

DOI:10.11686/cyxb2024110

http://cyxb.magtech.com.cn

麦晶晶, 冯琦胜, 王瑞泾, 等. 基于机器学习的高精度耕地识别模型构建——以甘肃省张掖市为例. 草业学报, 2025, 34(2): 149-162.

MAI Jing-jing, FENG Qi-sheng, WANG Rui-jing, et al. Construction of a high-precision cultivated land identification model based on machine learning—using Zhangye City, Gansu Province as an example. Acta Prataculturae Sinica, 2025, 34(2): 149-162.

基于机器学习的高精度耕地识别模型构建 ——以甘肃省张掖市为例

麦晶晶¹, 冯琦胜^{1*}, 王瑞泾², 封森耀³, 金哲人⁴, 张忠雪¹, 梁天刚¹, 金加明⁵

(1. 兰州大学草种创新与草地农业生态系统全国重点实验室, 兰州大学草地农业科技学院, 兰州大学寒旱区生态环境遥感研究中心, 甘肃兰州 730020; 2. 北京师范大学地表过程与资源生态国家重点实验室, 北京师范大学地理科学学部, 北京 100875; 3. 清华大学深圳国际研究生院, 环境与生态研究院, 广东深圳 518055; 4. 苏州市吴江区七都镇农业服务中心, 江苏苏州 215200; 5. 甘肃省草原技术推广总站, 甘肃兰州 730070)

摘要:耕地是农业生产和保障粮食安全重要的物质基础, 耕地的准确识别对耕地资源的保护和农业生产可持续发展有着重要意义。为了构建高精度的耕地识别模型, 本研究基于空间云计算平台使用 Sentinel-1/2 数据, 构建不同特征类型组合, 通过特征重要性分析对耕地识别特征进行筛选, 形成最优特征集合, 使用随机森林 (random forest, RF)、支持向量机 (support vector machine, SVM) 和分类回归树 (classification and regression tree, CART) 模型对甘肃省张掖市 2021 年度的耕地进行识别, 同时对比分析了各分类器的分类精度。结果表明, 使用植被指数特征、雷达特征和地形特征的特征类型组合能够把分类精度提升到 91.32%; 在研究区耕地识别中表现较好的特征有海拔 (elevation)、雷达 VH 极化通道及归一化水指数 (normalized difference water index, NDWI) 等; 在张掖市耕地识别中, RF 算法优势明显, 总精度达 90.04%, Kappa 系数为 0.79, 基于 RF 模型得到的张掖市耕地面积为 58.5 万 hm^2 , 面积占比为 15.4%。本研究实现了张掖市耕地的精确识别, 可为该地区耕地制图提供参考。

关键词:耕地识别; 机器学习; 随机森林; 哨兵卫星

Construction of a high-precision cultivated land identification model based on machine learning—using Zhangye City, Gansu Province as an example

MAI Jing-jing¹, FENG Qi-sheng^{1*}, WANG Rui-jing², FENG Sen-yao³, JIN Zhe-ren⁴, ZHANG Zhong-xue¹, LIANG Tian-gang¹, JIN Jia-ming⁵

1. State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Grassland Industry, Ministry of Education, College of Pastoral Agriculture Science and Technology, Lanzhou University, Center for Remote Sensing of Ecological Environments in Cold and Arid Regions, Lanzhou University, Lanzhou 730020, China; 2. State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; 3. Institute of Environment and Ecology, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; 4. Agricultural Service Center of Qidu Town, Wujiang District, Suzhou City, Suzhou 215200, China; 5. Grassland Technology Extension Station of Gansu Province, Lanzhou 730070, Gansu, China

Abstract: Cultivated land is a vital foundation resource for agricultural production and ensuring food security. Accurate identification of cultivated land is of great significance for the conservation of cultivable land resources and

收稿日期: 2024-04-09; 改回日期: 2024-06-20

基金项目: 财政部和农业农村部: 国家现代农业产业技术体系 (CARS-34), 甘肃省林业和草原局科技创新项目 (kjcx2022010), 2023 年提前批中央财政林业改革发展资金草原科技支撑项目 (甘林草发[2023]211 号) 和近自然恢复技术在退化草地修复中的应用与示范项目资助。

作者简介: 麦晶晶 (1998—), 女, 广东广州人, 在读硕士。E-mail: 220220902470@lzu.edu.cn

* 通信作者 Corresponding author. E-mail: fengqsh@lzu.edu.cn

the sustainable development of agricultural production. In order to construct a high-precision cultivated land identification model, this study used Sentinel-1/2 data together with the spatial cloud computing platform and built combinations of different feature types. Through feature importance analysis, cultivated land identification features were then evaluated to identify the optimal feature set. Random Forest (RF), support vector machine (SVM), and classification and regression tree (CART) models were employed to identify the cultivated land in Zhangye City, Gansu Province for the year 2021. Simultaneously, the classification accuracy of each classifier was compared and analyzed. The results show that using a combination of vegetation index features, radar features, and topographic features improved the classification accuracy to 91.32%; Features that performed well in cultivated land identification in the study area included elevation, radar polarization channel VH, and normalized difference water index (NDWI). In the cultivated land identification of Zhangye City, RF algorithm demonstrates clear advantages, with an overall accuracy of 90.04% and a Kappa coefficient of 0.79. Based on the RF model, the cultivated land area associated with Zhangye City is estimated to be 585000 ha, accounting for 15.4% of the total area. The methodology developed in this study achieves accurate identification of cultivated land in Zhangye City and offers a tool for cultivated land mapping in the region.

Key words: identification of cultivated land; machine learning; random forest; Sentinel

耕地是保障粮食安全的物质基础,快速、准确地提取耕地的面积及分布能为耕地变化监测、粮食政策制定等提供数据支撑,对耕地资源保护和国家粮食安全的保障具有重要意义。遥感技术与机器学习算法的结合,为耕地信息提取提供了新的技术手段。基于遥感数据使用机器学习算法准确快速地提取耕地分布信息,对保障粮食安全、生态安全及对该地区草地农业可持续发展有重要的现实意义。

卫星遥感技术在获取数据方面具有宏观、精确、快速等优点,能够有效地获取土地覆盖信息,国内外许多学者使用MODIS(中分辨率成像光谱仪 moderate-resolution imaging spectroradiometer)^[1-2]、Landsat^[3-4] TM/ETM+/OLI和Sentinel-2^[5-7]等卫星遥感数据进行耕地信息的提取。近年来随着智慧农业的发展,中低分辨率的光学遥感数据逐渐无法满足日益严苛的研究需求^[8],高分辨率的光学遥感数据虽然在一定程度上能够提高识别精度,但容易受到云等的影响,并且无法完全解决“同物异谱”和“同谱异物”的问题^[9]。合成孔径雷达数据(synthetic aperture radar, SAR)具有穿透云层的能力,然而由于耕地的斑块化和高度复杂性,使用雷达数据在耕地提取的过程中会产生较多的噪声^[10]。

相比于单源数据,多源数据的结合使用能够增强耕地区域在物候、空间及纹理等特征的差异信息^[11],从而优化分类结果,因此,许多研究融合多源数据对耕地进行识别^[8],雷达数据和光学数据的融合使用是一种常见的方法^[12],随着高空间分辨率的Sentinel-2数据的公开,许多学者采用该数据结合雷达数据进行研究,Sun等^[13]使用Sentinel-1、Sentinel-2和Landsat-8对亚热带农业种植区作物进行识别;郭交等^[14]基于Sentinel-1和Sentinel-2数据使用支持向量机(support vector machine, SVM)分类算法对农作物分类,精度达到90.25%。然而由于耕地具有异质性强、分布破碎等特点,且存在易与草地、林地等土地利用类型的光谱信息混淆的情况^[15],虽然多源遥感数据融合在耕地识别上的可行性已经被证明,但研究耕地识别特征集合的构建对于识别模型精度的影响依然具有挑战。

