引文格式：刘腾,庞新华,朱秀芳,等. 基于SHAP可解释特征优选的乌克兰作物制图研究[J]. 地理空间信息, 2026, 24(3): 92-98. doi:10. 3969/j. issn. 1672-4623.2026.03.019

基于SHAP可解释特征优选的乌克兰作物制图研究

刘　腾[1]，庞新华[2*]，朱秀芳[3]，曹建荣[1]，姬忠林[1]
（1. 聊城大学 地理与环境学院，山东 聊城 252000；2. 中航勘察设计研究院有限公司，北京 100098；
3. 北京师范大学 遥感和数字地球全国重点实验室，北京 100875）

摘　要：精确且高效地绘制冬小麦、玉米、大豆等作物的分布图，对于农业资源管理、作物种植规划和农业政策制定等都具有重要作用。聚焦于乌克兰农业区域，基于Sentinel-2遥感影像，结合Spearman相关系数和SHAP特征优选方法，优化了冬小麦、玉米和大豆等主要作物的制图流程；并评估了7种特征组合策略与3种机器学习模型在区分主要作物类型上的效能。结果表明：①Spearman相关系数与SHAP相结合的特征选择方法效果最佳；②基于Spearman相关系数与SHAP相结合得到的最优特征，LightGBM算法的分类精度最高，冬小麦制图的总体精度和Kappa系数为95.42%和90.83%，玉米、大豆制图的总体精度和Kappa系数为96.19%和94.28%，分类结果与统计年鉴数据高度一致，冬小麦、大豆和玉米的面积提取精度分别达到97.31%、97.55%和99.61%。

关键词：Sentinel-2；机器学习；SHAP；特征优选；作物制图
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2026）03-0092-07

Research on Ukrainian Crop Mapping Based on SHAP Explainable Feature Selection

LIU Teng[1], PANG Xinhua[2], ZHU Xiufang[3], CAO Jianrong[1], JI Zhonglin[1]
(1. School of Geography and Environment, Liaocheng University, Liaocheng 252000, China;
2. AVIC Geotechnical Engineering Institute Co., Ltd., Beijing 100098, China;
3. State Key Laboratory of Remote Sensing and Digital Earth, Beijing Normal University, Beijing 100875, China)

Abstract: The precise and efficient delineation of distribution of winter wheat, corn, and soybeans is of vital importance for agricultural resource management, crop planting planning, and agricultural policy formulation. Focused on the agricultural regions of Ukraine, based on Sentinel-2 remote sensing images, we combined the feature selection method of Spearman correlation coefficient with SHAP, to optimize the mapping process of major crops such as winter wheat, corn, and soybeans. We explored the effectiveness of seven combination strategies with three machine learning models in distinguishing major crop types. The results showed that ①the feature selection method combining Spearman correlation coefficient with SHAP demonstrated the best performance. ②Based on the optimal features derived from Spearman correlation coefficient and SHAP, the LightGBM algorithm achieved the highest classification accuracy with overall accuracy and Kappa coefficient of 95.42% and 90.83% for winter wheat mapping, and overall accuracy and Kappa coefficient of 96.19% and 94.28% for corn, soybeans mapping, respectively. The classification results demonstrated high consistency with statistical yearbook data, achieving area extraction accuracy of 97.31% for winter wheat, 97.55% for soybean, and 99.61% for corn.

Key words: Sentinel-2, machine learning, SHAP, feature selection, crop mapping

粮食安全是全球经济社会稳定与持续发展的基石，对维护国家安全和促进社会稳定具有至关重要的作用[1]。冬小麦、玉米和大豆等主要粮食作物的种植不仅直接关系到粮食供应的稳定性和自给自足能力，而且是农业资源高效利用和可持续发展的重要组成部分。因此，精确且高效地绘制作物分布图，对于指导农业生产、优化资源配置和制定科学合理的农业政策具有深远意义[2-4]。随着全球气候变化、城市化进程加速和农业结构调整，作物种植环境日益复杂多变，给作物制图带来了前所未有的挑战[5]。

遥感技术因其覆盖范围广、信息量大和更新频率高等优势，在作物制图领域展现出巨大的应用潜力。因此，探索新的作物制图技术和方法，提高作物分类的准确性和效率，已成为当前农业遥感研究的热点[6]。随着遥感数据源的不断丰富和机器学习算法的快速发展，基于多源遥感数据和先进机器学习算法的作物制

项目来源：国家重点研发计划资助项目（2023YFB3906201）。
第一作者简介：刘腾（2000—），硕士研究生，研究方向为农业遥感，E-mail：2310140210@stu.lcu.edu.cn。
通信作者：庞新华（1982—），硕士，研究方向为遥感与地理信息应用，E-mail：xhpang05@126.com。

