

# 基于CNN-SVR网络的黄渤海近岸海域叶绿素a浓度预测

王晓霞<sup>1,2,3</sup>, 汪健平<sup>1,3</sup>, 王佳莹<sup>1,3\*</sup>, 孙珊<sup>1,3</sup>, 苏博<sup>1,3</sup>, 姜会超<sup>1,3</sup>, 朱明明<sup>1,3</sup>

(1. 山东省海洋资源与环境研究院, 山东 烟台 264006; 2. 自然资源部空间海洋遥感与应用重点实验室, 北京 100081; 3. 山东省海洋生态修复重点实验室, 山东 烟台 264006)

**摘 要:** 利用海洋卫星观测数据和黄渤海近岸海域实测生态水质数据, 建立了一种基于卷积神经网络结合支持向量回归 (Convolutional Neural Network-Support Vector Regression, CNN-SVR) 的深度学习网络模型的叶绿素a浓度预测方法。采用皮尔逊方法对叶绿素a与环境动力因子和生态水质因子作相关分析, 发现营养盐因子大多与叶绿素a有显著相关性, 水质因子如pH、溶解氧、盐度等与叶绿素a的相关性不大; 将黄渤海近岸海域划分为渤海南部与黄海北部、黄海中部, 进行春夏、秋冬两个时期1×1和2×2两种卷积核大小的CNN-SVR网络模型实验以及单因子敏感性分析试验。结果显示: 卷积核大小为2×2时, CNN-SVR网络模型对训练数据的学习和对测试样本的预测检验效果都更优; 渤海南部与黄海北部近岸海域模型预测效果更好。营养盐因子对模型预测能力的影响更显著, 悬浮物等水质因子的影响相对较弱。单变量对模型预测的敏感性较弱, 多变量整合具有互补性, 改善了模型的预测效果。

**关键词:** 卷积神经网络结合支持向量回归模型; 叶绿素a浓度预测; 单因子敏感性分析; 海洋卫星; 海洋生态水质因子

**中图分类号:** X145 **文献标识码:** A **文章编号:** 1003-0239(2024)04-0077-11

## 0 引言

叶绿素a (Chlorophylls a, Chl-a) 是表征浮游生物量分布情况和水体富营养化程度的基本指标之一。一般地, Chl-a浓度增加说明浮游植物数量增加, 藻华爆发最显著的表征量是Chl-a浓度的急剧增加<sup>[1]</sup>。在浒苔绿潮爆发期间与前后, 水温<sup>[2-3]</sup>、叶绿素a<sup>[4]</sup>、盐度<sup>[5]</sup>、溶解氧、营养盐<sup>[6-7]</sup>等海洋生态环境因子会相应地发生一定的变化, 其中, 绿潮的时空分布与Chl-a浓度具有较大的相关性<sup>[8]</sup>。生物量指标Chl-a既可以表征水体浮游植物生物量大小, 也能反映水体水质状况<sup>[9-10]</sup>。Chl-a浓度的时空分布对于浒苔绿潮程度、赤潮生态灾害预警和水质富营养化等级分类等起重要的指示作用。

Chl-a浓度与其他生态环境因子之间具有复杂的非线性关系, 应用传统方法预测Chl-a浓度已经不能满足当下业务工作的精度要求。随着人工智能的发展, 越来越多的机器学习方法被应用到生态环境因子浓度预测反演研究中。NGUYEN等<sup>[11]</sup>基于水质参数和水体表面反射率数据比较了随机森林模型 (Random Forest, RF)、支持向量机 (Support Vector Machine, SVM)、高斯过程 (Gaussian Process, GP)、极限梯度提升算法 (Extreme Gradient Boosting, XGB) 和 CatBoost (CB) 决策树预测Chl-a浓度的性能表现, 结果表明GP和CB的预测效果较好。MAIER等<sup>[12]</sup>基于高光谱数据, 应用RF、SVM和人工神经网络 (Artificial Neural Networks, ANN) 3种机器学习回归模型估算内陆水域的Chl-a浓度,

收稿日期: 2023-04-19。

基金项目: 自然资源部空间海洋遥感与应用重点实验室开放基金(202102002)。

作者简介: 王晓霞(1995-), 女, 研究实习员, 硕士, 主要从事海洋卫星在海洋生态环境监测、生态灾害预警等的应用研究。E-mail: 1457549281@qq.com

\*通信作者: 王佳莹(1991-), 女, 工程师, 硕士, 主要从事海洋生态环境监测相关研究。E-mail: 290448889@qq.com

发现ANN模型的预测效果较好。

遥感技术与机器学习算法的结合应用为水质监测提供了一种科学实时的手段。ZHANG<sup>[13]</sup>基于陆地卫星 Landsat8 影像数据,开展了波段反射率与 Chl-a 浓度的相关性分析,并利用 SVM 算法预测 Chl-a 浓度,同时分析了季节性因素、不同抽样方法、噪声点对模型性能的影响。JIN 等<sup>[14]</sup>利用卷积神经网络 (Convolutional Neural Network, CNN) 建立起卫星观测的生态水质因子(如溶解有机物、悬浮物、透明度)数据和水动力模型数据(海流、水位、盐度和温度)与 Chl-a 浓度之间的函数关系模型,发现溶解有机物对 Chl-a 的预测影响最大。APTOULA 等<sup>[15]</sup>基于哨兵 Sentinel-2 卫星影像数据,利用 CNN 回归模型反演了 Chl-a 浓度,模型取得了优越的预测效果。XIA 等<sup>[16]</sup>对比分析了基于 SVM 的多种优化算法,包括基于粒子群算法优化支持向量机 (Particle Swarm Optimization - Support Vector Machine, PSO-SVM)、人工蜂群算法优化支持向量机 (Artificial Bee Colony - Support Vector Machine, ABC-SVM)、布谷鸟算法优化支持向量机 (Cuckoo Search - Support Vector Machine, CS-SVM) 和灰狼优化器算法优化支持向量机 (Grey Wolf Optimizer - Support Vector Machine, GWO-SVM) 在 Chl-a 浓度预测和水质富营养化等级分类中的模型效果,模型中输入多个水质变量进行预测,结果发现 pH、总氮、总磷、氨氮在预测模型中起关键作用。本文将利用海洋水色卫星参数数据与动力要素数据及实测生态水质数据,尝试建立一种基于卷积神经网络结合支持向量回归 (Convolutional Neural Network-Support Vector Regression, CNN-SVR) 网络的 Chl-a 浓度预测模型。