随着遥感数据源的增多及人工智能的发展,机器学习算法因其能够有效处理高维数据的优势在遥感图像的分类中被广泛应用。与传统参数分类器比较,对于具有高维特征空间的数据,使用机器学习方法进行分类更具潜力^[16]。由于耕地的高度复杂性,对其进行识别时往往需要构建多维度的特征空间,因此中外许多学者基于机器学习算法对耕地识别进行研究^[16-18]。然而不同机器学习分类算法在不同研究区的表现可能有所差异,比较不同算法在耕地识别方面的能力差异是当前研究的重点。Savitha等^[19]在印度Tadepalligudem流域的耕地识别研究中发现SVM算法的分类效果优于随机森林(random forest, RF)算法;我国地形地貌丰富,耕地类型多样,面对国内耕地复杂的情况,Wang等^[20]基于Sentinel-2数据在张掖山丹马场地区耕地识别时对比了RF、SVM和分类回归树

(classification and regression tree, CART) 3 种分类算法的精度, 发现在该地区耕地识别中 RF 算法更具优势; 对于西北旱区土地分类的研究中, 张静等^[21]发现结合标准植被指数和纹理特征的 SVM 分类算法比最大似然法和 BP (back propagation) 神经网络法分类精度更高。因此针对不同地区的耕地识别使用何种机器学习算法能够构建更高精度的分类模型仍然有待探究。

随着空间大数据云计算平台的发展, 以其可以提供海量遥感数据和支持科学分析及计算等优点^[22-23], 被许多学者用于进行耕地识别的研究^[24-26], 如 Teluguntla 等^[24]基于 Landsat-8 OLI 数据, 使用 RF 算法, 得到澳大利亚和中国 30 m 分辨率的耕地范围产品。

张掖市作为全国重要的商品粮基地和畜牧大市^[27], 有着丰富的耕地资源, 实现该地区耕地的准确识别和制图, 对耕地资源的保护及农业的可持续发展有着重要的现实意义。为构建高精度的耕地识别模型, 本研究基于空间云计算平台, 针对研究区耕地特性, 通过融合 Sentinel-1 和 Sentinel-2 数据集, 运用机器学习的方法训练分类器, 对比不同分类算法的精度, 以期构建针对本研究区的高精度耕地识别模型, 为耕地识别提供技术与理论支持。

1 材料与方法

1.1 研究区概况

张掖地处甘肃省西北部(37°28′-39°57′ N, 97°20′-102°12′ E), 位于青藏高原东北部河西走廊中段, 该市地处甘肃省的西北角, 与青海省、内蒙古自治区和陕西省接壤, 是连接中国西部和中亚的重要交通枢纽, 也是河西地区中心城市及重要的商贸中心之一(图 1)。张掖市总面积 3860000 万 m², 辖区内分为甘州区、临泽县、高台县、山丹县、民乐县、肃南裕固族自治县。张掖属温带大陆性气候, 具有丰富的自然地貌, 是我国西部的重要生态安全屏障, 光照条件充足, 昼夜温差大, 是我国重要的商品粮基地和畜牧大市, 有着丰富的耕地资源, 盛产小麦 (*Triticum aestivum*)、玉米 (*Zea mays*)、甜菜 (*Beta vulgaris*)、苜蓿 (*Medicago sativa*)、油菜 (*Brassica rapa* var. *oleifera*) 等作物。

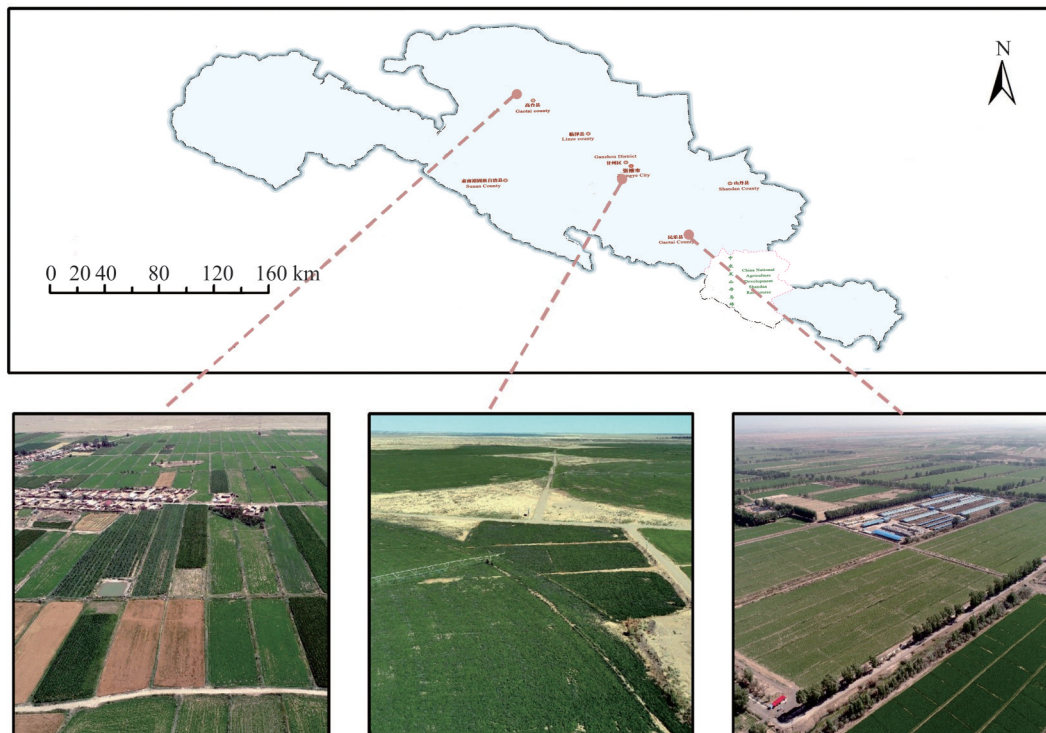


图 1 研究区概况

Fig. 1 Overview of the study area

基于自然资源部标准地图服务网站甘 S(2023)91 号标准地图制作, 底图边界无修改。The map was based on the standard map service website of the Ministry of Nature Resources with the drawing review No. gan S(2021)91, and the base map borders were not modified.

1.2 数据介绍及预处理

1.2.1 Sentinel-1及 Sentinel-2 Sentinel-1卫星由 Sentinel-1A 和 Sentinel-1B 双星组成,搭载有 C 波段合成孔径雷达传感器,具有全天候、全天时和高分辨率的特点。本研究基于空间云计算平台对 Sentinel-1 数据进行去除热噪声、辐射校准和正射校正等预处理,对研究区范围进行裁剪,使用其干涉宽幅(interferometric wide swath, IW)模式下的双极化方式(VV+VH)数据作为本研究主要数据集,VV:垂直发送,垂直接收(vertical transmit and receive),VH:垂直发送,水平接收(vertical transmit, horizontal receive)。

Sentinel-2卫星由 Sentinel-2A 和 Sentinel-2B 双星组成,采用推扫式成像模式,成像幅宽为 290 km。Sentinel-2 搭载的多光谱成像仪(multispectral imager, MSI)可以同时获取 13 个波段的数据,空间分辨率涵盖 10、20 和 60 m,双星时间分辨率可达 5 d。本研究基于空间云计算平台提供的 Sentinel-2 Level-2A 数据,按照研究区范围对数据进行裁剪和计算形成研究区内 Sentinel-2 光谱及植被指数数据集。选取的影像时间节点涵盖 2021 年时期的影像,使用 QA 60 波段进行了去云操作,进行影像数据的最大值合成。

1.2.2 其他数据 高程数据是由航天飞机雷达地形测绘任务(shuttle radar topography mission, SRTM)^[28]提供的近全球范围内的数字高程模型(digital elevation model, DEM)数据。本研究基于 NASA JPL 提供 SRTM V3 产品(SRTM Plus)DEM 数据。

土壤数据是由国际土壤参考和资料中心(international soil reference and information centre, ISRIC)提供的全球尺度 250 m 分辨率的土壤数据^[29](<https://soilgrids.org/>),本研究使用的土壤数据为 2020 年 5 月发布的版本。

1.2.3 样本数据 本研究基于空间云计算平台提供的 Sentinel-2 高清影像结合实测样本点获得张掖市 2021 年度的样本数据集,将研究区分为荒地、太阳能发电站、水体、道路、城市绿地、耕地、建筑、草地、森林、山地、冰川 11 类,共 12200 个样本点,并将样本以 8:2 随机抽取作为训练集和验证集(图 2)。