图方法逐渐成为研究主流[7]，能综合运用光谱、纹理、地形、时序变化等多种特征，提高作物分类的精度和鲁棒性[8]。然而，机器学习模型的黑箱特性限制了其在实际应用中的可解释性，使得模型的决策过程难以被理解和信任[9]。因此，本文提出了一种结合SHAP和Spearman相关系数的特征选择方法，旨在识别出对作物分类影响最大的特征组合，从而优化机器学习模型的输入，提高作物制图的精度和效率。SHAP是一种先进的特征重要性评估工具，能量化模型中各变量对预测结果的具体贡献，进而揭示模型的决策过程[10]。

乌克兰作为世界重要粮食生产国之一，其作物种植结构和产量对全球粮食市场具有重要影响[11]。因此，本文以乌克兰为研究区，基于Sentinel-2遥感影像构建光谱波段、光谱指数、纹理、邻域和地形等多种特征，设计多种特征组合方案，并结合先进的机器学习算法进行作物分类。通过深入分析不同特征组合和算法对作物分类的影响，探索最佳特征组合和算法配置，为乌克兰及类似地区的作物制图提供一种新的思路，为农业资源管理和政策制定提供科学依据。

1　研究区概况与数据来源

1.1　研究区概况

乌克兰地理位置优越，东接俄罗斯、南濒黑海、北与白俄罗斯相邻、西与多个国家接壤；地形地貌多样，从北部的平原延伸至南部的黑海沿岸低地，再到东部的丘陵和山地，为农业生产提供了丰富的土地资源；属温带大陆性气候，夏季温暖湿润、冬季寒冷干燥，适宜多种农作物生长。乌克兰的黑土区以其肥沃的土壤和丰富的有机质而闻名，为农业生产提供了得天独厚的自然条件。乌克兰主要种植小麦、大麦、油菜、玉米和大豆等作物[12]。城市化进程的加速、劳动力迁移的加剧以及地形复杂等因素对乌克兰的农业生产提出了诸多挑战。

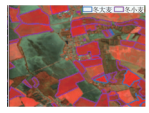
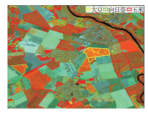
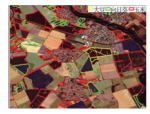
1.2　数据来源与预处理

本文采用的数据见表1，其中ESA World Cover 10 m v2土地覆盖图为ESA基于Sentinel-1和Sentinel-2数据生产的10 m分辨率全球土地覆盖产品，包含11个土地覆盖类别，并在EOEP-5的框架下生成；DEM数据采用欧盟哥白尼计划发布的Copernicus DEM GLO-30，分辨率为30 m，基于TanDEM-X卫星任务（2010—2015年）的合成孔径雷达数据生成，高程精度可达±2 m，覆盖乌克兰全境，提供地形高程、坡度、坡向等关键地理信息；EuroCrops是欧盟各个国家和地区公开可用的作物报告数据集。结合夏粮（冬小麦）和秋粮（玉米、大豆）的生长周期以及目视解译的难度，本文最终选取2024年4月15日—5月15日和2024年7月10日—8月10日两个时间段开展研究。在Google Earth Engine平台中对原始影像进行裁剪、镶嵌和去云等预处理，再利用ESA数据提取合成影像中的耕地，最终得到乌克兰10 m分辨率的夏粮和秋粮合成影像。通过参考EuroCrops数据集相同时间段的作物在Sentinel-2影像不同波段组合下的颜色（表2），目视解译2024年乌克兰地区夏粮和秋粮相应时间段的Sentinel-2影像。

表1　数据来源

数据名称	时间	来源
Sentinel-2遥感影像	2024年	https://code.earthengine.google.com/
DEM	2024年	https://code.earthengine.google.com/
乌克兰行政矢量地图	2024年	https://gadm.org/data.html/
乌克兰统计年鉴数据	2024年	https://www.ukrstat.gov.ua/
ESA World Cover 10 m v2	2021年	https://code.earthengine.google.com/
EuroCrops数据	2024年	https://zenodo.org/records/14094196

表2　目视解译

目视要点	Sentinel-2图像	
冬大麦和冬小麦在B4、B3、B2波段组合中均呈深绿色，在B8、B5、B4波段组合中则呈亮红色和酒红色		
	B4、B3、B2	B8、B5、B4
玉米、大豆和向日葵在B8、B11、B4波段组合中呈棕色、橘黄色和亮黄色，在B12、B4、B3波段组合中则呈深棕色、深棕色和深青色		
	B8、B11、B4	B12、B4、B3

2　研究方法

本文首先基于预处理后的遥感合成影像和DEM数据，提取光谱波段、光谱指数、纹理、邻域和地形5类41个特征；再通过Spearman相关系数初步筛选特征，剔除纹理、邻域和地形特征中相关性较大的特征；然后利用SHAP对特征的重要性进行排序，并设计了7种特征组合方案，采用3种分类算法进行地物分类；最后对结果进行精度评价，以确定最佳分类方案和最佳分类器。