## 1 数据资料

### 1.1 区域海洋环境特征

渤海、黄海是我国典型的半封闭型陆架浅海,平均水深 31 m,最深处为 140 m。黄海海底地形比较平缓,总体由西北向东南倾斜。黄海东西宽约 556 km,南北长约 870 km,平均水深 44 m,面积约为 38 万平方公里<sup>[17]</sup>,灌河、新沂河、沐河、苏北灌溉总渠、鸭绿江、汉江和大同江等河流流入黄海<sup>[18]</sup>。渤

海是我国北部的半封闭型内海,水深较浅,平均水深 18 m,最深处为 83 m<sup>[19]</sup>,沿岸有辽河、滦河、海河和黄河等 40 余条河流携带大量营养物质流入渤海。

研究区域处于季风区,海表风场具有明显的季节性特征。夏季盛行东南风和偏南风,南部沿岸城市如青岛、日照等常受到浒苔绿潮侵袭的影响;冬季以西北风为主,风力较强。水体在风、浪、潮汐的驱动作用下混合,海域水质在一定范围内具有季节性和区域性特征。夏季温度较高,水动力条件弱,入海径流营养盐输入丰富,浮游植物生长繁殖旺盛,Chl-a 浓度相对较高,这与赤潮、绿潮灾害的频发时间相稳合;冬季温度较低,水动力条件强,浮游植物生命活动弱,Chl-a 浓度相对较低。从空间上看,渤海、黄海近岸海域的 Chl-a 浓度较高,离岸区域较低。

### 1.2 研究数据与研究范围

#### 1.2.1 HY-1C/HY-2B 卫星数据

海洋水色业务卫星可以实时获得西北太平洋海域海表可溶有机物、悬浮泥沙含量、海表温度、Chl-a 浓度等数据。海洋一号 C (HY-1C) 水色卫星搭载了 3 个成像载荷:海洋水色水温扫描仪 (Chinese Ocean Color and Temperature Scanner, COCTS)、海岸带成像仪 (Coastal Zone Imager, CZI) 和紫外成像仪 (Ultraviolet Imager, UVI)。本文使用的是 2019 年 5 月—2022 年 7 月 COCTS 观测的悬浮物数据。COCTS 载荷能够实现每天一次的全球覆盖,星下点分辨率优于 1 100 m,扫描幅宽大于 2 900 km。

海洋动力环境卫星主要用于监测和调查海洋环境,获得包括海面风场、浪高、海面高度、海面温度等多种海洋动力环境参数,它集主动、被动微波遥感器于一体,具有高精度测轨和定轨能力以及全天候、全天时、全球的探测能力。卫星搭载的有效载荷主要是雷达高度计、微波散射计、扫描微波辐射计、校正辐射计。本文使用的是 2019 年 5 月—2022 年 7 月海洋二号 B (HY-2B) 卫星微波散射计观测的风矢量场数据。卫星微波散射计 1~2 d 内在全球海域的覆盖率不小于 90%,地面分辨率在 25 km,风速精度为 2 m/s,风向精度为 20°。

### 1.2.2 实测生态水质数据

本文使用的是2019年5月—2022年7月山东省近海生态趋势性调查数据。该数据来自每年4个时间段(3月、5月、8月、10月)的航次调查,调查范围为黄渤海近岸海域,一共布设575个站位(见图1)。本研究使用的是水文气象数据中的水温数据,水体环境数据中的活性磷酸盐、活性硅酸盐、亚硝酸盐、硝酸盐、氨氮、总氮、总磷、盐度、pH、溶解氧和Chl-a数据。

### 1.2.3 研究范围

考虑到海域水质、海面风场具有区域性特征,将研究区域分成A、B两部分。A为渤海南部与黄海北部近岸海域,B为黄海中部近岸海域。

## 1.3 数据预处理

### 1.3.1 异常值、缺失值处理

原始数据中存在一些不完整和偏离的数据,因而需要进行数据预处理来识别离群值、剔除异常值和缺失值,改善数据集的质量,保障算法模型的执行效率和结果。异常值的剔除采用 $3\sigma$ 准则:

$$P(|x - \mu| > 3\sigma) \leq 0.003 \quad (1)$$

式中: $x$ 表示观测要素变量; $\sigma$ 、 $\mu$ 分别为标准差和均值;观测要素值分布超出区间 $[(\mu-3\sigma), (\mu+3\sigma)]$ 的离

群点只占0.3%,以异常值处理。

由于异常值和缺失值数量占数据集样本总量的比例较小,本文直接剔除实测生态水质数据中的异常值和缺失值,经过处理后的实测生态水质数据记录共有6410条。

### 1.3.2 数据整合

匹配同一时空窗口下的卫星数据与实测生态水质数据,时间尺度为6 h,空间尺度为 $0.125^\circ$ 。由于卫星观测空白和设备异常等,卫星观测数据中存在一些异常数据。剔除异常值后,最终获得有效数据1085条。考虑到生态水质要素浓度变化具有季节性特征,因此将数据集分成两类,即以5月、8月航次为代表的春夏时期数据集,以及以10月、3月航次为代表的秋冬时期数据集。

### 1.3.3 相关性分析

本文采用皮尔逊(Pearson)方法分析研究区域A和B在春夏和秋冬时期Chl-a浓度与环境动力因子和生态水质因子的相关关系。Pearson相关系数公式<sup>[20]</sup>如下:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}} \quad (2)$$

$X$ 的离均差平方和为:

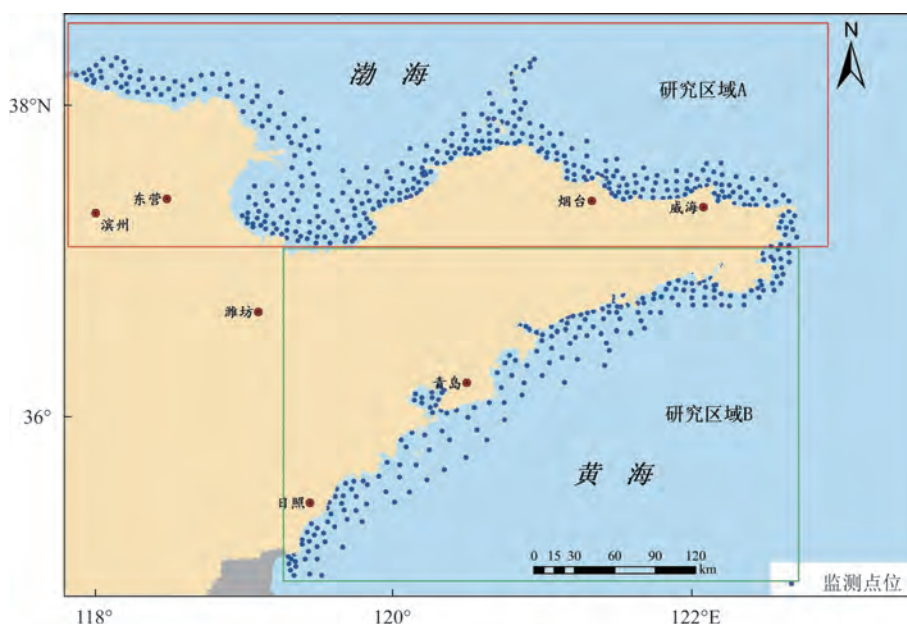


图1 山东省近海生态趋势性调查站位布设

Fig.1 Distribution map of ocean ecological monitoring stations in shandong coastal area



$$l_{xx} = \sum (X - \bar{X})^2 \quad (3)$$

式中:  $X$  为样本数据变量;  $\bar{X}$  为样本数据均值。

$Y$  的离均差平方和为:

$$l_{yy} = \sum (Y - \bar{Y})^2 \quad (4)$$

式中:  $Y$  为样本数据变量;  $\bar{Y}$  为样本数据均值。

$X$  与  $Y$  间的离均差积和为:

$$l_{xy} = \sum (X - \bar{X})(Y - \bar{Y}) \quad (5)$$

在显著性检验中, 总体相关系数用  $\rho$  表示; 样本相关系数用  $r$  表示。假设  $\rho=0$  为无关,  $\rho \neq 0$  为相关, 确定显著性水平为  $\alpha=0.05$ 。

Chl-a 与水质要素、营养盐要素和环境动力要素的相关性结果见表 1 和表 2。研究区域 A 春夏时期 Chl-a 与亚硝酸盐-氮呈显著正相关(见表 1), 且通过了 0.05 显著性水平检验; 秋冬时期 Chl-a 与总氮呈显著相关, 且通过了 0.05 显著性水平检验。研究区域 B 春夏时期 Chl-a 与亚硝酸盐-氮呈显著正相关, 水温、悬浮物、亚硝酸盐-氮、硝酸盐-氮、硅酸盐和总氮均通过了 0.05 显著性水平检验; 秋冬时期与各因子的相关性较弱。营养盐因子大多与 Chl-a 有显著相关性(见表 2), 水质因子如 pH、溶解氧、盐度等与 Chl-a 的相关性较弱。一般地, pH 受海水  $\text{CO}_2$  含量影响, 除了浮游植物活动会对海水  $\text{CO}_2$  含量产生影响外, 陆源入海有机物的氧化分解也会造成海水  $\text{CO}_2$  含量的变化<sup>[21]</sup>, 进而影响海水 pH 变化, 受入海径流和海表水汽蒸发影响, 近岸海域盐度也会发生相应的变化。溶解氧主要来源于大气中氧气的溶入和浮游植物光合作用释放的游离氧, 生物呼

表 1 Chl-a 与水质要素的相关性

Tab.1 Correlation between water quality elements and chlorophylls a

水质要素	研究区域 A		研究区域 B	
	春夏时期	秋冬时期	春夏时期	秋冬时期
pH	0.104	0.009	0.120	0.018
水温	0.222*	0.059	0.215*	-0.021
溶解氧	0.033	0.011	0.025	0.019
盐度	-0.056	0.028	0.045	0.017
悬浮物	0.254*	0.139	0.217*	0.168

注: \*表示在 0.05 的水平上显著。

表 2 Chl-a 与营养盐要素和环境动力要素的相关性

Tab.2 Correlation between chlorophylls a elements of nutrient salt and environmental dynamics

营养盐要素	研究区域 A		研究区域 B	
	春夏时期	秋冬时期	春夏时期	秋冬时期
亚硝酸盐-氮	0.664*	0.369*	0.646*	0.234*
硅酸盐	0.550*	0.434*	0.335*	-0.042
硝酸盐-氮	0.579*	0.425*	0.463*	0.253*
氨-氮	0.232	0.029	0.225	0.095
总氮	0.485*	0.550*	0.423*	0.309*
总磷	0.046	0.053	0.384	-0.028
磷酸盐	0.034	-0.301	0.330	-0.104
风速	0.059	0.036	-0.124	0.007
风向	0.065	-0.032	-0.027	0.012

注: \*表示在 0.05 的水平上显著。

吸作用与有机物质的氧化分解作用等会消耗溶解氧<sup>[22]</sup>。因此, 影响 pH、溶解氧、盐度的因素较为复杂, 推测是造成本研究中这 3 种水质因子与 Chl-a 相关性较小的重要原因。为保证模型的预测效果, 选择研究区域 A 和 B 在两个时期中相关性较大的 6 个因子输入模型。

## 2 方法原理

### 2.1 CNN-SVR 模型结构

本文 CNN-SVR 模型结构见图 2。输入数据包括水温、悬浮物、亚硝氮、硅酸盐、硝氮等, 输出数据为 Chl-a 数据。CNN 用于特征提取, 可以提高模型的预测准确率; SVR 用于模型预测, 适用于小样本数据训练, 来降低算法的整体训练难度。卷积神经网络由输入层、卷积层(2 Dimension Convolution Layer, Conv2D)、池化层(Max Pooling)和全连接层(Full Connection)组成, 各层神经元之间通过权值连接。卷积层主要用于数据特征的提取; 池化层通过下采样运算减少特征维度, 并保留有效信息; 全连接层将提取的特征合并, 并输入到输出层。

支持向量回归算法是基于统计学习的神经网络技术, 通过非线性变换将样本转化为高维特征空间。该模型在高维向量空间里用线性函数去拟合样本, 以所有样本实际位置到该线性函数的综合距离