1.3 研究方法

本研究主要包括以下步骤(图 3):1)基于空间云计算平台的高清卫星图像及 Sentinel-2 影像,通过目视解译法及实地调查获取 2021 年度的样本点数据;2)构建耕地识别的初始特征并对其进行筛选;3)在空间云计算平台通过机器学习算法识别出研究区域内的耕地空间分布;4)对不同分类器进行精度评价。

1.3.1 特征筛选方法 由于研究区的耕地面积及分布在同一年内不会产生较为明显的变化,因此特征变量的构建均以年为单位。本研究选择了 5 种类型的特征变量作为耕地识别的初始特征:1)光谱特征:12 个 Sentinel-2 光谱波段,包括 Aerosols(B1)、Blue(B2)、Green(B3)、Red(B4)、Red Edge 1(B5)、Red Edge 2(B6)、Red Edge 3(B7)、Near infrared(B8)、Red Edge 4(B8A)、Water vapour(B9)、SWIR 1(B11)以及 SWIR 2(B12);2)雷达特征: Sentinel-1 数据的 VV 和 VH 波段;3)植被指数特征:耕地的人类活动度高,其植被覆盖程度随着物候而变化^[30],据实地调查,研究区内耕地具有较强的耕作节律,其植被覆盖具有一定的物候规律。本研究选择 6 种常用于进行耕地及作物分类的指数数据,包括归一化植被指数(normalized difference vegetation index, NDVI)、增强型植被指数(enhanced vegetation index, EVI)、归一化水指数(normalized difference water index, NDWI)、归一化物候指数(normalized difference phenology index, NDPI)、土壤调整植被指数(soil adjusted vegetation index, SAVI)以及简单比值(sample ratio, SR),它们能较好地反映植被覆盖的状态,具体计算公式如下所示:

$$NDVI = \frac{NIR - R}{NIR + R}$$

$$EVI = 2.5 \times \frac{NIR - R}{NIR + 6 \times R - 7.5 \times B + 1}$$

$$NDWI = \frac{G - NIR}{G + NIR}$$

$$NDPI = \frac{NIR - (0.74 \times R + 0.26 \times SWIR)}{NIR + (0.74 \times R + 0.26 \times SWIR)}$$

$$SAVI = \frac{(NIR - R) \times (1 + 0.5)}{NIR + R + 0.5}$$

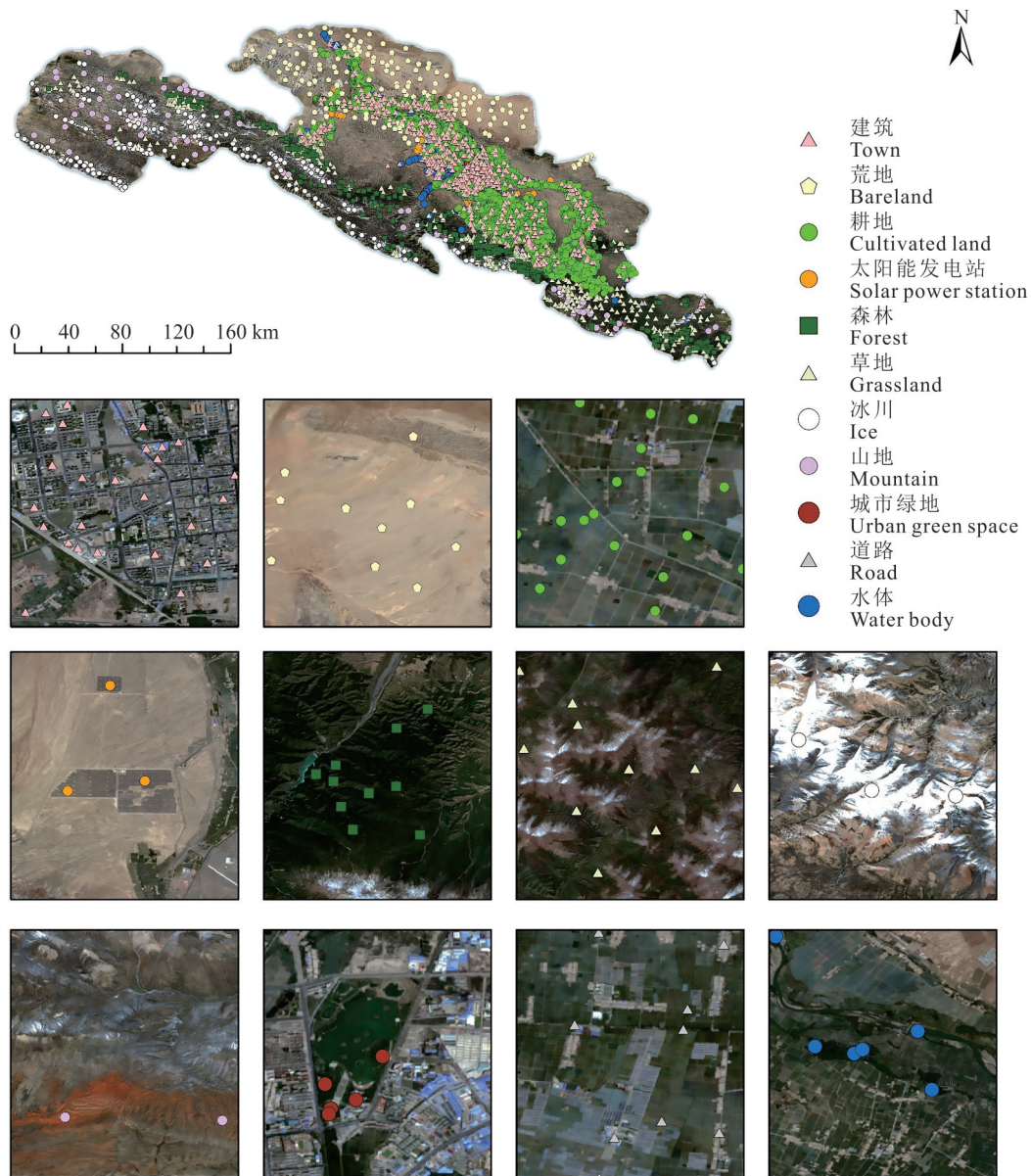


图 2 研究区样本点分布

Fig. 2 Samples distribution of study area

基于自然资源部标准地图服务网站甘 S(2023)60 号标准地图制作, 底图边界无修改。The map was based on the standard map service website of the Ministry of Nature Resources with the drawing review No.gan S(2023)60, and the base map borders were not modified.

$$SR = \frac{NIR}{R}$$

式中: B(blue)为蓝光波段, G(green)为绿光波段, R(red)为红光波段, NIR(near infrared)为近红外波段, SWIR(short-wave length infrared band)为短波红外波段, 分别对应 Sentinel-2A 影像的第 2、3、4、8 及 11 波段。

4) 地形特征: 由于耕地应该具有一定规律性的平坦土地^[31], 有别于山地、林地等具有较为突出地形特征的土地类型。本研究选择了海拔(elevation)、坡度(slope)、坡向(aspect)以及山体阴影(hill shade)。

5) 土壤特征: 由于耕地的土壤理化性质可能有别于其他土地利用类型, 因此本研究加入了土壤数据作为识别特征, 包括容重(bulk density, BD)、总氮含量(total nitrogen, TN)、pH 值、有机碳密度(organic carbon density, OCD)、有机碳含量(organic carbon stock, OCS)、砂粒含量(sand)、粉粒含量(silt)、黏粒含量(clay)。

