

孙新锋, 王晶, 王晶晶, 等. 基于 Sentinel-2 影像与参数优化的随机森林水稻种植分类提取方法[J]. 江苏农业学报, 2026, 42(2): 325-336.

doi: 10.3969/j.issn.1000-4440.2026.02.011

基于 Sentinel-2 影像与参数优化的随机森林水稻种植分类提取方法

孙新锋^{1,2,3}, 王晶^{2,3}, 王晶晶^{2,3}, 李楠^{2,3}, 任妮^{2,3}

(1.淮阴工学院计算机与软件工程学院, 江苏 淮安 223003; 2.江苏省农业科学院农业信息研究所, 江苏 南京 210014; 3.农业农村部长三角智慧农业技术重点实验室, 江苏 南京 210014)

摘要: 及时准确地获取大田水稻的种植面积有利于掌握水稻的种植规模、分布, 为水稻生长动态监测、预测产量提供基础数据。有效挖掘水稻与其他地物的光谱波段特征是卫星遥感提取水稻种植面积的重要依据。本研究以江苏省镇江市为研究区, 选用 Sentinel-2 卫星遥感影像, 将递归特征消除和贝叶斯优化方法结合, 分别构建了基于随机森林的水稻单时相(分蘖期)影像和两时相(分蘖期和扬花期)影像的种植面积提取模型, 并进行了精度检验。结果表明, 水稻单时相影像和两时相影像的种植面积提取模型的总体精度分别为 97.39% 和 98.33%, Kappa 系数分别为 0.96 和 0.97。模型提取的水稻面积与实际水稻面积的对比结果表明, 本模型可实现扬花期水稻面积高精度提取, 同时使用两时相遥感影像构建的模型能够进一步提高分类以及水稻面积提取的精度。

关键词: 多光谱遥感; 水稻面积; 随机森林; Sentinel-2

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-4440(2026)02-0325-12

A method for rice cultivation classification and extraction using Sentinel-2 imagery and parameter-optimized random forest

SUN Xinfeng^{1,2,3}, WANG Jing^{2,3}, WANG Jingjing^{2,3}, LI Nan^{2,3}, REN Ni^{2,3}

(1. Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an 223003, China; 2. Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China; 3. Key Laboratory of Smart Agricultural Technology (Yangtze River Delta), Ministry of Agriculture and Rural Affairs, Nanjing 210014, China)

Abstract: Timely and accurate acquisition of paddy rice planting area is essential for assessing the cultivation scale and spatial distribution, thereby providing fundamental data for monitoring growth dynamics and predicting yield. The effective extraction of spectral band characteristics of rice and other ground objects is a significant basis for mapping rice planting areas using satellite remote sensing. This study focused on Zhenjiang City, Jiangsu province as the research area. Using Sentinel-2 satellite remote sensing imagery and combining recursive feature elimination (RFE) with Bayesian optimization, extraction

收稿日期: 2025-05-21

基金项目: 江苏省重点研发计划(现代农业)项目(BE2023303); 江苏省农业科技自主创新基金项目[CX(23)3006]; 高分辨率对地观测系统国家科技重大专项(74-Y50G12-9001-22/23)

作者简介: 孙新锋(1999-), 男, 河南周口人, 硕士研究生, 研究方向为农业遥感信息技术及应用。(E-mail) 19937696739@163.com

通讯作者: 王晶, (E-mail) wjnj1108@jaas.ac.cn

models for rice planting area were constructed based on random forest, employing both single-temporal (tillering stage) and dual-temporal (tillering and flowering stages) imagery. Accuracy assessments were conducted. The results showed that the overall accuracies of the extraction models using single-temporal and dual-temporal imagery were 97.39% and 98.33%, with Kappa coefficients of 0.96 and 0.97, respectively. A comparison between the model-extracted rice area and the actual rice area indicated that this

model can achieve high-precision extraction of rice area at the flowering stage. Furthermore, the model constructed using dual-temporal remote sensing imagery can further improve classification accuracy and the precision of rice area extraction.

Key words: multispectral remote sensing; rice planting area; random forest; Sentinel-2

水稻是中国重要的粮食作物之一,年播种面积约 $3.0 \times 10^7 \text{ hm}^2$,年产稻谷超 $2 \times 10^7 \text{ t}$,在农业生产中占主导地位^[1],因此及时、准确、大范围地监测水稻的种植面积和空间分布,对提高农业生产管理效率和保障国家粮食安全具有重大意义^[2]。传统的依赖大量实地调查的水稻面积统计方法不仅费时费力、时效性差,而且无法全面反映水稻在大范围内的空间分布情况^[3-5]。近几十年来,卫星遥感技术以实时、快速、准确等优势成为大范围农作物面积监测的重要手段之一,高时空分辨率的遥感影像在农作物监测中的应用越来越广泛,其中哨兵 2 号(Sentinel-2)卫星因具有较高的空间分辨率、丰富的光谱信息以及较短的重访周期,已成为农业资源遥感监测的重要数据来源^[6-8]。基于多时相 Sentinel-2 数据,已有研究通过植被指数时序特征精准刻画作物生长曲线对作物实现高精度分类。孙庆松等^[9]利用最大值合成法将 63 幅 Sentinel-2 影像合成为 30 幅时序影像,构建了植被指数时序集和时序谐波特征,对南宮市冬小麦和夏玉米等主要作物进行精细化分类,总体精度最高达 94.95%。娄轶峰等^[10]利用水稻完整生长期内的 Sentinel-2 数据,构建了一种新型水稻指数,实现了 4 个不同区域的高精度水稻分布制图,总体精度达到 94.50%。Ni 等^[11]将水稻生长期划分为 4 个不同物候期,基于 Sentinel-2 时序数据提出了一种增强像元级物候特征组合方法,在 Google Earth Engine 平台上实现了东北地区高精度水稻制图,总体精度超过 98.00%。Nie 等^[12]利用 Landsat 与 Sentinel-2 的时间序列数据,对不同种植序列的水稻进行识别,结果表明,仅使用覆盖水稻关键物候事件的遥感数据就可实现高精度的水稻种植序列映射,总体精度可达 92.80%。Jiang 等^[13]利用 Sentinel-1/Sentinel-2 时序遥感数据,通过提取水稻在移栽和成熟等关键物候阶段的光谱和极化特征,实现了洞庭湖区水稻种植面积的高精度制图,总体分类精度为 95.44%,成功突破了我国南方多云地区水稻种植面积的监测瓶颈。