2.1　样本构建

结合冬小麦、玉米和大豆影像颜色以及研究区实际情况，将夏粮地物类型分为冬麦和其他耕地2类，将秋粮地物类型分为玉米、大豆和其他耕地3类。基于夏粮和秋粮合成影像分别进行目视解译，最终选取

35 046个样本点（冬小麦7 094个、夏粮其他耕地7 107个、玉米7 158个、大豆6 787个、秋粮其他耕地6 900个），其他耕地包括合成影像中的其他植物和裸地。随机选取70%的样本点作为训练集，剩余30%作为测试集。

2.2 特征优选

Sentinel-2遥感影像波段众多，信息丰富，选择不同的分类特征是提高模型分类精度的关键[13]。本文选取的5类41个特征包括：①11个光谱波段，即海洋气溶胶（B1）、蓝光（B2）、绿光（B3）、红光（B4）、红边1（B5）、红边2（B6）、红边3（B7）、近红外（B8）、水汽（B9）、短波红外1（B11）和短波红外2（B12）波段；②8个纹理特征为由第一主成分经计算灰度共生矩阵得到的均值（Mea）、方差（Var）、同质性（Hom）、对比度（Con）、差异性（Dis）、熵（Ent）、二阶矩（Sec）和相关性（Cor）；③7个领域特征，即中值滤波（Med）、高斯低通滤波（Gal）、高斯高通滤波（Gah）、罗伯特梯度（Rob）、索伯尔梯度（Sob）、拉普拉斯算子（Lap）、方向滤波（Dir）；④7个地形特征，即高程（DEM）、坡度（Slo）、坡向（Asp）、平面曲率（Pla）、剖面曲率（Pro）、地形起伏度（Rel）和地表粗糙度（Rou）；⑤8个光谱指数，即归一化水体指数（NDWI）、归一化植被指数（NDVI）、比值植被指数（RVI）、差值植被指数（DVI）、土壤调节植被指数（SAVI）、增强型植被指数（EVI）、绿光归一化差值植被指数（GNDVI）、归一化差异绿度指数（NDGI）。纹理特征处理窗口大小为3×3，步长为1；邻域特征的变换核大小为3×3；Slo、Asp、Pla、Pro、Rel、Rou基于DEM数据计算得到。各植被指数的计算公式见表3。

表3 地表粗糙度和植被指数的计算公式

特征	计算公式
NDWI	(B3−B8)/(B3+B8)[14]
NDVI	(B8−B4)/(B8+B4)[15]
RVI	B8/B4[16]
DVI	B8−B4[17]
SAVI	(B8−B4) ×(1+0.5)/(B8+B4+0.5)[18]
EVI	2.5×(B8−B4)/(B8+6×B4−7.5×B2+1)[19]
GNDVI	(B8−B3)/(B8+B3)[20]
NDGI	(B3−B8)/(B3+B8)[21]

2.3 基于Spearman相关系数和SHAP的特征优选

处理影像数据时，有效整合光谱与空间信息至关重要。特征间的高度相关性可能会引发模型过拟合、稳定性下降和泛化能力受限等问题[22]。因此，本文提出了一种结合相关系数与SHAP的特征优选策略。Spearman相关系数更适用于不满足正态性要求的数据，而Pearson相关系数则适用于正态分布的数据。通过IBM SPSS Statistics 27软件分析夏粮和秋粮的样本数据发现，Mea、Var和Con等特征不符合正态分布条件，因此采用Spearman相关系数进行相关性分析。本文首先采用Spearman相关系数剔除同类特征中高度相关的项，以减少信息冗余[23]；再利用SHAP进一步去除对模型贡献较小的特征，从而优化特征集合，增强模型的稳定性和泛化性[24]。

1）Spearman相关系数是非参数统计量，可衡量两个变量间可能是非线性的相关性[25]，计算公式为：

$$d_i = X_i - Y_i \tag{1}$$

$$R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{2}$$

式中，X_i和Y_i为第i组数据两个变量的观测值，d_i为X_i和Y_i在各自数据集中排名之差的绝对值，反映两个变量等级的不一致程度；n为样本总数；$n(n^2-1)$为标准化因子，确保R落在−1~1，若R接近1，则两个变量存在强烈的正相关关系，若接近−1，则存在强烈的负相关关系，若接近0，则没有明显的单调相关性。

本文将相关系数强度划分为5个区间，即0.8~1.0为极强相关，0.6~0.8为强相关，0.4~0.6为中等相关，0.2~0.4为弱相关，0~0.2为极弱或无相关[26]。为优化模型性能并降低冗余信息，本文设定相关性阈值为0.4，若同类特征之间的相关系数超过阈值，则剔除方差较小的特征。