为了构建最优耕地识别模型并避免冗余特征对模型精度的影响, 本研究基于上述初始特征根据分类精度与特征重要性进行筛选, 筛选的过程分为两个阶段。

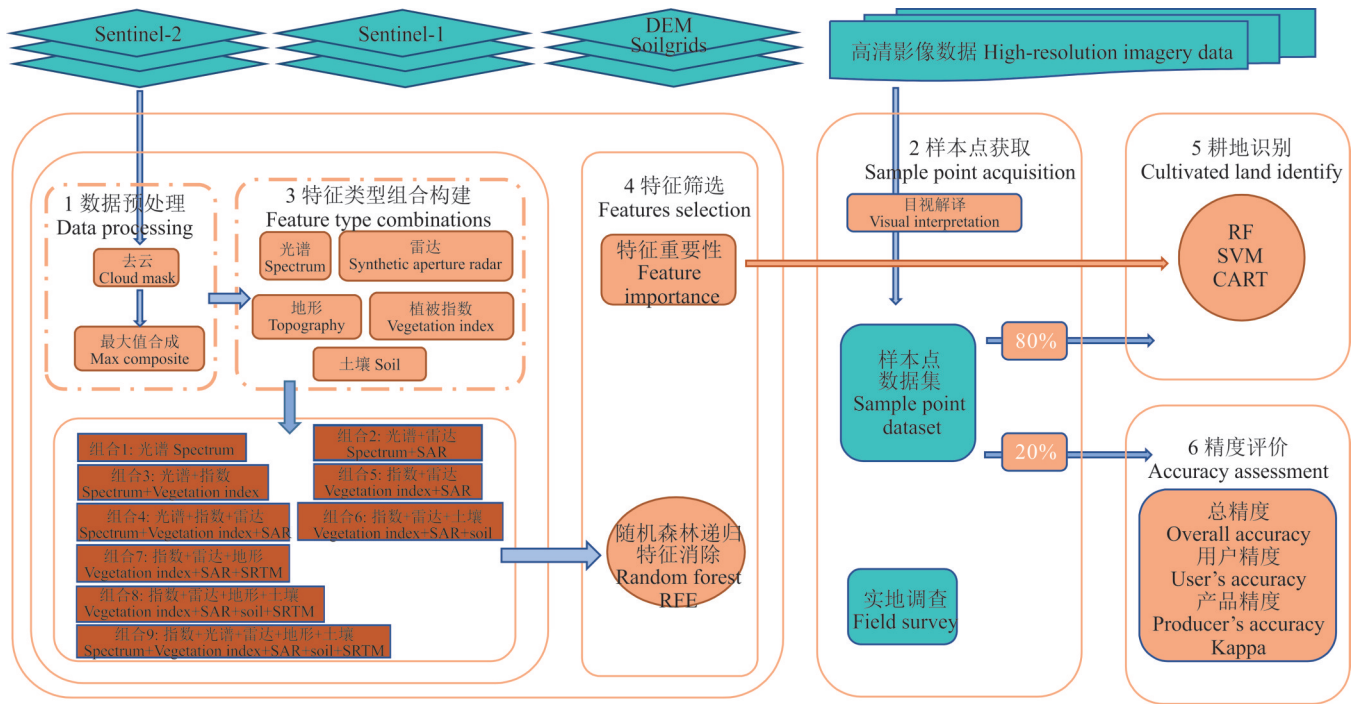


图3 工作流程

Fig. 3 Workflow of the cultivated land identification in Zhangye City

RF: 随机森林 Random forest; SVM: 支持向量机 Support vector machine; CART: 分类回归树 Classification and regression tree; RFE: 递归特征消除 Recursive feature elimination; Kappa: 卡帕系数.

1)把以上5个特征类型进行组合设计出9组特征类型组合(表1),根据这些特征组合在RF分类器下表现出来的精度,对特征组合进行筛选。

表1 9组不同特征类型组合

Table 1 9 groups of different feature type combinations

编号 Number	特征类型组合 Feature type combinations
1	光谱 Spectrum
2	光谱+雷达 Spectrum+synthetic aperture radar (SAR)
3	光谱+植被指数 Spectrum+vegetation index
4	植被指数+雷达 Vegetation index+SAR
5	光谱+植被指数+雷达 Spectrum+vegetation index+SAR
6	植被指数+雷达+土壤 Vegetation index+SAR+soil
7	植被指数+雷达+地形 Vegetation index+SAR+topography
8	植被指数+雷达+地形+土壤 Vegetation index+SAR+topography+soil
9	全部特征类型组合 All feature types

其中光谱指 Sentinel-2 光谱特征,植被指数为根据光谱特征计算得到的特征,雷达指 Sentinel-1 的 SAR 特征,土壤指 Soilgrids 数据特征,地形指由 SRTM V3 产品 DEM 数据得到的特征。

2)对于确定为最优的特征类型组合,使用随机森林递归特征消除的方法,根据随机森林算法输出的特征重要性对组合内的特征进行筛选,以得到最优的特征集合。

1.3.2 分类器 本研究分别使用随机森林(RF)^[32-34]、支持向量机(SVM)^[35-36]和分类回归树(CART)^[37]3种不同的分类器进行耕地识别,并对不同分类器的分类效果进行对比。为了确定随机森林分类器中的决策树数量,本研究分别测试了10、30、50、100、300和500,并根据精度最终选择了决策树数量为500,其他参数如

minleafPopulation、variablesPerSplit、bagFraction 和 seed 均为默认设置,同时对于 SVM 分类器和 CART 分类器的参数设置均为默认值。参考 Wang 等^[20]的研究,随机抽取样本数据以 8:2 的比例分为训练集(80%)和验证集(20%)对分类器模型进行训练。

1.3.3 精度评价 混淆矩阵(confusion matrix)常用于图像精度评价,主要用于比较分类结果和实际测得值,可以把分类结果的精度显示在一个($n \times n$)的混淆矩阵中。本研究通过构建混淆矩阵来评价分类算法的精度。评价的指标包括总精度(overall accuracy, OA)、生产者精度(producer's accuracy, PA)、用户精度(user's accuracy, UA)、Kappa 系数^[38]。

$$OA = \frac{\sum_{i=1}^n X_{i+}}{N}$$

$$PA = \frac{X_{ii}}{X_{+i}}$$

$$UA = \frac{X_{ii}}{X_{i+}}$$

$$Kappa\ coefficient = \frac{N \sum_{i=1}^n X_{ii} - \sum_{i=1}^n (X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} \times X_{+i})}$$

式中: n 为混淆矩阵的总行列数, N 为验证集的样本总数, X_{ii} 为混淆矩阵第*i*行第*i*列土地类型样本正确分类的个数; X_{i+} 和 X_{+i} 分别为第*i*行和第*i*列土地类型样本总数。

2 结果与分析

2.1 特征筛选结果

通过对比不同特征类型组合的分类精度,可初步判断各特征类型对模型精度的贡献。如图 4 所示,多种特征类型的组合使用可以使模型精度得到提升,然而考虑到精度在 90% 以上的组合之间差异不大,为了避免冗余特征对特征模型精度的干扰与其导致的运算时间过长,因此本研究综合考虑确定使用组合 7 (植被指数特征+雷达特征+地形特征)作为最优特征组合。

基于确定的最优特征组合(组合 7),本研究根据 RF 算法对特征进行递归消除,当使用组合 7 内全变量作为特征时总精度为 91.32%,根据 RF 模型得到特征重要性顺序由低到高逐个剔除 *Hillshade*、*Slope*、*Aspect*、*NDVI*、*EVI* 和 *SAVI*,总精度分别变为 91.19%、90.61%、90.24%、90.49%、90.04% 和 87.84%,剔除 *SAVI* 后模型精度出现了较为明显的下降,表明 *SAVI* 对研究区耕地识别有着积极作用。如表 2 所示,本研究选择了前 7 个特征并对其在 2021 年研究区耕地识别中的重要性进行排名,表现最好的是 *Elevation*,其次是 *VH*、*NDWI*、*SR*、*VV*、*NDPI*、*SAVI*。

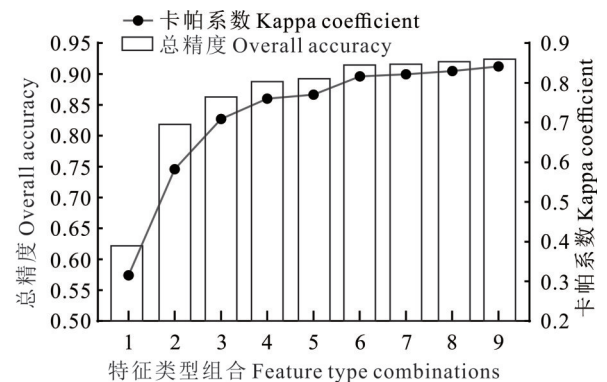


图 4 9 种特征类型组合的分类精度和 Kappa 系数比较

Fig. 4 Accuracy and Kappa coefficient comparison of 9 feature type combinations

图中特征类型组合 1 为光谱特征,组合 2 为光谱和雷达特征组合,组合 3 为光谱和植被指数的特征组合,组合 4 为植被指数和雷达的特征组合,组合 5 为光谱、植被指数及雷达的特征组合,组合 6 为植被指数、雷达及土壤的特征组合,组合 7 为植被指数、雷达及地形的特征组合,组合 8 为植被指数、雷达、地形及土壤的特征组合,组合 9 为全特征组合。The combinations of feature types in the figure are as follows: Combination 1 consists of spectral features; Combination 2 consists of a combination of spectral and synthetic aperture radar features; Combination 3 consists of a combination of spectral features and vegetation indices; Combination 4 consists of a combination of vegetation indices and synthetic aperture radar features; Combination 5 consists of a combination of spectral, vegetation indices, and synthetic aperture radar features; Combination 6 consists of a combination of vegetation indices, synthetic aperture radar features, and soil features; Combination 7 consists of a combination of vegetation indices, synthetic aperture radar features, and topography features; Combination 8 consists of a combination of vegetation indices, synthetic aperture radar features, topography features, and soil features; Combination 9 consists of all features.