现有研究多数使用较长时序的遥感数据进行作物监测,这种方法虽然能够捕捉作物不同生育期的

光谱特征,提高作物监测的精度,但长时间序列数据的获取和处理周期长,时效性差,在作物收获之后甚至收获后第 2 年才能获取分类结果,难以及时反映作物生长现状,无法满足实时田间管理的需要^[14]。在提取方法方面,随机森林模型凭借自主特征优选机制与抗过拟合能力,成为作物分类的主流模型。经典模型通过 Bootstrap 抽样构建决策树群,有效筛选植被指数、物候相位及纹理特征中的关键变量,已有研究表明,随机森林结合先验物候特征可进一步提升分类稳健性,比如 You 等^[15]利用水稻移栽期的光谱特征,结合随机森林算法绘制了黑龙江省的水稻分布图,尽管使用移栽期的影像能够实现良好的制图效果,但该方法易将水稻与湿地、沼泽或其他含水的土地覆盖区混淆^[16]。分蘖期内,水稻迅速生长,形成多数分蘖,导致冠层面积扩大,同时红光和近红外等波段的反射率显著增强;扬花期内,随着花穗的逐渐显现,冠层光谱特征发生显著变化,从而使水稻在光谱响应上更易与其他地物区分^[17]。鉴于此,本研究缩短了监测时序窗口,选择水稻分蘖期与扬花期 2 个关键生育期作为分类时间窗口,利用水稻分蘖期的一景 Sentinel-2A 数据,通过构建基于特征优选和参数优化的随机森林分类模型实现研究区内水稻面积的早期提取,在此基础上,为了进一步提升提取精度,引入水稻扬花期内的一景 Sentinel-2A 数据,通过融合水稻 2 个生育期的时相信息,捕捉其生长过程中的光谱特征变化,增强模型的分类能力和面积提取的准确性,以期为农业生产管理提供更为可靠的数据支撑。

1 材料与方 法

1.1 研究区概况

本研究以江苏省镇江市为研究区。镇江市位于江苏省中南部,地处长江下游南岸,东临长江,西接南京,南望常州,北隔长江与扬州相望,属亚热带季风气候,年平均气温约 $15.5 \text{ }^\circ\text{C}$,年均降水量约为 $1\ 100 \text{ mm}$ 。镇江市土地资源丰富,土壤肥沃,农业生产条件良好。水稻是镇江市最重要的粮食作物之一,当地的水稻种植采用先进的农业技术和管理方法,种植面积大。

1.2 数据采集与处理

Sentinel-2 卫星是欧洲空间局 (ESA) 为哥白尼计划发射的 2 颗高分辨率地球观测光学卫星, 分别为 Sentinel-2A 和 Sentinel-2B, 其搭载的核心传感器是多光谱成像仪 (MSI), 幅宽为 290 km, 集成了 13 个波段, 其中关键波段如蓝、绿、红和近红外波段具有 10 m 的高空间分辨率, 其他波段的分辨率分别为 20 m 或 60 m。根据本研究分类需求, 选择研究区云量小于 10%、成像时间为 2024 年 8 月 23 日和 9 月 22 日的两景 Sentinel-2A 的 L2A 级数据, 使用 SNAP 软件对两景数据的 20 m 和 60 m 空间分辨率波段采用三次卷积插值法重采样至 10 m 分辨率, 然后使用环境影像可视化软件 (ENVI) 对 10 m 分辨率的波段进行合成, 合成后的 Sentinel-2A 数据包含 12 个波段, 分辨率为 10 m。

根据研究区实地观测调研, 本研究将地物划分为 4 类: 水稻、其他植被 (树木、草地、经济作物等)、建筑 (房屋、道路等) 和 水体 (河流、池塘等)。结合谷歌高清地图对 4 种地物类型进行目视解译并在 ENVI 中进行标注。标注得到的样本集共包含 91 178 个像元, 将样本集按照 7:3 划分为训练集和验证集, 地物类型样本标注结果如表 1 所示。

表 1 训练样本与测试样本的像元数目

Table 1 The number of pixels of the training samples and test samples

样本名称	像元数	训练样本数	验证样本数
水稻	39 086	27 360	11 726
其他植被	29 980	20 986	8 994
建筑	11 307	7 915	3 392
水体	10 805	7 563	3 242

1.3 分类特征的构建

挖掘 Sentinel-2A 数据中地物对不同波段的响应特征是提升水稻识别精度的重要手段, 参考国内外作物遥感监测的相关文献^[18-25], 选择归一化差值植被指数 (NDVI)、差值植被指数 (DVI)、增强型植被指数 (EVI)、土壤调整植被指数 (SAVI)、归一化差异水分指数 (NDMI)、改进型归一化水体指数 (mNDWI)、红边归一化差值植被指数 (Ndrel)、简单比值植被指数 (SRI) 8 个植被指数。其中 NDVI 是最常用的植被指数, 它能够反映不同作物的长势; mNDWI 改进了传统的水体指数, 能更加准确地识别

和提取影像中的水体信息; SAVI 能有效减少土壤亮度对植被指数的影响; 使用红边波段计算的 Ndrel 对植被叶绿素含量有较高的敏感性, 可用于估算植被的叶面积指数和叶绿素含量等。

为了实现水稻的早期识别与面积提取, 本研究构建了 2 种特征空间。第一种特征空间仅基于水稻分蘖期的 Sentinel-2A 数据, 从中提取了 12 个波段特征和 8 个植被指数特征, 构成的 20 维特征向量反映了水稻在单一时相下的静态光谱信息和生长状态。第二种特征空间则通过融合水稻分蘖期和扬花期的两景 Sentinel-2A 数据, 分别提取各自的波段特征和植被指数特征后, 在特征维度上采用横向拼接的方式构建了 40 维特征向量, 不仅包含了两时相影像各自的静态光谱信息, 还融入了水稻生长过程中的动态变化信息。

1.4 特征变量优选

为了剔除冗余信息, 降低模型的复杂度, 本研究采用基于随机森林的递归特征消除与交叉验证 (RFECV) 方法来筛选最优特征。RFECV 方法首先构造全部候选特征集合, 通过交叉验证递归剔除对模型贡献较低的特征, 直至获得能使交叉验证性能最佳的特征子集^[26]:

$$S^* = \frac{1}{K} \sum_{k=1}^K \text{Score} [M_S, D_{\text{valid}}^{(k)}] \quad (1)$$

式中, K 为交叉验证的折数, 本研究设 $K=5$, 将全部样本划分为 5 个折, 以保证各折样本类别比例基本一致, 从而降低因样本不平衡带来的偏差; M_S 为仅使用特征子集 S 训练得到的模型, 本研究为随机森林模型; $D_{\text{valid}}^{(k)}$ 为第 k 个折时的验证集; $\text{Score}(\cdot)$ 为性能指标函数, 本研究选用总体准确率和 Kappa 系数衡量模型在验证集上的性能。

为进一步提高模型的可解释性与训练效率, 对 RFECV 方法筛选得到的特征子集按照随机森林特征重要性进行二次排序, 并最终保留排名靠前的关键特征, 用于后续构建水稻分类模型。

1.5 贝叶斯优化随机森林模型

本研究选用的贝叶斯优化 (BO) 方法能够自动搜索和调整随机森林的超参数, 提高模型性能。相比于传统的随机森林模型依赖经验手动或网格搜索的调参方法, BO 方法能够更高效地在超参数空间内找到更佳的参数组合, 进而提升贝叶斯优化随机森林 (BORF) 模型的泛化能力, 并减少过拟合风险。其核

