

协同 Sentinel-2 和 GF-3 多特征优选的农作物识别

张青松，王金鑫^{*}，赫晓慧

(郑州大学地球科学与技术学院，郑州 450001)

摘要：农作物识别是精准农业的重要研究领域。在时空大数据和智能计算时代，如何充分挖掘和综合应用各种数据、方法和模型的优势是提高遥感农作物识别精度的有效途径。该研究以安徽省颍上县为例，采用 Sentinel-2 和 GF-3 卫星影像数据，提取了包括光谱、指数、纹理和极化等在内的 58 个特征指标；随后分别选取 3 种特征优选算法和 3 种机器学习方法进行组合，设计了 3 种试验方案，探索特征选择和机器学习方法对农作物分类的影响；通过对比特征维度和分类精度，对各种分类方案进行评价。研究结果显示：红边特征在农作物识别中具有重要作用，同时纹理特征的加入也适当提高了分类精度；3 种特征优选算法分别和随机森林方法组合时，分类精度均为最优；其中 Relief F 与随机森林组合在遥感农作物识别分类中效果最好，总体精度达到了 93.39%，Kappa 系数为 0.893 3，F1 得分为 93.31%；比 Relief F 结合极限梯度提升和支持向量机分类方法的总体精度、Kappa 系数、F1 得分分别提高 1.36 个百分点、0.021 和 1.31 个百分点，8.81 个百分点、0.131 2 和 8.78 个百分点；在随机森林分类方法下，Relief F 特征选择维度为 28 维，比随机森林的递归特征消除和卡方检验特征优选算法分别低 4 和 22 维，证明了 Relief F 结合随机森林分类方法的有效性和先进性。该研究为精准农作物识别提供了新的技术思路。

关键词：农作物；分类；特征；优选；随机森林；Sentinel-2；GF-3

doi: 10.11975/j.issn.1002-6819.202406135

中图分类号：TP79

文献标志码：A

文章编号：1002-6819(2025)-04-0153-12

张青松，王金鑫，赫晓慧. 协同 Sentinel-2 和 GF-3 多特征优选的农作物识别[J]. 农业工程学报, 2025, 41(4): 153-164.

doi: 10.11975/j.issn.1002-6819.202406135 <http://www.tcsae.org>

ZHANG Qingsong, WANG Jinxin, HE Xiaohui. Crop identification by synergistic Sentinel-2 and GF-3 multi-feature optimization[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(4): 153-164. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202406135 <http://www.tcsae.org>

0 引言

全球气候变化给农业生产和粮食安全带来了严峻挑战。近年来，中国对遥感技术的发展给予了高度重视，特别是在航天遥感领域取得了显著成就，已跻身世界先进行列^[1]。中国作为农业强国，加强农业遥感监测应用，及时准确获取作物种类种植结构信息，对农业普查、农情监测、作物估产、灾情评价等意义重大^[2]。随着遥感技术的不断发展，基于卫星遥感时空大数据的定量遥感监测是当代精准农业的重要发展方向^[3-4]。

尽管研究人员积累了丰富的遥感数据，提出了各式各样的农作物识别方法，但目前的农作物遥感识别分类方法在准确性上还有待提高^[5]，传统的使用单源遥感影像进行农作物分类研究往往不能满足应用需求。由于“同物异谱，同谱异物”现象的存在，单一遥感影像数据无法真实反应农作物的分布情况^[6]。在此背景下，有学者开始融合多种遥感数据来对地物种类进行识别分类。如 MOSLEH 等^[7]基于光学和微波遥感数据的协同，对水稻进行识别分类，克服了单一遥感数据的缺陷，取得良好的效果。

有研究表明融合光学与雷达影像可以有效提高作物

识别精度^[8]，但是多数研究把注意力放在了遥感影像质量处理上，很大程度上忽略了雷达影像的纹理特征和极化特征^[9]。充分利用多源遥感数据的多重特征参与作物分类，对提高分类精度十分重要和必要。然而，不加筛选地将所有特征信息用于分类，不仅会导致数据冗余，降低计算效率，进而会影响结果的可靠性。鉴于此，有学者开始重点关注遥感影像特征对农作物的敏感度，并注意应用先进的智能算法。通过采用特征工程技术对特征进行优选，选择相对重要的特征，在减少数据量的同时，显著提升分类精度^[10]。张悦琦等^[11]利用高分六号影像数据来构建相关参数，基于特征优选对盘锦市水稻进行识别分类研究，在移栽期影像的 20 个特征中选择 8 个特征时，总体精度和 Kappa 系数最高，分别为 97.67% 和 0.97。李利娟等^[12]基于 Sentinel-2 遥感影像构建最优特征数据集，利用随机森林分类算法对农耕区的土地利用进行分类，相比使用全部特征，其精度有所提高。卢献健等^[13]基于 Sentinel-1/2 影像，使用随机森林优选方法对构建的特征集进行优选，结果表明当加入地形特征、纹理特征和极化特征时更有利于区分甘蔗和其他植被，分类精度比单一特征和全部特征都要高。

综上所述，多源遥感、多特征协同结合智能分类算法能够充分发挥多方优势，逐渐成为遥感农作物识别的主流方法，但是存在多源数据、特征选择算法与智能分

收稿日期：2024-06-19 修订日期：2024-12-22

基金项目：国家自然科学基金面上项目（42371358）

作者简介：张青松，研究方向为农业遥感。Email: qsz9805@163.com

※通信作者：王金鑫，博士，教授，研究方向为空间信息网络，3S 技术及其应用。Email: jxwang@zzu.edu.cn

类算法的适配性问题。在已有的类似研究中,大多关注多源数据融合或多特征优选的策略,对不同特征优选算法和分类器的适配性研究较少。本研究将多源数据、多特征、多种特征优选算法、多种分类器进行排列组合,系统地评估其适配性,以期筛选出设定背景下最优的农作物识别方法,有效提高识别精度。

1 研究区概况与数据源

1.1 研究区概况

阜阳市颍上县地处安徽省西北部, N32°27′—32°54′、E115°56′—116°38′之间,位于淮河与颍河交汇处、淮北平原最南部。颍上县是最典型的淮北冲积平原,地势平坦,最高海拔为 28.5 米。该地区属于北温带与亚热带之间的过渡型气候,由于受西伯利亚和太平洋、印度洋冷暖气流的交互影响,夏热冬冷,四季分明,雨水充沛。颍上县农业发达,在夏秋季主要种植水稻、玉米、大豆 3 种农作物,研究区域如图 1 所示。

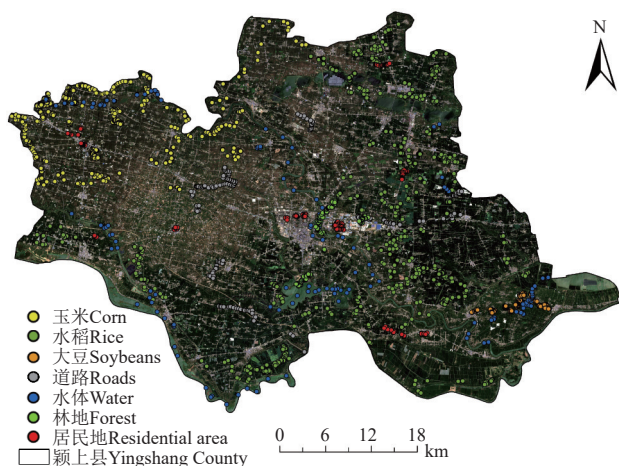


图 1 研究区域及样本点分布图

Fig.1 Map of study area and sample points

表 1 颍上县典型农作物物候期

Table 1 Phenological periods of typical crops in Yingshang County

时间 Time	六月 June			七月 July			八月 August			九月 September			十月 October		
	上 Early	中 Mid	下 Late	上 Early	中 Mid	下 Late	上 Early	中 Mid	下 Late	上 Early	中 Mid	下 Late	上 Early	中 Mid	下 Late
玉米 Corn	播种	播种	三叶	三叶	拔节		开花	开花		成熟	成熟				
水稻 Rice		播种	播种		分蘖	分蘖	开花	开花					成熟	成熟	
大豆 Soybean		播种	播种		三叶	三叶	开花	开花		结瘤	结瘤		成熟	成熟	

本文采用的 GF-3 卫星数据来源于中国资源卫星应用中心,选取了 4 景影像数据,产品等级为 L1A 级,成像模式是精细条带成像,空间分辨率为 5 m,影像获取时间为 2023 年 8 月 23 日和 2023 年 8 月 29 日。利用 PIE-SAR 雷达影像处理软件对这 4 景遥感影像数据进行复数转换、多视、滤波、地理编码,分贝 (DB) 转换、镶嵌融合和裁剪等处理,并将 GF-3 空间分辨率重采样为 10 m。