表2 不同识别特征的重要性排名

Table 2 Rank of feature importance

识别特征 Feature	特征重要性 Feature importance
海拔 <i>Elevation</i>	8771.81
VH, 雷达特征中的 VH 极化 Synthetic aperture radar features VH polarization	8210.86
归一化水指数 Normalized difference water index, <i>NDWI</i>	8128.18
简单比值 Sample ratio, <i>SR</i>	7979.18
VV, 雷达特征中的 VV 极化 Synthetic aperture radar features VV polarization	7978.85
归一化物候指数 Normalized difference phenology index, <i>NDPI</i>	7947.32
土壤调整植被指数 Soil adjusted vegetation index, <i>SAVI</i>	7896.25

2.2 耕地识别结果

如图5所示,3种分类器提取出的耕地分布与卫星图中的耕地分布基本重合,本研究选择了两处典型区域进行结果展示。RF、SVM和CART分类算法识别出耕地面积分别为58.5万、52.3万和63.9万 hm^2 ,耕地面积分别占土地总面积的15.4%、13.5%、16.5%,在地形平坦,耕地连续性强的区域,基于RF模型提取出来的耕地轮廓清晰,内部完整,能够较好地地区分出建筑、道路及水体等,基于SVM模型提取出的耕地图斑边界也较为清晰,但对耕地内部区域的建筑和道路的区分效果不如RF模型,CART模型分类结果噪声现象明显,错误地形成很多细小的斑块。在耕地形状不规则且破碎程度严重的区域,RF模型仍能够较为准确地识别出耕地的边界。

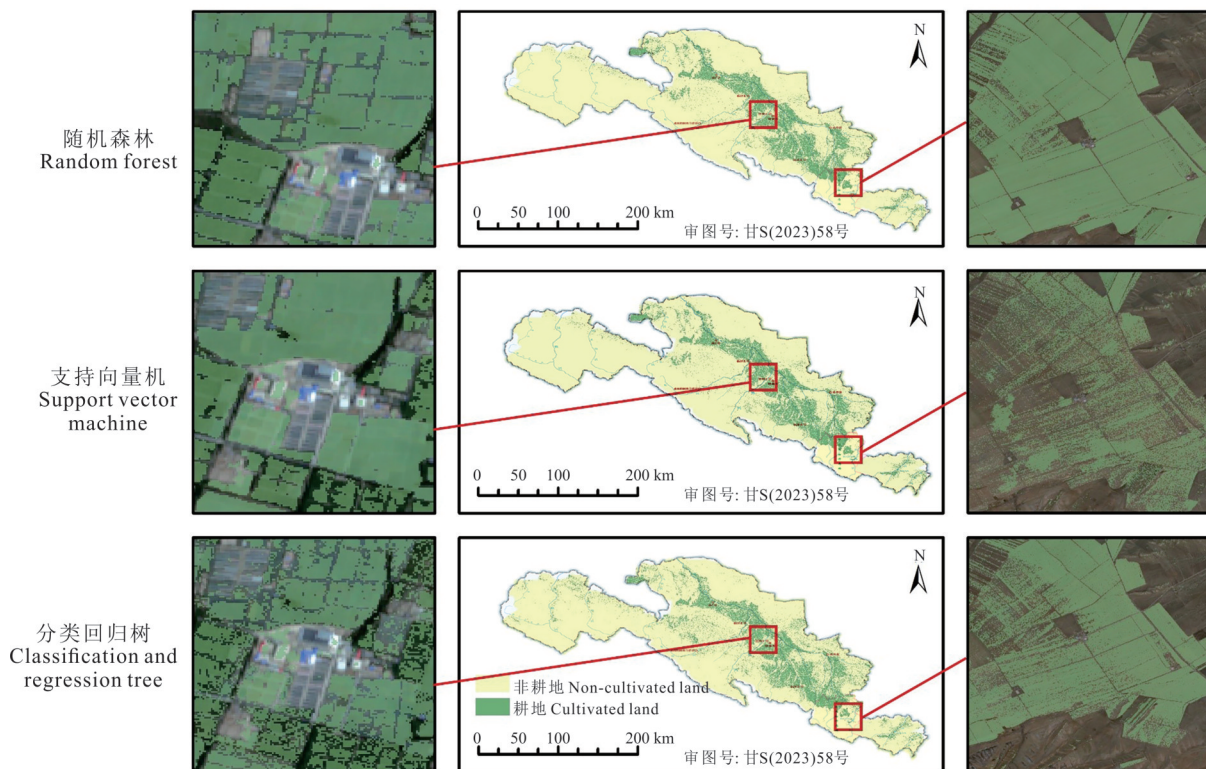


图5 不同模型的耕地识别结果

Fig. 5 Cultivated land identification result of random forest, support vector machine and classification and regression tree respectively

基于自然资源部标准地图服务网站甘S(2023)58号标准地图制作,底图边界无修改。The map was based on the standard map service website of the Ministry of Nature Resources with the drawing review No.gan S(2023)58, and the base map borders were not modified.

由表 3 可知,在研究区耕地识别中,使用 RF 分类器的总精度、用户精度、生产者精度及 Kappa 系数均为最高。综上,在这 3 种机器学习模型中,RF 模型在研究区域内耕地识别中表现出最优的平均总精度和 Kappa 系数。

表 3 RF、SVM、CART 模型分类精度比较

Table 3 Comparison of accuracy of random forest, support vector machine, and classification and regression tree

模型 Model	总精度 Overall accuracy	用户精度 User's accuracy	生产者精度 Producer's accuracy	卡帕系数 Kappa coefficient
随机森林 Random forest, RF	0.9004	0.9540	0.9747	0.7944
支持向量机 Support vector machine, SVM	0.8868	0.9466	0.9694	0.7660
分类回归树 Classification and regression trees, CART	0.8547	0.9467	0.9494	0.7054

3 讨论

3.1 耕地识别结果评价

本研究基于空间云计算平台使用机器学习方法提取了张掖市耕地的空间分布,最优算法是 RF 算法,基于该算法的耕地识别模型 OA 和 Kappa 系数分别为 90.04% 和 0.7944,识别出研究区耕地面积为 58.5 万 hm^2 ,占比为 15.4%。为了更好地评价研究结果的真实性与准确性,本研究将 RF 模型的耕地识别结果与 ESRI 公司提供的 2021 年度全球尺度下 10 m 分辨率的土地类型图(Sentinel-2 10-Meter Land Use/Land Cover)^[39]及欧洲空间局 WorldCover 项目^[40]提供的 2020 年全球尺度下 10 m 分辨率的土地利用产品(ESA WorldCover 10)进行了结果对比(图 6)。在比较中发现,本研究识别出的耕地分布与 Sentinel-2 10-Meter Land Use/Land Cover 产品识别的耕地范围大致相同,然而该产品对分类结果进行了去噪平滑处理,导致识别出来的耕地图斑边界有所改变,相对而言本研究识别的耕地图斑形状更符合实际情况。而和 ESA WorldCover 10 的比较中,该产品在部分区域未能准确识别耕地,且本研究识别出的耕地具有更好的细节,识别效果更符合真实情况。此外,在本研究耕地识别结果及上述产品中,均有“椒盐现象”的存在,可能是由于研究区内的耕地破碎且异质性强,耕地之间的光谱差异较大所导致,在 Feng 等^[41]的甘南地区土地利用研究中也印证了这一观点。总体来说,在研究范围内,相比较于 Sentinel-2 10-Meter Land Use/Land Cover 和 ESA WorldCover 10,本研究识别出耕地面积占比较大,可能因为不同土地利用分类系统存在一定区别、耕地定义不同或分类精度存在差异。为了进一步探究本研究识别耕地面积占比偏大的原因,将表现最优的 RF 模型得到的耕地识别结果与 2019 年全国土地利用状况主要数据中的张掖市各区县数据进行对比,结果表明差异较大的区县为山丹县、民乐县及肃南裕固族自治县,相较于平原种植区,这些地区的地形地貌更加多样化,同时存在山地丘陵梯田区和林草区。根据分类结果及混淆矩阵,本研究在张掖市东南部耕地与山地交界的地方存在错分的现象,部分耕地被误判为山地及草地,这可能是本研究耕地识别面积偏大的原因,这与上述识别结果与调查数据的对比得出的结论相符。