2）SHAP可解释框架。SHAP是一种源自博弈论合作收益分配理念的框架，专门用于解析机器学习模型的输出结果，提供从局部到全局的深入理解。通过评估每个特征在不同特征组合下对预测结果的边际贡献，量化这些特征的平均影响力。特征对预测的贡献度越大，其重要性排序越靠前。SHAP在处理高维特征时尤为有效，能在维持模型预测性能的同时，减少模型的复杂度并降低过拟合的风险，从而增强模型的泛化能力和计算效率[27]。其计算公式为：

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!}[f(S \cup \{i\}) - f(S)] \tag{3}$$

式中，f为以一个特征子集为输入，返回该子集对应的模型预测值的函数；n为特征总数；S为特征集合中除X_i之外的所有可能子集。

2.4　分类方案设计和精度评价

根据特征优选时包含的波段类别，本文设计了 7 种特征组合方案（表 4）。需要说明的是，本文旨在区分耕地内部的作物，因此在方案六中保留了 11 个光谱波段和 8 个光谱指数作为分类的关键特征，再根据方差与相关系数对其余 22 个特征进行优选。方案七在方案六的基础上进一步结合 SHAP 进行特征筛选。

表 4　特征组合方案

组合方案	组合方式	特征数量
方案一	纹理	8
方案二	纹理+邻域	15
方案三	纹理+邻域+地形	22
方案四	纹理+邻域+地形+光谱指数	30
方案五	纹理+邻域+地形+光谱指数+光谱	41
方案六	初选后的特征（Spearman）	根据方差与相关系数确定
方案七	Spearman 与 SHAP 相结合优化后的特征	根据模型确定

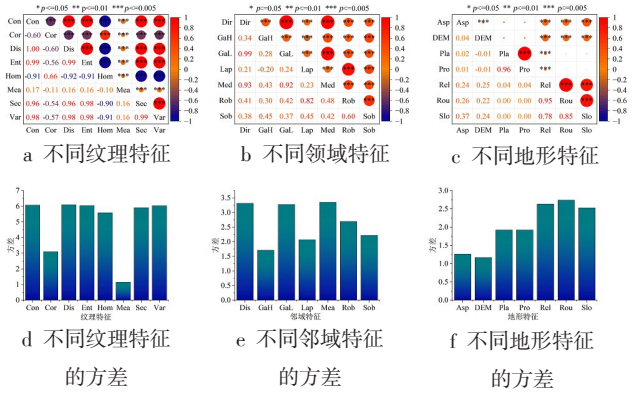
在土地利用分类研究中，随机森林（RF）、轻量级梯度提升机（LightGBM）以及类别特征梯度提升算法（CatBoost）的机器学习模型在二分类和多分类问题上表现优异。因此，本文采用上述分类算法对 2024 年乌克兰的夏粮和秋粮的合成影像进行分类，并采用生产者精度（*PA*）、用户精度（*UA*）、总体精度（*OA*）以及 Kappa 系数[28]等指标评估模型的分类精度。

3　研究结果与分析

3.1　夏粮制图结果

3.1.1　夏粮分类特征优选结果

首先对夏粮的纹理、邻域和地形特征进行 Spearman 相关性分析，筛选出相关系数低于 0.4 的 3 个指标，即 DEM、Mea、Asp；再从相关性较强的指标中提取出方差最大的 5 个特征，即 Dis、Dir、Rob、Pla 和 Rou；最后将这些特征与光谱波段、光谱指数进行组合，形成经过优选的 27 个夏粮分类特征（图 1）。



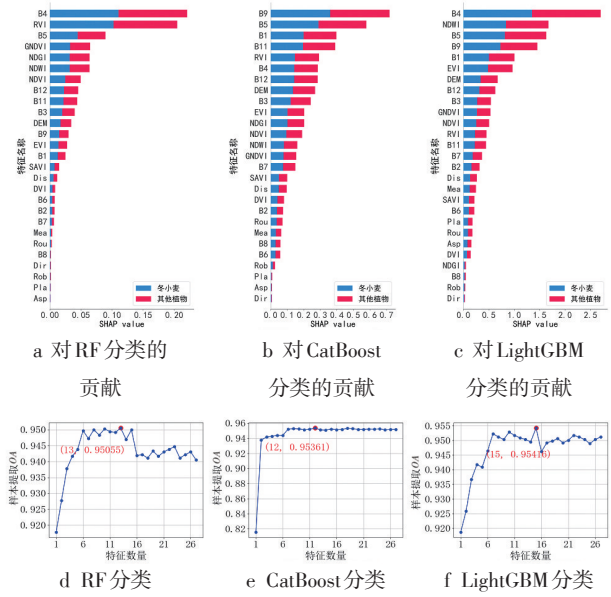
a 不同纹理特征　b 不同领域特征　c 不同地形特征

d 不同纹理特征的方差　e 不同邻域特征的方差　f 不同地形特征的方差

图 1　夏粮不同特征之间的相关系数以及不同特征的方差

为进一步提升特征集的质量和模型性能，本文采用 SHAP 值对 RF、LightGBM 和 CatBoost 三个模型的特征重要性进行深入分析，结果见图 2，色带越长，贡献越大；反之，贡献越小。例如，在夏粮的 RF 模型中，B4 的色带最长，表明其影响力最大；而 B11 及其后续特征的 SHAP 值均低于 0.05，说明它们对模型的贡献较小。此外，Asp 和 Pla 的贡献度近乎为零，这可能是因为不同作物在同一耕地上的地形特征差异并不显著。值得注意的是，同一特征在不同模型中的重要性可能有所不同，这或许与模型处理特征的具体方式和赋予的权重有关。总体而言，同一季节粮食作物的特征在不同模型中的重要性排序具有较高的一致性，说明这些特征在同一季节内具有稳定的影响力和预测效能。