1.2.2 样本数据

样本质量的高低直接关系到农作物分类的精度,本文借助 Google Earth Engine 高分辨率影像采用目视解译的方法,依据相对均匀分布、覆盖主要地表覆盖类型、重点突出典型农作物的原则对样本点进行采样,并进行实地调查 (2023 年 9 月 28 日—2023 年 9 月 30 日,实际

1.2 数据来源与预处理

1.2.1 影像数据

本文研究对象是颍上县秋季典型农作物——玉米、水稻和大豆,根据当地农作物的物候期^[14] (表 1),八月上旬到九月上旬为作物发育的中后期,此期间 3 种农作物具有高辨识度,因此选择的影像时间为 2023 年 8 月份。

(1) Sentinel-2 影像数据

Sentinel-2 是由欧洲空间局和欧洲委员会共同推出,具有高分辨率、多光谱成像的光学卫星。同一轨道上的双卫星相距 180°,使得重访周期由单颗 10 d 缩短为 5 d。Sentinel-2 所搭载的成像光谱仪具有 13 个波段:4 个波段位于 10 m (B2、B3、B4、B8)、6 个波段位于 20 m (B5、B6、B7、B8A、B11、B12)、3 个波段位于 60 m (B1、B9、B10) 的空间分辨率处^[15]。Sentinel-2 目前主要应用于土地管理、农业、林业、灾害监测、救援行动、风险测绘和安全等领域。

本研究采用的 Sentinel-2 数据来源于哥白尼数据下载中心 (<https://search.asf.alaska.edu/>),选取研究区域云量较少,且质量较好 (2023 年 8 月 12 日) 的两景数据,产品等级为 L2A 级。使用 SNAP 软件对数据进行处理,同时将波段空间分辨率重采样为 10 m,并对其进行镶嵌融合和裁剪。

(2) GF-3 雷达数据

高分三号 (GF-3) 卫星于 2016 年 8 月发射,是中国首颗分辨率达到 1 m 的 C 频段多极化合成孔径雷达 (Synthetic Aperture Radar, SAR) 卫星,GF-3 具有高分辨率、成像带宽大、辐射精度高、多成像模式的特点,能够全天候实现对陆地海洋的实时监测。GF-3 成像模式有滑块聚束 (SL)、条带成像 (UFS、FSI、FSII、SS、QPSI、QPSII)、扫描成像 (NSC、WSC、GLO)、波成像 (WAV)、扩展入射角 (EXT) 等 12 种^[16]。

工作时间约 1 天) 验证,随机抽取了五十铺乡、建颍乡、黄桥镇、谢桥镇、江店孜镇、黄坝乡 6 个乡镇,共验证了样本点 853 个,其准确率达到 98% 以上。本文将研究区域地表覆盖分为玉米、水稻、大豆、居民地、道路、水体、林地等 7 类。其中玉米 554 个、水稻 995 个、大豆 539 个、居民地 313 个、道路 168 个、水体 292 个、林地 139 个样本点。样本分布情况如图 1 所示。将样本数据按照 3:7 的比例随机分为验证样本和训练样本。

1.2.3 特征提取

本文主要选取研究区 4 类特征变量:光谱、指数、纹理和极化,构建了共计 58 个特征数据集,其中光谱特征 10 个、指数特征 14 个,纹理特征 32 个、极化特征 2 个,见表 2。

表 2 遥感影像特征提取
Table 2 Remote sensing image feature extraction

类型 Type	名称 Name	缩写 Abbreviation	公式及说明 Formula and explanation	特征个数 Number of features	
光谱特征 Spectral features	波段	Band (B)	B2、B3、B4、B5、B6、B7、B8、B8A、B11、B12	10	
指数特征 Exponential features	归一化植被指数	NDVI	$(B8-B4) / (B8+B4)$	14	
	归一化植被指数红边 1	NDVIre1	$(B8A-B5) / (B8A+B5)$		
	归一化植被指数红边 2	NDVIre2	$(B8A-B6) / (B8A+B6)$		
	归一化植被指数红边 3	NDVIre3	$(B8A-B7) / (B8A+B7)$		
	增强型植被指数	EVI	$2.5(B8-B4) / (B8+6B4-7.5B2+1)$		
	陆地表面水分指数	LSWI	$(B8-B11) / (B8+B11)$		
	土壤调整植被指数	SAVI	$(B7-B3) / (B7+B3+0.5)(1+0.5)$		
	比值植被指数	RVI	$B7 / B3$		
	归一化耕作指数	NDTI	$(B11-B12) / (B11+B12)$		
	新型倒红边叶绿素指数	IRECI	$(B7-B4) / (B5/B6)$		
	红边叶绿素指数	Cire	$B7 / B5 - 1$		
	地面叶绿素指数	MTCI	$(B6-B5) / (B5 -B4)$		
	归一化秸秆指数	NDRl	$(B4-B12) / (B4+B12)$		
	简单耕作指数	STI	$B11 / B12$		
纹理特征 Texture features	对比度	ST2_B1_Contrast,ST2_B2_Contrast, GF3_HH_Contrast,GF3_HV_Contrast		32	
	二阶矩	ST2_B1_Second_Moment, ST2_B2_Second_Moment, GF3_HH_Second_Moment, GF3_HV_Second_Moment			
	相关性	ST2_B1_Correlatlon, ST2_B2_Correlatlon, GF3_HH_Correlatlon, GF3_HV_Correlatlon			
	均值	ST2_B1_Mean,ST2_B2_Mean, GF3_HH_Mean,GF3_HV_Mean			
	熵	ST2_B1_Entropy,ST2_B2_Entropy, GF3_HH_Entropy,GF3_HV_Entropy			
	方差	ST2_B1_Variance, ST2_B2_Variance, GF3_HH_Variance, GF3_HV_Variance			
	同质性	ST2_B1_Homogenelty, ST2_B2_Homogenelty, GF3_HH_Homogenelty, GF3_HV_Homogenelty			
	不相似性	ST2_B1_Dissimilarity, ST2_B2_Dissimilarity, GF3_HH_Dissimilarity, GF3_HV_Dissimilarity			
	极化特征 Polarization features	HH 极化后向散射系数	GF3_HH_B		2
		HV 极化后向散射系数	GF3_HV_B		

注：B2、B3、B4、B5、B6、B7、B8、B8A、B11、B12 分别是指 Sentinel-2 遥感影像中的波段名称；ST2_B1_xx 和 ST2_B2_xx 分别是指 Sentinel-2 影像进行主成分分析后的第一主成分和第二主成分所提取的纹理特征；GF3_HH_xx 和 GF3_HV_xx 分别是指 GF-3 遥感影像 2 种不同极化方式所提取的纹理特征，其中 HH 和 HV 分别表示 GF-3 的 2 种极化方式；GF3_HH_B 和 GF3_HV_B 分别是指 GF-3 的 2 种极化后向散射系数。

Note: B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 refer to the specific bands of the Sentinel-2 remote sensing imagery. ST2_B1_xx and ST2_B2_xx represent the texture features extracted from the first and second principal components of the Sentinel-2 imagery after Principal Component Analysis (PCA). GF3_HH_xx and GF3_HV_xx refer to the texture features extracted from GF-3 remote sensing imagery under two different polarization modes, where HH and HV correspond to the two polarization modes of GF-3, respectively. GF3_HH_B and GF3_HV_B refer to the backscatter coefficients of GF-3 under the two polarization modes.

光谱特征：选择 Sentinel-2 原始遥感影像的 10 个波段。

指数特征：计算常见的植被指数，考虑到 Sentinel-2 拥有红边波段的特点，本文特加入了与农作物生长有关的红边指数。共计 14 个指数特征。

纹理特征：利用灰度共生矩阵提取影像的纹理特征。由于 Sentinel-2 拥有多个波段，因此首先对原始影像进行主成分分析^[17]。第一主成分累计贡献率为 93.51%，第二主成分累计贡献率为 98.76%，选取这 2 个主成分来提取 Sentinel-2 的纹理特征。由于 GF-3 雷达影像（FSI 成像模式）本身只有 HH 和 HV2 个极化，因此直接对 GF-3 进行提取纹理特征。利用 GLCM 提取常见的 8 个相关参数：均值 (mean)、方差 (variance)、同质性 (homogeneity)、对比度 (contrast)、不相似性 (dissimilarity)、熵 (entropy)、二阶矩 (second moment)、相关性 (correlation)，共提取

32 个纹理特征。

极化特征：选取的 GF-3 是精细条带 1（FSI）成像模式，极化方式具有双极化特点，利用 GF-3 提取 HH、HV 极化后向散射系数，共 2 个极化特征。

2 方案设计及模型构建

2.1 方案设计

本文将所涉及的关于特征优选算法、分类方法、特征优选算法和分类方法组合分别统称为算法、方法和模型。为提高农作物分类精度，共设计了 3 种试验方案，如表 3 所示。方案 1 是基于不同特征优选算法和不同机器学习分类方法进行组合试验，探究不同优选算法与分类方法之间的关联；方案 2 是基于相同特征优选算法的不同分类方法集成试验，分析在相同特征算法下使用不同分类方法集成模型的表现；方案 3 是基于相同分类方