心原理如公式(2)所示,通过统计学中的高斯过程不断优化目标函数 $f(x)$,寻求 $f(x)$ 的最小值^[27]。

$$x^* = \arg \min f(x); x \in U \quad (2)$$

式中, x^* 是待优化的超参数集; U 为输入训练的数据集。

1.6 精度评价方法

为了评价 BORF 模型的分类表现,利用基于像元的混淆矩阵进行分类评价,混淆矩阵能够直观且便捷地展示出各个类别之间的混淆情况,基于混淆矩阵,本研究通过计算总体精度(OA)、用户精度(UA)、生产者精度(PA)和 Kappa 系数(Kappa)评价模型精度:

$$OA = \frac{\sum_{i=1}^n x_{ii}}{N} \quad (3)$$

$$UA = \frac{x_{ii}}{x_{i+}} \quad (4)$$

$$PA = \frac{x_{ii}}{x_{+i}} \quad (5)$$

$$Kappa = \frac{N \sum_{i=1}^n x_{ii} - \sum_{i=1}^n x_{i+} x_{+i}}{N^2 - \sum_{i=1}^n x_{i+} x_{+i}} \quad (6)$$

式中, N 为验证样本总数; x_{ii} 代表混淆矩阵中第*i*行、第*i*列的样本数,代表真实类别为第*i*类的样本中,被正确分类的样本数量; n 为混淆矩阵的行列数,代表分类类别数; x_{ij} 为混淆矩阵中第*i*行、第*j*列上的样本数,代表真实类别为*i*的样本中,有多少被预测为类别*j*; x_{i+} 表示混淆矩阵中第*i*行所有元素之和,即真实类别为*i*的样本总数; x_{+i} 为混淆矩阵中第*i*列的样本总数,即预测为*j*类的样本总数。

1.7 基于随机森林提取水稻种植面积的技术路线图

本研究选取研究区内水稻分蘖期和扬花期内的两景 Sentinel-2A 数据,在进行重采样、波段合成等预处理后,对水稻等地物进行样本标注,分别构建基于单时相与两时相影像的特征空间,使用 RFECV 方法初步筛选最优分类特征,按照随机森林特征重要性排序对得到的特征子集进行二次筛选,确定分类的特征变量组合用于 BORF 模型的超参数优化和训练,最后将训练好的 BORF 模型用于研究区内水稻的分类与面积提取,并进行精度评价,技术路线如图 1 所示。

2 结果与分析

2.1 卫星遥感影像光谱特征变量优选

使用 RFECV 方法对单时相影像和两时相影像

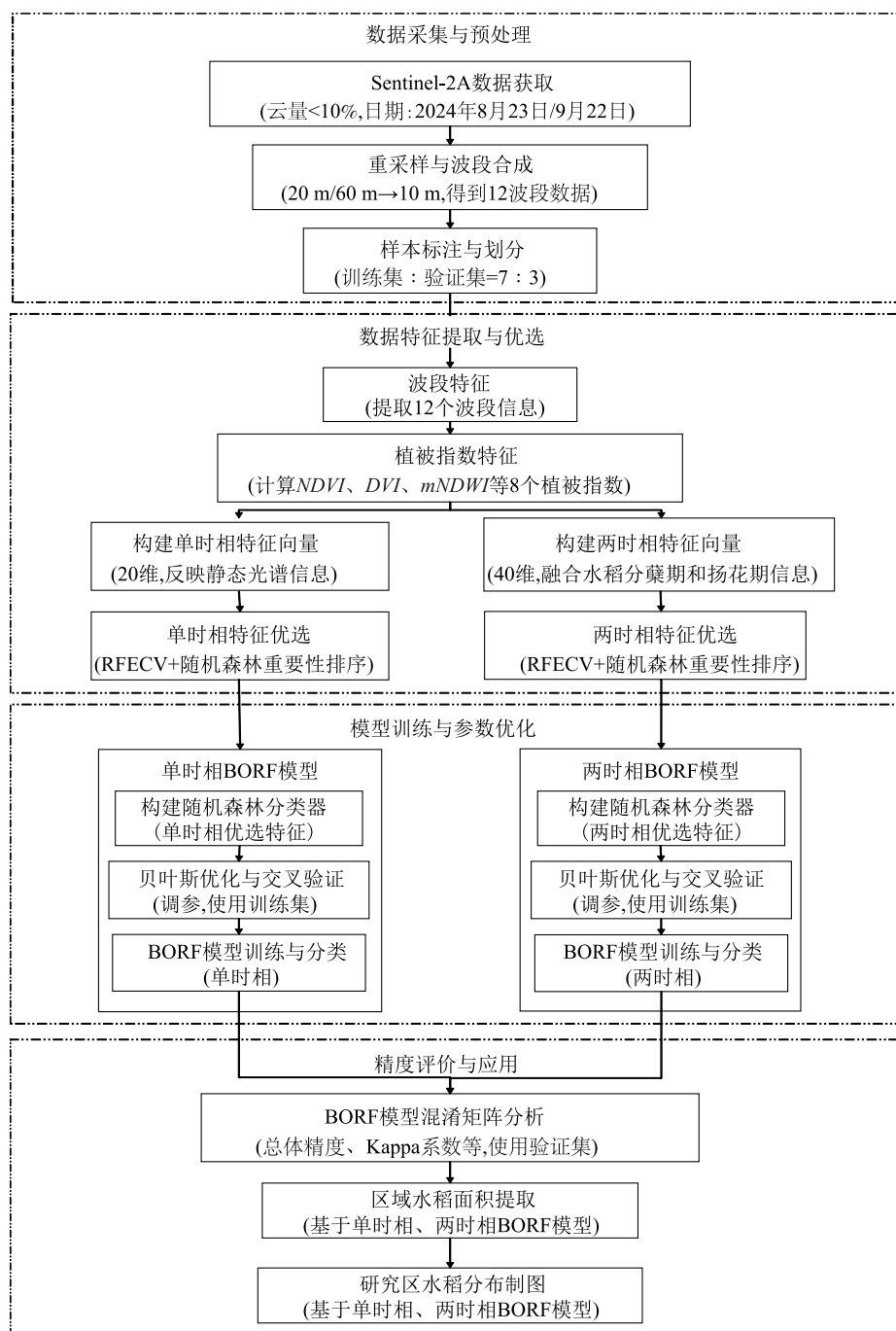
进行特征优选的过程如图 2 所示,图 2 展示了模型的总体精度和 Kappa 系数随特征数量变化的趋势。由图 2 可知,随着特征数量的增加,模型的精度也迅速增加,当特征数量增加到 10 个左右时,精度趋于平稳,此阶段后继续增加特征数量对提高模型性能的贡献有限,因此,选择单时相和两时相模型分类的最佳特征数量为 10,这样能够在保持模型的预测性能的前提下剔除冗余特征,降低模型复杂度。

对 RFECV 方法初步筛选得到的候选特征集进行随机森林特征重要性分析,进一步筛选出对水稻分类具有显著贡献的 10 个关键特征和其重要性(图 3),图 3 可以直观地反映各特征对模型判别能力的相对贡献,为最终分类特征的组合提供明确依据,这些特征既综合反映了分蘖期水稻迅速生长、冠层扩展等早期生长信息,也捕捉了扬花期水稻冠层逐渐呈现花穗特征时的光谱变化,使得整体分类精度和 Kappa 系数均有提升。某些在分蘖期表现突出的特征在扬花期也具有较高的重要性,表明时相融合能为模型提供更丰富的判别信息,筛选出的关键特征显示出互补优势。这种联合筛选策略在剔除冗余特征的同时,保留了对水稻生长关键阶段高度敏感的信息,降低了模型复杂度与过拟合风险,增强了模型的稳定性和泛化能力,为水稻面积的早期提取提供了坚实的数据支撑。

