

DOI: 10.13718/j.cnki.xdzk.2026.07.003

王克晓, 周蕊, 姚强, 等. 基于空间迁移对象的 SAR 特征优化及西南山区冬油菜遥感识别 [J]. 西南大学学报(自然科学版), 2026, 48(7): 29-42.

基于空间迁移对象的 SAR 特征优化及 西南山区冬油菜遥感识别

王克晓^{1,2}, 周蕊^{1,2}, 姚强^{1,2}, 栗超^{1,2}

1. 重庆市农业科学院农业科技信息研究所, 重庆 401329; 2. 国家数字农业区域创新分中心(西南), 重庆 401329

摘要: 我国西南丘陵山区地块破碎、冬季多云雾, 给油菜的精准遥感识别带来了极大挑战。为提高油菜遥感识别精度, 并满足相对精细的地块尺度应用要求, 本研究联合油菜全生育期的 VH 和 VV 双极化哨兵 1 号(Sentinel-1)数据后向散射、关键生育期极化分解、雷达植被指数等特征以及同年夏季 GF-2 耕地影像分割单元, 基于空间迁移对象构建了一种适用于我国西南山区冬油菜 SAR 遥感识别的 mRMR-PSO (Minimum Redundancy and Maximum Relevance-Particle Swarm Optimization) 特征优化方法, 并分析了关键生育期植被指数和极化分解特征对油菜识别的影响。结果表明: ① 优选特征下, 对 SVM(Support Vector Machine)模型油菜识别贡献较大的特征依次为绿熟期 VH 后向散射、收获期 VV 后向散射、绿熟期 VV 后向散射和现蕾期的 VH 后向散射, 且绿熟期 VH 和 VV 后向散射的高特征值更有利于油菜的识别。② 优选特征下, SVM 油菜识别模型的总体精度 OA(Overall Accuracy)为 88.76%, Kappa 系数为 77.57%, AUC(Area Under the Curve)最高, 为 0.963; 其总体精度 OA 和 Kappa 值均明显高于 LASSO(Least Absolute Shrinkage and Selection Operator)、SVM-RFE(Support Vector Machine-Recursive Feature Elimination)和 XGBoost (Extreme Gradient Boosting)等特征优选结果精度, 其特征与 XGBoost 的特征优选集有较高的相似度, Jaccard 相似度系数为 0.51。③ 相对于关键生育期, 全生育期的散射特征组合能够提高油菜识别精度, F1 值从 85.29% 提高到 89.23%; 油菜关键期植被指数和极化分解特征的加入, 更有助于提高油菜的查准率和预测可信度, 其识别精确率从 85.29% 提高到 90.00%, 而本研究优选特征下的油菜识别精确率为最高 92.86%。

关键词: 冬油菜; 分布提取; 西南丘陵山区; 哨兵 1 号;

特征优选

中图分类号: S127; S565.4

文献标识码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1673-9868(2026)07-0029-14

Optimization of SAR features and remote sensing identification of winter rapeseed in mountainous areas of Southwest China based on spatially migration objects

收稿日期: 2025-11-06

基金项目: 重庆市自然科学基金项目(CSTB2024NSCQ-MSX0804); 重庆市市级财政项目(KYLX20250600058)。

作者简介: 王克晓, 硕士, 助理研究员, 主要从事农业遥感方面的研究。

Wang Kexiao^{1,2}, Zhou Rui^{1,2}, Yao Qiang^{1,2}, Su Chao^{1,2}

1. Institute of Agricultural Science and Technology Information, Chongqing Academy of Agricultural Sciences, Chongqing 401329, China;

2. National Digital Agriculture Regional Innovation Sub-Center (Southwest), Chongqing 401329, China

Abstract: China's southwestern hilly and mountainous regions are cloudy and foggy in winter, coupled with fragmented plots, which pose challenges for the precise remote sensing identification of winter rapeseed. To improve the accuracy of winter rapeseed identification and meet the requirements of fine plot-scale applications, this study integrated backscatter characteristics from dual-polarization Sentinel-1 VH and VV data throughout the entire growth stage, polarization decomposition and radar vegetation indices during key growth stages, and segmentation units derived from GF-2 summer cropland imagery of the same year. Based on spatial migration objects, this study proposed the Minimum Redundancy and Maximum Relevance-Particle Swarm Optimization (mRMR-PSO) feature optimization method suitable for SAR remote sensing identification of winter rapeseed in the mountainous regions of Southwest China, and further analyzed the impact of vegetation indices and polarization decomposition features during key growth stages on winter rapeseed identification. The results showed that: ① Among the preferred features of mRMR-PSO, the features that contributed most significantly to rapeseed identification in the Support Vector Machine (SVM) model were, in descending order of contribution, VH backscatter during the green ripening stage, VV backscatter during the harvest stage, VV backscatter during the green ripening stage, and VH backscatter during the budding stage. Furthermore, higher feature values for VH and VV backscatter during the green ripening stage were more conducive to the identification of winter rapeseed. ② Under optimized features, the Overall Accuracy (OA) of the SVM model reached 88.76%, with a Kappa value of 77.57% and an Area Under the Curve (AUC) of 0.963. Both its OA and Kappa values significantly outperformed feature selection results from Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine-Recursive Feature Elimination (SVM-RFE), and Extreme Gradient Boosting (XGBoost). Its feature set showed substantial similarity to XGBoost's, with a Jaccard coefficient of 0.51. ③ Compared with the critical growth stage, the combination of scattering characteristics across the entire growth period could improve the rapeseed identification accuracy, with F1 increasing from 85.29% to 89.23%. The incorporation of vegetation indices and polarization decomposition features during the critical growth period could further enhance the winter rapeseed identification precision and prediction reliability, with precision rising from 85.29% to 90.00%. Moreover, the precision of winter rapeseed identification achieved a maximum of 92.86% using the optimized feature set proposed in this study.

Key words: winter rapeseed; distribution extraction; hilly and mountainous areas of Southwest China; Sentinel-1; feature optimization

合成孔径雷达(Synthetic Aperture Radar, SAR)遥感不受云雾限制、穿透力强,能够同时捕捉地物的散射特征与极化信息。哨兵1号(Sentinel-1)影像为欧洲航天局的双极化 SAR 数据集,具有周期短、数据源稳定且可免费获取等特点,已广泛应用于农作物遥感监测中^[1-3]。Sentinel-1 影像用于作物识别多作为独立数据源或与光学影像相结合使用,如基于时序 Sentinel-1 散射特征的平原地区小麦种植信息遥感提

取^[4]、融合多时相 Sentinel-1 和 Sentinel-2 影像的大规模农场多种农作物分类^[5]、多时相 Sentinel-1 散射特征结合 Landsat-8 影像的都市农业区冬小麦识别^[6]等。在以上地块较为规整的规模化作物种植区,作物识别总体精度(Overall Accuracy, OA)均为 90%以上。在丘陵山区作物遥感识别应用方面,研究多集中在在地表切割较缓的我国东南低山丘陵地区,如融合 Sentinel-1 时序散射特征与 Sentinel-2 影像的江西南昌县水稻种植识别^[7]、Sentinel-1 影像辅助 Sentinel-2 数据的安徽宣州区多类型农作物识别^[8]、亚米级光学影像语义分割结合时序 Sentinel-1 的福建省浦城县多季水稻地块级识别^[9]等,都融入了作物生育期内的光学影像数据,但也影响了其在多云雾无光学影像可用地区的实用性。而在 Sentinel-1 影像为独立数据源的相关研究中,利用作物关键物候期 Sentinel-1 多时相散射特征与最佳时相组合优化的南昌蒋巷镇水稻种植分布提取分类 OA 可达到 94.30%^[10],但由于其空间分辨率的限制,其识别结果在小尺度上仍难以满足应用需求。另外,在识别方法上,机器学习仍是 SAR 作物识别的主流方法,如极端梯度提升树模型(Extreme Gradient Boosting, XGBoost)^[8, 11]、支持向量机(Support Vector Machines, SVM)^[6, 12]、随机森林(Random Forest, RF)^[3, 7, 10, 13]、人工神经网络^[14]等,都取得了令人满意的效果。近年来,随着计算机技术的发展,各种深度学习方法在 SAR 作物识别相关研究中脱颖而出,如卷积神经网络(Convolutional Neural Network, CNN)^[15-16]、卷积自编码神经网络(Convolutional-Autoencoder Neural Network, C-AENN)^[17]及高分辨率神经网络(High Resolution Net, HRNet)^[18]等,但高质量 SAR 样本集的匮乏,阻碍了其进一步应用^[19]。

