业在混凝剂智能控制领域的技术发展相对滞后，长期以来主要依赖人工经验进行药剂投加决策[3－4]。

由于原水水质与处理效果间存在复杂的非线性关系，常规数学建模方法难以构建精准预测模型。在此背景下，基于深度学习的智能控制技术展现出独特优势。该技术

*基金项目：福建省级科技创新重点项目（2022G02011）

依托海量运行数据，通过自主特征提取和模式识别，无需预设固定模型结构即可完成动态优化，已在多个工业场景中验证了其适应复杂系统控制的准确性与可靠性。近年来，水质预测与优化研究通过深度学习和机器学习技术的创新应用不断突破。Im 等人基于韩国 33 家净水厂五年高分辨率时序数据，构建覆盖全国供水系统的深度学习模型，其平均预测准确率达 98.78%，最大预测准确率接近 99.98%[5]。Cai 等人提出的 TWQ-TPN 网络通过时序特征提取与长期波动建模，在 pH、浊度和余氯预测中实现行业领先性能，并通过消融实验验证了模型设计的有效性[6]。Torky 等人开发的混合机器学习框架在饮用水安全分类任务中达到 94.7% 平均准确率，其中随机森林和光梯度提升机模型表现最佳，而水质指数预测任务中轻量梯度提升回归模型以 0.99 测试准确率和低误差率显著优于传统方法[7]。Saroja 等人采用 LSTM 和 CNN 分别构建水质指数预测与分类系统，其中 LSTM 模型以 97% 准确率实现水质指数精准预测，CNN 分类器则将错误率降至 0.02[8]。Sv 等人设计的 CNN-ELM 混合异常检测模型通过 0.92 的 $F1$ 分数显著提升传感器数据可靠性[9]。Mousavi 等人通过小波去噪与 ANFIS 融合建模，使浊度预测精度提升 12%[10]。Trejo-Zuniga 等人利用 CNN 突破传统浊度测量局限，在实验室和实际水体分类中分别取得 97% 与 85% 准确率，实现理论与实践的平衡[11]。Zhu 等人开创的絮凝张量图深度学习系统以 98% 分类准确率实现污染物快速识别，将絮凝过程反馈延迟缩短至实时水平[12]。针对原水水质动态变化对混凝效果的影响，深度学习技术通过构建动态关联模型，能够有效解析水质指标与药剂投量的非线性关系。该方法利用海量水质监测数据，结合大数据挖掘技术，突破传统固定建模的局限性，实现水质波动下的精准预测。

本研究基于深度学习算法，建立水质特征与投药量的自适应映射机制，通过实时解析浊度、pH 值等关键参数的变化规律，动态生成最优投加策略。这种自主优化机制不仅降低了水质时变性对混凝工艺的干扰，同时依托多维数据融合分析，显著提升了混凝剂调控的时效性与精准度。

1　基于 CNN 的混凝剂投放量预测模型

本研究在东南某百万级人口城市的水厂实施，监测期为 2021 年 1 月至 2023 年 12 月。监测指标如下：原水流量、原水浊度、pH 值、耗氧量、色度、菌落总数、氨氮、温度、投放混凝剂沉淀后原水的浊度及当天的混凝剂投放量。在监测期内共收集了 911 条数据。所收集的数据信息如表 1 所示。这些数据被预处理并提取特征后将用于模型的训练和构建。

表 1　水质数据统计分析表

水质参数	最大值	最小值	平均值
原水流量	199 450.00	6 224.00	104 740.58
原水浊度	348.80	2.40	19.43
pH 值	7.43	6.51	6.92
耗氧量	3.68	1.26	2.11
色度	35	10	12.10
菌落总数	4 000	200	943.39
氨氮	0.098	0.025	0.038
温度	30.6	11.0	21.49

1.1　CNN 基本理论

CNN 是深度学习领域中的标志性算法之一，借助权值共享与局部连接的策略，显著降低了模型参数的数量，进而提升计算效率，并且能够高效地从数据中提取特征。CNN 的核心网络架构涵盖输入层、卷积层、展平层、全连接层以及输出层等部分。卷积层主要负责对输入数据实施特征提取，其内部配置了多个卷积核，本文采用的是 ReLU 激活函数。由于全连接层的需求是输入数据为一维形式，因此展平层会将卷积层输出的所有特征拉伸成一个单一的长向量，向量中的每个元素对应一个特征或特征组合。全连接层在整个 CNN 结构中发挥着输出预测的关键作用，通常被设置在模型的最后几层，并且能够将卷积层学习到的分布式特征表示映射到样本空间中。卷积的计算公式如下：

$$\boldsymbol{T}(i,j)=\sum_{k=1}^{n}(\boldsymbol{X}_k\boldsymbol{W}_k)(i,j) \tag{1}$$