3.2 特征筛选对分类精度的影响

特征变量的选择对基于机器学习算法的遥感图像分类精度和结果及运算效率有重要影响^[42],过多的特征输入会导致“维度灾难”,如何从丰富的遥感数据源中确定能够平衡分类精度和运算效率的特征是一个关键问题。由于研究区内生长季阶段云量增加,为减少由于天气原因导致的光学影像数据不可用的情况,本研究使用 Sentinel-2 光谱特征、Sentinel-1 雷达特征及指数特征作为初始特征,除此以外还添加了地形特征与土壤特征。耕地一般地块形状规则且平坦,地形特征也是重要的分类依据之一,海拔、坡度等特征对耕地分类有着重要意义^[31],如钱鑫等^[43]使用光谱特征、纹理特征和地形特征的组合,获得黑河流域的耕地分布;此外,由于耕地上人类活动的强度较大,耕地土壤的理化性质有别于其他地物类型^[44],本研究还加入了土壤数据作为初始特征。

本研究尝试对不同特征进行组合并根据精度进行筛选,结果表明在耕地识别前进行特征类型集合的构建和特征筛选可以有效提升耕地识别的精度并减少特征的冗余使得运算效率得到提升。

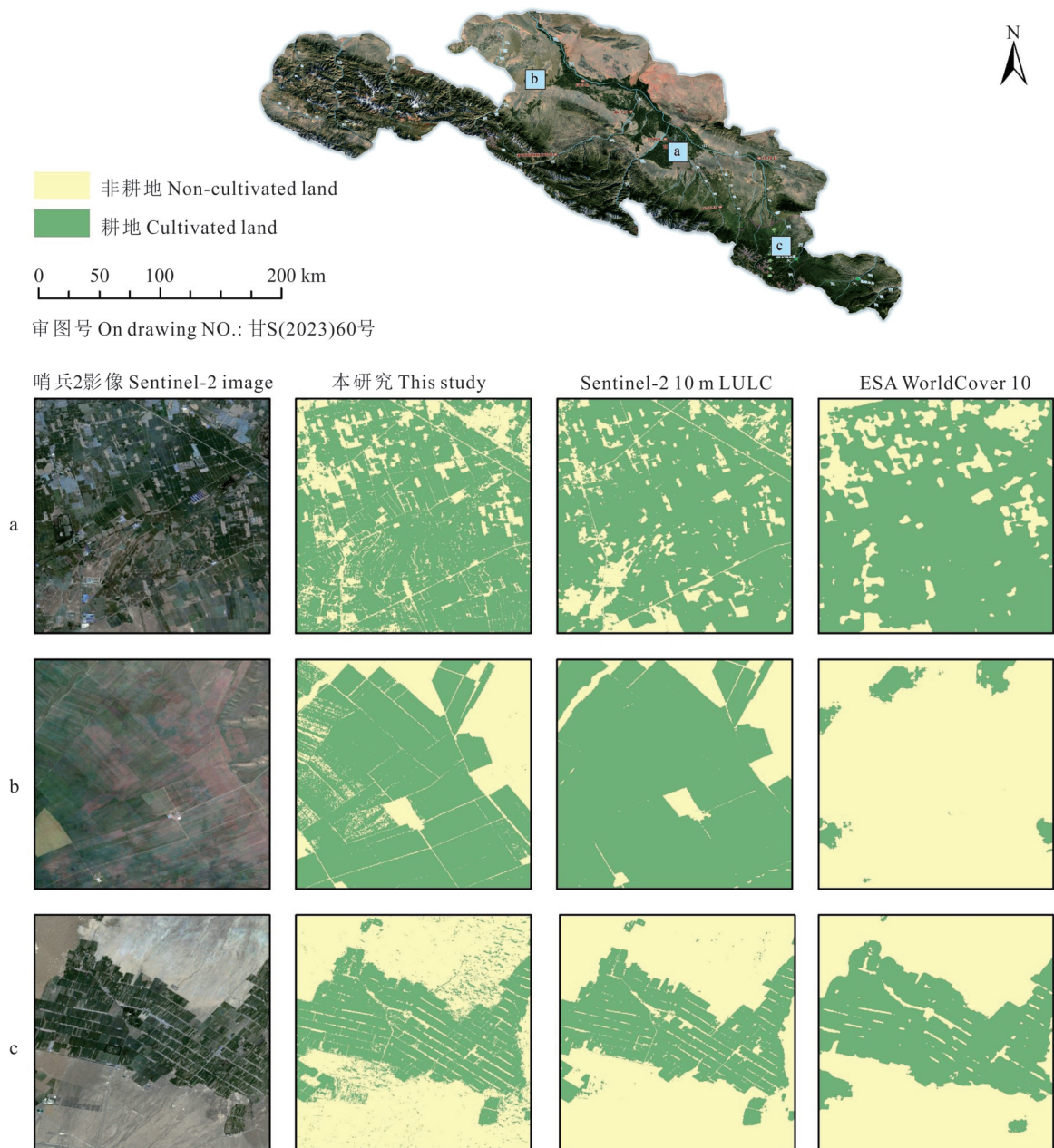


图6 本研究与 Sentinel-2 10 m LULC 和 ESA WorldCover 10 产品耕地识别结果比较

Fig. 6 Comparison of this study, Sentinel-2 10 m LULC and ESA WorldCover 10

基于自然资源部标准地图服务网站甘S(2023)60号标准地图制作,底图边界无修改。The map was based on the standard map service website of the Ministry of Nature Resources with the drawing review No.gan S(2023)60, and the base map borders were not modified. a, b, c 分别代表图中三个展示区域 a, b and c represent the three display areas in the figure respectively.

本研究中使用S2光谱数据(组合1)作为特征的分类精度仅有62.38%,而结合S2光谱数据和S1雷达数据的特征类型组合(组合2)分类精度达到81.73%,分类精度相较单独使用S2光谱数据有着大幅提升,相比于单源数据,多源数据的组合使用在耕地识别中具有一定优势^[11],这与Chakhar等^[45]和Blickensdörfer等^[46]的研究结果相符。由于耕地的植被覆盖程度存在一定的人为节律,因此在耕地识别中,植被指数也是一种常用的特征^[47],在组合3的基础上,本研究添加了植被指数特征(组合5),分类精度提升到89.0%,由于指数特征是由S2光谱数据计算得到的,因此本研究尝试在加入植被指数特征后,去除光谱特征(组合4),发现仅下降了0.6%。为了避免特征冗余,本研究认为可以使用植被指数特征替代光谱特征的贡献。由于耕地一般是具有一定规律性且较为平坦的土地,同时耕地的土壤理化性质可能有别于其他土地利用类型,且目前较少有使用土壤数据进行土地分类的研

究,本研究尝试添加地形数据和土壤数据作为初始特征类型,拟探讨二者对模型精度的提升效果。在确定使用雷达和指数特征的基础上,分别单独添加土壤或地形数据(组合6和组合7),精度分别提升了2.06%和2.27%,加入地形特征的精度略优于加入土壤特征的精度,但是二者差异不大。而将地形和土壤数据同时添加(组合8),精度却几乎没有提升(0.41%)。此外,本研究还尝试了全特征类型的组合,精度有所提高(92.3%),但提升十分有限,说明并不是特征变量越多,模型精度提升就越大,综合考虑模型的复杂性和精度,在进行分类前对特征类型进行组合和筛选是有必要的,而本研究最终选择了雷达特征、植被指数特征和地形特征的组合(组合7)。