2.2 影像光谱特征超参数优化分析

基于特征优选后的特征子集,采用贝叶斯优化方法对随机森林模型的超参数进行调优,图 4 展示了 5 折交叉验证下,进行 30 次迭代时单时相和两时相模型的总体精度和 Kappa 系数的变化情况。从图中可知,贝叶斯收敛曲线随迭代次数增加呈现出较为平稳的收敛趋势,但在某一迭代区间内出现明显波动,如总体精度突然降低,这是由于贝叶斯优化高度依赖于高斯过程对未知目标函数的建模,而该过程对超参数的选择十分敏感^[28],当贝叶斯优化在迭代时动态探索到敏感区域时,因关键超参数如决策树最大深度的微小变化导致模型捕捉数据特征的能力下降,反映为在超参数空间中的某些敏感拐点处(如图 5 中决策树最大深度较小时),因参数小幅调整引起预测的不确定性波动,导致评估指标暂时下降。由于超参数空间的复杂性或噪声的存在,在超参数调优的探索阶段常会遇到这种波动现象,随着贝叶斯优化对超参数空间的全局搜索更加充分,模型最终将会收敛于性能较

优且稳定的参数组合^[29]。



NDVI:归一化差值植被指数; DVI:差值植被指数; mNDWI:改进型归一化水体指数。RFECV:递归特征消除与交叉验证法; BORF:贝叶斯优化随机森林模型。

图1 基于随机森林提取水稻种植面积的技术路线图

Fig.1 The technical roadmap for extracting rice cultivation areas based on random forest

图5展示了贝叶斯优化过程中各个超参数的变化趋势及其对应的模型分类精度变化情况,其中决策树最大深度对模型精度的影响最大,其作用在于

控制单棵决策树的最大分裂深度,从而决定模型捕捉数据特征的复杂程度。在贝叶斯优化过程中,找到“最佳深度”是确保模型既能充分拟合数据又保

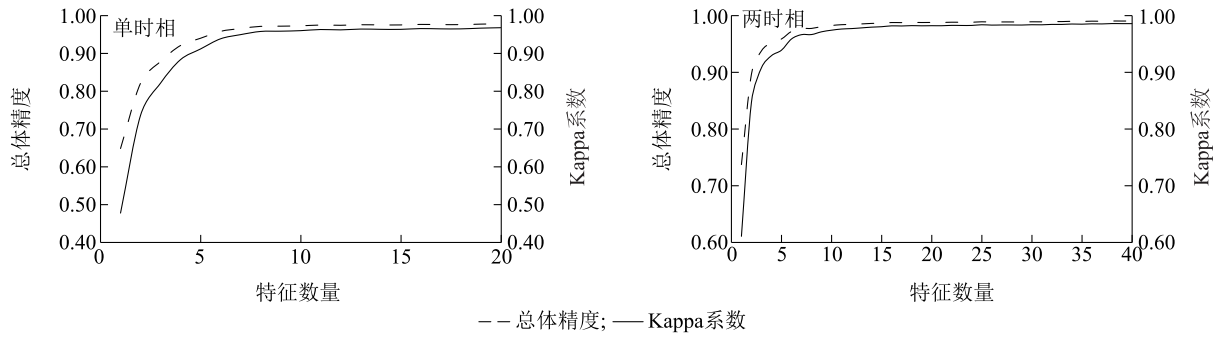
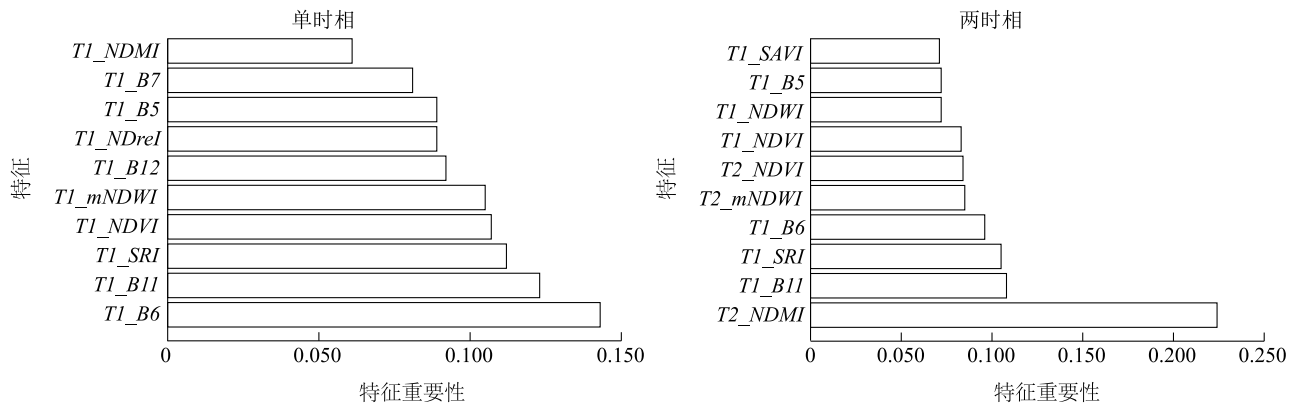


图 2 递归特征消除与交叉验证法 (RFECV) 特征优选过程

Fig.2 Feature selection process using recursive feature elimination and cross-validation (RFECV)



T1_NDMI: 基于分蘖期影像计算的归一化差异水分指数; *T1_B7*: 分蘖期影像 B7 波段的反射率值; *T1_B5*: 分蘖期影像 B5 波段的反射率值; *T1_NDrel*: 基于分蘖期影像计算的归一化差值植被指数; *T1_B12*: 分蘖期影像 B12 波段的反射率值; *T1_mNDWI*: 基于分蘖期影像计算的改进型归一化水体指数; *T1_NDVI*: 基于分蘖期影像计算的归一化差值植被指数; *T1_SRI*: 基于分蘖期影像计算的简单比值植被指数; *T1_B11*: 分蘖期影像 B11 波段的反射率值; *T1_B6*: 分蘖期影像 B6 波段的反射率值; *T1_SAVI*: 基于分蘖期影像计算的土壤调整植被指数; *T1_NDWI*: 基于分蘖期影像计算的归一化水体植被指数; *T2_NDVI*: 基于扬花期影像计算的归一化差值植被指数; *T2_mNDWI*: 基于扬花期影像计算的改进型归一化水体指数; *T2_NDMI*: 基于扬花期影像计算的归一化差异水分指数。

图 3 特征优选结果

Fig.3 Results of feature selection

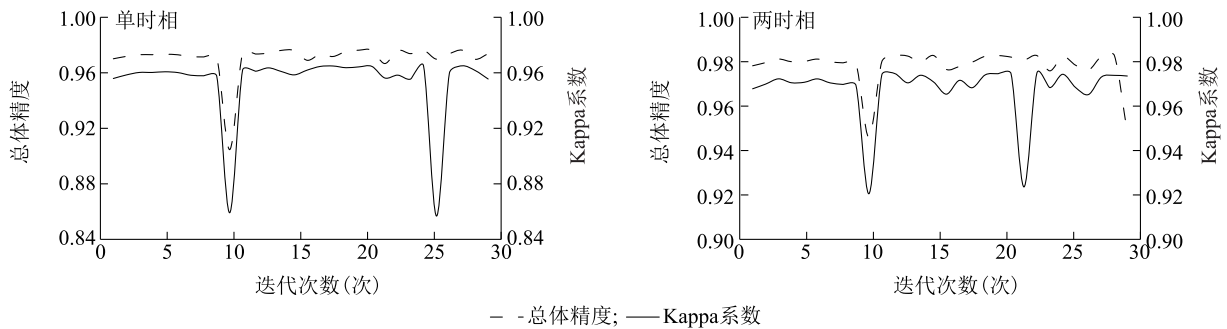


图 4 贝叶斯优化迭代过程

Fig.4 Bayesian optimization iterative process

持良好泛化能力的关键。从图 5 中可以看出,当决策树最大深度较小时,模型的总体精度明显偏低,随着决策树最大深度增大,分类精度迅速上升,直至趋于平稳,这表明在适当的深度范围内,模型既能表达