西南丘陵山区是我国重要的冬油菜主产区,准确掌握其种植面积和空间分布,对优化区域粮油作物种植结构,加强油菜生产、管理有重要意义。但是西南丘陵山区是我国典型的山地农业区,油菜产季多云雾的气候特点及分散零碎的“巴掌田”油菜种植分布结构,给油菜精准遥感识别带来了极大挑战。在已有的诸多研究中,关于西南丘陵山区 SAR 油菜识别的研究相对较少,且效果也不甚理想。虽然基于时序 SAR 散射特征能够有效提高作物识别精度,但全生育期时序 Sentinel-1 散射特征下的冬油菜像元级分布识别的生产精度(Producer's Accuracy, PA)和用户精度(User's Accuracy, UA)也仅为 61.81%和 71.43%^[20]。以 Sentinel-2 影像分割单元为空间辅助约束后,目标作物油菜的 PA 和 UA 有所改善,能够分别提高到 70.93%和 75.86%^[21],但与 85%的作物识别可靠性标准^[22]仍有较大的差距,且其分布图斑的分辨率也无法满足精细尺度要求。因此,有必要以精细的地块单元为空间约束,探索一种适宜西南多云雾破碎种植区的冬油菜 SAR 遥感精准识别方法。同时考虑到研究区油菜生育期内的高分辨率光学数据难以保障,油菜产季地块信息获取困难。因此,本研究利用夏季高分二号(GF-2)影像分割对象和油菜全生育期双极化 Sentinel-1 后向散射、关键生育期极化分解和植被指数等特征,基于空间迁移对象构建了一种适用于西南山区冬油菜 SAR 遥感识别的 mRMR-PSO 特征优化方法,旨在为我国西南丘陵山区乃至其他多云雾丘陵山区或平原地区的冬季作物遥感提取提供有益参考。

1 研究区及作物情况

1.1 研究区概况

研究区为重庆市合川区东北部的香龙镇,区域总面积 59.05 km²,属于亚热带湿润季风气候区,降雨充沛,多云少日照。区域内地形复杂多样,地势起伏较大,作物种植结构复杂且地块细碎,是中国西南地区典型的山地农业区(图 1)。粮油作物以水稻、玉米和油菜为主,其中水稻和玉米的物候期主要集中在 4 月下旬至 8 月中旬,而油菜的物候期主要集中在 11 月中下旬至次年 4 月下旬,油菜作物主要物候期:移栽期为 11 月中旬至 11 月下旬,越冬期为 12 月上旬至 12 月下旬,现蕾期为 1 月上

旬至 1 月中旬，抽薹期为 1 月下旬至 2 月中旬，开化期为 2 月下旬至 3 月中旬，绿熟期为 3 月下旬至 4 月上旬，黄熟期为 4 月中旬至 4 月下旬。

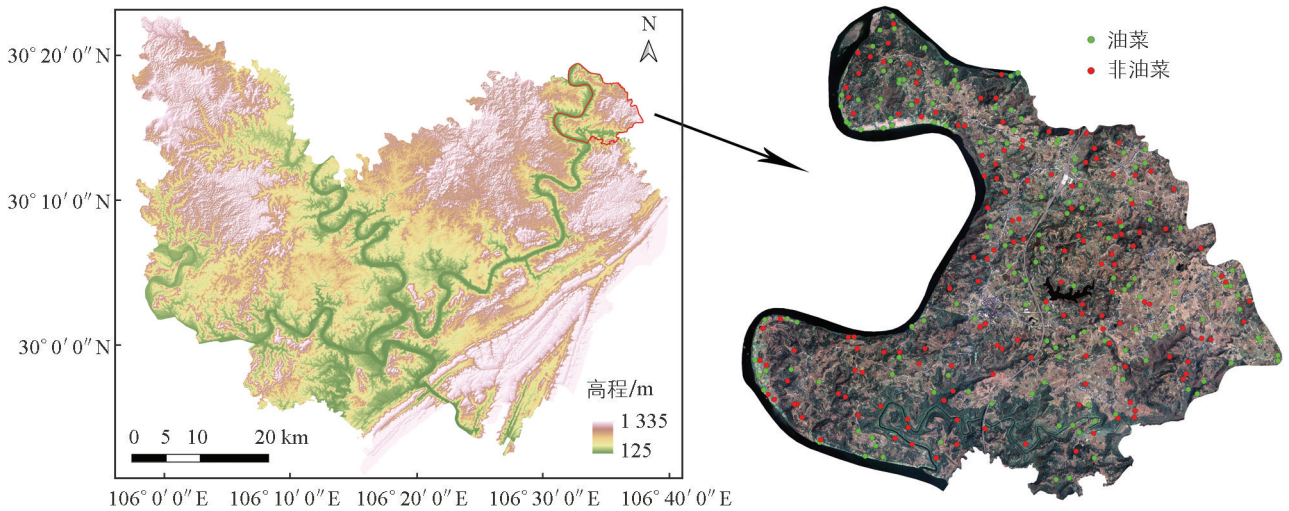


图 1 研究区位置及样本数据分布

1.2 SAR 数据及其预处理

采用的 SAR 影像为双极化 Sentinel-1A 雷达影像，主要选用干涉宽幅(IW, Interferometric Wide)模式的地距多视产品(GRD, Ground Range Detected)，像元分辨率为 10 m，极化方式为垂直发送接收 (VV)和垂直发送水平接收(VH)，成像时间范围为 2018 年 12 月—2019 年 5 月，基本涵盖了研究区油菜生长发育主要过程。如表 1 所示，其中 T_4 、 T_7 、 T_{10} 和 T_{11} 分别为油菜抽薹期、开花期、角果绿熟期及黄熟期等 4 个关键生育期，对应的双极化 Level-1 单视复数(SLC, Single Look Complex)产品用于提取油菜关键生育期的雷达植被指数与双极化分解特征。本研究主要利用 SNAP (Sentinel Application Platform)对 Sentinel-1 双极化 SAR 数据进行处理，其中 GRD 数据预处理过程包括精密轨道校正、辐射定标、滤波、地形校正以及分贝化数据变换等过程，SLC 数据预处理包括轨道校正、辐射定标、条带拼接、极化矩阵生成、多视处理、滤波、地形校正与地理编码等。

表 1 SAR 影像信息

传感器	模式	极化方式	产品类型	成像时间
Sentinel-1A	IW	VV、VH	GRD	2018-12-23 (T_1), 2019-01-04 (T_2), 2019-01-16 (T_3), 2019-01-28 (T_4), 2019-02-09 (T_5), 2019-02-21 (T_6), 2019-03-05 (T_7), 2019-03-17 (T_8), 2019-03-29 (T_9), 2019-04-10 (T_{10}), 2019-04-22 (T_{11}), 2019-05-04 (T_{12})
			SLC	2019-01-28(T_4), 2019-03-05 (T_7), 2019-04-10 (T_{10}), 2019-04-22 (T_{11})

1.3 其他数据源

辅助数据主要为 2018 年 7 月 31 日成像的空间分辨率 0.8 m 的 GF-2 PMS 正射影像数据、2019 年 3 月 11 日成像的空间分辨率约为 0.6 m 的 19 级 Google 数据和研究区第三次全国国土调查(简称“三调”)成果。其中“三调”图斑数据用于叠加 GF-2 正射影像获取研究区耕地影像，正值油菜花期的 Google 数据用于油菜样本数据的获取。为使上述不同传感器影像上相同地物的地理位置能够对齐，以 GF-2 正射影像为基准，利用 ENVI 自动配准工具进行地理配准，配准误差在 2 个像元以内。