法的不同特征优选算法模型集成试验, 分析在相同分类方法下使用不同特征优选算法集成模型的表现。主要目标是通过不同试验设计, 探索并确定最优的特征优选算法与分类器组合, 以期找到针对既定农作物分类任务的最优解决方案。

本研究首先对 Sentinel-2 光学影像和 GF-3 雷达影像进行预处理, 从中分别提取光谱、指数、纹理和极化等特征。随后采用典型的特征优选方法对多维遥感特征进行降维优选, 分析不同特征对农作物识别的重要性, 并

以其重要程度进行排序, 最终选取最重要的特征对农作物进行分类, 采用混淆矩阵的方法评估不同优选特征对农作物分类的精度。同时将 RF、SVM、XGBoost 等机器学习分类结果进行对比, 分析不同分类方法相对不同优选算法所构建模型性能的优劣。具体技术路线如图 2 所示。采用 2023 年全球 10 m 精度土地覆盖数据 (<https://livingatlas.arcgis.com/landcoverexplorer/>) 中耕地范围对分类结果做掩膜处理, 只分析研究区域内的作物识别结果。

表 3 不同特征优选算法和机器学习方法组合试验方案

Table 3 Experimental scheme for combining different feature selection algorithms and machine learning methods

试验方案 Experimental scheme	方案说明 Scheme description	模型名称 Model name
方案 1 Scheme 1	基于不同特征优选算法和不同机器学习分类方法进行组合试验, 探究不同优选算法与分类器之间的关联	RF_RFE+RF、RF_RFE+SVM、RF_RFE+XGBoost CST+RF、CST+SVM、CST+XGBoost Relief F+RF、Relief F+SVM、Relief F+XGBoost
方案 2 Scheme 2	基于相同特征优选算法的不同分类方法集成试验, 分析在相同特征算法下使用不同分类方法集成模型的表现	RF_RFE+RSX (集成 RF_RFE+RF、RF_RFE+SVM、RF_RFE+XGBoost) CST+RSX (集成 CST+RF、CST+SVM、CST+XGBoost) Relief F+RSX (集成 Relief F+RF、Relief F+SVM、Relief F+XGBoost)
方案 3 Scheme 3	基于相同分类方法的不同特征优选算法模型集成试验, 分析在相同分类方法下使用不同特征优选算法集成模型的表现	RCR+RF (集成 RF_RFE+RF、CST+RF、Relief F+RF、) RCR+SVM (集成 RF_RFE+SVM、CST+SVM、Relief F+SVM) RCR+XGBoost (集成 RF_RFE+XGBoost、CST+XGBoost、Relief F+XGBoost)

注: RF_RFE、CST 和 Relief F 分别是指随机森林的递归特征消除算法、卡方检验特征选择算法和基于 Relief 算法扩展的 Relief F 特征选择算法; RF、SVM 和 XGBoost 分别是指随机森林、支持向量机、极限梯度提升机器学习方法; RSX 和 RCR 分别是将 3 种机器学习方法和 3 种特征优选算法集成后的简称。

Note: RF_RFE, CST, and Relief F refer to the Recursive Feature Elimination (RFE) algorithm of Random Forest, the Chi-Square (CST) feature selection algorithm, and the Relief F feature selection algorithm based on the extension of the Relief algorithm, respectively. RF, SVM, and XGBoost refer to the Random Forest, Support Vector Machine, and Extreme Gradient Boosting machine learning methods, respectively. RSX and RCR are abbreviations for the integration of these three machine learning methods with three feature selection algorithms.

2.2 特征优选算法

2.2.1 随机森林的递归特征消除算法

随机森林的递归特征消除算法 (random forest-recursive feature elimination, RF_RFE) 结合了随机森林 (random forest, RF) 和递归特征消除 (recursive feature elimination, RFE) 的思想来选择最优特征子集。首先使用 RF 对所有特征进行训练, 并获取每个特征的重要性得分, 随后通过递归应用 RFE 算法, 不断减少特征集大小, 直到达到所需 (最优) 特征数。在每轮应用 RFE 时, 都会通过 RF 对特征重要性进行评估, 最终得到个最佳特征集。RF_RFE 算法的优点是充分利用了随机森林对特征重要性评估和随机递归消除算法对特征选择的能力, 不仅可以减少数据冗余和噪声, 而且还能提高模型的泛化能力, 保留最重要的特征。在 RF 算法随机抽取样本的过程中会随机抽取部分数据集作为样本用来训练决策树, 未被抽中的数据称为袋外数据 (out of bag, OOB), OOB 计算模型预测的错误率称为 OOB 误差。其中特征变量重要性计算式^[18]为

$$X(A) = \frac{1}{N} \sum_{t=1}^N (B_{nt}^A - B_{ot}^A) \quad (1)$$

式中 X 为特征变量的重要性; N 为树的个数; B_{ot}^A 为某个特征变量 A 未加入噪声时第 t 棵决策树的 OOB 误差; B_{nt}^A 为某个特征变量 A 加入噪声时第 t 棵决策树的 OOB 误差。如果袋外数据准确率大幅下降 (B_{nt}^A 变大), 则说明这个特征对于样本预测结果有很大影响, 从而证明此特征重要性较高。

2.2.2 卡方检验特征选择算法

卡方检验 (chi-squared test, CST) 是基于统计学中

的卡方统计量, 通过观察频数与期望频数之间的差异来判断变量之间是否存在显著性关联。卡方检验特征选择是基于过滤式模型的特征选择方法, 属于非参数检验的范畴。在遥感分类特征选择时与 SelectKBest 类结合, 可以用来计算特征对目标类别的相关性, 相关性分数越高说明该特征对影像分类越重要, 贡献率越高。卡方检验的计算式为:

$$X^2 = \sum ((O - E)^2 / E) \quad (2)$$

式中 X^2 用于衡量实际值与理论值的差异程度, O 为实际值, E 为理论值。在特征选择中, 卡方检验算法可以降低特征维度、提高模型性能并加快模型训练速度, 由于它属于非参数检验, 因此该算法简单有效且能够发现数据之间非线性等优点^[19]。

2.2.3 Relief F 算法

Relief F 算法的核心思想是近邻差异比较、特征权重更新与特征选择。通过比较样本实例之间特征差异性来评估特征重要性。首先从样本中随机选取样本 X , 分别从同类和不同类中找到最近邻的样本并进行比较, 如果不同类之间的距离较大, 则认为该特征对于分类的贡献值是较大的, 增加其权重值; 反之降低其权重。最后重复上述步骤, 将 n 次计算结果的均值作为每个特征的最终权重, 计算式^[20]为

$$W(A_{i+1}) = W(A_i) - \frac{1}{nk} \sum_{h \in H} |R_i - h_i| + \frac{1}{nk} \sum_{m \in M} |R_i - m_i| \quad (3)$$

式中 $W(A_i)$ 为特征 i 的权重值, R_i 为随机抽取的样本, $\sum_{h \in H} |R_i - h_i|$ 为 k 个同类最近邻样本与样本 R 在特征上的距离之和, $\sum_{m \in M} |R_i - m_i|$ 为 k 个不同类最近邻样本与样本 R 在特征上的距离之和, n 是样本总数, H 和 M 分别为