数据特征,又避免了计算资源浪费和过拟合风险,在这一阶段,决策树最大深度足以捕获大部分关键特征,而进一步增大深度仅带来微小的精度提升,甚至可能引入噪声,导致过拟合现象。

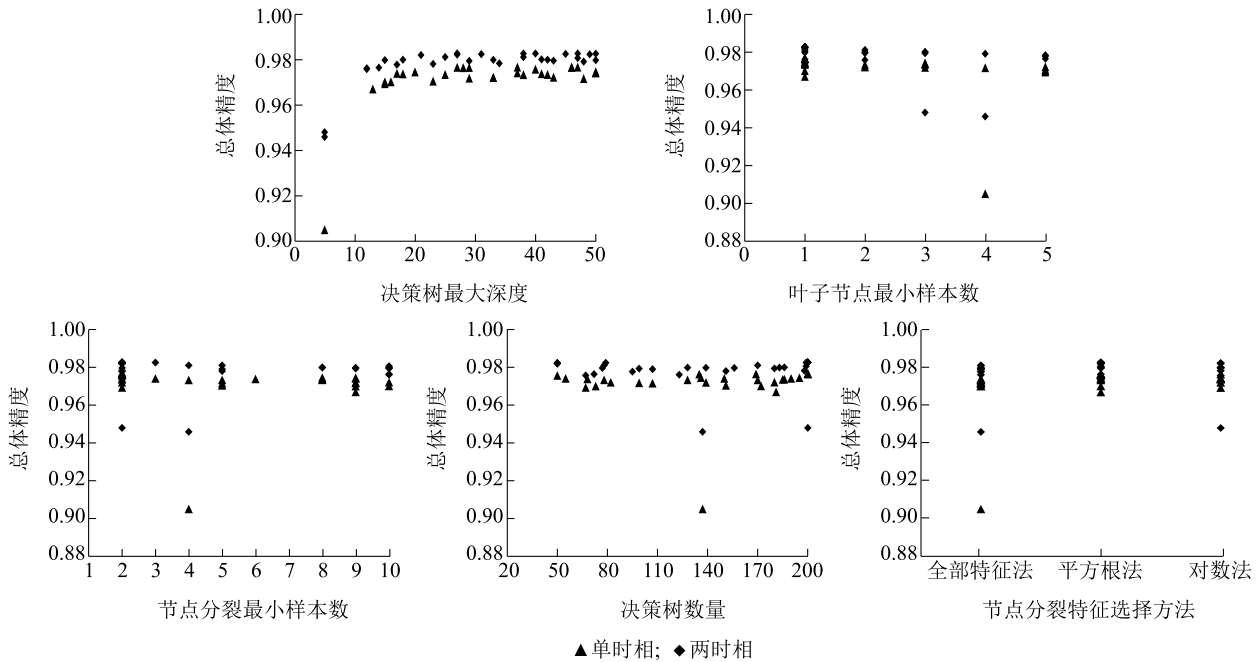


图5 超参数变化趋势

Fig.5 Trend of hyperparameter changes

表2展示了在5折交叉验证下,当贝叶斯优化的迭代次数设为30次时,分别迭代了25次和23次时,得到的单时相和两时相模型的超参数调优结果,此时模型取得了最优的总体精度和Kappa系数。无论使用单时相还是两时相数据,模型的超参数配置均趋于一致,获得了较为稳定的调参结果,都具有较深的决策树最大深度、较小的节点分割约束(节点分裂所需最小样本数)、较多的决策树数量以及均采用了基于平方根的特征采样方法。这一结果与前期的特征优选有关,通过特征优选剔除冗余变量,输

入数据的维度得到大幅降低,使得模型仅需关注10个高判别性特征,从而使得贝叶斯优化能集中在一个较小的参数空间内找到稳定的和最优的配置,减少了超参数调优时需要探索的参数组合。这体现了特征优选对减少噪音、提升模型鲁棒性的积极作用。通过将特征优选与贝叶斯优化相结合,不仅提高了超参数调优的效率和稳定性,也为后续基于随机森林构建水稻分类模型提供了坚实且高效的参数基础,确保了模型在早期水稻面积提取任务中的高精度和良好泛化能力。

表2 水稻种植面积提取模型的参数调优结果

Table 2 Results of parameter tuning for the rice cultivation area extraction model

参数名	意义说明	参数空间	单时相模型参数	两时相模型参数
max_depth	决策树最大深度	[5,50]	37	28
min_samples_leaf	叶子节点最小样本数	[1,5]	1	1
min_samples_split	节点分裂最小样本数	[2,10]	2	2
n_estimators	决策树数量	[50,200]	200	200
max_features	节点分裂特征选择方法	['sqrt','log2','none']	sqrt	sqrt

sqrt:平方根法;log2:对数法;none:全部特征法。

2.3 水稻种植面积提取模型精度验证

为了定量评估使用单时相数据和两时相数据构建的BORF模型的水稻分类效果,基于验证样本计算了模型的混淆矩阵,从中可以直观看出各类别样

本的正确分类数与误分类情况。如图6所示,2种时相数据下的分类模型总体表现均较为理想,但两者之间仍存在一定差异,对于“水稻”类别,在单时相下,其正确分类像元数量为11 521,而在两时相数