2 研究方法

2.1 主要技术路线

技术路线如图 2 所示, 主要分为以下 4 部分: ① 对油菜全生育期的 Sentinel-1 GRD 数据和关键生育期的 Sentinel-1 SLC 数据进行处理, 提取研究区全生育期时序后向散射特征、关键生育期极化分解特征和雷达植被指数。② 利用“三调”耕地图斑裁剪研究区 GF-2 数据获取耕地影像, 随后进行面向对象分割, 并与上述 SAR 特征进行空间分析获取基于耕地对象单元的 SAR 特征集。③ 构建 5 种特征方案组合 A~E, 同时采用 mRMR-PSO 算法对全特征组合方案 E 进行特征优选, 构建特征方案 F。④ 开展不同算法优选特征集对比, 并基于优化特征方案 F 进行最优分类器筛选; 进一步通过不同特征组合油菜识别精度对比, 分析关键期植被指数和极化分解特征对油菜识别的影响, 以及优选特征对模型识别的贡献。本研究基于 Python 3.9.16 实现上述特征优选、模型构建、精度计算及特征贡献分析。

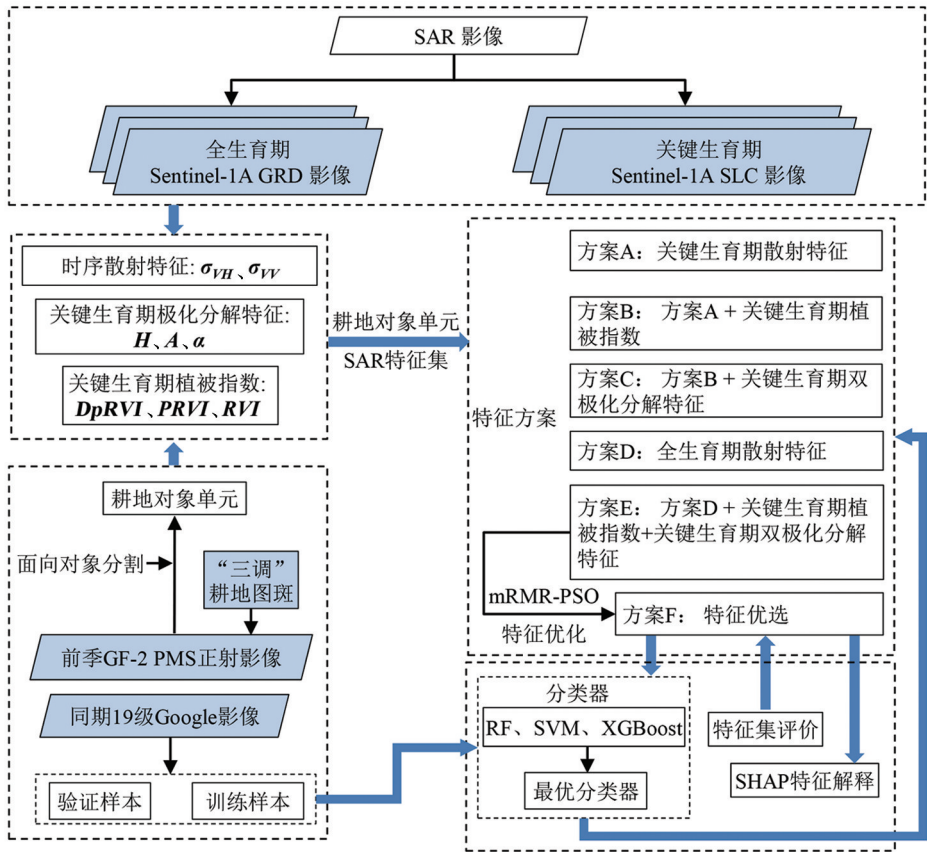


图 2 技术路线图

2.2 样本点选取

样本数据一般通过实地采集或影像解译两种方式获取。本研究样本点选取以 2019 年 3 月 11 日成像的 19 级 Google 影像为基础, 通过人工解译方式采集样本点 300 个, 其中油菜和非油菜样本点各 150 个, 并采用分层随机抽样方法选取 70% 用于模型训练, 剩余的 30% 作为测试集样本用于精度分析。

2.3 耕地对象单元 SAR 特征获取

2.3.1 双极化 SAR 特征提取

SAR 后向散射特征是描述地物目标对雷达波反射性能的重要指标, 在一定程度上反映地物目标的特性; 而极化目标分解则能够深层次地反映地物目标的散射机理, 表征地物散射机制的复杂性和散射的类型。两类特征已经成为作物遥感识别和长势监测的主要雷达指标^[23]。同时, 近年来一些学者也提出了

多种对作物长势较为敏感的雷达植被指数用于作物识别和监测。不同作物以及相同作物不同生育期所对应的 SAR 特征可能有所差异,这些特征差异正是实现作物有效区分的基础。本研究基于 Sentinel-1A 影像包含的 VV 与 VH 两种极化方式,提取其时间序列后向散射特征、关键生育期的双极化分解特征和雷达植被指数 3 类特征共计 48 个。其中,后向散射系数 σ_{VV} 和 σ_{VH} 在 GRD 产品数据预处理后可直接提取,极化分解特征采用适用于双极化 SAR 的 H/ α 分解法获取地物目标的散射熵 H、各向异性度 A 和平均散射角 α 等特征,而雷达植被指数则选用双极化 SAR 影像可用的 $DpRVI$ ^[24]、 $PRVI$ ^[25] 和 RVI ^[26] 3 种雷达植被指数。

2.3.2 耕地对象 SAR 特征集构建

利用 ENVI 5.3 软件对研究区耕地 GF-2 正射影像进行面向对象图像分割,采用基于边缘信息的多尺度分割算法,经多次调整与比对,在设定分割尺度为 50、合并尺度为 45 的组合下,各对象间的异质性及地物分割对象的纯度能够得以保证,分割效果较好。随后以影像分割单元为空间边界约束对 SAR 影像特征进行区域统计,以其特征均值作为对象单元特征值,获取研究区以分割对象为单元的耕地 SAR 特征集。

2.4 mRMR-PSO 特征优选

本研究中 SAR 特征优化主要包括 mRMR 特征过滤和粒子群优化特征选择两个过程。其中,基于 mRMR 的特征过滤能够选择出与样本类别相关度高且相互之间冗余度低的 SAR 特征组合,而基于 PSO 的特征选择方法能够有效筛选出预测精度较高的 SAR 特征子集。

2.4.1 mRMR 特征过滤

互信息是信息论中一种用于衡量两个变量间关联程度的指标^[27]。最小冗余最大关联 (Minimum Redundancy and Maximum Relevance, mRMR) 可利用互信息值衡量不同特征间的相关性和冗余度,并通过代价函数优选出与目标类别高度相关且冗余度较小的特征子集^[28]。mRMR 作为一种高效的过滤式特征选择方法,已广泛应用于机器学习图像分类领域。本研究在利用 mRMR 算法对 48 个 SAR 特征进行过滤时,通过构建不同数量的特征组合,选取测试集上最高预测准确率所对应的特征集作为候选特征组合,便于后续 PSO 算法进行特征优化,确定最佳特征组合。

2.4.2 PSO 特征优化

粒子群优化 (Particle Swarm Optimization, PSO) 算法是一种模拟鸟群搜索食物行为的经典群体智能优化算法,具有参数相对较少、全局搜索能力强和收敛速度快等特点,已广泛用于遥感图像分类领域^[29]。在 PSO 算法迭代优化中,粒子的每个位置代表特征子集组合的一组潜在解,其“优劣”程度可以通过适应度函数值来确定^[30]。本研究基于 PSO 算法对 mRMR SAR 特征过滤结果进行持续优化,其适应度函数采用训练样本分类准确率 $Accuracy$ 和 SAR 特征选中率 $Feature_{rate}$ 加权值。 $Fitness$ 值越大,表明粒子位置对应的特征子集越优。

$$Fitness = \alpha \times Accuracy + (1 - \alpha) \times (1 - Feature_{rate}) \quad (1)$$

$$Feature_{rate} = \frac{SF}{B} \quad (2)$$

式中: $Accuracy$ 为分类模型的样本预测准确率; α 为其权重因子; $Feature_{rate}$ 为 SAR 特征选中率; SF 表示选出的 SAR 特征个数; B 表示备选 SAR 特征总数。