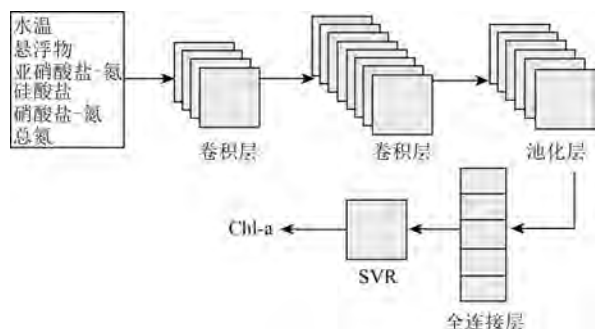


图2 CNN-SVR模型结构

Fig.2 Structure of CNN-SVR

为损失,通过最小化损失来求取线性函数的参数。

支持向量回归公式为:

$$f(x) = w_i \phi_i(x) + b \quad (6)$$

式中:  $i = 1, 2, \dots, l$ ,  $l$  为样本数量;  $x$  为输入向量;  $w_i$  为权重向量;  $b$  为偏置;  $\phi_i(x)$  是将输入向量  $x$  映射到多维特征空间的特征向量。支持向量回归模型优化方程为:

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (7)$$

$$\text{s.t.} \begin{cases} ((w \cdot x_i) + b) - y_i \leq \varepsilon + \xi_i, i = 1, \dots, l \\ y_i - ((w \cdot x_i) + b) \leq \varepsilon + \xi_i^*, i = 1, \dots, l \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (8)$$

式中:  $\xi_i$  和  $\xi_i^*$  为松弛变量;  $C$  为正则化参数,用于防止模型过拟合,  $C$  值越小,对模型的惩罚越低,泛化能力越强;  $\varepsilon$  为损失函数参数。

最终模型为:

$$y = f(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x_i, x) + \bar{b} \quad (9)$$

式中:  $\bar{\alpha}_i^*$  和  $\bar{\alpha}_i$  为拉格朗日系数;  $k$  为核函数。

## 2.2 建模实现过程

CNN-SVR模型构建主要包括:

①参数初始化。设置网络权值变量和模型偏置参数为固定小数值 0.1,以使得修正线性单元(Rectified Linear Unit, ReLU)函数在训练初期输出灵敏,提高模型的运算效率。

②网络信号的传递:前向传播与反向传播。前向传播部分使用权重矩阵进行加权求和,并加上偏置卷积运算后输入到激活函数<sup>[23]</sup>。反向传播部分采用梯度下降算法进行优化,以最小损失函数为目标进行多次迭代,调整参数权重和阈值。此时,全连

接层的输出数据特征以向量形式存储。

③SVR模型拟合与预测。设置正则化系数为均匀分布的随机小数,选择线性核函数,将数据特征输入模型作训练并预测。

## 2.3 单因子敏感性分析

本文仅使用单个输入变量作为深度学习模型的训练数据,以此来评价每个输入变量对模型预测结果的影响。单个生态环境因子输入至CNN-SVR模型中,评估模型预测的准确性,确定Chl-a浓度预测模型最大影响因子。

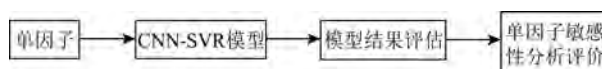


图3 单因子敏感性分析试验

Fig.3 Sensitivity analysis of single element

## 2.4 模型效果评价

为确保模型更好地适应实际应用研究,本文采用科学标准来评价预测模型的性能,主要包括均方误差(Mean Square Error, MSE)和决定系数(Coefficient of Determination,  $R^2$ )。MSE衡量的是模型预测值与Chl-a浓度实测值之间的偏差,用于评估CNN-SVR模型预测的准确性。 $R^2$ 衡量的是预测值与真实值的拟合程度。

MSE的计算公式为:

$$f = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (10)$$

式中:  $i = 1, 2, \dots, n$ ,  $n$  为样本数量;  $y_i$  是Chl-a浓度实测数据值;  $\hat{y}_i$  是模型预测值。

决定系数计算公式为:

$$f(y) = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

式中:  $i = 1, 2, \dots, n$ ,  $n$  为样本数量;  $y_i$  是Chl-a浓度真实值;  $\bar{y}$  是平均值;  $\hat{y}_i$  是模型预测值。

## 3 模型实验与结果分析

### 3.1 CNN-SVR模型试验结果

选取与Chl-a浓度具有显著相关性的生态环境

因子作为模型预报因子,Chl-a浓度作为模型预测变量。模型预报因子包括水温、悬浮物、亚硝酸盐-氮、硅酸盐、硝酸盐-氮和总氮,各因子对应样本数量为1 085个。本文将数据集的80%作为训练集,20%作为测试集,并在数据输入模型前采用标准差作z-score(zero-mean normalization)归一化处理。归一化公式为:

$$x^* = \frac{x - \mu}{\sigma} \quad (12)$$

式中: $x$ 为样本数据变量; $\mu$ 为样本数据均值; $\sigma$ 为样本数据的标准差。

神经网络卷积核的大小影响网络特征提取效果和模型运算效率。一般地,卷积核越大,特征提取效果越好,但参数和计算量较多,模型运算效率会下降;卷积核越小,参数和计算量较少,网络特征学习更深,但需要多卷积层数才可以达到大卷积核的特征提取效果。本文进行了两种卷积核的网络试验:卷积核 $1 \times 1$ ,滑动步长为1,全0填充,池化窗口为 $2 \times 2$ ,滑动步长为2,全0填充;卷积核 $2 \times 2$ ,滑动步长为1,全0填充,池化窗口为 $2 \times 2$ ,滑动步长为2,全0填充。

模型预测结果显示(见表3),当网络卷积核大小为 $2 \times 2$ 时,研究区域A秋冬时期的训练误差最小,MSE为 $0.144 \mu\text{g/L}$ ,训练和测试阶段的决定系数最大, $R^2$ 分别为0.887和0.668;春夏时期的测试误差最小,MSE为 $0.374 \mu\text{g/L}$ 。秋冬时期渤海南部与黄海北部近岸海域模型的预测效果优于春夏时期。

研究区域B春夏时期的训练误差最小,MSE为 $0.137 \mu\text{g/L}$ ,训练和测试阶段决定系数最大, $R^2$ 分别为0.852和0.632;秋冬时期的测试误差最小,MSE为 $0.193 \mu\text{g/L}$ 。春夏时期黄海中部近岸海域模型的预测效果优于秋冬时期。研究区域B秋冬时期的预测结果最差,测试阶段的 $R^2$ 为0.326,这与秋冬时期数据集的数量有关,由于存在一部分数据缺失,模型习得特征有限,模型预测效果较差。