据中,正确分类的像元数上升到11 563,“其他植被”类别的正确分类像元数由单时相模型的8 641提升

至两时相模型的8 734,体现了多时相数据在区分水稻与其他植被时的优势。

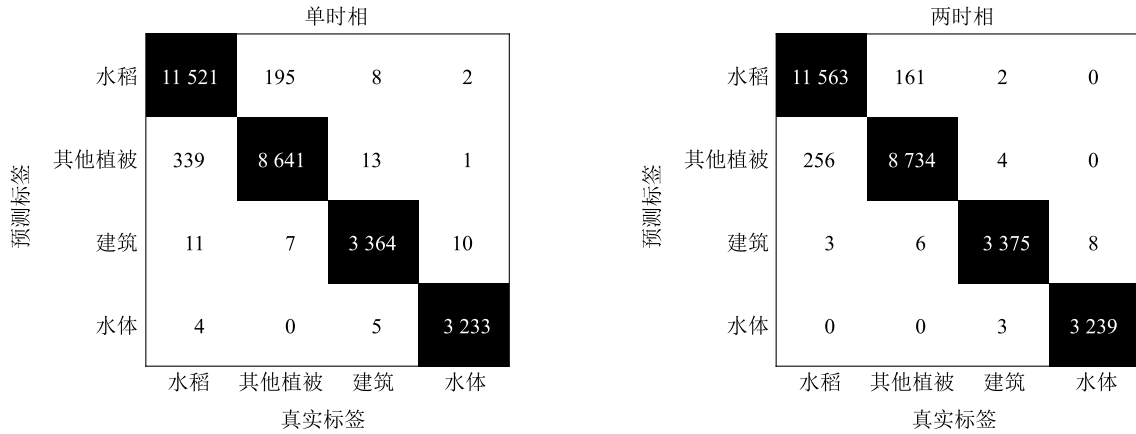


图 6 混淆矩阵结果

Fig.6 Confusion matrix results

基于混淆矩阵计算得到的模型总体精度、Kappa 系数、水稻的生产者和用户精度结果如表 3 所示,单时相模型总体精度为 97.39%,Kappa 系数为 0.96,而两时相模型总体精度提升到 98.33%,Kappa 系数提高至 0.97。这表明在引入时序信息后,模型对各类别间混淆的抑制作用更加明显。在“水稻”类别上,单时相和两时相数据的 BORF 模型分类表现均十分优异,生产者精度由 98.29%增加至 98.70%,用户精度则从 97.15%提升至 97.71%,均保持在较高水平。

表 3 水稻种植面积提取单、两时相模型精度对比

Table 3 Accuracy comparison of rice cultivation area extraction between single-temporal and dual-temporal models

精度指标	单时相 BORF 模型	两时相 BORF 模型
总体精度 (%)	97.39	98.33
水稻生产者精度 (%)	98.29	98.70
水稻用户精度 (%)	97.15	97.71
Kappa 系数	0.96	0.97

BORF:贝叶斯优化随机森林模型。

使用单时相和两时相数据构建的 BORF 模型对镇江全市范围内的水稻进行分类和面积提取,提取结果如表 4 所示,镇江市 2024 年全市水稻种植面积约为 72 000 hm²,主要种植区分布在丹阳市、句容市和丹徒区,扬中市和京口区也有少量水稻种植区分布。单时相和两时相 BORF 模型水稻面积提取结果分别为 82 579 hm²和 64 408 hm²。在除了句容市外的区域,结合分蘖期与扬花期数据的两时相模型显

示出更高的面积提取精度,两时相数据融合了 2 个时相的水稻表征信息,不仅提供了稳定的光谱特征,还弥补了单时相数据在复杂环境下难以捕捉动态变化的不足。这种方法在混合、零散分布的稻田面积提取中表现尤为突出,能有效区分水稻与邻近的山体、森林等地物,减少混淆,提高总体提取结果。然而,句容市的提取结果却未能体现出两时相数据的优势,面积误差明显较大。由于句容市山体和森林覆盖较多,土地利用类型呈现高度杂乱,稻田分布零散,这种复杂的地表特征使得模型在提取水稻面积时,容易将部分稻田误判为森林、山地或者其他过渡型地物,从而造成较大的误差。同时,稻田的光谱信号也可能受到阴影、大气散射等多重因素影响,进一步加剧了模型对细微目标的提取难度,降低了模型对该区域稻田的判别能力。

表 4 研究区域水稻种植面积提取结果对比

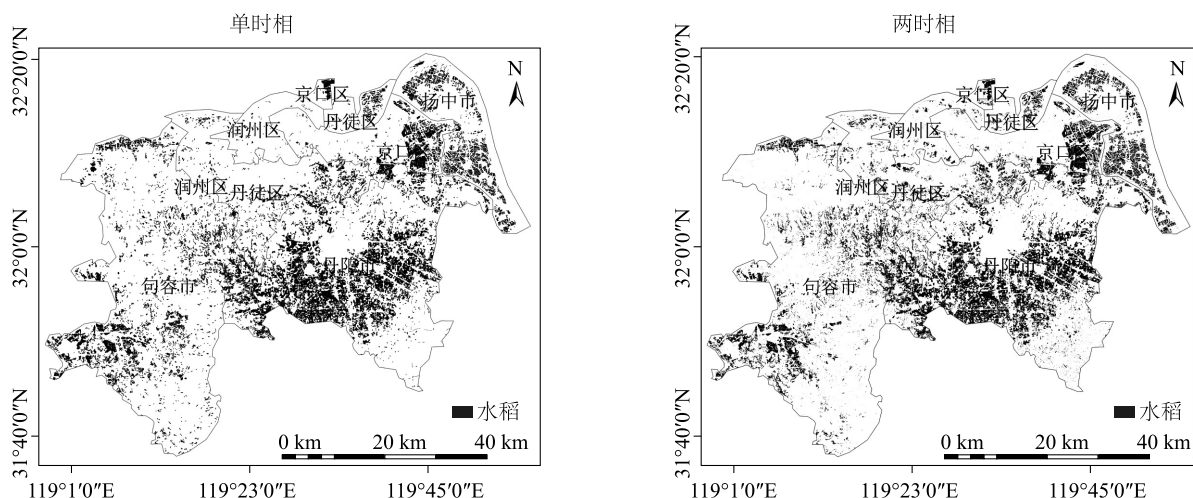
Table 4 Comparison of rice cultivation area extraction results in the study area

区域	水稻种植面积 (hm ²)		
	统计数据	单时相模型预测数据	两时相模型预测数据
镇江市	72 000	82 579	64 408
丹徒区	10 000	12 898	8 444
丹阳市	30 000	33 583	28 759
京口区	5 000	9 011	6 119
句容市	22 000	19 078	13 749
扬中市	5 000	8 009	7 337

2.4 研究区域水稻种植面积提取及制图

图 7 展示了使用单时相和两时相构建的 BORF 模型的镇江市水稻分类结果,其中单时相分类结果存在过提取现象,原因可能是部分非稻田区域(例如周边的其他作物,如大豆等)被错误归类为水稻,导致水稻面积被高估。此外当模型在提取破碎零散地块时,单时相数据的提取在边界处也较为模糊,缺乏足够的动态信息去细化和区分不同地物的边缘,即使在较小的分类误差下,也会在总面积上产生较

大影响。此外混合像元的存在使得水稻光谱特征容易受到干扰,导致背景被误包含,进而加剧模型欠提取和过提取的现象,增大了面积提取的误差。相比之下,两时相数据的提取结果显示了一定程度的改进。融合了扬花期的补充信息,两时相数据能更好地捕捉水稻在不同生育阶段的光谱变化,从而减少了非目标区域的误提,同时使水稻区域的边界更加清晰、准确。



审图号为 GS(2024)0650 号,底图边界无修改。

图 7 基于随机森林的研究区域水稻种植面积的提取和制图

Fig.7 Mapping of rice cultivation areas in the study area based on random forest

图 8 展示了单时相和两时相数据的局部水稻面积提取对比结果,从图 8 中可以明显地看出,在采用两时相数据进行水稻面积提取时,模型能更准确地捕捉到目标区域的细微特征,比如道路、水体边界等区域,其轮廓表现得更加细致、边界更加平滑、分类更加清晰。相比之下,单时相模型由于仅依赖单一时间点的数据,其特征区分能力受限,在一些地物较为复杂的区域,例如地物边界模糊、地块分布零散的区间,由于背景杂乱、地物混合以及“异物同谱”以及“同谱异物”现象的影响,模型在提取时容易受到噪声和干扰的影响,导致难以区分真实的水稻区域与周边水体或杂草等异质结构,模型往往出现误提取和过提取现象,进而影响整体分类的效果。