基于初选特征在不同模型中的重要性排序，本文逐步增加特征数量，并利用受试者工作特征曲线下的面积（AUC）来评估样本提取的精度，结果见图 2，可以看出，随着特征数量的增加，样本提取的精度逐渐攀升至一个峰值，之后趋于平稳或略有下降，RF、LightGBM 和 CatBoost 模型在特征数量为 13、15 和 12 时精度最高。



a 对 RF 分类的贡献　b 对 CatBoost 分类的贡献　c 对 LightGBM 分类的贡献

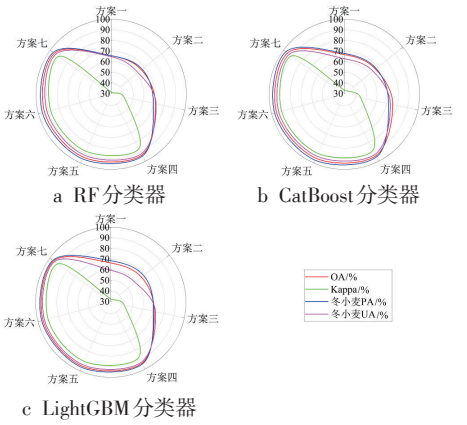
d RF 分类　e CatBoost 分类　f LightGBM 分类

图 2　各特征在不同分类器夏粮分类中的 SHAP 值和 *OA* 随特征数量的变化

3.1.2　夏粮分类方案对比结果

由图 3 可知，方案一的平均 *OA* 和平均 Kappa 系数最低，分别为 66.2%和 32.37%；随着邻域、地形、光谱指数和光谱波段的逐步加入，方案二~五的分类效果逐步提升，方案五的平均 *OA* 达到了 94.72%，平均 Kappa 系数增至 89.44%；经过特征初选后，方案六在减少特征数的基础上分类效果与方案五基本保持一

致；方案七的表现最佳，平均 *OA* 和 Kappa 系数分别为95.28%和90.55%，相较于方案一，平均 *OA* 提高了29.08%，平均Kappa系数增长了58.18%。从各地物类型的 *PA* 和 *UA* 来看，随着特征数量的增加，大多数算法的平均 *PA* 和 *UA* 呈上升趋势，方案七的精度最高，平均 *PA* 为93.99%， *UA* 为96.82%。在方案七中，LightGBM 算法的 *OA* 和 Kappa 系数最高，分别为95.42%和90.83%，与RF和CatBoost算法相比，*OA* 分别提高了0.36%和0.05%，Kappa 系数分别提高了0.72%和0.11%；LightGBM算法的平均 *PA* 和 *UA* 分别为95.44%和95.41%。综上所述，方案七LightGBM算法的各类精度几乎均为最高，因此该算法为区分冬小麦的最佳算法。



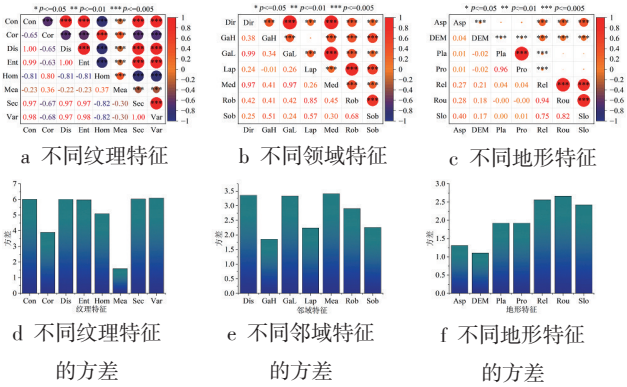
a RF分类器　b CatBoost分类器　c LightGBM分类器

图3　不同特征筛选方案与不同分类器组合提取的夏粮精度

3.2　秋粮制图结果

3.2.1　秋粮分类特征优选结果

首先对秋粮的纹理、邻域和地形特征进行Spearman相关性分析，筛选出相关系数低于0.4的3个指标，即DEM、Mea和Asp；再从相关性较强的指标中提取方差最大的5个特征，即Var、Dir、Rob、Pro和Rou；最后将这些特征与光谱波段和光谱指数进行组合，形成经过优选的27个秋粮分类特征（图4）。由图5可知，RF、LightGBM和CatBoost模型在特征数量分别为21、17和24时精度最高。