与目标样本同类和不同类样本的最邻近样本集合, h_i , m_i 分别为同类邻近样本 h 与不同类邻近样本 m 在第 i 个

特征上的值。该算法具有一定的随机性, 因此求取各个特征值权重的平均值作为各特征的最最终权重。

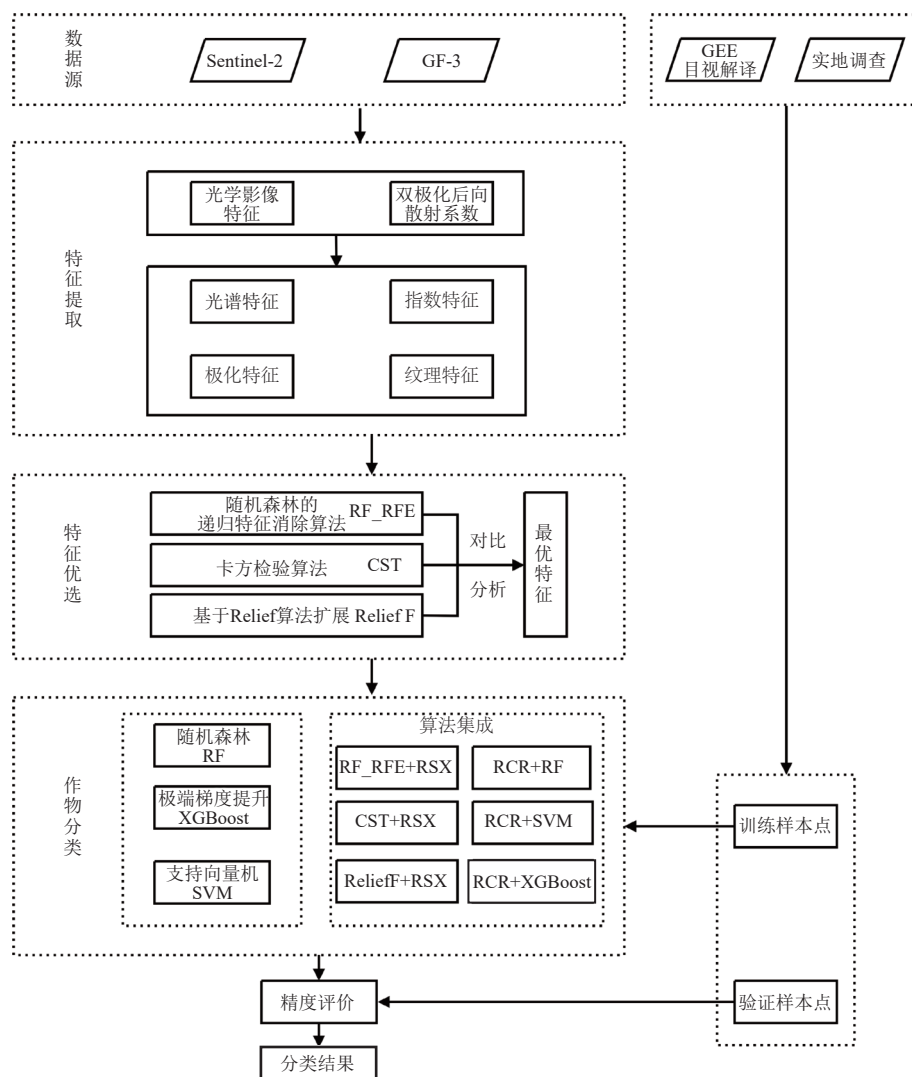


图2 技术路线图

Fig.2 Technology roadmap

2.3 分类方法

2.3.1 随机森林

RF 是一种由决策树并行构成的集成算法, 结果通过投票或取均值使得整体模型具有较高的精确度和泛化性能, 同时具有良好的稳定性, 被广泛应用于遥感图像分类。随机森林是一种有监督的机器学习方法, 在学习过程中依赖于分类标签, 来识别图像中不同的类别。因其运算速度快, 准确率高、抗噪能力强, 已经成为最广泛使用的分类算法之一^[21]。

2.3.2 极限梯度提升

极限梯度提升 (extreme gradient boosting, XGBoost) 是一种基于梯度提升决策树且可扩展的机器学习方法。通过集成多个弱分类器, 利用梯度提升算法逐步优化模型, 进而提高分类精度。在机器学习和数据挖掘方面有着极为广泛的应用。XGBoost 作为一种强大的集成学习模型, 在分类任务中具有较强的灵活性。它引入了正则化, 通过控制模型复杂度, 以提高模型的泛化能力和过拟合能力, 同时也可以对特征重要性进行评估^[22]。

2.3.3 支持向量机

支持向量机 (support vector machine, SVM) 在本质上是一种二分类模型, 是在特征空间中定义最大间隔的线性分类器。基本原理是通过构造最优分割超平面, 以此来实现训练样本的分类^[23]。SVM 是非参数方法, 因此数据即使不满足标准概率密度分布, 该方法仍然适用。SVM 具有较好的鲁棒性和高泛化能力, 不需要太多数据仍然能实现较好的分类, 特别适合用来研究遥感图像分类问题^[24,25]。目前, 支持向量机分类器已经被广泛应用于农作物种植信息提取和分类中^[26]。但是在处理多分类问题仍需进一步的研究, 来克服传统二分类的缺点^[27]。

2.4 精度评价方法

文章基于混淆矩阵的计算结果, 从 5 个方面对分类结果进行精度评价: 总体精度 (overall accuracy, OA)、Kappa 系数、F1 得分、用户精度 (user accuracy, UA) 和制图精度 (producer accuracy, PA)^[28-30]。其中 OA 是被正确分类的像元数和总类别像元数的比值, 反映了算法总体分类效果; Kappa 系数是指分类与完全随机分类产生错误减少的比例, 表示地面实际数据与预测值之间

的一致性程度；F1 得分是评估不平衡数据分类时模型性能的重要指标，是精度和召回率的调和平均数；UA 是指被该类别的样本点被正确划分的比率，反映了分类过程中被错分的误差；PA 是指验证样本点被正确划分的概率，反映了分类过程中的漏分误差。

3 结果与分析

3.1 特征重要性分析

本文共选择了 58 个特征参与颍上县秋季农作物遥感识别分类，利用 RF_RFE、CST、Relief F 3 种特征优选算法对所有特征进行优选，并对所选特征进行重要性评估。为了避免人为因素的影响，本文有关方法中的参数利用 Optuna 算法和交叉验证的方式对参数进行比较分析，确定最终参数。

针对特征个数的选择，每次剔除重要性得分最低的特征，剩余的特征参与分类，并计算每次分类精度。重复以上过程，将每次结果进行输出，最终确定特征个数与分类精度之间的关系。为了便于比较，本文统一使用 RF 作为特征优选后的分类器，对比 RF_RFE、Relief F 和 CST 算法优选特征个数对分类精度的影响，不同特征个数对应的 F1 得分如图 3 所示。

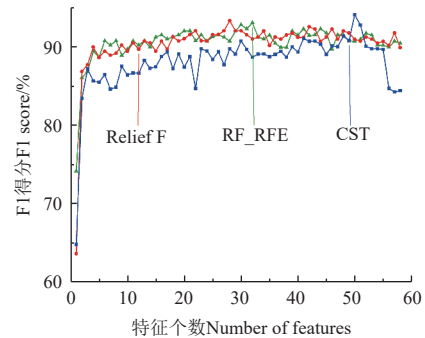


图 3 不同特征个数下 F1 得分

Fig.3 F1 score under different number of features

RF_RFE、Relief F 和 CST 特征优选算法精度最高时对应的特征如表 4 所示。由表中可以得出，每一个优选算法中，红边波段都被优选了出来。对于 Relief F 特征优选算法，优选出来的特征重要性得分（仅展示了排名靠前的前 20 个）如图 4 所示。短波红外波段的加入（B11、B12），对于农作物分类的贡献值较大，相关红边波段 B5、B6、B7 的加入，在一定程度上对于农作物分类精度的提高也比较有效。

表 4 不同特征优选算法下最优遥感影像特征集

Table 4 Optimal Remote Sensing Image Feature Set under Different Feature Selection Algorithms

算法 Algorithm	名称 Name	Number
RF_RFE	GF3_HH_Dis, ST2_B2, ST2_B3, ST2_B4, ST2_B5, ST2_B6, ST2_B7, ST2_B8, ST2_B8A, ST2_B11, ST2_B12, ST2_B1_Ent, ST2_B1_Mea, ST2_B1_Sec, ST2_B2_Dis, ST2_B2_Ent, ST2_B2_Mea, ST2_B2_Sec, ST2_B2_Var, ST2_Cire, ST2_IReCI, ST2_LSWI, ST2_MTCI, ST2_NDRI, ST2_NDTI, ST2_NDVI, ST2_NDVIr2, ST2_NDVIr3, ST2_RVI, ST2_STI, GF3_HH_B, GF3_HV_B	32
Relief F	ST2_B6, ST2_B11, ST2_B1_Mea, ST2_B7, ST2_NDVIr2, ST2_B8A, ST2_B8, ST2_B12, ST2_LSWI, ST2_IReCI, ST2_B5, ST2_EVI, ST2_SAVI, ST2_RVI, ST2_B3, ST2_B2_Mea, ST2_Cire, ST2_B4, ST2_NDRI, ST2_STI, ST2_NDVI, ST2_NDTI, ST2_B2, ST2_NDVIr1, GF3_HV_Cor, GF3_HV_Sec, ST2_B2_Sec, ST2_B1_Ent	28
CST	GF3_HH_Con, GF3_HH_Cor, GF3_HH_Dis, GF3_HH_Ent, GF3_HH_Mea, GF3_HH_Sec, GF3_HH_Var, GF3_HV_Con, GF3_HV_Cor, GF3_HV_Dis, GF3_HV_Ent, GF3_HV_Mea, GF3_HV_Sec, GF3_HV_Var, ST2_B3, ST2_B4, ST2_B5, ST2_B6, ST2_B7, ST2_B8, ST2_B8A, ST2_B11, ST2_B12, ST2_B1_Con, ST2_B1_Cor, ST2_B1_Dis, ST2_B1_Ent, ST2_B1_Hom, ST2_B1_Mea, ST2_B1_Sec, ST2_B1_Var, ST2_B2_Con, ST2_B2_Cor, ST2_B2_Dis, ST2_B2_Ent, ST2_B2_Mea, ST2_B2_Sec, ST2_B2_Var, ST2_Cire, ST2_EVI, ST2_IReCI, ST2_LSWI, ST2_MTCI, ST2_NDVI, ST2_NDVIr1, ST2_NDVIr2, ST2_NDVIr3, ST2_RVI, ST2_STI, ST2_SAVI	50

