

基于特征筛选和粒子群优化的花生生物量估算

刘 涛^{1,2}, 杨奉源¹, 刘 望¹, 张 寰², 殷冬梅³, 张全国², 焦有宙^{2,4*}

(1. 河南财经政法大学城乡规划学院, 郑州 450046; 2. 河南农业大学机电工程学院, 郑州 450002; 3. 河南农业大学农学院, 郑州 450002; 4. 河南工程学院机械工程学院, 郑州 451191)

摘 要: 为解决花生植株生物量估算精度低、破坏性大等问题, 该研究提出一种无人机低空遥感技术结合高光谱特征筛选的花生生物量估算方法。通过无人机搭载高光谱成像仪, 获取田块尺度多个花生品种的高光谱影像数据, 首先对获取的影像进行拼接、辐射定标、大气校正等预处理, 提取出地面采样点位置的光谱反射率, 计算光谱反射率的一阶微分和植被指数, 使用变量投影重要性 (variable importance in projection, VIP) 方法对光谱反射率、一阶微分和植被指数等三种数据进行特征筛选, 利用筛选后的特征和地面实测数据构建支持向量机回归 (support vector regression, SVR)、反向传播神经网络回归 (back propagation neural network, BPNN) 和随机森林回归 (random forest regression, RFR) 模型, 并使用粒子群优化算法 (particle swarm optimization, PSO) 进行模型优化。结果表明: 相比原始光谱反射率和植被指数, 一阶微分光谱反射率与花生生物量具有较好的相关性; 使用一阶微分光谱反射率与植被指数组合的 RF 回归模型精度最高 (决定系数 R^2 为 0.754, 均方根误差 RMSE 为 0.085 kg/m²), 使用粒子群优化后的 PSO-RF 模型可进一步提高模型精度 (R^2 为 0.80, RMSE 为 0.076 kg/m²)。该研究为花生生物量精准估算提供了有效的方法, 为智慧乡村建设中的精细化农田管理提供技术支持。

关键词: 花生; 生物量; 智慧乡村; 特征筛选; 机器学习; 粒子群优化

doi: 10.11975/j.issn.1002-6819.202407251

中图分类号: S252

文献标志码: A

文章编号: 1002-6819(2025)-01-0238-10

刘涛, 杨奉源, 刘望, 等. 基于特征筛选和粒子群优化的花生生物量估算[J]. 农业工程学报, 2025, 41(1): 238-247. doi:

10.11975/j.issn.1002-6819.202407251 <http://www.tcsae.org>

LIU Tao, YANG Fengyuan, LIU Wang, et al. Estimation of peanut biomass based on feature selection and particle swarm optimization[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(1): 238-247. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202407251 <http://www.tcsae.org>

0 引 言

花生是全球广泛种植的主要油料作物之一^[1]。中国是世界上最大的花生生产国、消费国和进口国。随着人民生活水平的提高, 油料的需求显著增加。2023 年, 中国花生产量达到 1 640 万 t, 占全球总量的 37%。然而, 许多粮油出口国加强了对出口的管控, 这对中国油料供应的安全来了挑战^[2]。由此可见, 提高花生产量对于促进国民经济和农业高质量发展具有重大意义。快速、准确、无损地监测花生生长状况和估算其生物量, 对于提高花生产量、保障国家油料供应安全和提高农业生产效率至关重要^[3-4]。

生物量是指在某一时间单位面积内所有植株的总干质量, 包括植物在地上及地下部分的质量。作为监测和

估测作物产量的基本因子, 生物量直接反映了植物的生长状况和光合作用的效果, 是评估作物长势的关键参数之一^[5]。花生作为一种典型的根系植物, 其果实主要生长在地下。传统的花生产量计算方法多侧重于测定地下果实的鲜质量。然而, 花生植株的地上部分也同样具有一定的经济和生态价值, 花生植株地上部分不仅可以作为家禽牲畜的饲料^[6], 还可以用于生物质发酵底物制备氢气等清洁能源^[7]。因此, 全方位估算花生植株的生物量, 不仅能够更精准地估测花生果产量, 还能为农业生产管理、资源高效利用以及碳足迹核算提供重要依据。全面的生物量估算有助于优化花生种植策略, 提高资源利用效率, 降低农业生产对环境的负面影响。

传统的生物量估测方法需要进行破坏取样, 不仅费时耗力, 且结果在空间和时间上存在局限性^[8-9]。遥感技术的发展为作物生物量估测提供了高效的方法。基于卫星的遥感探测技术被人们逐渐运用于生物量估测以及植被长势监测^[10-12]。然而, 卫星遥感因存在空间和时间分辨率低、易受天气影响等缺陷, 其在作物监测方面的应用仍受到一定限制^[13-14]。近年来, 无人机遥感技术凭借其轻便、灵活、低成本和高精度等优势, 在农业农情监测中得到了广泛应用^[15-16]。然而, 无人机搭载的 RGB 相机和多光谱相机所获取的影像波段较少, 难以充分反映作物冠层的光谱信息^[17-18]。相比之下, 高光谱相机因其