本研究表明在张掖地区基于RF算法使用植被指数特征、地形特征和雷达特征构建特征组合进行耕地识别的精度达到90%以上。可见针对不同地区耕地特点存在差异的问题上,根据研究区性质选择分类精准的特征仍然有待研究。

3.3 分类器评价

由于耕地的高度复杂性,不同地区的优势分类器也存在差异,因此探究针对不同地区的耕地识别使用何种算法能够构建更高精度的分类模型仍具有挑战。Savitha等^[19]认为在印度Tadepalligudem流域的耕地识别中,分别基于Sentinel-2和Landsat数据使用SVM算法的精度均优于RF算法,原因是SVM算法在二分类问题中具有更好的分类能力;而Wang等^[20]在张掖山丹马场地区耕地识别中则认为相较于SVM和CART,该区域更适合使用RF算法,林陈捷等^[15]在广州市耕地识别的研究中认为RF算法表现更优。此外,也有学者如Xie等^[48]比较深度学习算法中的一些深度语义分割算法(U-Net、U-Net++、U-Net3+、MPSPNet)和传统机器学习算法(RF)的分类效果,发现在华南地区耕地识别中,深度学习算法较RF算法得到的识别精度有着显著提升。

在本研究中,相较于SVM和CART算法,RF算法在张掖市耕地识别中表现更好,分类精度和Kappa系数达到了90.04%和0.7944,在使用相同样本数据和特征情况下,SVM算法对耕地的分类存在一定漏分的现象,而CART算法分类的耕地内部存在很多噪声,相较下RF算法实现了该地区耕地较为精准的识别。研究区内耕地的高度复杂性导致使用机器学习算法准确地识别出耕地需要具有高维度的数据,RF是一种集成型算法,它对样本的数量和质量敏感度较低,在训练模型时表现出较好的泛化能力,能够有效减少过拟合的风险,同时RF对于数据存在的噪声和异常值有着较好的鲁棒性^[31,49-50],在该地区的耕地识别中表现出更好的分类效果。可见在耕地类型复杂的地区,随机森林分类器在耕地识别上具有一定的优越性,这与Wang等^[20]、林陈捷等^[15]的研究结果相符。

尽管基于像元对耕地进行识别的方法已经很成熟,但在一定程度上忽略了耕地丰富的纹理特征^[8],本研究使用高清分辨率影像仅采用基于像元的分类方法,而Xiong等^[7]的研究表明以RF算法作为主要分类器,局部使用SVM分类器弥补噪声的缺陷,再采用递归层次分割(recursive hierarchical image segmentation, RHSeg)算法,从相同的输入数据中基于光谱和空间属性生成面向对象的分割层,为耕地识别提供新思路。在今后的研究中可以参考该思路,把面向像元和面向对象的分类算法结合使用,或许能得到更理想的分类结果。

4 结论

以张掖市为研究区,在空间云计算平台基于Sentinel-1和Sentinel-2遥感影像数据,分别使用RF、SVM、CART算法对2021年研究区内耕地进行识别,并进行精度评价,研究主要获得以下结论:

1) 比起单源数据,多源数据的组合使用能够有效地提高耕地识别模型的精度。本研究中,对比单独使用S2数据作为特征类型,使用植被指数特征、雷达特征和地形特征的特征类型组合能够使精度达到91.32%。同时,在精度达到90%以上时,继续加入新的特征类型组合对模型的精度提升较少,表明虽然组合使用特征类型能有效提升模型精度,但加入过多的特征类型并不能使分类精度大幅提升。

2) 在分类之前对特征进行筛选,能够避免冗余特征对模型精度和运算时间的影响,在2021年张掖市耕地识别模型的特征中,按照重要性程度表现最好的是Elevation,其次是VH、NDWI、SR、VV,最后是NDPI、SAVI。

3) 本研究中,RF模型的总精度为90.04%,比SVM模型和CART模型的精度分别提升2%和5%,并且在分类效果的比较中,RF算法识别出的耕地图斑最接近实际情况,表明该算法在本研究区域的耕地识别中有着更强

的优越性,能够有效地提高识别精度和识别效率。

参考文献 References:

- [1] Wardlow B D, Egbert S L. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U. S. Central Great Plains. *Remote Sensing of Environment*, 2008, 112(3): 1096–1116.
- [2] Clark M L, Aide T M, Grau H R, *et al.* A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment*, 2010, 114(11): 2816–2832.
- [3] Johnson D M. Using the Landsat archive to map crop cover history across the United States. *Remote Sensing of Environment*, 2019, 232: 111286.
- [4] Gong P, Wang J, Yu L, *et al.* Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 2013, 34(7): 2607–2654.
- [5] Du B J, Zhang J, Wang Z M, *et al.* Crop mapping based on Sentinel-2A NDVI time series using object-oriented classification and decision tree model. *Journal of Geo-information Science*, 2019, 21(5): 740–751.
杜保佳, 张晶, 王宗明, 等. 应用 Sentinel-2A NDVI 时间序列和面向对象决策树方法的农作物分类. *地球信息科学学报*, 2019, 21(5): 740–751.
- [6] Valero S, Morin D, Inglada J, *et al.* Processing Sentinel-2 image time series for developing a real-time cropland mask//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Milan, Italy: Institute of Electrical and Electronics Engineers, 2015: 2731–2734.
- [7] Xiong J, Thenkabail P, Tilton J, *et al.* Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. *Remote Sensing*, 2017, 9(10): 1065.
- [8] Xiong X L, Hu Y M, Wen N, *et al.* Progress and prospect of cultivated land extraction research using remote sensing. *Journal of Agricultural Resources and Environment*, 2020, 37(6): 856–865.
熊曦柳, 胡月明, 文宁, 等. 耕地遥感识别研究进展与展望. *农业资源与环境学报*, 2020, 37(6): 856–865.
- [9] Wu B, Zhang M, Zeng H, *et al.* Challenges and opportunities in remote sensing-based crop monitoring: a review. *National Science Review*, 2023, 10(4): nwac290. <https://academic.oup.com/nsr/article/10/4/nwac290/6939854>.
- [10] Liu C, Chen Z, Shao Y, *et al.* Research advances of SAR remote sensing for agriculture applications: A review. *Journal of Integrative Agriculture*, 2019, 18(3): 506–525.
- [11] Song Q, Zhou Q B, Wu W B, *et al.* Recent progresses in research of integrating multi-source remote sensing data for crop mapping. *Scientia Agricultura Sinica*, 2015, 48(6): 1122–1135.
宋茜, 周清波, 吴文斌, 等. 农作物遥感识别中的多源数据融合研究进展. *中国农业科学*, 2015, 48(6): 1122–1135.
- [12] Orynbaikyzy A, Gessner U, Conrad C. Crop type classification using a combination of optical and radar remote sensing data: a review. *International Journal of Remote Sensing*, 2019, 40(17): 6553–6595.
- [13] Sun C, Bian Y, Zhou T, *et al.* Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors*, 2019, 19(10): 2401.
- [14] Guo J, Zhu L, Jin B, *et al.* Crop classification based on data fusion of Sentinel-1 and Sentinel-2. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(4): 192–198.
郭交, 朱琳, 靳标, 等. 基于 Sentinel-1 和 Sentinel-2 数据融合的农作物分类. *农业机械学报*, 2018, 49(4): 192–198.
- [15] Lin C J, Liu Z H, Hu Y M, *et al.* Cultivated land extraction based on Google Earth Engine multisource data. *Southwest China Journal of Agricultural Sciences*, 2022, 35(10): 2372–2378.
林陈捷, 刘振华, 胡月明, 等. 基于 Google Earth Engine 多源数据的耕地提取研究. *西南农业学报*, 2022, 35(10): 2372–2378.
- [16] Maxwell A E. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 2018, 39(9): 2784–2817.
- [17] Pazúr R, Huber N, Weber D, *et al.* A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data. *Earth System Science Data*, 2022, 14(1): 295–305.
- [18] Rawat S, Saini R. Cropland mapping using single date Sentinel-2 imagery using machine learning classifiers//2022 International Conference on Advances in Computing, Communication and Materials (ICACCM). Dehradun, India: Institute of Electrical and Electronics Engineers, 2022: 1–7.