其中，$\boldsymbol{T}$ 为卷积后的特征序列，n 为输入数据 $\boldsymbol{X}_k$ 的长度，$\boldsymbol{W}_k$ 为卷积的核函数序列。

在 CNN 训练过程中，通过持续优化损失函数，使损失值不断减小，直到达到稳定值或基本稳定状态，这时网络训练即告完成。在此过程中，采用自适应矩估计（Adam）算法来优化损失函数，该算法运用不同的学习率参数以改进网络训练效果，并且能够依据优化进程自动调整适应正在优化的损失函数。上述过程可表述为：

$$\boldsymbol{\theta}_{n+1}=\boldsymbol{\theta}_n-\frac{\alpha\boldsymbol{m}_n}{\sqrt{\boldsymbol{v}_n}+\varepsilon} \tag{2}$$

$$\boldsymbol{m}_n=\beta_1\boldsymbol{m}_{n-1}+(1-\beta_1)\nabla J(\boldsymbol{\theta}_n) \tag{3}$$

$$\boldsymbol{v}_n=\beta_2\boldsymbol{v}_{n-1}+(1-\beta_2)[\nabla J(\boldsymbol{\theta}_n)]^2 \tag{4}$$

其中，n 为迭代次数；α 为学习率；$\boldsymbol{\theta}$ 为参数向量；$J(\boldsymbol{\theta}_n)$为损失函数；$\boldsymbol{m}_n$、$\boldsymbol{v}_n$ 分别为梯度的一阶矩估计和二阶矩估计；ε 为防止分母为 0 的安全系数；β_1、β_2 为移

动平均线的衰减率。

1.2 数据预处理

水质监测站点由于传感器故障或操作失误可能会导致数据缺失，数据集中共有 31 条数据出现缺失。为了提高预测准确性并为预测模型提供干净、准确的数据，需预先对实验数据进行处理。本研究利用线性插值法补齐缺失值，并对数据进行划分后做归一化处理。

首先是划分数据，将 728 组历史数据作为训练集用于模型参数优化，剩余 183 组作为独立测试集执行最终效能验证。为消除各参数量纲差异并加速梯度收敛，对 pH 值、浊度等输入特征实施极差标准化处理，通过“最大最小标准化”将原始数据线性映射至［0，1］区间。归一化的处理公式如下：

$$X_n = \frac{X - X_m}{X_M - X_m} \tag{5}$$

其中，X 为真实数据，X_n 为归一化数据，X_M 为最大值，X_m 为最小值。

在模型训练结束后，采用式（6）对数据进行反归一化，其公式如下：

$$X = X_n\ (X_M - X_m) + X_m \tag{6}$$

本研究采用均方根误差（Root Mean Square Error，RMSE）、平均绝对误差（Mean Absolute Error，MAE）以及拟合优度 R^2 作为评价标准对模型的预测结果进行误差评估。R^2 是统计学中用于衡量回归模型拟合优度的常用指标之一[13]，它表示因变量的变异中可被模型解释的部分的比例。MAE 是预测值与真实值之间绝对误差的平均值。当模型的 MAE、RMSE 值越低，拟合优度越趋近于 1 时，模型越精确[14]。具体公式如下：

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N} |y_t - \hat{y}_t| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2} \tag{8}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2} \tag{9}$$

其中，$\hat{y}_t$ 为预测值，y_t 为真实值，$\bar{y}$ 为真实值的平均值。

1.3 特征选取

特征选择通过筛选关键变量并剔除冗余及无关参数实现数据降维。该过程可减少计算负荷，避免噪声干扰，同时通过保留强关联性指标降低过拟合风险，从而提升模型训练效率和预测精度。本研究使用信息增益比率作为特征选择的依据，从八个特征中选取高相关性特征作为模型输入。信息增益比率是评估特征选择的一种指标，旨在克服信息增益对“具有多个取值”的特征有偏好的缺陷[15]。计算信息增益比率需要先计算各特征信息熵。信息熵的计算如下：

$$H_{(X)} = -\sum_{k=1}^{K}\frac{|N_K|}{|X|}\log_2\frac{|N_K|}{|X|} \tag{10}$$

其中，$H_{(X)}$ 为信息熵，X 为水质数据集，K 为数据的取值类别，$|N_K|/|X|$ 为每个类别的概率，$|N_K|$ 为类别区的样本个数，$|X|$ 为样本总数。