收稿日期: 2024-07-30 修订日期: 2024-11-27

基金项目: 河南省科技研发计划联合基金青年科学家项目 (225200810089); 河南省高校科技创新人才支持计划 (23HASTIT028); 河南省青年人才托举工程项目 (2024HYTP018); 河南省博士后科研项目 (202101041); 河南省中原科技创新领军人才项目 (234200510029)

作者简介: 刘涛, 博士, 副教授, 研究方向为农业资源与环境遥感监测、空间位置服务。Email: liutao@huel.edu.cn

*通信作者: 焦有宙, 博士, 教授, 博士生导师, 研究方向为农业废弃物高效与清洁转化。Email: jiaoyouzou@henau.edu.cn

较高的光谱分辨率，可以获取丰富的作物反射率光谱信息，有利于构建光谱指数、精准提取作物生长信息^[19-20]。目前，国内外已有学者利用无人机高光谱技术监测作物生长情况。唐子竣等^[21]使用一阶微分光谱反射率和光谱指数，通过机器学习算法构建大豆产量估算模型。结果显示，全莢期光谱指数构建的模型验证集 R^2 达到 0.85。齐浩等^[22]利用无人机高光谱数据估测小麦品系的产量，实验发现基于 Stacking 算法构建的多模型集成组合效果优于传统的线性模型。FENG 等^[23]结合地面和无人机高光谱数据，计算出不同植被指数和红边参数组合，构建了不同生育期冬小麦产量估算模型，结果表明，PLSR 相比 ANN 方法具有更好的产量预测效果，结合使用植被指数和红边参数可以获得更好的产量估算精度。CHEN 等^[24]利用无人机高光谱影像对关中地区冬小麦孕穗期、抽穗期、开花期、灌浆期及整个生育期氮素浓度进行了预测，结果表明，机器学习回归方法可以显著提高冬小麦氮含量的估计精度。XU 等^[25]利用无人机搭载高光谱相机采集了 5 种氮素水平下水稻生育早期的稻田图像，实现了动态估计水稻不同生长阶段的地上生物量。

此外，为了提高监测精度，学者们引入了特征筛选方法和机器学习的优化方法。例如，SAHOO 等^[26]利用混合 R^2 和变量投影重要性（variable importance in projection, VIP）的特征筛选方法，筛选出 13 个最优的敏感光谱指数，并在人工神经网络算法中建模，生成试验麦田植株氮素的空间分布图。结果表明使用相关性较高的特征变量构建的模型具有较高的估算精度和可靠性。江凯伦等^[27]设计了一种基于正则近邻成分分析的光谱特征选择方法，通过优化损失函数和正则化参数，获得水稻叶绿素不同含量的特征波段，并以此为输入，构建粒子群优化极限学习机叶绿素含量反演模型，该模型的叶绿素含量反演结果的均方根误差（root mean square error, RMSE）和 R^2 分别为 9.549 mg/L 和 0.891。上述研究表明，结合特征筛选与优化的机器学习方法可以显著提高高光谱数据在农业监测中的应用效果，具有较高的实际应用价值。

尽管无人机高光谱遥感在农作物生长监测中的应用前景广阔，但当前的研究主要集中于小麦、水稻、玉米等粮食作物，且较多关注叶绿素、氮浓度等长势参数。花生作为一种重要的经济作物，在其生长过程中生物量的遥感监测研究相对缺乏，特别是缺少针对不同品种间冠层光谱特性差异的深入分析。本文以河南荥阳花生种植试验田中处于饱果成熟期的 11 个不同品种花生为研究对象，通过无人机高光谱影像提取地面采样点的光谱反射率数据，并构建光谱一阶微分和植被指数作为特征输入。随后，使用变量投影重要性（variable importance in projection, VIP）方法对这些特征进行筛选和排序。采用支持向量机（support vector machine, SVM）、反向传播神经网络（back propagation neural network, BPNN）和随机森林（random forest, RF）三种机器学习方法分别建立生物量估测模型，并利用粒子群优化算法对模型

进行优化，最终对不同模型的精度进行分析和评价，旨在构建一种快速、准确预测花生生物量的模型，为花生遥感监测提供理论参考。

1 材料与方法

1.1 研究区域概况与试验设计

试验于河南省荥阳市河南农业大学花生种植试验田（113°25'27"E，34°51'11"N）进行（图 1）。试验田位于河南省中西部，地处黄河中下游平原，地势平坦、土壤肥沃，平均海拔为 110 m，气候属温带季风气候，年均降水量为 640 mm，年均温度约 14℃，适宜种植小麦、玉米、花生等作物。