当网络卷积核大小为 $1 \times 1$ 时,研究区域A秋冬时期训练过程和测试过程的 $R^2$ 最大,分别为0.665和0.513,预测效果最好;研究区域A春夏时期的预测效果次之;研究区域B秋冬时期的预测结果最差,训练过程和测试过程的 $R^2$ 分别为0.343和0.206。研究区域A的预测效果优于研究区域B,这与卷积核大小为 $2 \times 2$ 时的模型预测结果一致。

同一区域在同一时期内,卷积核 $2 \times 2$ 的模型对训练数据的学习和对测试样本的预测检验效果都更优。这也说明卷积核 $2 \times 2$ 的CNN-SVR模型对信息特征的学习能力比卷积核 $1 \times 1$ 的更强,预测更准确,泛化能力更强。相同时期内研究区域A测试过程的决定系数大于研究区域B,预测效果更优,这说明了本文的方法更适用于渤海南部与黄海北部近岸海域。

### 3.2 单因子敏感性分析试验

本文使用单个生态环境因子变量(包括水温、悬浮物、亚硝酸盐-氮、硅酸盐、硝酸盐-氮和总氮数

表3 不同大小卷积核CNN-SVR模型预测效果对比

Tab.3 MSE and  $R^2$  of CNN-SVR models with different kernels

研究范围	研究时段	CNN-SVR 模型类别	MSE		$R^2$	
			训练误差	测试误差	训练过程	测试过程
研究区域A	春夏时期	卷积核 $1 \times 1$	0.363	0.420	0.644	0.424
		卷积核 $2 \times 2$	0.154	0.374	0.862	0.650
研究区域A	秋冬时期	卷积核 $1 \times 1$	0.353	0.884	0.665	0.513
		卷积核 $2 \times 2$	0.144	0.856	0.887	0.668
研究区域B	春夏时期	卷积核 $1 \times 1$	0.366	0.468	0.644	0.298
		卷积核 $2 \times 2$	0.137	0.323	0.852	0.632
研究区域B	秋冬时期	卷积核 $1 \times 1$	0.761	0.317	0.343	0.206
		卷积核 $2 \times 2$	0.493	0.193	0.601	0.326



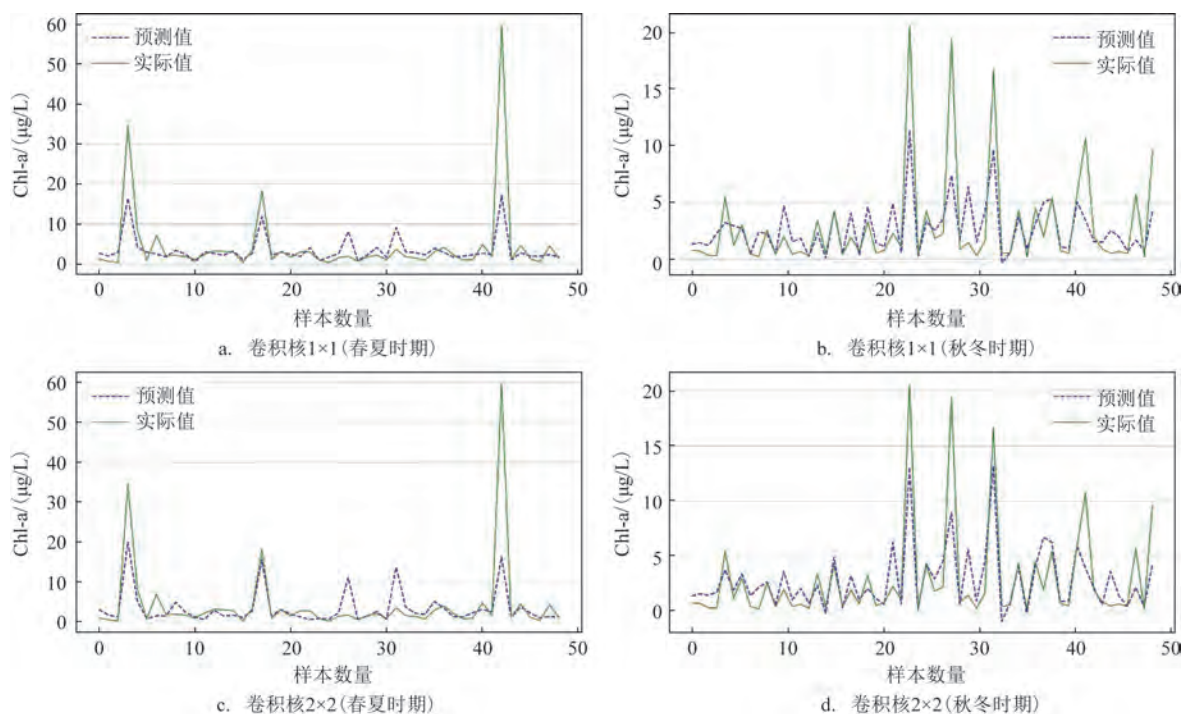


图4 研究区域A春夏与秋冬时期CNN-SVR模型预测结果

Fig.4 Results predicted by CNN-SVR models during spring and summer in contrast to autumn and winter in region A