- [19] Savitha C, Talari R. Mapping cropland extent using Sentinel-2 datasets and machine learning algorithms for an agriculture watershed. *Smart Agricultural Technology*, 2023(4): 100193.
- [20] Wang R, Feng Q, Jin Z, *et al.* Identification and area information extraction of oat pasture based on GEE—A case study in the Shandan racecourse (China). *Remote Sensing*, 2022, 14(17): 4358.
- [21] Zhang J, Zhang X, Tian L, *et al.* The support vector machine method for RS images' classification in northwest arid area. *Science of Surveying and Mapping*, 2017, 42(1): 49—52, 58.
张静, 张翔, 田龙, 等. 西北旱区遥感影像分类的支持向量机法. *测绘科学*, 2017, 42(1): 49—52, 58.
- [22] Fu D J, Xiao H, Su F Z, *et al.* Remote sensing cloud computing platform development and earth science application. *National Remote Sensing Bulletin*, 2021, 25(1): 220—230.
付东杰, 肖寒, 苏奋振, 等. 遥感云计算平台发展及地球科学应用. *遥感学报*, 2021, 25(1): 220—230.
- [23] Tamiminia H, Salehi B, Mahdianpari M, *et al.* Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 164: 152—170.
- [24] Teluguntla P, Thenkabail P S, Oliphant A, *et al.* A 30-m Landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 144: 325—340.
- [25] Pan L, Xia H M, Wang R M, *et al.* Mapping of the winter crop planting areas in Huaihe river basin based on Google Earth Engine. *Transactions of the Chinese Society of Agricultural Engineering*, 2021, 37(18): 211—218.
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. *农业工程学报*, 2021, 37(18): 211—218.
- [26] Han J, Zhang Z, Cao J, *et al.* Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sensing*, 2020, 12(2): 236.
- [27] Chen X Y, Zhang H P, Hu J L. Current status and promotion measures of Zhangye City's grain-forage transformation work. *China Dairy Cattle*, 2022(1): 40—42.
陈晓燕, 张和平, 胡江林. 张掖市粮改饲工作现状及推进措施. *中国奶牛*, 2022(1): 40—42.
- [28] Farr T G, Rosen P A, Caro E, *et al.* The shuttle radar topography mission. *Reviews of Geophysics*, 2007, 45(2): RG2004. doi:10.1029/2005RG000183.
- [29] Poggio L, de Sousa L M, Batjes N H, *et al.* SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil*, 2021, 7(1): 217—240.
- [30] Cao X, Chen X H, Zhang W W, *et al.* Global cultivated land mapping at 30 m spatial resolution. *Scientia Sinica (Terrae)*, 2016, 46(11): 1426—1435.
曹鑫, 陈学泓, 张委伟, 等. 全球 30 m 空间分辨率耕地遥感制图研究. *中国科学: 地球科学*, 2016, 46(11): 1426—1435.
- [31] Dong J, Xiao X, Menarguez M A, *et al.* Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sensing of Environment*, 2016, 185: 142—154.
- [32] Breiman L. Random forests. *Machine Learning*, 2001, 45: 5—32.
- [33] Fang K N, Wu J B, Zhu J P, *et al.* A review of technologies on random forests. *Journal of Statistics and Information*, 2011, 26(3): 32—38.
方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述. *统计与信息论坛*, 2011, 26(3): 32—38.
- [34] Zhang B H, Zhang Y L, Gu C J, *et al.* Land cover classification based on random forest and feature optimism in the Southeast Qinghai—Tibet Plateau. *Scientia Geographica Sinica*, 2023, 43(3): 388—397.
张炳华, 张德锂, 谷昌军, 等. 基于随机森林与特征选择的藏东南土地覆被分类方法及精度评价. *地理科学*, 2023, 43(3): 388—397.
- [35] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995, 20(3): 273—297.
- [36] Yang X F, Tureniguli•Amuti. Study of SVM classification method optimized by artificial bee colony algorithm: A case study of the Ancient Manasi Lake Basin. *Geography and Geo-Information Science*, 2018, 34(4): 40—45.
杨雪峰, 吐热尼古丽•阿木提. 基于人工蜂群算法优化的 SVM 遥感分类方法——以玛纳斯湖古湖盆为例. *地理与地理信息科学*, 2018, 34(4): 40—45.
- [37] Breiman L, Friedman J, Olshen R A, *et al.* Classification and regression trees. Monterey, California, U. S. A. : Wadsworth International Group, 1984.
- [38] Congalton R G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*,

- 1991, 37(1): 35—46.
- [39] Karra K, Kontgis C, Statman-Weil Z, *et al.* Global land use/land cover with Sentinel 2 and deep learning//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Brussels, Belgium: Institute of Electrical and Electronics Engineers, 2021: 4704—4707.
- [40] Zanaga D, Van De Kerchove R, Daems D, *et al.* ESA WorldCover 10 m 2021 v200. Zenodo. 2022, <https://doi.org/10.5281/zenodo.7254221>.
- [41] Feng S, Li W, Xu J, *et al.* Land use/land cover mapping based on GEE for the monitoring of changes in ecosystem types in the upper Yellow River Basin over the Tibetan Plateau. *Remote Sensing*, 2022, 14(21): 5361.
- [42] Palanisamy P A, Jain K, Bonafoni S. Machine learning classifier evaluation for different input combinations: A case study with Landsat 9 and Sentinel-2 data. *Remote Sensing*, 2023, 15(13): 3241.
- [43] Qian X, Li P X, Xie H Q, *et al.* Crop information extraction and water demand analysis based on Sentinel-2 remote sensing image. *Water Saving Irrigation*, 2022(5): 33—38, 46.
钱鑫, 李培显, 谢宏全, 等. 基于 Sentinel-2 遥感影像的作物信息提取与需水量分析研究. *节水灌溉*, 2022(5): 33—38, 46.
- [44] Li H Y, He R B, Xie M D, *et al.* Influence of natural and anthropogenic factors on soil organic matter content in farmland. *Chinese Journal of Soil Science*, 2023, 54(5): 1050—1059.
李洪义, 贺任彬, 谢模典, 等. 自然和人为因素对耕地土壤有机质含量影响的研究. *土壤通报*, 2023, 54(5): 1050—1059.
- [45] Chakhar A, Hernández-López D, Ballesteros R, *et al.* Improving the accuracy of multiple algorithms for crop classification by integrating Sentinel-1 observations with Sentinel-2 data. *Remote Sensing*, 2021, 13(2): 243.
- [46] Blickensdörfer L, Schwieder M, Pflugmacher D, *et al.* Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sensing of Environment*, 2022, 269: 112831.
- [47] Mu Y X, Wu M Q, Niu Z, *et al.* Method of remote sensing extraction of cultivated land area under complex conditions in southern region. *Remote Sensing Technology and Application*, 2020, 35(5): 1127—1135.
牟昱璇, 邬明权, 牛铮, 等. 南方地区复杂条件下的耕地面积遥感提取方法. *遥感技术与应用*, 2020, 35(5): 1127—1135.
- [48] Xie D, Xu H, Xiong X, *et al.* Cropland extraction in southern China from very high-resolution images based on deep learning. *Remote Sensing*, 2023, 15(9): 2231.
- [49] Wang M, Zhang X C, Wang J Y, *et al.* Forest resource classification based on random forest and object-oriented method. *Acta Geodaetica et Cartographica Sinica*, 2020, 49(2): 235—244.
王猛, 张新长, 王家耀, 等. 结合随机森林面向对象的森林资源分类. *测绘学报*, 2020, 49(2): 235—244.
- [50] Gu H Y, Yan L, Li H T, *et al.* An object-based automatic interpretation method for geographic features based on random forest machine learning. *Geomatics and Information Science of Wuhan University*, 2016, 41(2): 228—234.
顾海燕, 闫利, 李海涛, 等. 基于随机森林的地理要素面向对象自动解译方法. *武汉大学学报(信息科学版)*, 2016, 41(2): 228—234.