假设按照特征 A 对水质数据集 X 进行划分，则特征 A 的信息熵的计算如下：

$$H_{A(X)} = \sum_{i=1}^{n}\frac{|X_i|}{|X|}H_{(X_i)} \tag{11}$$

其中，$H_{A(X)}$ 为特征 A 的信息熵，n 为水质特征 A 的取值。

定义水质数据集特征在某时刻前后信息上的差值为信息增益，其计算如下：

$$I_{G_{(A)}} = H_{(X)} - H_{A(X)} \tag{12}$$

其中，$I_{G_{(A)}}$ 为信息增益。

本研究采用信息增益比指标修正传统信息增益对多值特征的倾向性偏差。该指标在信息增益计算中引入惩罚因子，其数值与特征离散程度呈负相关。通过将信息增益与惩罚因子进行加权处理，可有效平衡特征分类能力与取值分布的关联性，从而消除因特征取值数量差异导致的评估失真问题[16]。惩罚参数和信息增益比率的计算公式如下：

$$P_{I_{A(X)}} = -\frac{1}{\sum_{i=1}^{n}\frac{|X_i|}{|X|}\log_2\left(\frac{|X_i|}{|X|}\right)} \tag{13}$$

$$I_{\mathrm{GR}_{(A)}} = \frac{I_{G_{(A)}}}{P_{I_{A(X)}}} \tag{14}$$

其中，$P_{I_{A(X)}}$ 为惩罚参数，$I_{\mathrm{GR}_{(A)}}$ 为信息增益比率。

基于式（14）对水质数据集各特征的信息增益比率进行量化评估，通过降序确定特征权重排序。本研究筛选出信息增益比率最高的 4 项指标作为预测模型的核心输入变量，分别是流量、浊度、色度和氨氮。水质的 8 项指标特征信息如表 2 所示。

表 2 水质数据各特征信息增益比率

特征	信息增益比率
流量	0.477
浊度	0.305
pH	0.088
耗氧量	0.070
色度	0.287
菌落总数	0.079
温度	0.077
氨氮	0.165

1.4 模型的结构与实现

本研究提出了将 CNN 应用于混凝剂投放量的预测任务，具体针对水质的四项指标进行分析。将该问题转换为多元回归问题，其中混凝剂投放量作为目标变量，原水浊度的变化作为样本权重用于模型训练。

CNN 模型通过多层卷积操作，逐步提取水质数据中不同指标的局部特征及其变化模式，从而有效捕捉影响混凝剂投放量的关键特征。其结构包括两层一维卷积层、两层最大池化层、一层展平层以及三层全连接层，并使用 ReLU 激活函数，如图 1 所示。由于原水水质指标和混凝剂的投放量是非线性关系并具有一定的耦合性，需要利用卷积层中的滤波器提取四项水质数据中的高阶空间特征，尤其是不同指标的变化趋势。然而，由于相邻卷积核可能提取出相似和冗余的信息，因此使用最大池化层对特征图进行下采样，在减少冗余的同时保留重要信息。将四项水质数据作为输入层数据输入，第一层卷积层处理后的输出大小为 32 × 7 × 64，卷积操作中通过零填充保持特征图大小不变。接下来的最大池化层对其进行下采样，输出的大小缩减为 32 × 3 × 64。在第二层卷积层中进一步提取水质特征，输出大小为 32 × 2 × 64。经过第二个最大池化层的下采样后，输出尺寸进一步缩减为32 × 1 × 64。接着，数据被展平为 32 × 64 的向量格式，便于输入全连接层。最终的输出层为大小为 1 × 32 的向量，对应混凝剂投放量的预测结果。