2.5 其他特征组合方案

为评价上述优化特征组合的油菜识别性能,同时分析油菜关键生育期的双极化 SAR 特征参数对油菜分布识别的影响,本研究依据双极化 SAR 特征对应的生育期、特征类型等构建了 A、B、C、D 和 E 共 5 种特征组合,详见表 2。

表 2 特征组合表

组合	特征描述	特征集	特征数
A	关键生育期散射特征	$\sigma_{VH}^{T(i)}, \sigma_{VV}^{T(i)} (i = 4, 7, 10, 11)$	8
B	A+关键生育期植被指数	$\sigma_{VH}^{T(i)}, \sigma_{VV}^{T(i)}, DpRVI_{T(i)}, RVI_{T(i)}, PRVI_{T(i)} (i = 4, 7, 10, 11)$	20
C	B+关键生育期极化分解特征	$\sigma_{VH}^{T(i)}, \sigma_{VV}^{T(i)}, DpRVI_{T(i)}, RVI_{T(i)}, PRVI_{T(i)}, H_{T(i)}, A_{T(i)}, \alpha_{T(i)} (i = 4, 7, 10, 11)$	32
D	全生育期散射特征	$\sigma_{VH}^{T(j)}, \sigma_{VV}^{T(j)} (j = 1, 2, \cdots, 12)$	24
E	D+关键生育期植被指数+关键生育期极化分解特征	$\sigma_{VH}^{T(j)}, \sigma_{VV}^{T(j)}, DpRVI_{T(i)}, RVI_{T(i)}, PRVI_{T(i)}, H_{T(i)}, A_{T(i)}, \alpha_{T(i)} (i = 4, 7, 10, 11; j = 1, 2, \cdots, 12)$	48

2.6 分类器选择与特征优选集评价

2.6.1 分类器优选

RF 是一种由决策树构成的集成算法, 因其在应对样本异常值和噪声时具有良好的稳定性和泛化能力而被广泛应用于遥感作物识别。SVM 是一种基于统计学习理论和核函数的高效机器学习方法, 具有较好的鲁棒性和泛化能力, 能够在保证模型精度的同时, 有效提升模型构建效率, 在小样本机器学习和高维特征优化及分类场景中表现尤为突出; XGBoost 作为一种高效的梯度提升树算法, 通过在目标函数中加入正则化项能够有效控制模型的复杂度, 提高模型的预测准确性和泛化能力。上述机器分类模型凭借各自独特的优势, 已被广泛应用于 SAR 遥感图像分类中^[31]。本研究基于贝叶斯优化 (Bayesian Optimization, BO) 算法开展上述分类器参数优化, 在训练集样本上进行模型训练, 并在测试集样本上进行验证, 以 ROC 曲线下面积 (Area Under the Curve, AUC)、总体分类精度 (OA)、Kappa 系数、精确率、召回率和 F1 值为指标开展 SAR 优选特征组合油菜识别精度评价, 实现最优分类器获取, 并开展不同特征组合油菜识别效果的精度对比。

2.6.2 特征优选集评价

Jaccard 系数通常表示为两个特征集合间的交集数与并集数的比值, 可以成对评估特征的相似性和多样性^[32]。它的取值范围为 0~1, 其值越大, 表示两个集合的相似度越高。通常情况下, 大于 0.5 时可认为两个集合有较强的相似性。为分析优选 SAR 特征的可靠性, 本研究在分类精度评价的同时, 引入 Jaccard 特征相似度来比较 LASSO (Least Absolute Shrinkage and Selection Operator)、SVM-RFE (Support Vector Machine-Recursive Feature Elimination)、XGBoost 和 mRMR-PSO 4 种 SAR 特征集优选结果。其中, XGBoost 算法采用默认的 gain 方法获取特征的重要性得分, 依据特征重要性降序构建特征组合, 选择最高准确率对应的特征组合作为特征选择结果。

2.7 SHAP 方法

SHAP (SHapley Additive exPlanations) 方法为所有复杂的机器学习模型提供了一个统一的解释框架。该方法基于博弈论通过计算 Shapley 值来量化每个输入特征变量对单个样本预测的边际贡献。它不仅可以评估特征变量的全局重要性, 还可以量化它们的局部重要性, 从而为模型解释提供更多细节^[33]。本研究将基于最优分类器, 利用 SHAP 方法实现 SAR 优选特征对油菜识别模型特征重要性解释和对油菜样本的贡献解析。

3 结果与分析

3.1 SAR 特征优选结果

准确率是确定 mRMR 和 PSO 最佳优选特征集的重要参考指标, 基于不同分类器进行特征优选的准确

率和效率可能存在明显差异。为提高 SAR 特征优化效率,本研究采用小样本学习能力较强的 SVM 算法作为分类器进行模型训练、预测,计算不同特征组合下的分类准确率。SVM 分类器采用默认径向基核函数(Radial Basis Function, RBF),超参数主要为惩罚系数 C 和核系数 gamma。mRMR 特征过滤下不同特征数量组合下的测试集验证总体准确率 OA 变化如图 3a 所示,当选用前 45 个特征组合时,分类精度达到最高值 0.876。故本研究选取 45 个特征对应的特征组合作为候选特征进入 PSO 特征优化选择过程。在 PSO 特征优选中,设定粒子个数 N 为 30,惯性权重 ω 为 0.8,认知参数 c_1 和社会参数 c_2 均为 1.8,粒子最大飞行速度为 0.5,迭代次数为 50 次。设置核函数为 RBF、 $C=10$ 和 $gamma=1$ 的 SVM 分类器计算粒子群迭代搜索适应度值,其收敛曲线如图 3b 所示。可以看出,随着迭代次数的增加,粒子群搜索过程能够快速收敛,粒子群适应度值达到稳定值 0.833。基于此,本研究最终确定 26 个特征为最优特征组合,即 $\sigma_{VH}^{T(1)}$ 、 $\sigma_{VH}^{T(2)}$ 、 $\sigma_{VH}^{T(3)}$ 、 $\sigma_{VH}^{T(6)}$ 、 $\sigma_{VH}^{T(8)}$ 、 $\sigma_{VH}^{T(9)}$ 、 $\sigma_{VH}^{T(10)}$ 、 $\sigma_{VH}^{T(12)}$ 、 $\sigma_{VV}^{T(1)}$ 、 $\sigma_{VV}^{T(4)}$ 、 $\sigma_{VV}^{T(9)}$ 、 $\sigma_{VV}^{T(10)}$ 、 $\sigma_{VV}^{T(12)}$ 、 $A_{T(7)}$ 、 $H_{T(10)}$ 、 $A_{T(10)}$ 、 $\alpha_{T(10)}$ 、 $\alpha_{T(11)}$ 、 $RVI_{T(4)}$ 、 $PRVI_{T(4)}$ 、 $DpRVI_{T(4)}$ 、 $PRVI_{T(7)}$ 、 $DpRVI_{T(10)}$ 、 $RVI_{T(10)}$ 、 $PRVI_{T(10)}$ 和 $PRVI_{T(11)}$ 组成的特征集合。