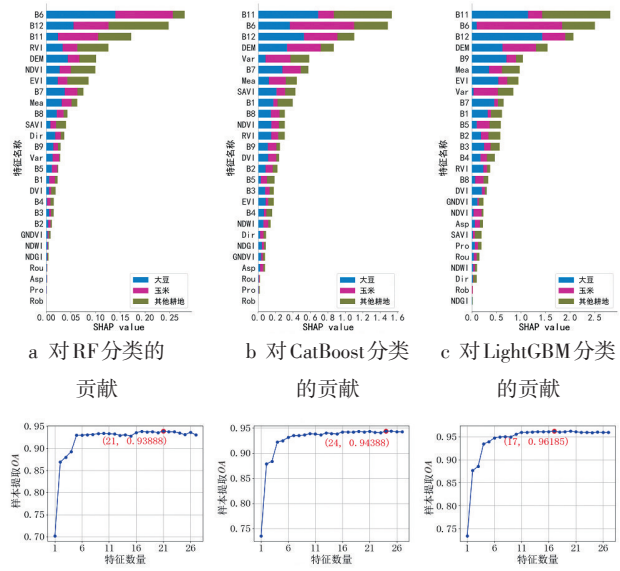
a 不同纹理特征　b 不同领域特征　c 不同地形特征

d 不同纹理特征的方差　e 不同邻域特征的方差　f 不同地形特征的方差

图4　秋粮不同特征之间的相关系数以及不同特征的方差

3.2.2　秋粮分类方案对比结果

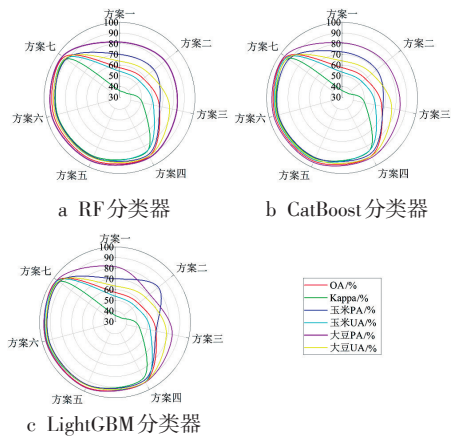
由图6可知，方案一的平均 *OA* 和平均Kappa系数同样最低，分别为57.52%和36.09%；随着特征的逐步加入，方案五的平均 *OA* 达到了94.18%，平均Kappa系数为91.28%；经过特征初选后，方案六的分类效果进一步优化，平均 *OA* 为94.41%，平均Kappa系数提升至91.61%；方案七的表现最佳，平均 *OA* 和Kappa系数分别为94.82%和92.23%，相较于方案一，平均 *OA* 提高了34.71%，平均Kappa系数增长了56.14%。



a 对RF分类的贡献　b 对CatBoost分类的贡献　c 对LightGBM分类的贡献

d RF分类　e CatBoost分类　f LightGBM分类

图5　各特征在不同分类器秋粮分类中的SHAP值以及 *OA* 随特征数量的变化



a RF分类器　b CatBoost分类器　c LightGBM分类器

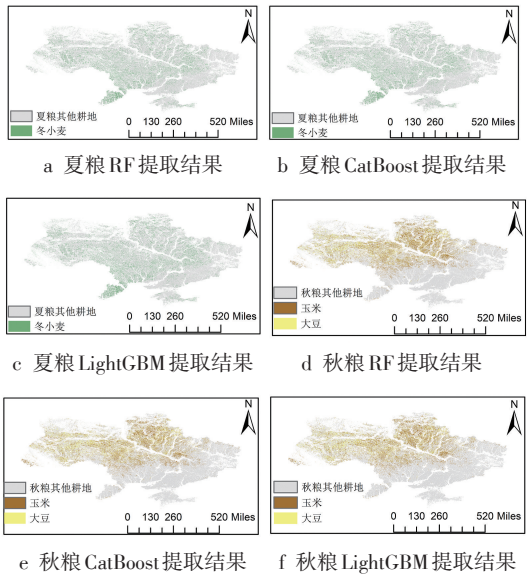
图6　不同特征筛选方案与不同分类器组合提取的秋粮精度

从各地物类型的 *PA* 和 *UA* 来看，随着特征数量的增加，大多数算法的平均 *PA* 和 *UA* 呈上升趋势，方案七的精度最高，平均 *PA* 为93.85%， *UA* 为93.35%。在方案七中，LightGBM算法的 *OA* 和Kappa系数最高，分别为96.19%和94.28%，与RF和CatBoost算法相比，*OA* 分别提高了2.29%和1.79%，Kappa系数分

别提高了3.45%和2.69%；LightGBM算法的平均 *PA* 和 *UA* 分别为96.21%和96.21%。综上所述，方案七LightGBM算法的各类精度几乎均为最高，即该算法为区分玉米和大豆的最佳算法。