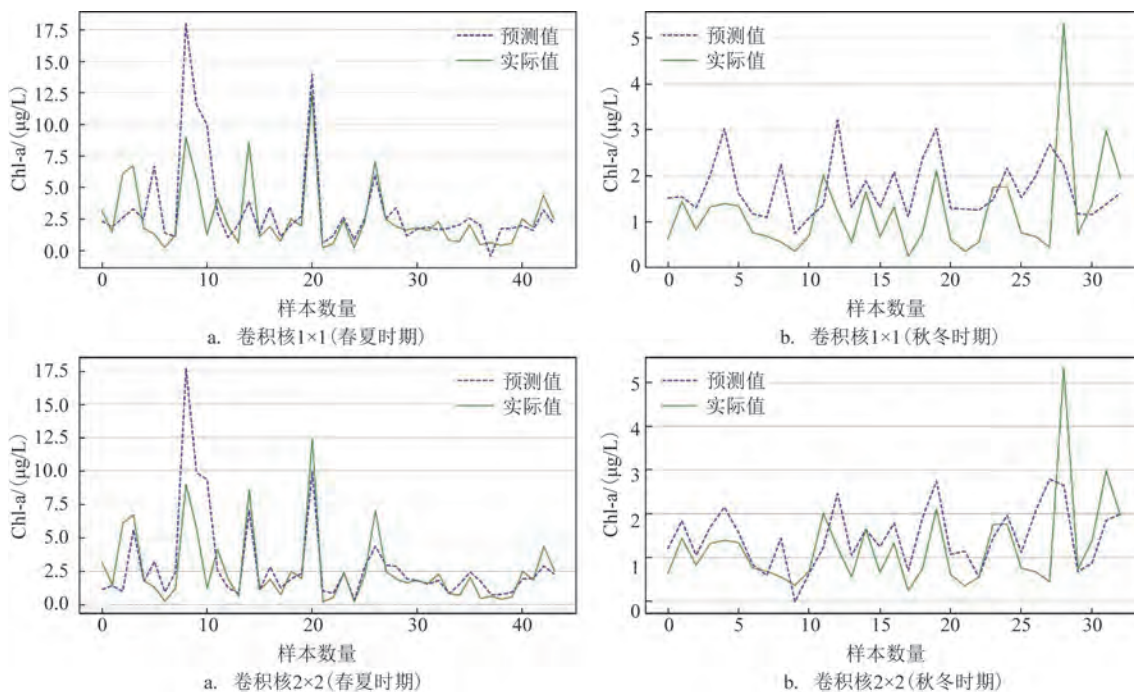


图5 研究区域B春夏与秋冬时期CNN-SVR模型预测结果

Fig.5 Results predicted by CNN-SVR models during spring and summer in contrast to autumn and winter in region B

据)作为模型的训练数据。选择卷积核大小 $2 \times 2$ 的CNN-SVR模型作为单因子敏感性分析试验模型,

计算MSE以评估各因子对模型预测结果的影响。结果显示(见表4),在研究区域A中,春夏时期亚硝

表 4 研究区域 A 春夏时期单因子敏感性分析试验结果

Tab.4 Results of sensitivity analysis of single element during spring and summer in region A

误差类别	水温/℃	悬浮物/ (μg/L)	亚硝酸盐-氮/ (μg/L)	硅酸盐/ (μg/L)	硝酸盐-氮/ (μg/L)	总氮/ (μg/L)	全部因子/ (μg/L)
训练误差	0.976	0.987	0.882	0.963	0.965	0.952	0.154
测试误差	2.614	2.691	2.322	2.551	2.559	2.547	0.374

表 5 研究区域 A 秋冬时期单因子敏感性分析试验结果

Tab.5 Results of sensitivity analysis of single element during autumn and winter in region A

误差类别	悬浮物/ (μg/L)	亚硝酸盐-氮/ (μg/L)	硅酸盐/ (μg/L)	硝酸盐-氮/ (μg/L)	磷酸盐/ (μg/L)	总氮/ (μg/L)	全部因子/ (μg/L)
训练误差	0.992	0.982	0.866	0.906	0.998	0.865	0.144
测试误差	2.087	2.032	1.828	1.874	2.051	1.870	0.856

酸盐-氮对 Chl-a 浓度的预测有最显著的影响,测试误差最小,MSE 为 2.322 μg/L,其次是总氮,MSE 为 2.547 μg/L;秋冬时期硅酸盐对 Chl-a 浓度的预测有最显著的影响(见表 5),测试误差最小,MSE 为 1.828 μg/L,其次是总氮和硝酸盐-氮。悬浮物对 Chl-a 浓度的预测影响最弱,MSE 分别为 2.691 μg/L 和 2.087 μg/L。

研究区域 B 中,春夏(见表 6)和秋冬(见表 7)时期总氮对 Chl-a 浓度的预测有最显著的影响,测试误差最小,MSE 分别为 0.324 μg/L 和 0.311 μg/L。春夏时期,磷酸盐对 Chl-a 浓度的预测影响最弱,测试

误差最大,MSE 为 0.435 μg/L;秋冬时期,悬浮物对 Chl-a 浓度的预测影响最弱,测试误差最大,MSE 为 0.336 μg/L。试验结果发现,模型预测过程中营养盐因子的敏感性相对更强,悬浮物等水质因子的敏感性比其他因子弱。

同一区域在同一时期内,单个输入变量敏感性分析的试验结果相差不大,对模型性能的影响较小,但整合后的变量具有互补性,可以改善模型的预测效果。研究区域 A 和 B 在两个时期的全部因子测试误差分别为 0.374 μg/L、0.856 μg/L、0.323 μg/L、0.193 μg/L,准确度都高于相同区域相同时期内的单

表 6 研究区域 B 春夏时期单因子敏感性分析试验结果

Tab.6 Results of sensitivity analysis of single element during spring and summer in region B

误差类别	亚硝酸盐-氮/ (μg/L)	硅酸盐/ (μg/L)	硝酸盐-氮/ (μg/L)	总氮/ (μg/L)	磷酸盐/ (μg/L)	总磷/ (μg/L)	全部因子/ (μg/L)
训练误差	0.831	0.967	0.950	0.907	0.987	0.969	0.137
测试误差	0.337	0.416	0.411	0.324	0.435	0.346	0.323

表 7 研究区域 B 秋冬时期单因子敏感性分析试验结果

Tab.7 Results of sensitivity analysis of single element during autumn and winter in region B

误差类别	悬浮物/ (μg/L)	亚硝酸盐-氮/ (μg/L)	硝酸盐-氮/ (μg/L)	氨-氮/ (μg/L)	磷酸盐/ (μg/L)	总氮/ (μg/L)	全部因子/ (μg/L)
训练误差	0.993	0.995	0.965	0.998	0.999	0.960	0.493
测试误差	0.336	0.325	0.312	0.328	0.332	0.311	0.193



因子输入模型。将所有变量全部输入模型作预测,模型的预测效果最佳。

## 4 结论与讨论

Chl-a浓度是赤潮、绿潮等海洋生态灾害发生的重要指示因子之一,其发展演变与海洋环境动力因子、海水生态水质条件等密切相关。利用CNN-SVR网络映射海洋生态环境因子与Chl-a浓度之间的函数关系,可以较好地预测海洋Chl-a浓度,为海洋生态灾害的早期预警提供科学指导。本文利用海洋卫星观测数据和黄渤海近岸海域实测生态水质数据,基于CNN-SVR网络模型尝试建立一种Chl-a浓度的预测方法,将黄海、渤海划分为渤海南部与黄海北部、黄海中部近岸海域两个区域,并区分春夏和秋冬两个时期,对两种卷积核大小的网络模型预测效果做对比。采用Pearson方法对Chl-a与环境动力因子和生态水质因子作相关分析,优选出与Chl-a浓度相关性较大的变量。利用优选变量的数据对Chl-a浓度进行CNN-SVR网络模型实验,并对各生态环境因子进行了单因子敏感性分析试验。结果表明:

①在同一区域同一时期内,卷积核 $2 \times 2$ 的CNN-SVR模型对训练数据的学习和对测试样本的预测检验效果都更优,其信息特征学习能力比卷积核 $1 \times 1$ 的网络模型更强,预测更准确,泛化能力更强。

②秋冬时期模型对渤海南部与黄海北部近岸海域的预测效果优于春夏时期。春夏时期模型对黄海中部近岸海域的预测效果优于秋冬时期。相同时期内研究区域A的预测效果相对更好,本文的方法更适用于渤海南部与黄海北部近岸海域。

③在研究区域A中,春夏时期亚硝酸盐-氮对Chl-a浓度的预测有最显著的影响,秋冬时期硅酸盐对Chl-a浓度的预测有最显著的影响;在研究区域B中,两个时期总氮对Chl-a浓度的预测均有最显著的影响。营养盐因子对模型预测能力的影响更显著,悬浮物等水质因子的影响相对较弱。单变量对模型预测的敏感性较弱,多变量整合具有互补性,可以改善模型的预测效果。

基于深度学习模型的Chl-a浓度预测方法既探

究了深度学习方法在海洋生态环境领域和卫星遥感的应用性,又促进了人工智能技术与海洋生态环境领域和卫星遥感的结合与发展。但是本文的网络模型预测能力还有待提升,考虑有以下几方面原因:一是冗余信息可能会影响模型预测的准确性和速度,需充分考虑各个生态环境因素对模型实际的影响。本文选取了与Chl-a相关性较大的前6个因子输入模型,但是存在部分相关性较弱的因子被用于模型训练,比如研究区域B秋冬时期Chl-a与各因子的相关性都比较弱,因此会在一定程度上影响模型预测的准确性。水体在风、浪、潮汐的驱动作用下混合,海域水质呈现季节性和区域性分布特点。下一步可以考虑加入海浪、潮汐等环境动力因子,并结合区域海洋环境特征,充分评估各生态环境因子对模型的影响。二是用于模型训练的数据量不足,仅有1 000多条数据,网络模型习得的数据特征过少,影响模型预测效果。三是研究区域的划分应综合考量区域海洋环境特征,比如叶绿素高值区、海洋灾害高发区等具有代表性海洋环境特征的区域,并开展分区分析,进而改善模型预测效果。

## 参考文献:

- [1] 刘建萍, 张玉超, 钱新, 等. 太湖叶绿素a浓度预测模型初探[J]. 环境保护科学, 2009, 35(4): 46-49.
- [2] LIU J P, ZHANG Y C, QIAN X, et al. Preliminary studies on the prediction model of chlorophyll-a in Taihu Lake[J]. Environmental Protection Science, 2009, 35(4): 46-49.
- [3] AHN O, PETRELL R J, HARRISON P J. Ammonium and nitrate uptake by *Laminaria saccharina* and *Nereocystis luetkeana* originating from a salmon sea cage farm[J]. Journal of Applied Phycology, 1998, 10(4): 333-340.
- [4] KIM J H, KANG E J, PARK M G, et al. Effects of temperature and irradiance on photosynthesis and growth of a green-tide-forming species (*Ulva linza*) in the Yellow Sea[J]. Journal of Applied Phycology, 2011, 23(3): 421-432.
- [5] 苑克磊, 侯查伟, 丁一, 等. 我国黄海浒苔绿潮暴发海域环境因子季节性变化及年际变化分析[J]. 广西科学院学报, 2018, 34(3): 204-209.
- [6] YUAN K L, HOU C W, DING Y, et al. Analysis of the seasonal and inter-annual changes of environmental factors in *Enteromorpha prolifera* green tide outbreak in the Yellow Sea[J]. Journal of Guangxi Academy of Sciences, 2018, 34(3): 204-209.
- [7] 夏斌, 马绍赛, 崔毅, 等. 黄海绿潮(浒苔)暴发区温盐、溶解氧和营养盐的分布特征及其与绿潮发生的关系[J]. 渔业科学进展,