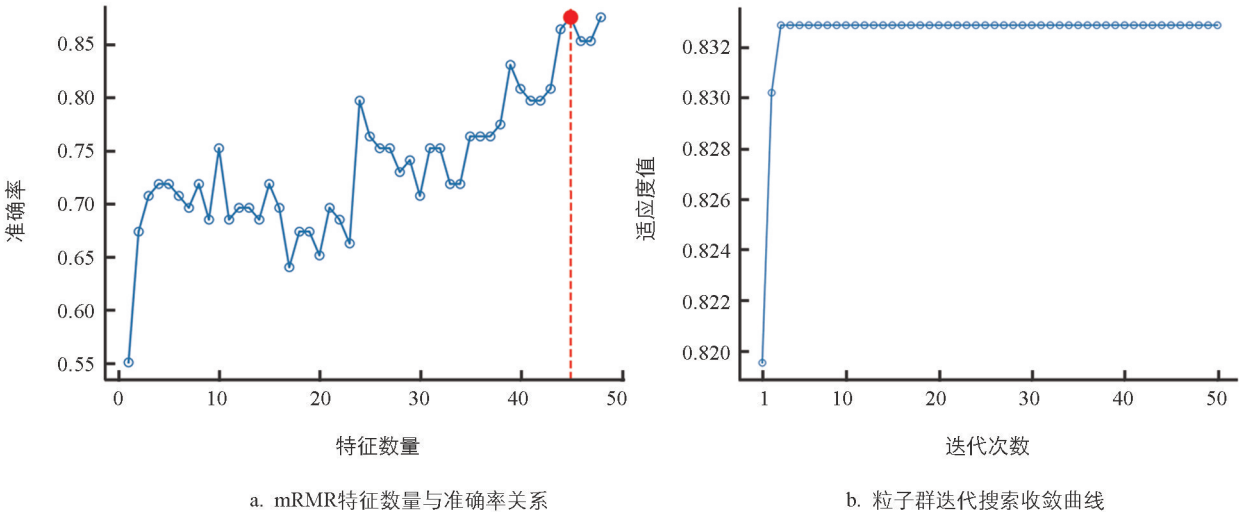


图 3 粒子群搜索收敛曲线

3.2 优选特征下的不同分类器性能对比

基于优选特征组合,利用 RF、SVM 和 XGBoost 3 种分类器在训练样本上开展油菜识别模型构建。其中,各模型参数优化过程采用 BO 优化结合样本 10 折交叉验证。图 4 展示了 3 种分类器下的识别精度指标,SVM 分类器的各项精度值均为最高,XGBoost 分类器次之,RF 分类器最低。整体识别效果上,SVM 分类器的总体分类精度 OA 为 88.76%,Kappa 系数为 77.57%,分别高于 RF 对应指标值 2.25%和 4.45%;目标油菜识别方面,SVM 分类器的识别精确率为 92.86%,召回率为 84.78%,二者调和值 F1 值达到 88.64%。对比 3 种分类模型的 ROC 曲线,SVM 模型的 AUC 最高(0.963),XGBoost 模型最低(0.871)。对比表明,在本研究特征优选组合下,SVM 油菜识别模型最优。其原因可能是前述特征优化中采用了 SVM 进行样本准确率预测,因此在 SVM 下能够展现出较高的精度。另外,RF、XGBoost 等集成学习模型高度依赖训练样本量,在小样本量下弱分类器多样性的不足导致集成后的精度难以提升,这也可能是其性能低于 SVM 的重要原因之一。

3.3 优选特征对模型识别的贡献分析

利用 SHAP 方法量化最优 SVM 模型中各特征变量在油菜识别中的重要性和 Shapley 贡献值,前 15 个重要特征的排序及蜂窝图如图 5 所示。从全局重要性看,有利于油菜识别的散射特征中,平均绝对 Shapley 值为 0.125 的绿熟期 $\sigma_{VH}^{T(10)}$ 最为重要,其次为 Shapley 值均在 0.087 左右的收获期 $\sigma_{VV}^{T(12)}$ 、绿熟期 $\sigma_{VV}^{T(10)}$ 和现蕾期的 $\sigma_{VH}^{T(2)}$ 3 个特征。现蕾期油菜开始进入营养与生殖生长两旺阶段,冠层水分积累增加,散射作用增强;

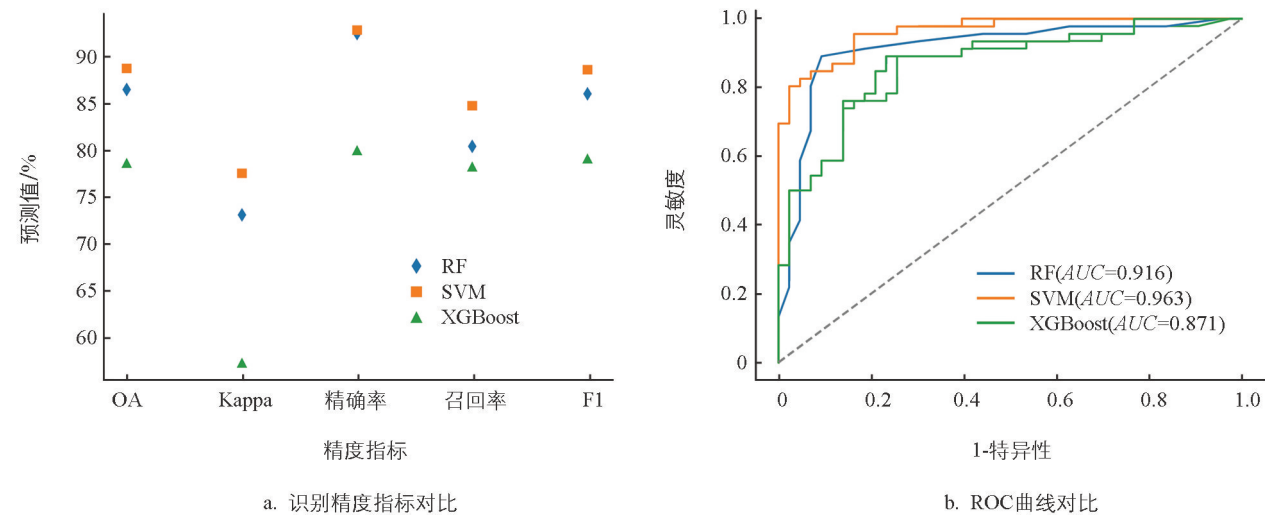


图 4 不同分类模型下的识别精度与 ROC 曲线

花期叶片退化,冠层水分积累开始降低,后期随着角果生长,水分逐步转移至角果;待角果逐渐成熟,冠层水分快速缺失。油菜冠层的这些变化与其他地物形成了较为明显的散射特征差异,并在油菜识别的特征重要性程度上有所体现。需要注意的是,油菜花期的 VV 后向散射特征会呈现出先随冠层水分降低而减小,后随角果绿熟期冠层水分增加而上升的变化。但对于油菜作物从花期到角果绿熟期冠层水分这一变化不太敏感的 VH 散射在特征重要性中处于第 9 位,而相对较为敏感的 VV 散射特征却并未出现在 SVM 模型的前 15 个重要特征中,其原因可能是花期油菜相对较低的 VV 散射特征与非油菜中样本中的部分其他作物太过接近以至于其可区分性较小。另外,在油菜 SAR 植被指数与极化分解特征中,角果绿熟期的平均散射角 $\alpha_{T(10)}$ 相对较为重要,其次为现蕾期的雷达植被指数 $RVI_{T(4)}$ 。从特征值与油菜样本的预测结果看,不同特征值大小对油菜识别结果的贡献不同。前 6 个特征中, $\sigma_{VH}^{T(10)}$ 、 $\sigma_{VV}^{T(10)}$ 和 $\sigma_{VH}^{T(9)}$ 3 个 SAR 变量的高特征值对油菜样本的预测有较好的正向作用。其值越高,样本类型越有利于油菜类型的判定,这与油菜角果绿熟期较高的冠层水分含量和后向散射特征值相一致。