3.3　作物提取与精度验证

在方案六中，LightGBM算法夏粮 *OA* 和Kappa系数为95.42%和90.83%，秋粮 *OA* 和Kappa系数为96.19%和94.28%，均为最高值。最优算法与其他两种算法相比，方案七中夏粮 *OA* 分别提高了0.36%和0.05%，Kappa系数分别提高了0.72%和0.11%；秋粮 *OA* 分别提高了2.29%和1.79%，Kappa系数分别提高了3.45%和2.69%。夏粮RF算法的平均 *PA* 和 *UA* 分别为95.44%和95.41%，秋粮LightGBM的平均 *PA* 和 *UA* 分别为96.21%和96.21%。基于此，本文最终选择特征组合方案七，并利用3种分类算法对夏粮、秋粮影像进行分类。乌克兰2024年冬小麦、玉米和大豆的分类影像见图7。



a 夏粮RF提取结果　b 夏粮CatBoost提取结果

c 夏粮LightGBM提取结果　d 秋粮RF提取结果

e 秋粮CatBoost提取结果　f 秋粮LightGBM提取结果

图7　乌克兰夏粮和秋粮空间分布图

审图号：GS(2025)4530

为验证遥感监测面积的准确性，本文将分类结果与乌克兰统计局发布的2024年统计年鉴数据进行对比（表5）。由于统计年鉴未包含克里米亚及东部冲突地区的数据，本文通过掩膜处理去除了该区域，最终计算得到冬小麦、大豆和玉米的相对精度为97.31%、97.55%和99.61%，其中冬小麦的监测面积为4 601.23千hm²，与年鉴值的绝对误差仅为127.27千hm²，表明模型对冬小麦拔节期的光谱特征具有极高的识别能力；玉米的监测面积为4 033.95千hm²，与年鉴值的误差率仅为0.39%，验证了B6、B11、B12波段对玉米抽穗期叶片叶绿素含量的敏感性；大豆的监测面积为2 590.43千hm²，与年鉴值的误差率为2.45%，主要受大豆与其他豆科作物（如豌豆）的光谱混淆影响。

表5　农作物面积提取精度

作物	分类器	遥感监测面积/千hm²	除未统计区域面积/千hm²	相对精度/%
冬小麦	RF	4 533.82	4 419.38	93.46
	CatBoost	4 675.18	4 548.35	96.19
	LightGBM	4 709.89	4 601.23	97.31
大豆	RF	2 583.60	2 582.96	97.27
	CatBoost	2 583.62	2 583.15	97.27
	LightGBM	2 590.94	2 590.43	97.55
玉米	RF	4 339.47	4 338.87	93.34
	CatBoost	4 130.38	4 129.92	98.05
	LightGBM	4 039.18	4 033.95	99.61

3.4　结果分析

本文基于SHAP和Spearman相关系数的特征优选方法，深入分析了乌克兰夏粮和秋粮的作物制图流程。结果表明，利用Sentinel-2遥感影像和DEM提取的多种特征有效提高了作物分类精度。在特征筛选过程中，利用Spearman相关系数剔除了相关性较高的特征，减少信息冗余，提高模型的稳定性；通过SHAP值进一步识别对模型预测贡献最大的特征，确保了模型输入的有效性，该策略不仅提升了分类精度，还为后续的模型优化提供了明确的方向。数据分析结果表明，方案七的分类性能最佳，夏粮平均 *OA* 达到了95.28%，Kappa系数为90.55%，相较于方案一，平均 *OA* 提高了29.08%，Kappa系数增加了58.18%；秋粮平均 *OA* 达到了94.82%，Kappa系数为92.23%，相较于方案一，平均 *OA* 提高了34.71%，Kappa系数增加了56.14%，说明特征优化在机器学习模型中的应用具有显著效益，能有效提高作物分类的准确性。

此外，RF、LightGBM和CatBoost等机器学习算法展现出强大的性能和适应性，特别是在处理高维特征和复杂数据模式时。LightGBM在夏粮与秋粮的分类中的精度最高，夏粮 *OA* 和Kappa系数分别为95.42%和90.83%，秋粮 *OA* 和Kappa系数分别为96.19%和94.28%。研究结果为未来农业遥感研究和应用提供了新的思路。

4　结　论

本文基于乌克兰2024年Sentinel-2遥感影像和统

计年鉴数据，通过引入融合Spearman相关系数与SHAP的特征优选框架，实现了高精度作物制图。

1）结合Spearman相关系数与SHAP的特征筛选策略，成功将41个特征缩减至27个，在保证计算效率的同时提升了模型的可解释性。SHAP分析表明，光谱特征（B4、B6、B11、B12和B8）和植被指数（NDVI、RVI）是作物分类的核心贡献因子。