- 2009, 30(5): 94-101.
- XIA B, MA S S, CUI Y, et al. Distribution of temperature, salinity, dissolved oxygen, nutrients and their relationships with green tide in *Enteromorpha prolifera* outbreak area of the Yellow Sea[J]. *Progress in Fishery Sciences*, 2009, 30(5): 94-101.
- [6] 庞秋婷, 李凤, 刘湘庆, 等. 围隔实验中浒苔在不同营养盐条件下的生长比较[J]. *环境科学*, 2013, 34(9): 3398-3404.
- PANG Q T, LI F, LIU X Q, et al. Compare the growth of *Enteromorpha prolifera* under different nutrient conditions[J]. *Environmental Science*, 2013, 34(9): 3398-3404.
- [7] 吴婷, 田小亭, 刘召辉, 等. 浒苔对不同氮源吸收特性的初步研究[J]. *海洋科学*, 2011, 35(4): 51-54.
- WU T, TIAN X T, LIU Z H, et al. Absorption of different nitrogen species by *Enteromorpha prolifera*[J]. *Marine Sciences*, 2011, 35(4): 51-54.
- [8] 孙晓. 黄海绿潮和叶绿素a浓度的遥感时空分异及两者响应机制[D]. 烟台: 鲁东大学, 2018.
- SUN X. Spatial and temporal variations and response mechanism of green tide and chlorophyll a concentrations based on remote sensing in southern Yellow Sea[D]. Yantai: Ludong University, 2018.
- [9] 汪益斌, 张维砚, 徐春燕, 等. 淀山湖浮游植物初级生产力及其影响因素[J]. *环境科学*, 2011, 32(5): 1249-1256.
- WANG Y P, ZHANG W Y, XU C Y, et al. Phytoplankton productivity and its influencing factors in Dianshan Lake[J]. *Environmental Science*, 2011, 32(5): 1249-1256.
- [10] 金相灿, 刘鸿亮, 屠清瑛, 等. 中国湖泊富营养化[M]. 北京: 中国环境科学出版社, 1990.
- JIN X C, LIU H L, TU Q Y, et al. Eutrophication in lake of China [M]. Beijing: China Environmental Science Press, 1990.
- [11] NGUYEN H Q, HA N T, NGUYEN-NGOC L, et al. Comparing the performance of machine learning algorithms for remote and in situ estimations of chlorophyll-a content: a case study in the Tri An Reservoir, Vietnam[J]. *Water Environment Research*, 2021, 93(12): 2941-2957.
- [12] MAIER P M, KELLER S. Estimating chlorophyll a concentrations of several inland waters with hyperspectral data and machine learning models[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, IV-2-W5: 609-614.
- [13] ZHANG T. Estimation of chlorophyll-a concentration of lakes based on SVM algorithm and Landsat 8 OLI images[J]. *Environmental Science and Pollution Research*, 2020, 27(13): 14977-14990.
- [14] JIN D, LEE E, KWON K, et al. A deep learning model using satellite ocean color and hydrodynamic model to estimate chlorophyll-a concentration[J]. *Remote Sensing*, 2021, 13(10): 2003.
- [15] APTOULA E, ARIMAN S. Chlorophyll-a retrieval from Sentinel-2 images using convolutional neural network regression[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 6002605.
- [16] XIA J J, ZENG J. Environmental factor assisted chlorophyll-a prediction and water quality eutrophication grade classification: a comparative analysis of multiple hybrid models based on a SVM [J]. *Environmental Science: Water Research & Technology*, 2021, 7(6): 1040-1049.
- [17] WANG S Q, HUAN Y, QIU Z F, et al. Remote sensing of particle cross-sectional area in the Bohai Sea and yellow sea: algorithm development and application implications[J]. *Remote Sensing*, 2016, 8(10): 841.
- [18] ZHANG M W, TANG J W, SONG Q J, et al. Backscattering ratio variation and its implications for studying particle composition: a case study in Yellow and East China seas[J]. *Journal of Geophysical Research: Oceans*, 2010, 115(C12): C12014.
- [19] 姜玲玲, 王龙霄, 王林, 等. 基于 Sentinel-3 OLCI 影像的渤海透明度遥感反演研究[J]. *光谱学与光谱分析*, 2022, 42(4): 1209-1216.
- JIANG L L, WANG L X, WANG L, et al. Research on remote sensing retrieval of Bohai sea transparency based on sentinel-3 OLCI image[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(4): 1209-1216.
- [20] 何恩业, 王丹, 黄莉, 等. 西太平洋副热带高压的变动对我国赤潮发生的影响分析[J]. *海洋预报*, 2015, 32(4): 83-89.
- HE E Y, WANG D, HUANG L, et al. Impact of the west pacific subtropical high on the red tide in the China seas[J]. *Marine Forecasts*, 2015, 32(4): 83-89.
- [21] 黄岁樑, 臧常娟, 杜胜蓝, 等. pH、溶解氧、叶绿素a之间相关性研究 I: 养殖水体[J]. *环境工程学报*, 2011, 5(6): 1201-1208.
- HUANG S L, ZANG C J, DU S L, et al. Study on the relationships among pH, dissolved oxygen and chlorophyll a I: aquaculture water[J]. *Chinese Journal of Environmental Engineering*, 2011, 5(6): 1201-1208.
- [22] 张建乐, 张心宇, 张永丰, 等. 北戴河海域夏季叶绿素a分布特征及其与pH、溶解氧的关系[J]. *海洋湖沼通报*, 2015(4): 155-160.
- ZHANG J L, ZHANG X Y, ZHANG Y F, et al. Distribution of chlorophyll a of the coastal water of beidaihe and its relationship with pH and dissolved oxygen in summer[J]. *Transactions of Oceanology and Limnology*, 2015(4): 155-160.
- [23] 吴浠. 基于卷积神经网络的推荐算法[D]. 广州: 华南理工大学, 2016.
- WU X. Recommendation algorithm based on convolutional neural network[D]. Guangzhou: South China University of Technology, 2016.

## Prediction of chlorophyll a concentration in nearshore waters of the Yellow and Bohai Seas based on CNN-SVR networks

WANG Xiaoxia<sup>1,2,3</sup>, WANG Jianping<sup>1,3</sup>, WANG Jiaying<sup>1,3\*</sup>, SUN Shan<sup>1,3</sup>, SU Bo<sup>1,3</sup>, JIANG Huichao<sup>1,3</sup>, ZHU Mingming<sup>1,3</sup>

(1. Shandong Marine Resource and Environment Research Institute, Yantai 264006, China; 2. Key Laboratory of Space Ocean Remote Sensing and Applications, Beijing 100081, China; 3. Shandong Key Lab of Marine Ecological Restoration, Yantai 264006, China)

**Abstract:** A chlorophyll-a (Chl-a) concentration prediction method based on the Convolutional Neural Network-Support Vector Regression (CNN-SVR) model is developed using satellite observations and in-situ ecological water quality measurements in near-shore waters of the Yellow and Bohai Seas. Firstly, we use Pearson method to establish correlation between Chl-a concentration and factors of environmental dynamics and ecological water quality. It is found that Chl-a concentration correlates significantly with nutrient salt factors, while poorly with water quality factors such as pH, dissolved oxygen, salinity. Then, we divide two regions, one is nearshore waters of the southern Bohai Sea and northern Yellow Sea, and the other one is nearshore waters of the central Yellow Sea. We also divide two periods: spring – summer and autumn – winter. We perform the CNN-SVR model experiments with two convolutional kernel sizes,  $1 \times 1$  and  $2 \times 2$ , as well as the single factor sensitivity analysis experiment. The results show that the CNN-SVR network model has better learning of the training data and better prediction of the test samples when the convolution kernel size is  $2 \times 2$ . The CNN-SVR network model performs better in nearshore areas of the southern Bohai Sea and northern Yellow Sea. Compared to water quality factors, the nutrient salt factors have larger impacts on the model's prediction ability. The sensitivity of single factor to model's prediction ability is weak, while multiple variables exhibit complementary feature which improves the model's prediction ability.

**Key words:** CNN-SVR; chlorophyll-a concentration; sensitivity analysis of single variable; ocean satellite; ecological factors of water quality in ocean