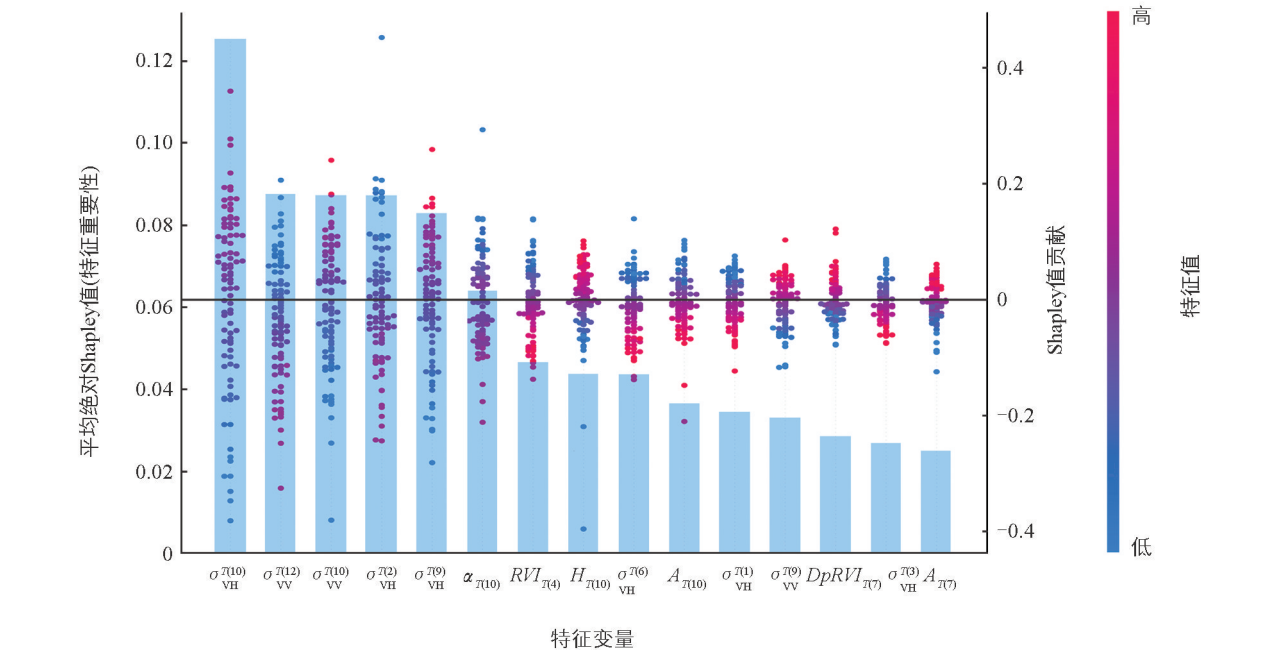


图 5 基于 SHAP 的优选特征重要性排序与 Shapley 贡献

3.4 不同特征选择算法优选结果对比

不同特征选择方法对应的特征优选结果可能不同，其油菜分布识别性能也有所差异。基于 LASSO、SVM-RFE、XGBoost 和 mRMR-PSO 4 种特征选择结果，在训练样本上开展 SVM 分类器格网搜索参数优化与模型训练，比对不同特征选择结果的油菜识别精度。如表 3 所示，LASSO 特征组合的测试集精度 OA 为 83.14%，Kappa 系数为 66.23%。SVM-RFE 的特征组合的测试集精度 OA 为 80.90%，Kappa 系数为 61.96%；XGBoost 的特征选择数量最多为 31 个，但其总体识别精度与 LASSO 方法的结果相当。mRMR-PSO 优选的特征子集支持下的 SVM 模型测试集精度 OA 和 Kappa 系数分别达到 88.76%和 77.57%，均明显高于 LASSO、SVM-RFE 和 XGBoost，说明 mRMR-PSO 算法能够得到较好的特征子集组合，并在参数优化的 SVM 分类器下取得较好的分类精度，其特征组合具有较好的可靠性。

上述不同特征选择方法优选特征结果的关系网络如图 6 所示。从特征集相似度来看，mRMR-PSO 方法与 LASSO 和 SVM-RFE 的 Jaccard 系数均为 0.23，与 XGBoost 的 Jaccard 系数为 0.51，说明 mRMR-PSO 的优选特征集与 XGBoost 的特征集有较高的相似度，即具有 XGBoost 中较多的重要特征。从特征构成来看，mRMR-PSO 所选特征中有 19 个特征与 XGBoost 特征重合，占其特征数量的 73.08%，说明其入选者大多数为 XGBoost 特征中的关键特征，在油菜识别中有着重要作用。

表 3 基于 SVM 的不同特征选择方法分类精度

方法	参数	特征数	OA/%	Kappa/%
LASSO	$\text{Alpha}=0.000\ 7, \text{max_iter}=10\ 000, C=200, \text{gamma}=1$	18	83.14	66.23
SVM-RFE	$\text{Step}=1, \text{CV}=10, C=10, \text{gamma}=0.10$	11	80.90	61.96
XGBoost	$\text{CV}=10, C=20, \text{gamma}=0.000\ 1$	31	83.15	66.43
mRMR-PSO	$\text{CV}=10, C=1, \text{gamma}=0.001$	26	88.76	77.57

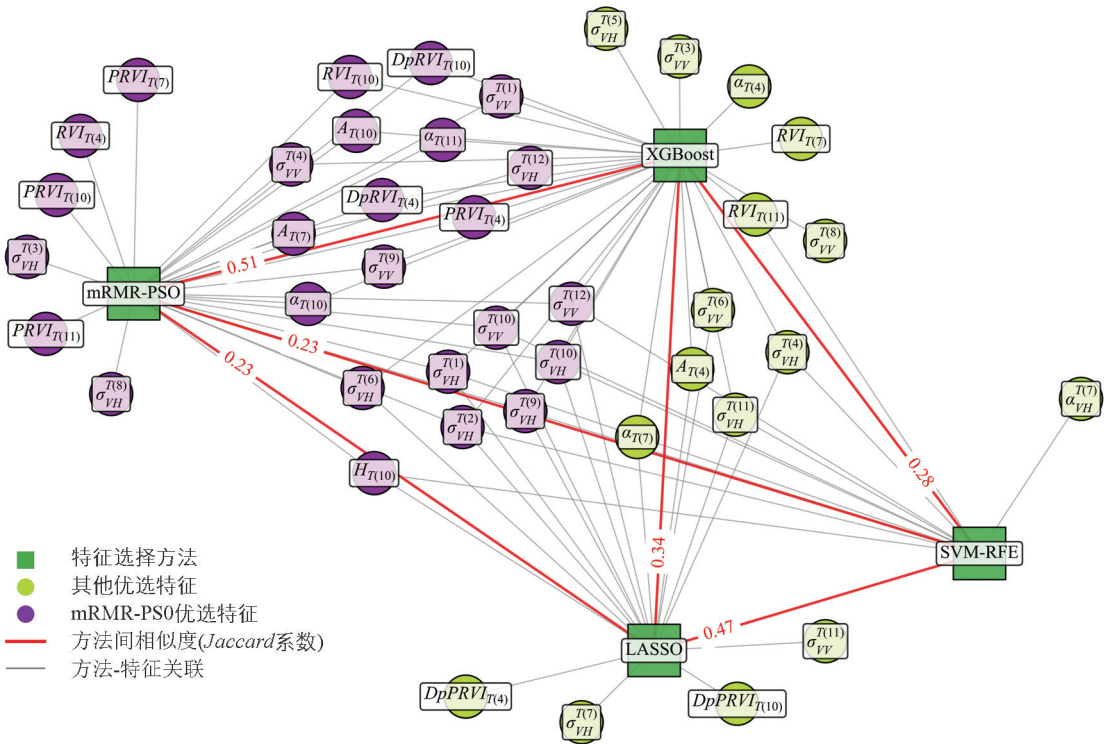


图 6 不同特征选择方法优选特征关系

3.5 不同特征组合下油菜识别精度对比

采用 BO 算法构建不同特征组合下的 SVM 油菜识别模型，测试集验证精度如表 4 所示。方案 A 中，

模型总体精度 OA 为 88.33%, Kappa 系数 66.40%, 精确率为 78.38%、召回率为 93.5%, F1 值为 85.29%。引入关键期的雷达植被指数(方案 B)和极化分解特征(方案 C)后, 识别模型的总体精度 OA 不变, Kappa 系数略微提高。相比于方案 A, 方案 C 中油菜的精确率提高 4.83%, 但召回率下降 8.09%, F1 下降 0.99%。方案 D 与方案 A 相比, OA 提升 0.23%, Kappa 系数提升 10.17%, F1 值提高 3.94%, 说明与关键生育期相比, 全生育期散射特征组合不仅有助于模型总体性能的改善, 还有助于提高油菜目标的识别精度。方案 E 相比于方案 D, 各指标呈现出与方案 A、C 一致的变化特点, 说明伴随着关键期植被指数 DpRVI、RVI、PRVI 和极化分解特征 H、A、 α 的加入, 模型识别结果与随机分类的差异有所增大, 对油菜的预测表现出了一定的保守性倾向, 油菜查准率提高, 预测可信度提升。方案 F 为基于本研究 mRMR-PSO 优选特征方法构建的 SAR 影像特征组合, 油菜识别模型的 OA 为 88.76%, Kappa 系数为 77.57%, 精确率为 92.86%、召回率为 85.78%, F1 值为 89.18%, 说明基于本研究 mRMR-PSO 的特征优选方法具有较好的有效性, 其优选特征组合在模型总体性能改善和油菜识别精度提升等方面都能取得较好的效果。