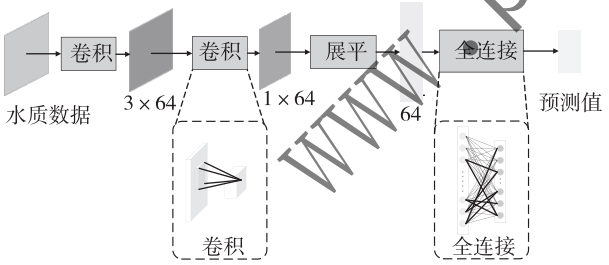

图 1　模型整体结构

2　结果与分析

在本研究中，共获取到了 8 个时序数据特征，分别是流量、浊度、pH、耗氧量、色度、菌落总数、温度和氨氮，其中流量指过去 24 h 的进水量。对 8 项水质数据指标进行特征选取和数据预处理，将通过特征选取和预处理的 4 项特征训练 CNN，让其充分学习，进行有效的混凝剂投放量预测。

2.1 流程设计

实验模型选取卷积核大小、学习率、隐藏层单元作为优化对象，并将原水浊度的变化作为样本权重输入模型训练。为了减少人为因素影响，对优化超参数的取值范围设置如下：隐藏层单元数范围是［1，200］，学习率范围是［0.001，0.01］，卷积核大小的范围是［1，64］。本研究选取 2021 年 1 月到 2023 年 6 月的水质数据作为数据集，在对数据集进行预处理和特征选择后，将数据集打乱后的 70% 作为训练数据，将原水出水浊度作为样本权重，与划分出的训练数据整合为训练集，用于模型训练。以 MAE 为损失函数，训练模型调整模型参数，并将数据集剩余的 30% 作为测试集进行训练后测试。若模型精度达到预期效果，则得到最优模型并输出预测值；否则更新权值继续调整模型参数。训练流程如图 2 所示。

2.2 结果呈现

2.2.1 不同参数对模型性能的影响

为探究参数对 CNN 模型预测性能的影响机制，本研究通过控制变量法分别测试了卷积核尺寸、学习率和隐藏层单元数的不同配置，实验结果如表 3 所示。

实验结果表明，参数调整对 CNN 模型的回归预测性能存在显著影响。随着卷积核数量从 16 逐步增加到 64，在保持学习率 0.001 和全连接层结构（100/50）的条件下，模型表现出持续的性能提升，R^2 从 0.922 上升至 0.926，MAE 和 RMSE 分别降低约 3.28 和 1.93，说明增大卷积核增强了水质特征的表征能力，有效提升了混凝剂投放量的预测精度。学习率对模型稳定性具有非线性

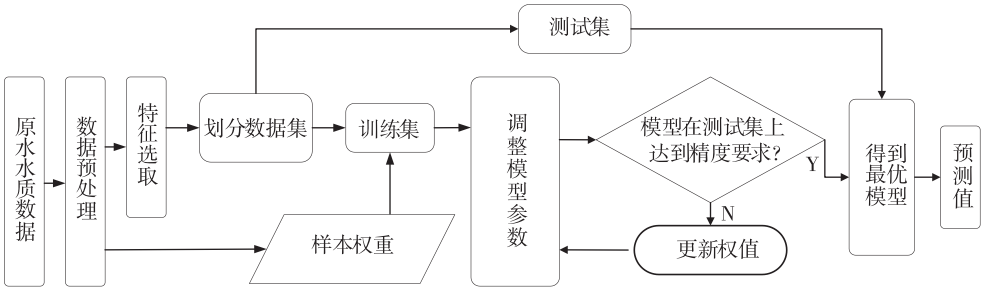

图 2　训练流程

表 3　不同参数对模型性能的影响

卷积核	学习率	全连接层 1 隐藏层	全连接层 2 隐藏层	R^2	MAE	RMSE
16	0.001	100	50	0.922	53.992	70.482
32	0.001	100	50	0.924	52.631	69.371
64	0.001	100	50	0.926	50.709	68.55
64	0.000 1	100	50	0.92	53.507	71.285
64	0.01	100	50	0.916	54.897	72.951
64	0.001	50	50	0.922	53.063	70.351
64	0.001	50	100	0.921	53.549	70.808

影响：当学习率维持在 0.001 时模型取得最佳平衡，此时 R^2 达到峰值 0.926；而将学习率降低到 0.000 1 或提升到 0.01 时，三项指标均出现明显退化，特别是过高学习率导致 RMSE 骤增 4.4，表明学习率偏离最优值时梯度更新过程出现震荡或收敛停滞。全连接层结构的影响相对温和但存在规律性，当首层隐藏单元从 50 增至 100 时，模型在保持其他参数不变的情况下 R^2 提升 0.4%，误差指标降低约 3%；而将第二层隐藏单元从 50 扩增至 100 反而轻微削弱性能，这可能与模型复杂度增加导致的局部过拟合有关。综合各参数交互作用，卷积核为 64、学习率为 0.001、全连接层（100/50）的组合展现出最优预测效能，其 0.926 的 R^2 和最低的误差指标表明该配置在特征提取能力和模型泛化性之间实现了良好平衡。