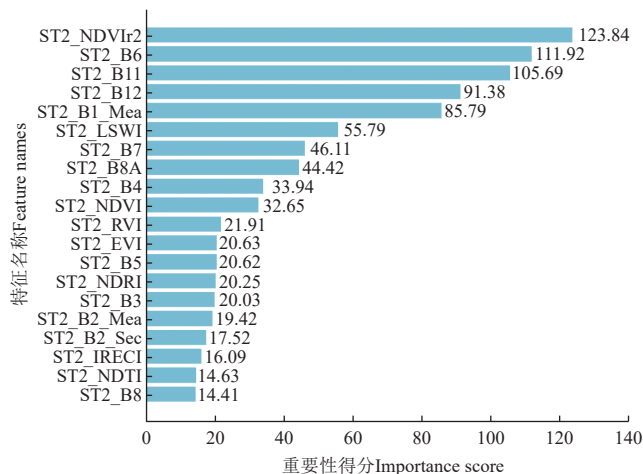


图 4 特征名称及其对应的重要性得分

Fig.4 Feature names and their corresponding importance scores

3.2 不同特征优选方法对比

由图 3 可知，在特征个数达到最优时，整体曲线呈现上升趋势，当达到最高点后，随着特征个数的增加，F1 得分整体呈下降趋势。表 5 中 3 种分类结果的 F1 得分均高于 93%，对于农作物都有较好的分类效果。由表 5 和表 6 知：RF_RFE 在特征个数为 32 时，总体精度 93.06%，Kappa 系数 0.888 5，F1 最高得分 93.03%；Relief F 算法在特征个数为 28 时，总体精度为 93.39%，Kappa 系数 0.893 3，F1 最高得分 93.31%；而 CST 算法在特征个数为 50 时，总体精度为 94.08%，Kappa 系数 0.902 5，F1 最高得分 94.06%。图 3 中可以看出 RF_RFE 分类精度最低。虽然 CST 算法 Kappa 系数稍高于 Relief F，但是在特征优选方面，CST 所选择的特征相对较多，而 Kappa 系数和 F1 得分仅比 Relief F 优选算法高 0.0092 和 0.75 个百分点。因此，Relief F 不仅在特征优选方面表现较好，

而且准确率较高，整体表现方面也较为优秀。而卡方检验优选通常会忽略特征之间的相关性，当特征维度不高时，对于特征选择表现不是很好。从用户精度和制图精度可以看出，不同特征优选算法在处理不同农作物时效果存在差异，RF_RFE、CST、Relief F 分别在大豆、水稻、玉米分类上表现最为突出，但整体来看，Relief F 优选算法效果最佳。

表 5 不同特征优选方法和 RF 组合地物分类的总体精度、Kappa 系数和 F1 得分

Table 5 Overall accuracy, Kappa coefficient, and F1 score of land cover classification with different feature selection methods combined with RF

算法 Algorithm	RF_RFE	CST	Relief F
特征个数	32	50	28
总体精度/%	93.06	94.08	93.39
Kappa 系数	0.888 5	0.902 5	0.893 3
F1 得分/%	93.03	94.06	93.31

为了更直观、清晰地评估分类结果的准确性，本文采用了 2023 年 8 月 2 日获取的 GF-1 卫星数据，融合全色与多光谱图像，得到 2 m 分辨率的高精度遥感影像，并对 3 种优选算法分类结果进行了可视化处理。如图 5

是 3 种特征优选算法基于 RF 分类器进行分类的结果，可以看出 3 种分类结果差异较小，符合研究区域当地种植分布。通过细节对比（表 7）分析可知，当存在混合种植时，CST 和 Relief F 特征优选算法能够将其分类出来，但是 RF_RFE 较难区别。因此，将选择 Relief F 作为最佳的特征优选算法，主要原因是：从选择特征个数来看，Relief F 选择特征最少，但精度相对较高；从精度来看，虽然 CST 特征优选算法比 Relief F 的 F1 得分高了 0.75 个百分点，但选择的特征较多，比 Relief F 选择的个数多 22 个，在特征降维方面表现不够好。

表 6 不同特征优选方法和 RF 组合地物分类的制图精度和用户精度

Table 6 Mapping accuracy and user accuracy of land cover classification with different feature selection methods combined %

指标 Index	PA	UA	PA	UA	PA	UA
玉米 Corn	90.82	89.00	96.72	88.06	90.29	93.00
水稻 Rice	95.34	92.00	95.03	94.44	96.79	90.50
大豆 Soybean	90.82	100.00	90.24	98.67	89.90	100.00

注：PA 和 UA 分别为玉米、水稻与大豆进行分类的制图精度和用户精度。
Note: PA and UA refer to the mapping accuracy and user accuracy for the classification of maize, rice, and soybean, respectively.

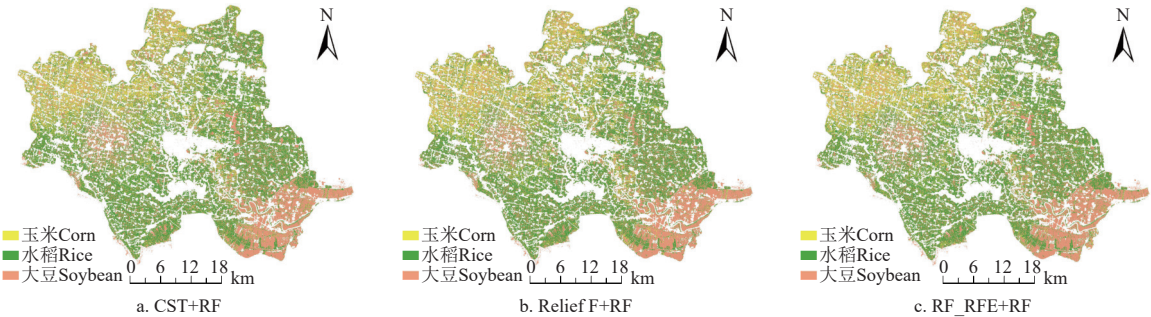
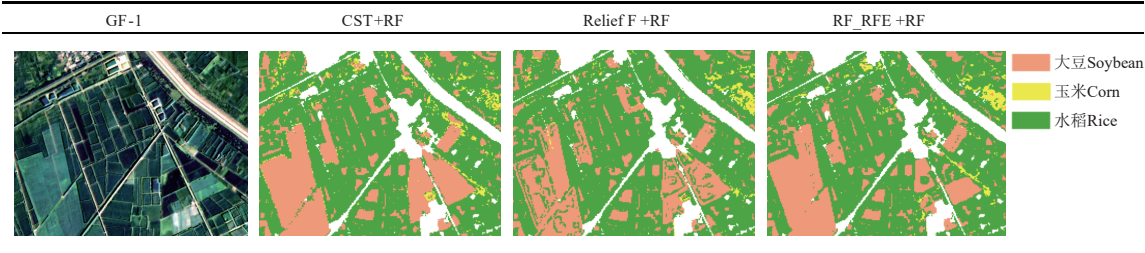


图 5 不同特征优选方法下的整体分类图
Fig.5 Overall classification chart under different feature selection methods

表 7 不同特征优选方法下的局部结果图
Table 7 Local result maps under different feature selection methods



3.3 不同分类方法对农作物识别影响

基于 ReliefF 特征优选算法，分别使用 RF、XGBoost、SVM 机器学习分类器对农作物进行训练和预测，将预测结果可视化如图所示。通过图 6 对比可知，RF（图 6a）和 XGBoost（图 6b）分类相差不大，但是 SVM（图 6c）明显能看出差异较大；SVM 在分类时，误把水稻分为玉米，并没有准确将玉米和水稻区分。从 RF 和 XGBoost 分类结果看，西北部玉米偏多，而 SVM 整个北部玉米都较多，且密度更大，这说明当存在作物混种时，SVM 更难将其进行区分；从分类器原理来看，SVM 更适合

二分类问题，对于多分类问题，没有前两种分类器表现好。

使用不同分类方法对农作物（玉米、水稻、大豆）进行分类，对比结果如表 8 和表 9 所示，可知，RF 的分类精度相比 XGBoost 和 SVM 表现较为优越，总体精度为 93.39%，比 XGBoost 和 SVM 分别高 1.36 个百分点和 8.81 个百分点；Kappa 系数比 XGBoost 和 SVM 分别高 0.021 0 和 0.131 2；F1 得分分别高 1.31 和 8.78 个百分点。分析结果可知，RF 比 XGBoost 和 SVM 更适合用于研究农作物分类。在 RF 分类器下玉米、水稻制图精度