表 4 SVM 分类器下不同特征组合验证精度

特征组合	特征数	OA/%	Kappa/%	精确率/%	召回率/%	F1/%
A	8	88.33	66.40	78.38	93.54	85.29
B	20	83.33	66.48	80.00	90.32	84.85
C	32	83.33	66.52	83.21	85.45	84.30
D	24	88.56	76.57	85.29	93.55	89.23
E	48	88.56	76.66	90.00	87.10	88.52
F	26	88.76	77.57	92.86	85.78	89.18

图 7 为优选特征组合下的研究区油菜分布影像与识别结果对比图。可以看出, 研究区油菜种植总体上呈现零散、破碎的分布趋势, 与西南丘陵山区破碎地貌相适应。对比显示, 基于油菜种植前季的 GF-2 高分辨率光学影像分割对象, 研究区较为破碎的油菜地块基本被识别, 但也无法避免一些错分和漏分现象。其原因除了模型自身外, 还可能与研究区复杂破碎的地形以及 Sentinel-1A 影像相对较低的空间分辨率等有关。

4 讨论与结论

4.1 讨论

光学卫星遥感技术能够高效、准确地识别作物分布信息。然而, 我国西南丘陵山区冬春季节多云雾, 油菜生育期内光学遥感影像数据可用性较差, 给油菜种植地块信息的获取带来了极大困难, 也使得较为成熟的基于多时相光学遥感的农作物分布信息精准提取技术因数据获取问题而较难开展。与光学遥感技术相比, 尽管 SAR 遥感数据存在空间分辨率先天不足的问题, 但其仍以独特的优势成为多云雾地区冬季作物遥感识别的主要技术手段。本研究利用油菜种植前季的亚米级耕地光学影像分割对象代替油菜种植地块单元, 并基于油菜花期高分辨率 Google 影像人工解译样本数据进行了 mRMR-PSO SAR 特征优化和油菜识别, 探索了一种适合西南多云雾丘陵山区的油菜精准识别方法, 其总体分类精度 OA 达到 88.76%, 油菜的精确率和召回率分别达到 92.86%和 85.78%, 表明了本方法在地形破碎种植区作物精准遥感识别中的有效性。

本研究对比 mRMR-PSO 优选特征下的 3 种机器学习分类方法(RF、SVM、XGBoost)及 4 种特征优选算法(LASSO、SVM-RFE、XGBoost、mRMR-PSO)下的 SVM 分类模型识别性能可知, 相同特征集下的不同机器学习方法、不同特征优选算法的结果在应对相同的机器学习分类方法时都会有不同的结果。另外,

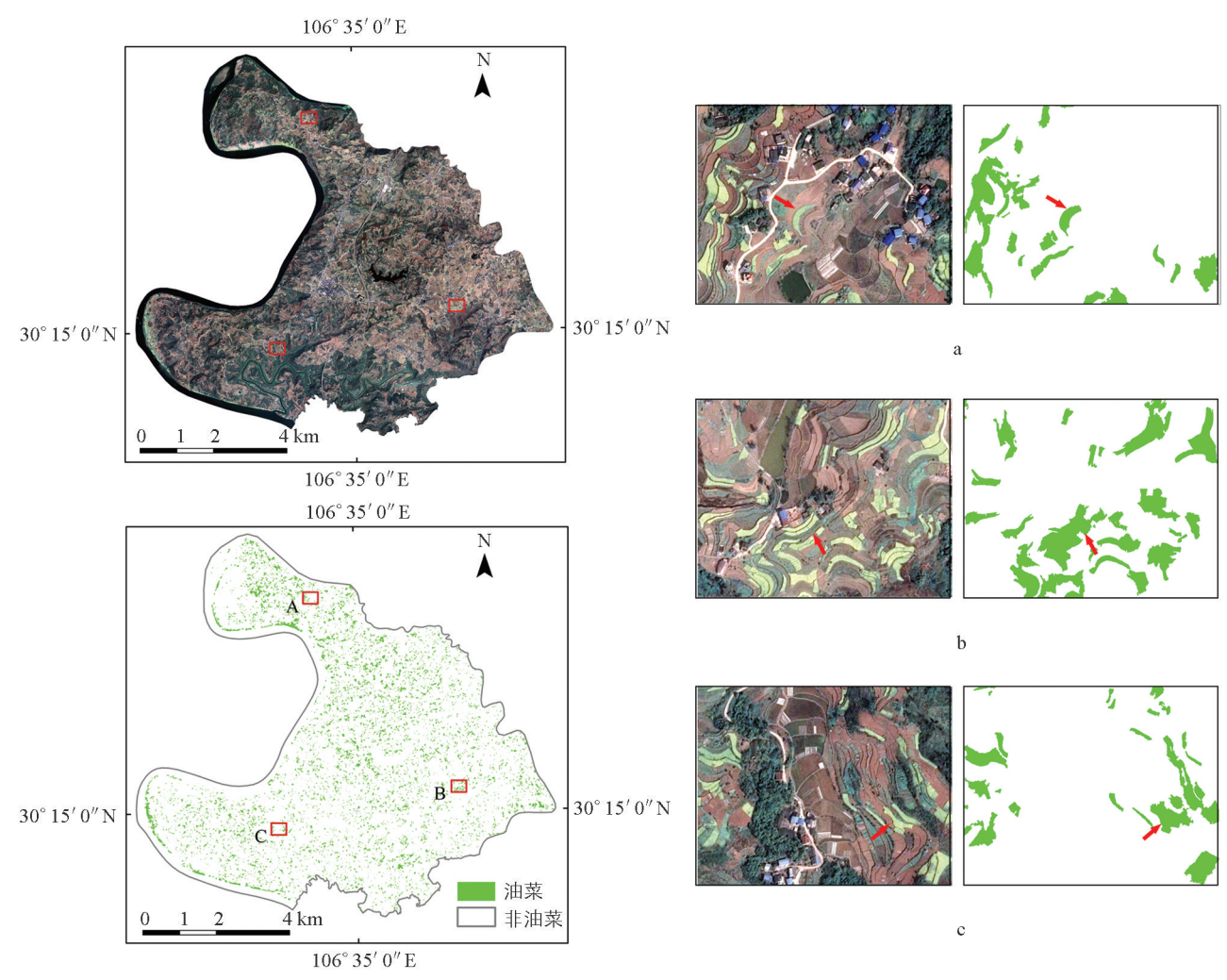


图 7 油菜分布影像与识别结果对比

不同特征组合下的 SVM 模型识别结果表明，不同特征支持下的相同机器学习分类方法的结果也会存在差异。基于特征优化的作物识别能够有效降低模型特征维度，同时提高模型的性能表现。因此，在基于特征优选进行农作物分类时，选择适合的特征优选算法和分类器进行研究具有重要意义。同时，本研究基于 SHAP 框架分析了 SVM 模型下各个优选特征对模型识别的贡献程度，获取了有助于油菜识别的双极化 SAR 特征重要性程度及其 Shapley 值与油菜的预测方向，可为今后利用双极化 Sentinel-1 SAR 数据进行油菜分布提取提供参考。

本研究提出的冬油菜识别方案具有较强的普适性和可推广性，可适用于其他多云雾的丘陵山区及平原地区开展冬季油菜作物的识别，主要体现在以下两个方面：① 数据获取方面，Sentinel-1A 数据可持续稳定获取，地块空间约束信息可基于作物种植前期高分辨率影像获取或者直接利用已有的地块图斑信息，样本数据可采用高分辨光学影像人工解译或者野外采样点替代；② 识别算法方面，本研究在联合 mRMR-PSO 算法进行高维 SAR 特征优化中采用了高效的 SVM 模型作为核心分类器进行准确率预测，能够同时兼顾优选特征集的质量和优选效率。另外，若有较高分辨率的可靠的光学遥感数据的融合，将更有助于冬油菜识别精度的提高。

4.2 结论

本研究引入空间单元迁移思想，以种植前期成像较好的夏季亚米级耕地光学影像分割单元为空间约束，基于油菜全生育期的双极化 Sentinel-1 数据后向散射、关键生育期极化分解及雷达植被指数等特征进行 mRMR-PSO SAR 特征优化及油菜提取，得到以下结论：① 以油菜种植前季的亚米级耕地光学影像分