3 讨论

本研究构建的基于 Sentinel-2A 影像的 BORF

模型在水稻面积早期提取中取得了较高精度,但在实际应用中仍存在一些局限性,首先,本研究主要依赖单一光学数据源,其在云影响、阴影以及混合像元问题较为严重的区域容易出现细节分割不够精细的情况。此外,尽管通过 RFECV 方法筛选关键特征和贝叶斯优化法确定超参数有效地降低了模型复杂度,提高了模型在研究区内的水稻面积提取精度,但在更大范围的区域应用时,需要注意计算量和模型精度的问题,大范围区域有更丰富的地物类型,区域的异质性增加,原本在局部区域内优化得到的关键特征组合和超参数设置有可能不完全适用于异质性更高的区域,从而导致模型的表现不如预期,并且特征选择和参数优化的计算量也会随着数据规模和特征复杂性的增加而显著上升,在大范围的应用场景中,模型的复杂度与普适性尚需进一步验证,针对这些局限性,未来研究可以从以下几个方向进行改进:

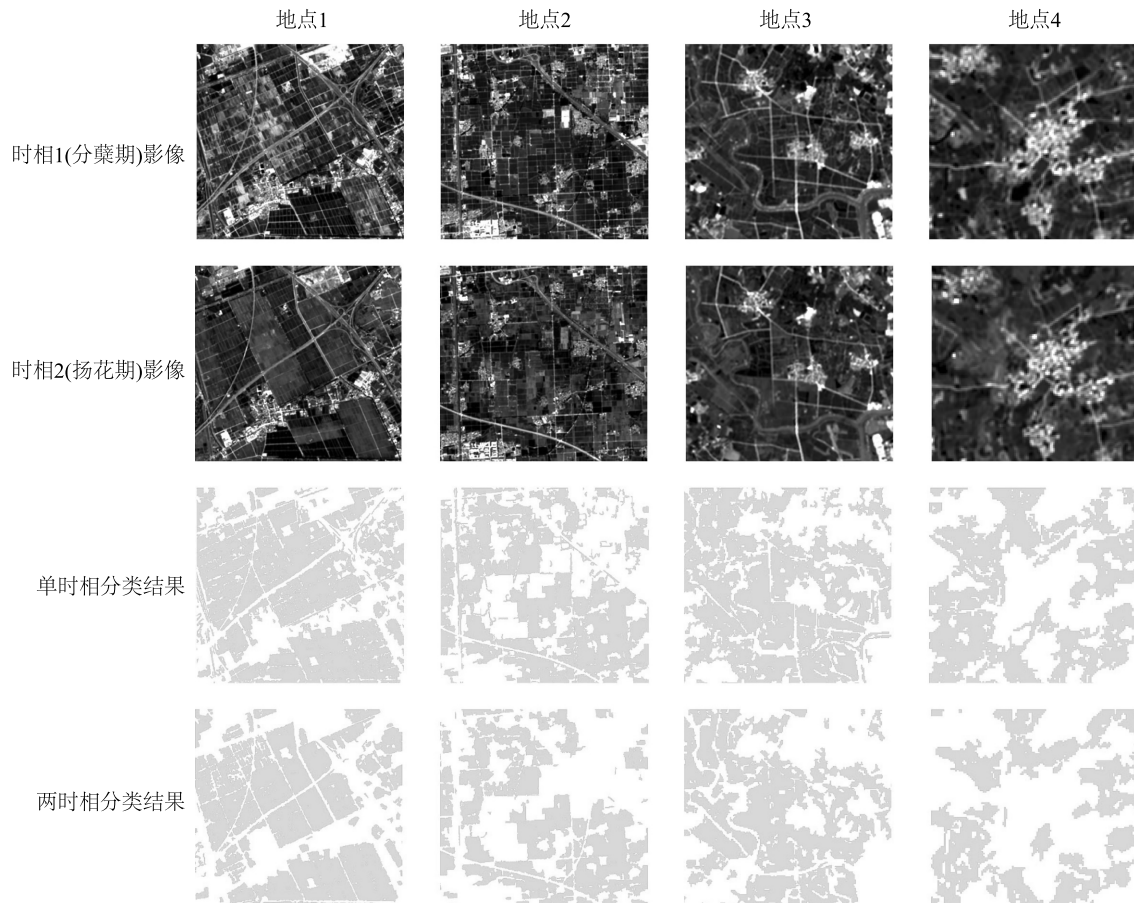


图8 BORF模型局部分类结果

Fig.8 Local classification results of BORF model

1) 多源数据融合:结合 Sentinel-1 的合成孔径雷达(SAR)数据、高分辨率卫星影像或激光雷达(LiDAR)数据,可以充分弥补单一光学数据在云影响、低反射状况下的信息不足。多传感器数据的互补优势已在农作物分类与生长监测中得到证明;可在保留高空间分辨率优势的同时,增强模型在恶劣气象条件下的适用性^[30]。

2) 面向对象的作物分类方法:目前基于像元的分类方法在处理混合像元和地块边界时存在一定局限性。应用面向对象的图像分析方法,通过影像分割得到具有相似光谱及纹理特征的区域,再结合形状和纹理信息进行分类,可以有效提高精细边界提取的准确性^[31]。

3) 深度学习与传统方法相融合:利用深度学习网络[如卷积神经网络(CNN)]进行多时相特征提取,再结合随机森林或其他经典机器学习分类器的集成方法,是提升模型对复杂场景识别能力的又一

可行途径。深度学习在捕捉多尺度、高层次特征方面的优势,有望进一步提升作物面积提取的泛化能力和实时性^[32]。

4 结论

本研究通过结合 RFECV 特征优选和贝叶斯参数优化技术,构建了基于单时相和两时相的 BORF 模型,用于实现水稻面积早期提取任务。研究以水稻分蘖期和扬花期为监测窗口,通过单时相和两时相数据构造不同特征空间,对研究区内的水稻进行了分类和面积提取,并评估了模型精度,主要结论如下:

1) 特征优选与模型稳健性:利用 RFECV 方法进行特征优选,有效剔除了特征空间中的冗余信息和噪声,保证了输入模型的特征具有较高的判别性与稳定性。这一过程不仅降低了模型复杂度,同时提供了更加一致的特征输入,为后续的贝叶斯参数

优化创造了良好条件。特征空间的缩减使得 BORF 模型能够在较小的搜索区域内精细定位最优超参数组合,从而进一步提高了模型的总体准确率和鲁棒性。

2) 单时相与两时相模型对比: 试验结果表明, 单时相模型能够在水稻面积早期提取中获得较高的分类精度和稳定性, 但在复杂背景区域, 如地块破碎或背景复杂的情况下, 模型较容易受相似光谱干扰而发生提取。相比之下, 两时相模型融合了分蘖期与扬花期的影像信息, 能够捕捉水稻生长过程中的动态光谱变化, 对“同物异谱”和“同谱异物”问题有更好的区分能力, 从而进一步优化了区域目标的边界提取。局部提取结果显示, 两时相模型在道路边界和破碎地块的提取上更为精细, 错误提取和过提取现象明显降低。