(PA) 和用户精度 (UA) 均大于 XGBoost 和 SVM, PA 精度分别为 90.29%、96.79%; UA 精度分别为 93.00%、90.50%; XGBoost 分类器在分类预测时, 大豆 PA 值比 RF 分类器下大 0.92 个百分点, 考虑可能是因为 XGBoost 在对大豆进行分类时, 表现更为出色, 但 RF 整体效果还是优于 XGBoost。从数值来看, SVM 分类器总体精度、Kappa 系数和 F1 得分均最低。综上所述,

Relief F 特征优选算法结合 RF 分类器效果较好, 分类精度高于 Relief F 和 XGBoost 或 SVM 分类器结合。在实际应用中, 由于遥感数据往往呈现出高度的非线性特征, 而 SVM 更倾向于处理线性数据, 在空间中寻找一个超平面来对样本进行分割。因此相比较 RF 算法, SVM 在本文研究中分类精度最低, 而 RF 在分类算法效果最好, 可以有效识别出农作物。

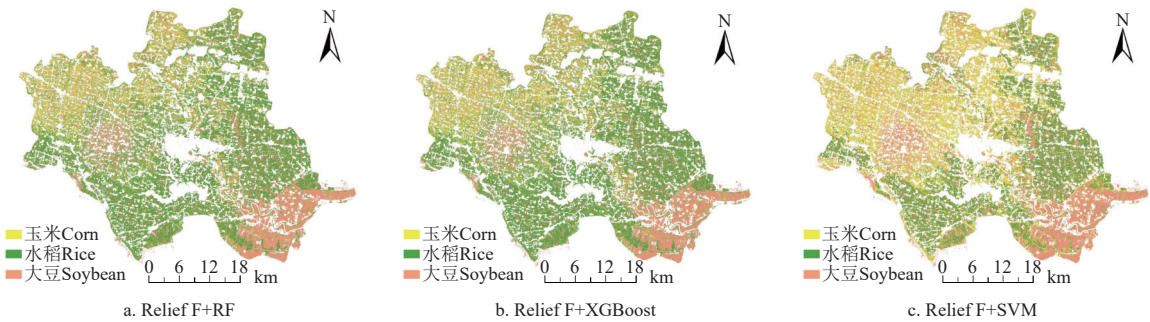


图 6 基于 Relief F 和不同机器学习方法组合分类结果图
Fig.6 Classification result chart based on combination of Relief F and different machine learning methods

表 8 基于 Relief F 和不同机器学习方法组合的地物分类总体精度、Kappa 系数和 F1 得分
Table 8 Overall accuracy, Kappa coefficient, and F1 score of land cover classification based on Relief F combined with different machine learning methods

分类算法 Classification algorithm	RF	XGBoost	SVM
总体精度/%	93.39	92.03	84.58
Kappa 系数	0.893 3	0.872 3	0.762 1
F1 得分/%	93.31	92.00	84.53

表 9 基于 Relief F 和不同机器学习方法组合的地物分类制图精度和用户精度
Table 9 The mapping accuracy and user accuracy of land cover classification based on Relief F combined with different machine learning methods

评价指标 Evaluation metrics	PA	UA	PA	UA	PA	UA
玉米 Corn	90.29	93.00	88.00	88.00	72.90	78.79
水稻 Rice	96.79	90.50	94.76	90.50	89.02	79.78
大豆 Soybean	89.90	100.00	90.82	100.00	88.98	98.13

为了验证结果的准确性以及方法的可行性, 本文通过计算方案 1、方案 2 以及方案 3 集成模型的总体精度和 Kappa 系数和 F1 得分, 结果如表 10。根据方案 1 输出结果, 从精度分析: CST 和 RF、CST 和 SVM、RF_RFE 和 XGBoost 是在各分类器中和 3 种特征优选算法的最优组合, 总体精度、Kappa 系数和 F1 得分分别为 94.08%、0.902 5 和 94.06%, 89.80%、0.833 9 和 89.73%, 93.06%、0.888 2 和 93.02%; 从特征维度分析: Relief F 和 RF 组合特征维度为 28 维, 总体精度、Kappa 系数和 F1 得分为 93.39%、0.893 3 和 93.31%, 比 CST 和 SVM、RF_RFE 和 XGBoost 组合分别高 3.59 个百分点、0.059 4 和 3.58 个百分点, 0.33 个百分点、0.005 1 和 0.29 个百分点, 且特征维度分别比他们低 22 维和 4 维。虽然 CST 和 RF 组合总体精度、Kappa 系数和 F1 得分比 Relief F 和 RF 组合高 0.69 个百分点、0.009 2 和 0.75 个百分点, 但特征

维度却比其高 22 维。综上, Relief F 和 RF 组合为最优结果。为了结果更具有说服力, 分别将各特征选择与各分类方法最优组合的特征维度都选择 28 维, 分析各组最优组合与 Relief F+RF 的总体精度、Kappa 系数和 F1 得分如表 11。Relief F+RF 的总体精度、Kappa 系数和 F1 得分最高, 分别比 CST+RF、CST+SVM、RF_RFE+XGBoost 高 3.59 个百分点、0.060 2 和 4.58 个百分点, 5.56 个百分点、0.093 5 和 6.70 个百分点, 0.96 个百分点、0.017 1 和 2.05 个百分点。

表 10 不同模型地物分类的总体精度、Kappa 系数和 F1 得分
Table 10 Overall Accuracy, Kappa Coefficient, and F1 Score of Object Classification for Different Models

模型名称 Model name	特征个数 Number of features	总体精度 Overall accuracy/%	Kappa 系数 Kappa coefficient	F1 得分 F1 Score/%
RF_RFE+RF	32	93.06	0.888 5	93.03
RF_RFE+SVM	32	84.83	0.749 1	83.98
RF_RFE+XGBoost	32	93.06	0.888 2	93.02
CST+RF	50	94.08	0.902 5	94.06
CST+SVM	50	89.80	0.833 9	89.73
CST+XGBoost	50	92.11	0.870 8	92.07
Relief F+RF	28	93.39	0.893 3	93.31
Relief F+SVM	28	84.58	0.762 1	84.53
Relief F+XGBoost	28	92.03	0.872 3	92.00
RF_RFE+RSX	32	93.06	0.888 3	93.02
CST+RSX	50	93.04	0.886 4	93.07
Relief F+RSX	28	92.03	0.872 8	92.02
RCR+RF	92.76	0.881 2	92.73	
RCR+SVM	90.13	83.85	89.32	
RCR+XGBoost	93.32	0.892 4	93.41	

结果表明 Relief F+RF 不仅在特征维度选择特征个数最少, 且精度也相对较高, 证明了该方法的可行性。

方案 2 和方案 3 的核心思想是, 分别固定一个特征优选算法或分类方法, 将集成时的 3 个子模型预测结果进行输出, 观察地物被分为哪一类。由于集成时使用了 3 个子模型, 因此每一个地物会被预测 3 次, 分别对 3 次预测结果取众数。即如果这个地物被两个子模型都

分为同一类，则认为分类结果是准确的，如果 3 个子模型预测结果都不一样，则以 RF 或 Relief F 分类结果为准。原因是通过上文试验表明，Relief F 和 RF 组合在进行识别分类时，效果最好。通过对比发现，集成模型的精度并没有 Relief F+RF 模型精度高。分析认为可能是作物混种，导致两种分类器（XGBoost、SVM）都错误识别；也可能 3 个子模型在集成预测结果时，不存在众数，则认为是 Relief F 或 RF 所分类别，即存在人为主观因素，导致结果不准确。

表 11 各分类器最优组合在特征维度为 28 时地物分类的总体精度、Kappa 系数和 F1 得分

Table 11 Overall accuracy, kappa coefficient, and F1 score of object classification for the optimal combination of classifiers at a feature dimension of 28

模型名称 Model name	CST+RF	CST+SVM	RF_RFE+XGBoost	Relief F+RF
特征个数	28	28	28	28
总体精度/%	89.80	87.83	92.43	93.39
Kappa 系数	0.833 1	0.799 8	0.876 2	0.893 3
F1 得分/%	89.73	87.61	92.26	94.31

4 讨论与结论

4.1 讨 论

本文利用 Sentinel-2 光学遥感数据和国产 GF-3 雷达遥感数据协同，通过特征优选算法和分类方法组合来研究农作物分类精度。分析特征选择算法优选后的结果发现，3 个优选特征集均包含了 GF-3 的相关特征，说明极化特征和纹理特征（GF-3）的加入对于提高分类精度有一定的作用，与王镭等^[31]研究结果相吻合。对比 3 种特征优选算法（Relief F、RF_RFE、CST）和 3 种机器学习分类方法（RF、XGBoost、SVM）可知，不同特征优选算法所选择的最优特征，应对不同机器学习分类方法会有不同的结果。在研究基于特征优选的农作物分类时，选择最适合特征优选算法的分类器进行研究具有重要意义。因此本文结合国内外遥感数据，利用特征优选方法对研究区域内典型农作物进行识别分类，分析得到 Relief F 和 RF 相结合时，所优选的特征维度最低，分类精度最高。这一结果与张东彦等^[32]基于 Sentinel-2 影像对大豆种植区进行识别分类，发现 Relief F 和 RF 结合分类精度最高这一结论相符。此外，本研究发现 XGBoost 方法对大豆识别精度也较高。