割对象为基本识别单元, 可以实现西南丘陵山区的冬油菜的小尺度精准识别, 整体提高冬油菜的识别精度。② 相对于关键生育期, 全生育期散射特征组合有助于模型总体性能的改善和油菜目标识别精度的提高; 融入关键生育期植被指数和极化分解特征能够提升油菜查准率和预测可信度, 但会因召回率下降导致油菜识别综合精度的降低; 基于 mRMR-PSO 的特征优化算法能够有效进行油菜识别关键特征筛选, 展现出了较强的可靠性。③ 最佳油菜识别模型中, 绿熟期和收获期的散射特征最为重要, 其次是绿熟期的平均散射角 α 和现蕾期的雷达植被指数 RVI, 绿熟期的高散射特征值最有利于油菜作物的预测。

但是, 本研究也存在以下问题或不足之处有待进一步研究: ① 本研究中的分割对象可能因不同时期的光谱差异与地块实际边缘有所差异, 且 SAR 特征优化及油菜识别方法的稳健性有待在更多样本的支持下及更大区域的应用上进一步验证。② 本研究在联合 mRMR-PSO 算法进行 SAR 特征优化中采用了 SVM 模型进行准确率预测, 所产生的优化特征组合在 SVM 模型评价中达到最优性能。而在采用 RF 或 XGBoost 作为优化算法的核心分类器时, 其衍生的特征优化组合在 SVM 下是否还能保持最优性能, 或者说优化特征组合下的最优分类器与 mRMR-PSO 优化算法的核心分类器是否具有 consistency 等问题有待进一步验证。③ 虽然本方法在研究区取得了较好的效果, 但是由于其地块破碎以及 SAR 数据分辨率的限制, 某些地物种植地块宽度或面积小于单位像元, 使得“混合像元效应”显著, 对识别结果的准确性必然造成一定影响^[31, 33], 需要进一步结合更高分辨率的 SAR 数据进行解决。

参考文献:

[1] 孙政, 周清波, 杨鹏, 等. 基于星载极化 SAR 数据的农作物分类识别进展评述 [J]. 中国农业资源与区划, 2019, 40(11): 63-71.

[2] 乔树亭, 叶回春, 黄文江, 等. 基于 Sentinel-1/2 影像的水稻种植面积提取方法研究: 以三江平原为例 [J]. 遥感技术与应用, 2023, 38(1): 78-89.

[3] 解毅, 王佳楠, 刘钰. 基于 Sentinel-1/2 数据特征优选的冬小麦种植区识别方法研究 [J]. 农业机械学报, 2024, 55(2): 231-241.

[4] Wu L, Qi W W, Guo Z W, et al. Winter wheat planting area extraction using SAR change detection [J]. Remote Sensing Letters, 2021, 12(10): 951-960.

[5] 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.

[6] 周涛, 潘剑君, 韩涛, 等. 基于多时相合成孔径雷达与光学影像的冬小麦种植面积提取 [J]. 农业工程学报, 2017, 33(10): 215-221.

[7] 李恒凯, 王利娟, 肖松松. 基于多源数据的南方丘陵山地土地利用随机森林分类 [J]. 农业工程学报, 2021, 37(7): 244-251.

[8] 张超, 陈畅, 徐海清, 等. 基于 XGBoost 算法的多云多雾地区多源遥感作物识别 [J]. 农业机械学报, 2022, 53(4): 149-156.

[9] 吴瑞姣, 陈光剑. 地块尺度的南方丘陵地区水稻种植识别方法 [J]. 中国农机化学报, 2024, 45(12): 181-186.

[10] 查东平, 蔡海生, 张学玲, 等. 基于多时相 Sentinel-1 水稻种植范围提取 [J]. 自然资源遥感, 2022, 34(3): 184-195.

[11] 钱丽沙, 姜浩, 陈水森, 等. 基于时空滤波 Sentinel-1 时序数据的田块尺度岭南作物分布提取 [J]. 农业工程学报, 2022, 38(5): 158-166.

[12] 夏俊, 苏涛, 刘丽娜, 等. 基于多时相 Sentinel-1A 数据的水稻面积提取 [J]. 江苏农业学报, 2022, 38(3): 666-674.

[13] Li Y C, Zhao R K, Wang Y. Mapping ratoon rice fields based on SAR time series and phenology data in cloudy regions [J]. Remote Sensing, 2024, 16(15): 2703.

[14] 古丽努尔·依沙克, 买买提·沙吾提, 马春玥. 基于多时相双极化 SAR 数据的作物种植面积提取 [J]. 作物学报, 2020, 46(7): 1099-1111.

[15] 郭交, 白清源, 李恒辉. 基于时变散射特征与 CNN 的双极化 SAR 作物分类 [J]. 上海航天(中英文), 2022, 39(3): 77-83.

- [16] 赵红伟, 陈仲新, 姜浩, 等. 基于 Sentinel-1A 影像和一维 CNN 的中国南方生长季早期作物种类识别 [J]. 农业工程学报, 2020, 36(3): 169-177.
- [17] Guo Z W, Qi W W, Huang Y B, et al. Identification of crop type based on C-AENN using time series sentinel-1A SAR data [J]. Remote Sensing, 2022, 14(6): 1379.
- [18] 蔡玉林, 刘照磊, 孟祥磊, 等. 基于 HRNet 和自注意力机制的多源遥感影像水稻提取 [J]. 农业工程学报, 2024, 40(4): 186-193.
- [19] 毕海霞, 况祖正, 李凡, 等. 极化 SAR 图像分类深度学习算法综述 [J]. 科学通报, 2024, 69(35): 5108-5128.
- [20] 王克晓, 周蕊, 欧毅, 等. 基于光学与雷达遥感的山地油菜分布提取研究 [J]. 西南大学学报(自然科学版), 2021, 43(6): 139-146.
- [21] Wang K X, Yu B, Huang X, et al. Identification of winter canola based on dual-polarized SAR datasets in hilly mountainous areas of southwest China [J]. Remote Sensing Letters, 2023, 14(12): 1283-1293.
- [22] Skakun S, Kussul N, Shelestov A Y, et al. Efficiency assessment of multitemporal C-band radarsat-2 intensity and landsat-8 surface reflectance satellite imagery for crop classification in Ukraine [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(8): 3712-3719.
- [23] 洪玉娇, 张硕, 李俐. 基于合成孔径雷达数据的农作物长势监测研究进展 [J]. 智慧农业(中英文), 2024, 6(1): 46-62.
- [24] Mandal D, Kumar V, Ratha D, et al. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data [J]. Remote Sensing of Environment, 2020, 247: 111954.
- [25] Chang J G, Shoshany M, Oh Y. Polarimetric radar vegetation index for biomass estimation in desert fringe ecosystems [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(12): 7102-7108.
- [26] Trudel M, Charbonneau F, Leconte R. Using RADARSAT-2 polarimetric and ENVISAT-ASAR dual-polarization data for estimating soil moisture over agricultural fields [J]. Canadian Journal of Remote Sensing, 2012, 38(4): 514-527.
- [27] 张婧, 曹峰, 董毓莹, 等. 基于互信息和遗传算法的特征选择算法 [J]. 山西大学学报(自然科学版), 2024, 47(1): 1-8.
- [28] 黄磊, 向泽君, 楚恒. 结合 mRMR 选择和 IFCM 聚类的遥感影像分类算法 [J]. 测绘通报, 2019(4): 32-37.
- [29] 余莎莎, 余洁, 朱腾, 等. 一种改进的 MCSM/H-PSO 全极化 SAR 影像分类方法 [J]. 测绘通报, 2018(11): 53-57.
- [30] 王克晓, 李波. 基于粒子群优化和互信息的特征优化与油菜倒伏识别 [J]. 测绘通报, 2025(4): 63-67, 74.
- [31] 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): 153-163.
- [32] 简开宇, 史涯晴, 黄松, 等. 业务流程模型相似度研究综述 [J]. 计算机科学, 2023, 50(6): 338-350.
- [33] 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.

责任编辑 王新娟