3) 水稻早期面积提取的优势: 由于水稻在早期生长阶段光谱特征较弱且容易与其他地物混淆, 采用单时相数据时可能难以充分反映目标信息。两时相模型融合了水稻在不同生育阶段的信息, 不仅提升了整体分类精度(总体精度和 Kappa 系数均有一定提升), 而且进一步强化了水稻与其他地物之间的区分性, 在实际分类中显示出更高的可靠性和适用性。利用关键影像在扬花期前即可实现快速、准确的水稻面积提取, 从而为农业生产管理提供及时的数据支持。

参考文献:

- [1] 徐春春, 纪 龙, 陈中督, 等. 中国水稻生产、市场与进出口贸易的回顾与展望[J]. 中国稻米, 2021, 27(4): 17-21.
- [2] 陈 岳, 寇卫利, 李 莹, 等. 农作物遥感灾损评估研究进展[J]. 江苏农业科学, 2022, 50(16): 12-20.
- [3] 李卫国, 赵春江, 王纪华, 等. 基于卫星遥感的冬小麦拔节期长势监测[J]. 麦类作物学报, 2007, 27(3): 523-527.
- [4] 潘 力, 夏浩铭, 王瑞萌, 等. 基于 Google Earth Engine 的淮河流域越冬作物种植面积制图[J]. 农业工程学报, 2021, 37(18): 211-218.
- [5] 张 宏, 李卫国, 张晓东, 等. 基于 HJ-1 星和 GF-1 号影像融合特征提取冬小麦种植面积[J]. 中国农业科技导报, 2024, 26(2): 109-119.
- [6] 金 晶, 毛 星, 张 欣, 等. 基于 Sentinel-2 影像的淡水养殖水生动物类型识别研究[J]. 河南农业科学, 2022, 51(4): 160-170.
- [7] 王美月, 万 红, 刘法军, 等. 基于哨兵二号红边波段特征的水稻分类[J]. 山东农业大学学报(自然科学版), 2025, 56(1): 11-20.
- [8] 于天祥, 樊 红. 基于 Sentinel-2 多时相遥感影像的冬小麦种植面积监测[J]. 江苏农业学报, 2024, 40(9): 1653-1661.
- [9] 孙庆松, 张晓楠, 陈利东, 等. 基于 Sentinel-2 时序谐波特征的县域农作物分类[J]. 江苏农业学报, 2022, 38(4): 967-975.
- [10] 姜轶峰, 黄 可, 杨 刚, 等. 基于 Sentinel-2 时间序列数据的水稻分布提取方法[J]. 遥感学报, 2025, 29(5): 1260-1272.
- [11] NI R G, TIAN J Y, LI X J, et al. An enhanced pixel-based phenological feature for accurate paddy rice mapping with Sentinel-2 imagery in Google Earth Engine[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 178: 282-296.
- [12] NIE H Z, LIN Y C, LUO W F, et al. Rice cropping sequence mapping in the tropical monsoon zone via agronomic knowledge graphs integrating phenology and remote sensing[J]. Ecological Informatics, 2025, 87: 103075.
- [13] JIANG Q, TANG Z G, ZHOU L H, et al. Mapping paddy rice planting area in Dongting lake area combining time series sentinel-1 and sentinel-2 images[J]. Remote Sensing, 2023, 15(11): 2794.
- [14] 朱 磊, 张伟业, 潘自林, 等. 基于 Sentinel-2 的青铜峡灌区春小麦和苜蓿早期识别[J]. 灌溉排水学报, 2024, 43(5): 86-94.
- [15] YOU N S, DONG J W, HUANG J X, et al. The 10-m crop type maps in NorthEast China during 2017-2019[J]. Scientific Data, 2021, 8(1): 41.
- [16] ZHAN P, ZHU W Q, LI N. An automated rice mapping method based on flooding signals in synthetic aperture radar time series[J]. Remote Sensing of Environment, 2021, 252: 112112.
- [17] CHEN Y P, HU J, CAI Z W, et al. A phenology-based vegetation index for improving ratoon rice mapping using harmonized Landsat and Sentinel-2 data[J]. Journal of Integrative Agriculture, 2024, 23(4): 1164-1178.
- [18] HUANG S, TANG L N, HUPY J P, et al. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing[J]. Journal of Forestry Research, 2021, 32(1): 1-6.
- [19] SAN BAUTISTA A, FITA D, FRANCH B, et al. Crop monitoring strategy based on remote sensing data (Sentinel-2 and Planet), study case in a rice field after applying glycinebetaine[J]. Agronomy, 2022, 12(3): 708.
- [20] MUNYATI C. Detecting the distribution of grass aboveground biomass on a rangeland using Sentinel-2 MSI vegetation indices[J]. Advances in Space Research, 2022, 69(2): 1130-1145.
- [21] 李元庆, 宋宏利, 刘 欢, 等. 基于 Sentinel-2 的江淮分水岭撂荒地提取[J]. 安徽农业科学, 2025, 53(3): 47-52, 76.
- [22] BADRUL HISHAM N H, HASHIM N, SARAF N M, et al. Monitoring of rice growth phases using multi-temporal sentinel-2 satellite image[J]. IOP Conference Series: Earth and Environmental Science, 2022, 1051(1): 012021.
- [23] 张征云, 江文渊, 张彦敏, 等. 基于哨兵 SAR 数据和多光谱数据的水稻识别研究[J]. 生态与农村环境学报, 2023, 39(4): 556-564.

- [24] SEKERTEKIN A. A survey on global thresholding methods for mapping openwater body using sentinel-2 satellite imagery and normalized difference water index [J]. Archives of Computational Methods in Engineering, 2021, 28(3):1335-1347.
- [25] 陈健, 李虎, 刘玉锋, 等. 基于 Sentinel-2 数据多特征优选的农作物遥感识别研究[J]. 自然资源遥感, 2023, 35(4):292-300.
- [26] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in Python [J]. The Journal of Machine Learning Research, 2011, 12:2825-2830.
- [27] 薛一阳, 竞霞, 叶启星, 等. 利用约束随机森林和贝叶斯优化算法的小麦条锈病遥感监测[J]. 遥感技术与应用, 2025, 40(1):69-76.
- [28] SNOEK J, LAROCHELLE H, ADAMS R P. Practical bayesian optimization of machine learning algorithms[J]. Advances in Neural Information Processing Systems, 2012, 25:2960-2968.
- [29] SHAHRIARI B, SWERSKY K, WANG Z Y, et al. Taking the human out of the loop: a review of Bayesian optimization [J]. Proceedings of the IEEE, 2016, 104(1):148-175.
- [30] SAMADZADEGAN F, TOOSI A, DADRASS JAVAN F. A critical review on multi-sensor and multi-platform remote sensing data fusion approaches: current status and prospects [J]. International Journal of Remote Sensing, 2025, 46(3):1327-1402.
- [31] 曹小鸿. 面向对象的遥感影像信息提取方法研究[J]. 测绘与空间地理信息, 2022, 45(5):116-119, 123.
- [32] BHATTI M A, SYAM M S, CHEN H F, et al. Utilizing convolutional neural networks (CNN) and U-Net architecture for precise crop and weed segmentation in agricultural imagery: a deep learning approach [J]. Big Data Research, 2024, 36:100465.

(责任编辑:陈海霞)