该方法在研究区虽然取得了较好的效果，但是由于地面采样数据、遥感影像质量以及数据源的限制，导致某些地物种植面积小于一个像元，使得“混合像元效应”显著，对分类结果的准确性造成一定影响^[33-35]。之后的研究考虑通过以下两种方法来解决这一问题：1) 融合多种遥感影像数据获得更高分辨率以及波段信息更加丰富的影像；2) 对遥感影像进行多尺度分割，基于面向对象方法进行作物的分类，从而提高分类精度。

4.2 结 论

本文基于 Sentinel-2 和 GF-3 遥感数据构建光谱、指数、纹理、极化 4 类特征作为特征变量集、利用特征优选算法（Relief F、RF_RFE、CST）对 58 个特征变量进行特征优选，构建最优特征集，利用 RF 机器学习方法

对研究区典型农作物样本点进行训练，进而实现对整个研究区的农作物进行预测，并采用 XGBoost 和 SVM 进行对比验证分析。研究结果表明：

1) 利用特征优选算法可以很好的实现特征维度降级、避免维数过高导致数据冗余。通过对比发现 Relief F 可以有效的实现降低特征维度，且能够有效对农作物实现分类。Relief F 优选特征个数为 28，比 RF_RFE 和 CST 分别少 4 个和 22 个。

2) 通过对比 3 种特征优选算法可以发现，红边特征 B5、B6、B7 全都参与了分类，说明红边特征对农作物识别分类具有重要作用，同时 B11、B12 波段以及纹理特征的加入也能够提高农作物识别的准确性。不同特征优选算法，对特征的选择也不尽相同，分析知 Relief F 更倾向于光谱特征和指数特征的选择。

3) 对比 3 种不同机器学习方法，发现 Relief F 与 RF 结合对于研究区的作物识别分类更加准确，而 XGBoost 和 SVM 表现较差。基于 Relief F 特征优选条件下，RF 的总体精度、Kappa 系数和 F1 得分分别是 93.39%、0.893 3 和 94.31%，比 XGBoost 和 SVM 分别高 1.36 个百分点、0.021 和 1.31 个百分点，8.81 个百分点、0.131 2 和 8.78 个百分点。

[参 考 文 献]

[1] 张悦琦. 基于国产卫星的水稻种植分布遥感提取研究[D]. 太原: 太原理工大学, 2022.
ZHANG Yueqi. Research on Remote Sensing Extraction of Rice Planting Distribution based on Domestic Satellite[D]. Taiyuan: Taiyuan University of Technology, 2022. (in Chinese with English abstract)

[2] CRAIG M Y B C. Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program[J]. *Geocarto International*, 2011, 26(5): 341-358.

[3] 杨迎港. 基于面向对象和随机森林算法的 GF-2 影像地物分类研究[D]. 焦作: 河南理工大学, 2022.
YANG Yinggang. Research on GF-2 Image Feature Classification based on Object-oriented and Random Forest Algorithm[D]. Jiaozuo: Henan Polytechnic University, 2022. (in Chinese with English abstract)

[4] NATALIJA K, MYKOLA L, SERGI S, et al. Deep learning classification of land cover and crop types using remote sensing data[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(5): 778-782.

[5] 任建强, 陈仲新, 周清波, 等. MODIS 植被指数的美国玉米单产遥感估测[J]. 遥感学报, 2015, 19(4): 568-577.
REN Jianqiang, CHEN Zhongxin, ZHOU Qingbo, et al. MODIS vegetation index data used for estimating corn yield in USA[J]. *Journal of Remote Sensing*, 2015, 19(4): 568-577. (in Chinese with English abstract)

[6] 程彬. 基于支持向量机的乾安县土地利用遥感分类研究[J]. 长春师范大学学报, 2017, 36(12): 86-88.
CHENG Bin. Land use information extraction based on support vector machine using multitemporal remote sensing in Qianan county[J]. *Journal of Changchun Normal University*, 2017, 36(12): 86-88. (in Chinese with English abstract)

[7] MOSLEH K M, HASSAN K Q, CHOWDHURY H E. Application of remote sensors in mapping rice area and forecasting its production: A review[J]. *Sensors*, 2015, 15(1): 769-791.

[8] 郭交, 朱琳, 靳标. 基于 Sentinel-1 和 Sentinel-2 数据融合的农作物分类[J]. 农业机械学报, 2018, 49(4): 192-198.

- GUO Jiao, ZHU Lin, JIN Biao. Crop classification based on data fusion of Sentinel-1 and Sentinel-2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(4): 192-198. (in Chinese with English abstract)
- [9] 王利花, 金辉虎, 王晨丞, 等. 基于合成孔径雷达的农作物后向散射特性及纹理信息分析——以吉林省农安县为例[J]. 中国生态农业学报(中英文), 2019, 27(9): 1385-1393. WANG Lihua, JIN Huihu, WANG Chencheng, et al. Backscattering characteristics and texture information analysis of typical crops based on synthetic aperture radar: A case study of Nong'an county, Jilin province[J]. Chinese Journal of Eco-Agriculture, 2019, 27(9): 1385-1393. (in Chinese with English abstract)
- [10] 王文静, 张霞, 赵银娣, 等. 综合多特征的 Landsat 8 时序遥感图像棉花分类方法[J]. 遥感学报, 2017, 21(1): 115-124. WANG Wenjing, ZHANG Xia, ZHAO Yindi, et al. Cotton extraction method of integrated multi-features based on multi-temporal Landsat 8 images[J]. Journal of Remote Sensing, 2017, 21(1): 115-124. (in Chinese with English abstract)
- [11] 张悦琦, 任鸿瑞. 融合特征优选与随机森林算法的 GF-6 影像东北一季稻遥感提取[J]. 遥感学报, 2023, 27(9): 2153-2164. ZHANG Yueqi, REN Hongrui. Remote sensing extraction of paddy rice in Northeast China from GF-6 images by combining feature optimization and random forest[J]. National Remote Sensing Bulletin, 2023, 27(9): 2153-2164. (in Chinese with English abstract)
- [12] 王李娟, 孔钰如, 杨小冬, 等. 基于特征优选随机森林算法的农耕地土地利用分类[J]. 农业工程学报, 2020, 36(4): 244-250. WANG Lijuan, KONG Yuru, YANG Xiaodong, et al. Classification of land use in farming areas based on feature optimization random forest algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(4): 244-250. (in Chinese with English abstract)
- [13] 卢献健, 张焕铃, 晏红波, 等. 协同 Sentinel-1/2 多特征优选的甘蔗提取方法[J]. 自然资源遥感, 2024, 36(1): 86-94. LU Xianjian, ZHANG Huanling, YAN Hongbo, et al. Sugarcane extraction method based on multi - feature optimal selection of Sentinel - 1/2[J]. Remote Sensing for Natural Resources, 2024, 36(1): 86-94. (in Chinese with English abstract)
- [14] 宋凯达. 江苏省典型农作物的遥感识别与分类研究[D]. 南京: 南京信息工程大学, 2023. SONG Kaida. Research on Remote Sensing Identification and Classification of Typical Crops in Jiangsu Province[D]. Nanjing: Nanjing University of Information Science and Technology, 2023. (in Chinese with English abstract)
- [15] 龚燃. 哨兵-2A 光学成像卫星发射升空[J]. 国际太空, 2015(8): 36-40. GONG Ran. Sentinel-2A satellite launches[J]. Space International, 2015(8): 36-40. (in Chinese with English abstract)
- [16] 张庆君. 高分三号卫星总体设计与关键技术[J]. 测绘学报, 2017, 46(3): 269-277. ZHANG Qingjun. System design and key technologies of the GF-3 satellite[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(3): 269-277. (in Chinese with English abstract)
- [17] 郑淑丹, 郑江华, 石明辉, 等. 基于分形和灰度共生矩阵纹理特征的种植型药用植物遥感分类[J]. 遥感学报, 2014, 18(4): 868-886. ZHENG Shudan, ZHENG Jianghua, SHI Minghui, et al. Classification of cultivated Chinese medicinal plants based on fractal theory and gray level co-occurrence matrix textures[J]. Journal of Remote Sensing, 2014, 18(4): 868-886. (in Chinese with English abstract)
- [18] 张磊, 宫兆宁, 王启为, 等. Sentinel-2 影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23(2): 313-326. ZHANG Lei, GONG Zhaoning, WANG Qiwei, et al. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images[J]. Journal of Remote Sensing, 2019, 23(2): 313-326. (in Chinese with English abstract)
- [19] 沈宇, 李强子, 杜鑫, 等. 玉米大豆生长中后期遥感辨识的指示性特征研究[J]. 遥感学报, 2022, 26(7): 1410-1422. SHEN Yu, LI Qiangzi, DU Xin, et al. Indicative features for identifying corn and soybean using remote sensing imagery at middle and later growth season[J]. National Remote Sensing Bulletin, 2022, 26(7): 1410-1422. (in Chinese with English abstract)
- [20] 李冰, 卢小平, 李新社, 等. 特征优选的 GF-2 影像湿地地表覆盖要素提取[J]. 测绘与空间地理信息, 2018, 41(9): 49-52. LI Bing, LU Xiaoping, LI Xinshe, et al. Extraction method of wetland surface factors using GF-2 images on optimal features[J]. Geomatics & Spatial Information Technology, 2018, 41(9): 49-52. (in Chinese with English abstract)
- [21] 姜伊兰, 陈保旺, 黄玉芳, 等. 基于 Google Earth Engine 和 NDVI 时序差异指数的作物种植区提取[J]. 地球信息科学学报, 2021, 23(5): 938-947. JIANG Yilan, CHEN Baowang, HUANG Yufang, et al. Crop planting area extraction based on Google Earth Engine and NDVI time series difference index[J]. Journal of Geo-information Science, 2021, 23(5): 938-947. (in Chinese with English abstract)
- [22] SHAO Z, AHMAD N M, JAVED A. Comparison of random forest and XGBoost classifiers using integrated optical and SAR features for mapping urban impervious surface[J]. Remote Sensing, 2024, 16(4): 665.
- [23] 刘志刚. 支撑向量机在光谱遥感影像分类中的若干问题研究[D]. 武汉: 武汉大学, 2004. LIU Zhigang. Key Problems of Applying Support Vector Machines to the Classification of Spectral Remote Sensing Imagery[D]. Wuhan: Wuhan University, 2004. (in Chinese with English abstract)
- [24] MALLETT C, BRETAR F, ROUX M, et al. Relevance assessment of full-waveform lidar data for urban area classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(6S): 71-84.
- [25] ALIMJAN G, SUN T, LIANG Y, et al. A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2018, 32(7): 1859012.
- [26] BAUMANN M, OZDOGAN M, KUEMMERLE T, et al. Using the Landsat record to detect forest-cover changes during and after the collapse of the Soviet Union in the temperate zone of European Russia[J]. Remote Sensing of Environment, 2012, 124: 174-184.
- [27] 刘志刚, 李德仁, 秦前清, 等. 支持向量机在多类分类问题中的推广[J]. 计算机工程与应用, 2004(7): 10-13, 65. LIU Zhigang, LI Deren, QIN Qianqing, et al. An analytical overview of methods for multi-category support vector machines[J]. Computer Engineering and Applications, 2004(7): 10-13, 65. (in Chinese with English abstract)
- [28] STEHMAN S V. Selecting and interpreting measures of thematic classification accuracy[J]. Remote Sensing of Environment, 1997, 62(1): 77-89.
- [29] 高国龙, 杜华强, 韩凝, 等. 基于特征优选的面向对象毛竹林分布信息提取[J]. 林业科学, 2016, 52(9): 77-85. GAO Guolong, DU Huaqiang, HAN Ning, et al. Mapping of moso bamboo forest using object-based approach based on the optimal features[J]. Scientia Silvae Sinicae, 2016, 52(9): 77-85. (in Chinese with English abstract)
- [30] 陈谌, 梁雪春. 基于基尼指标和卡方检验的特征选择方法