图 3 为 CNN 模型的预测结果图，横坐标表示实际的混凝剂投放量，纵坐标表示模型所预测的混凝剂投放量。由图 3 可见，利用已有数据训练 CNN 模型，经过训练后，能够较好地根据水质数据拟合出所需的混凝剂投放量，具有较高的准确性，对范围在 200 kg ~ 1 000 kg的混凝剂数据，该模型的预测效果尤为显著。综合评价指标和预测结果表明本研究提出的 CNN 模型能够较为准确地预测混凝剂的投放量，实现较好的预测效果。

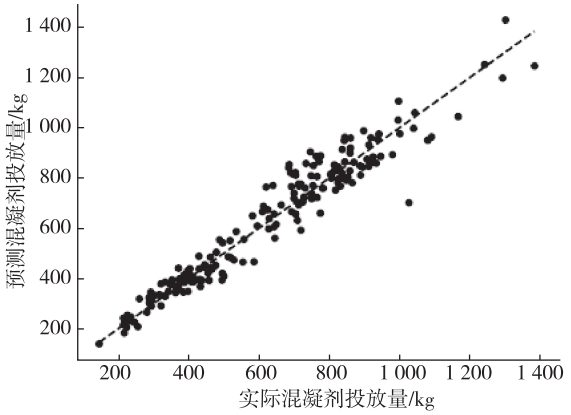

图 3　预测结果

2.2.2　样本权重对模型性能的影响

为了验证样本权重的作用，本文进行了对比试验。样本权重对模型性能的影响如表 4 所示。

表 4　样本权重对模型性能的影响

样本权重	RMSE	MAE	R^2
√	68.550	50.709	0.926
×	74.019	55.910	0.914

实验表明，引入样本权重能够提升模型性能。通过对比使用和未使用样本权重的实验结果，发现使用样本权重后，RMSE 降低了 7.4%，MAE 下降 9.3%，同时决定系数 R^2 提升至 0.926，表明模型的预测精度和拟合能力得到了优化。样本权重的引入能够引导模型更关注低出水浊度的样本，从而提升整体性能。

2.2.3　对比实验

为验证本文提出的 CNN 模型在多元回归预测任务中的有效性，本研究选取支持向量回归（SVR）、多层感知机（MLP）和 K 近邻回归（KNN）作为基线模型进行横向对比，各模型在相同数据集上的性能评价指标如表 5 所示。

表 5　各模型评价指标

模型	RMSE	MAE	R^2
CNN	68.550	50.709	0.926
SVR	83.142	50.661	0.892
MLP	130.244	96.891	0.737
KNN	83.225	62.222	0.892

从误差分析结果来看，本文提出的 CNN 模型在 RMSE 指标上表现最优，相较于 SVR 和 KNN 降低约 17.6%，较 MLP 降幅达 47.4%。在 MAE 指标方面，CNN 与 SVR 基本持平，较 KNN 降低 18.5%，较 MLP 显著减少 47.6%。值得注意的是，尽管 SVR 在 MAE 指标上略优

于 CNN，但其 RMSE 与 R^2 均弱于 CNN 模型，表明 SVR 对极端值的预测稳定性不及 CNN。

拟合优度方面，CNN 以 $R^2 = 0.926$ 的表现显著优于其他模型，较 SVR 提升 3.8%，较 MLP 提升 25.6%。这验证了 CNN 通过卷积层对局部特征的高效提取能力，能够更好地捕捉多维数据中的非线性关系和空间依赖特性。而传统机器学习模型受限于浅层结构，在复杂特征交互建模方面存在明显瓶颈。

实验结果表明，本文提出的 CNN 模型通过深度卷积结构实现了预测精度与泛化能力的同步提升，尤其在降低预测误差的波动性和提升模型解释力方面具有显著优势，验证了深度学习方法在多元回归任务中的适用性。