2）方案七LightGBM算法的作物识别精度最高。夏粮*OA*为95.42%、Kappa系数为90.83%；秋粮*OA*为96.19%、Kappa系数为94.28%，冬小麦、大豆和玉米提取的面积精度分别为97.31%、97.55%和99.61%。

参考文献

[1] 刘洪顺,冯竞略,邹利林. 广东省耕地“非粮化”的时空跃迁与整治路径研究[J]. 地理空间信息,2025,23(1):41-45

[2] 陈仲新,郝鹏宇,刘佳,等. 农业遥感卫星发展现状及我国监测需求分析[J]. 智慧农业,2019,1(1):32-42

[3] 何燕,王斌,江立庚,等. 基于GIS的广西水稻种植布局精细化气候区划[J]. 中国水稻科学,2013,27(6):658-664

[4] 韩文霆,张立元,牛亚晓,等. 无人机遥感技术在精量灌溉中应用的研究进展[J]. 农业机械学报,2020,51(2):1-14

[5] 董金玮,崔屹峰,邸媛媛,等. 大数据时代耕地红线管控监测的机遇与挑战[J]. 中国科学院院刊,2023,38(12):1781-1792

[6] 胡菡,厉芳婷,许文伟. 基于遥感影像的水资源提取精度评价方法研究[J]. 地理空间信息,2023,21(10):65-68

[7] 李颖,陈怀亮. 机器学习技术在现代农业气象中的应用[J]. 应用气象学报,2020,31(3):257-266

[8] 熊曦柳,胡月明,文宁,等. 耕地遥感识别研究进展与展望[J]. 农业资源与环境学报,2020,37(6):856-865

[9] 王敏,胡兆,徐晓巍,等. 融合知识驱动和数据驱动的混合决策模型构建:以室性心动过速病因诊断为例[J]. 协和医学杂志,2025,16(2):454-461

[10] 曾韬睿,王林峰,张俞,等. 基于CatBoost-SHAP模型的滑坡易发性建模及可解释性[J]. 中国地质灾害与防治学报,2024,35(1):37-50

[11] 王美知,魏凤. 哈萨克斯坦粮食生产效率动态演进及区域差异[J]. 自然资源学报,2021,36(3):594-605

[12] 浦男萍,赵晶,张颖,等. 大食物观统筹下的耕地保护协同治理体系研究[J]. 西南大学学报(自然科学版),2024,46(10):46-58

[13] 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区土地利用分类[J]. 农业工程学报,2020,36(4):244-250

[14] McFeeters S K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features[J]. International Journal of Remote Sensing, 1996, 17(7): 1425-1432

[15] Smith C B, Matthiesen R B. Vibration Testing of Full-scale Structures[J]. Nuclear Engineering and Design,1973,25(1): 17-29

[16] Jordan C F. Derivation of Leaf-area Index from Quality of Light on the Forest Floor[J]. Ecology,1969,50(4):663-666

[17] Richardson A J, Wiegand C L. Distinguishing Vegetation from Soil Background Information[J]. Photogrammetric Engineering and Remote Sensing,1977,43(12):1541-1552

[18] Huete A R. A Soil-adjusted Vegetation Index (SAVI)[J]. Remote Sensing of Environment,1988,25(3):295-309

[19] Huete A R, Liu H Q, Batchily K, et al. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS[J]. Remote Sensing of Environment, 1997, 59(3):440-451

[20] Gitelson A A, Kaufman Y J, Merzlyak M N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS[J]. Remote Sensing of Environment, 1996,58(3):289-298

[21] Revel C, Lonjou V, Marcq S, et al. Sentinel-2A and 2B Absolute Calibration Monitoring[J]. European Journal of Remote Sensing,2019,52(1):122-137

[22] 杨丹,李崇贵,张家政. 基于深度学习的多时相GF-1影像林分类型分类研究[J]. 林业资源管理,2022(1):142-149

[23] 李乐,张承云,裴男才,等. 基于被动声学监测技术的城市绿地景观格局与鸟类多样性关联分析[J]. 生物多样性,2024,32(10):157-171

[24] 李硕,艾丽菲拉·尔肯,罗文波,等. 基于AutoML-SHAP的超高性能混凝土抗压强度可解释预测[J]. 硅酸盐通报,2024,43(10):3634-3644

[25] 曾慧娴,杨之雨,柳东红,等. 医学研究中常见统计分析方法的应用[J]. 上海预防医学,2023,35(8):831-839

[26] 聂卉,吴晓燕. 结合梯度提升树算法与可解释机器学习模型SHAP的抑郁症影响因素研究[J]. 数据分析与知识发现,2024,8(3):41-52

[27] 张圣微,杨林,叶德成,等. 基于深度学习的内蒙古河套灌区2000—2021年种植结构提取与动态[J]. 农业工程学报,2023,39(18):142-150

[28] Chmura K H, Periyakoil V S, Noda A. Kappa Coefficients in Medical Research[J]. Statistics in Medicine, 2002,21(14): 2109-2129