- [J]. 计算机工程与设计, 2019, 40(8): 2342-2345, 2360.
CHEN Chen, LIANG Xuechun. Feature selection method based on Gini index and chi-square test[J]. Computer Engineering and Design, 2019, 40(8): 2342-2345, 2360. (in Chinese with English abstract)
- [31] 王镭, 赵红莉, 蒋云钟, 等. 月尺度农作物提取中 GF-1 WFV 纹理特征的应用及分析[J]. 自然资源遥感, 2021, 33(3): 72-79.
WANG Rong, ZHAO Hongli, JIANG Yunzhong, et al. Application and analyses of texture features based on GF-1 WFV images in monthly in-formation extraction of crops[J]. Remote Sensing for Natural Resources, 2021, 33(3): 72-79. (in Chinese with English abstract)
- [32] 张东彦, 杨玉莹, 黄林生, 等. 结合 Sentinel-2 影像和特征优选模型提取大豆种植区[J]. 农业工程学报, 2021, 37(9): 110-119.
ZHANG Dongyan, YANG Yuying, HUANG Linsheng, et al. Extraction of soybean planting areas combining Sentinel-2 images and optimized feature model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(9): 110-119. (in Chinese with English abstract)
- [33] 崔璐, 杜华强, 周国模, 等. 决策树结合混合像元分解的中国竹林遥感信息提取[J]. 遥感学报, 2019, 23(1): 166-176.
CUI Lu, DU Huangqiang, ZHOU Guomo, et al. Combination of decision tree and mixed pixel decomposition for extracting bamboo forest information in China[J]. Journal of Remote Sensing, 2019, 23(1): 166-176. (in Chinese with English abstract)
- [34] 周亚男, 陈绘, 刘洪斌. 基于多源数据和 Stacking-SHAP 方法的山地丘陵区土地覆被分类[J]. 农业工程学报, 2022, 38(23): 213-222.
ZHOU Yanan, CHEN Hui, LIU Hongbin. Land cover classification in hilly and mountainous areas using multi-source data and Stacking-SHAP technique[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(23): 213-222. (in Chinese with English abstract)
- [35] 林志坚, 姚俊萌, 苏校平, 等. 基于 MODIS 指数和随机森林的江西省早稻种植信息提取[J]. 农业工程学报, 2022, 38(11): 197-205.
LIN Zhiqiang, YAO Junmeng, SU Xiaoping, et al. Extracting planting information of early rice using MODIS index and random forest in Jiangxi Province, China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(11): 197-205. (in Chinese with English abstract)

Crop identification by synergistic Sentinel-2 and GF-3 multi-feature optimization

ZHANG Qingsong, WANG Jinxin^{*}, HE Xiaohui

(The school of Geo-Science & Technology, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Global climate has posed the significant challenges on agricultural production and food security in recent years. Alternatively, satellite remote sensing can be expected to timely and accurately monitor the crop types and planting structures in precision agriculture. An agricultural survey is then required to estimate the crop yields and disaster impacts. Existing remote sensing data and ground sample can also be fully utilized to combine with feature engineering techniques and machine learning. The accuracy of crop classification can be enhanced after quantitative remote sensing using satellite-derived spatiotemporal big data. However, there are the substantial variations in the feature selection algorithms and machine learning under different scenarios, leading to the significant fluctuations in the recognition and classification accuracy. Therefore, it is very necessary for the individual remote sensing imagery and multi-feature selection during crop recognition and classification. This study aims to realize the crop identification using synergistic Sentinel-2 and GF-3 multiple feature optimization. The study area was taken from the Ying Shang County in Anhui Province, China. Sentinel-2 and GF-3 satellite imagery data was employed to extract 58 feature indicators, including spectral, index, texture, and polarization characteristics. Subsequently, three algorithms of feature selection and three machine learning were combined to design three experimental schemes, in order to explore the effects of feature selection and machine learning techniques on the crop classification. Feature dimensions and classification accuracy were then compared to evaluate the effectiveness of various classification schemes. Several key insights revealed that: (1) The feature selection algorithms were effectively reduced the dimensionality to avoid the data redundancy associated with excessively high dimensions. After that, Relief F algorithm performed the best to reduce the feature dimensions with the effective performance of crop classification. Specifically, Relief F selected 28 features, which was 4 fewer than the number selected by RF_RFE and 22 fewer than those selected by CST. (2) A comparative analysis was performed on the three algorithms of feature selection. It was found that the red-edge feature bands (B5, B6, and B7) were emerged in three algorithms during crop classification. Additionally, the bands B11 and B12 with texture features were further enhanced the accuracy of crop recognition. Different algorithms were varied in the preferences during feature selection; Specifically, Relief F algorithm was tended to favor the spectral and index features. (3) Three machine learning revealed that the combination of Relief F with Random Forest (RF) was achieved in the highest accuracy of crop classification in the study area. While the XGBoost and Support Vector Machine (SVM) shared the inferior performance. In Relief F feature selection, the RF was achieved in an overall accuracy of 93.39%, a Kappa coefficient of 0.8933, and an F1 score of 94.31%. These results indicated that the RF outperformed the XGBoost and SVM by 1.36 percentage point, 0.021 and 1.31 percentage point, while by 8.81 percentage point, 0.1312 and 8.78 percentage point, respectively. This experiment was strongly validated the effectiveness and advanced nature of combined Relief F feature selection with RF classification. Feature selection with suitable machine learning can be expected to significantly enhance the classification performance under various agricultural contexts. These findings can greatly contribute to the effective utilization of remote sensing technologies in agricultural monitoring and decision-making. The relationship between feature selection and classification accuracy can also provide the technical approach for the recognition and classification in precision agriculture.

Keywords: crop; classification; feature; selection; random forest; Sentinel-2; GF-3