3 结论

本文以东南某百万级人口城市水厂为例，针对传统水质监测效率低下和混凝剂在投放前无法估计投放量的问题，介绍了一种基于卷积神经网络的混凝剂投放量预测模型，将水厂真实数据与 CNN 模型相结合搭建水质监测系统，以此系统为原型为后续水厂的监测系统提供支持。首先，通过对数据进行预处理，包括缺失值插补、数据归一化等步骤，确保了数据质量并加速了模型训练。然后，采用信息增益比率进行特征选择，从中选取了与混凝剂投放量预测最相关的特征。接着，使用卷积神经网络对水质数据进行特征提取，采用两层一维卷积层、两层最大池化层、展平层及三层全连接层结构，模型通过提取局部特征和模式，成功捕捉了影响混凝剂投放量的关键因素，最终得到了准确的预测结果。实验结果表明：CNN 预测模型的均方根误差 RMSE 为 68.550，平均绝对误差 MAE 为 50.709，拟合优度为 0.926，具有较好的预测效果。

由于本项目当前仍处于初期运行阶段，数据的收集和建模仍在进行中，后续研究将会持续收集数据，不断改进和优化算法，从而获得更有效的预测。

参考文献

[1] 侯立安，徐祖信，尹海龙，等. 我国水污染防治法综合评估研究 [J]. 中国工程科学，2022，24（5）：126－136.

[2] 常尧枫，谢嘉玮，谢军祥，等. 城镇污水处理厂提标改造技术研究进展 [J]. 中国给水排水，2022，38（6）：20－28.

[3] 袁卓异. 制水混凝加药控制的模糊 PID 算法研究 [J]. 计算技术与自动化，2022，41（3）：27－31.

[4] 李成. 智能控制在净水厂混凝投药过程中的应用研究 [J]. 新型工业化，2022，12（8）：217－220.

[5] IM Y，SONG G，LEE J，et al. Deep learning methods for predicting tap-water quality time series in South Korea [J]. Water，2022，14（22）：3766.

[6] CAI D，CHEN K，LIN Z，et al. Multi-step tap-water quality forecasting in South Korea with transformer-based deep learning model [J]. Urban Water Journal，2024，21（9）：1109－1120.

[7] TORKY M，BAKHIET A，BAKREY M，et al. Recognizing safe drinking water and predicting water quality index using machine learning framework [J]. International Journal of Advanced Computer Science and Applications，2023，14（1）：23－33.

[8] SAROJA，HASEENA，DHARSHINI S. Deep learning approach for prediction and classification of potable water [J]. Analytical Sciences，2023，39（7）：1179－1189.

[9] SV J R，RAMAKRISHNAN A，RISHIVARDHAN K. Improving water quality assessment through anomaly detection using hybrid convolutional neural network approach [J]. Global Nest Journal，2022：1－8.

[10] MOUSAVI S. Conjugation of deep learning and de noising data methods for short-term water turbidity forecasting [J]. Journal of Hydro-Environment Research，2024，52：26－37.

[11] TREJO-ZUÑIGA I，MORENO M，SANTANA-CRUZ R F，et al. Deep-learning-driven turbidity level classification [J]. Big Data and Cognitive Computing，2024，8（8）：89.

[12] ZHU G，LIN J，FANG H，et al. A flocculation tensor to monitor water quality using a deep learning model [J]. Environmental Chemistry Letters，2022，20（6）：3405－3414.

[13] SOLTANI A R. ARMA autocorrelation analysis：parameter estimation and goodness of fit test [J]. Journal of the Iranian Statistical Society，2022，21（2）：1－20.

[14] XIAO Y，GUO Y，LI M，et al. Machine learning to predict groundwater quality [J]. Journal of Beijing Normal University（Natural Science），2022，58（2）：261－268.

[15] ARAKI T，LUO Y，GUO M . Performance improvement validation of decision tree algorithms with non-normalized information distance in experiments [C]//Pacific Rim International Conference on Artificial Intelligence，2022：450－464.

[16] XIE X，ZHANG X，YANG J. Decision tree algorithm fusing information gain and Gini index [J]. Computer Engineering and Application，2022，58（10）：139－144.

（收稿日期：2025－03－18）

作者简介：

李泽楷（2000－），男，硕士研究生，主要研究方向：嵌入式系统、深度学习。

章杰（1979－），通信作者，男，硕士，副教授，主要研究方向：嵌入式系统、计算机通信。E-mail：13809507654@139. com。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部