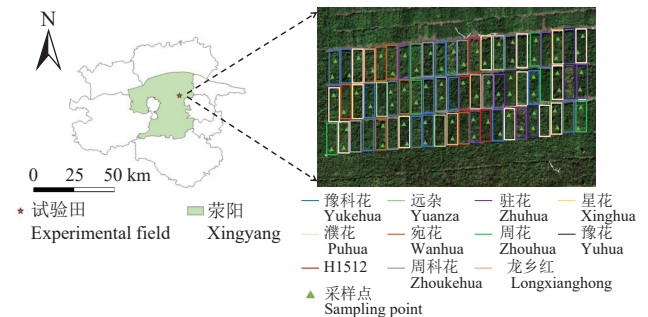


图 1 研究区位置与样地点分布图

Fig.1 Location of the study area and distribution of sampling sites

该试验样地为夏播花生育种试验田，花生种植时间为：2022 年 6 月—10 月，种植品种包括豫科花、龙乡红等在内的 11 个花生品种，每个品种重复 3 次。试验田共设计 66 个小区，每个小区面积为 6.7 m×2 m，各小区间距 0.5 m，行距 0.4 m，种 5 行，空 1 行，株距 13.9 cm。

1.2 高光谱影像获取与处理

本研究利用大疆六旋翼无人机 M600 Pro 搭载四川双利合谱有限公司的 GaiaSky-mini2-VN 高光谱成像仪进行数据采集。采集时间为 2022 年 10 月 10 日上午 11:00—12:00，天气晴朗微风，此时花生处于饱果成熟期。试验田于影像采集当天收获，因此所获取的高光谱数据能够很好的反映出植株整体生物量状况。使用 DJI A3 Pro 飞控系统，飞行高度设置为 50 m，影像的旁向和航向重叠率均设置为 80%，以确保数据的全面覆盖和高质量拼接。高光谱成像仪实时采集的影像波段范围为 400~1 000 nm，包含 176 个光谱通道，光谱分辨率为 3.5 nm，空间分辨率约为 1.8 cm/pixels，高光谱成像仪的详细参数见表 1。

在无人机起飞前，使用高光谱成像仪拍摄标准黑白板影像，以消除辐射误差和镜头几何畸变；起飞后，拍摄包含有反射率为 30% 和 50% 的灰布区域，并对所有影像进行辐射定标。高光谱影像采集完成后，使用成像仪自带 Specview 软件对所有影像数进行数据预处理，主要包括：影像拼接、镜头校正、几何校正、辐射定标和大气校正，预处理后的影像空间分辨率约为 5 cm。随后，使用 ENVI 软件提取影像的冠层光谱反射率。根据每个地面采样点位置坐标划分感兴趣区域（region of interest,

ROI), 提取感兴趣区域内所有像元的光谱反射率并计算平均值, 从而得到每个采样位置的冠层光谱反射率^[28]。

表 1 高光谱成像仪的详细参数

Table 1 Detailed parameters of the hyperspectral imager

参数名称	参数值
Parameter name	Parameter value
光谱范围	400~1 000
Spectral range/nm	
光谱分辨率	3.5
Spectral resolution/nm	
数值孔径	像方 F/2.8
Numerical aperture/nm	
光谱采样率	0.5
Spectral sampling rate/nm	
全幅像素	1936×1 456
Full frame pixels/像素	
像素间距	4.54
Pixel pitch/μm	
横向视角 (FOV)	30.25
Field of view/(°)	
单幅图像分辨率	1920×2 080
Single image resolution/像素	
空间分辨率	0.085
Spatial resolution/m	

1.3 地面数据获取

如图 1 所示, 在不同品种试验田中选取 70 个地面采样点, 使用 GNSS 接收机定位并记录每个点的空间位置坐标。在采样点位置, 收获 0.5 m×0.5 m 范围内的所有植株, 轻轻剔除泥土后, 将植株装袋封存并带回实验室清洗。清洗后的植株进行自然风干、放在烘箱中, 设定温度 75℃ 烘干 6~12 h 至恒质量, 测取质量并换算为单位面积后作为各采样点的生物量实测值。不同品种采样点植株生物量实测值分布如图 2 所示。为了进行模型训练和验证, 本文从这些样本中随机选取 80% 样本点作为训练集, 剩余的 20% 样本点作为验证集。

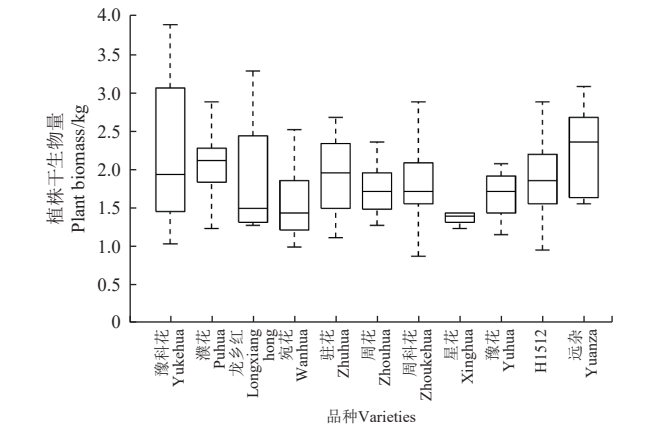


图 2 不同品种采样点植株生物量实测值分布

Fig.2 Distribution of measured plant biomass at sampling points of different varieties

1.4 特征选取

目前国内外对植被指数的研究已经相对成熟, 但关于哪些植被指数适合估算花生生物量尚无定论。因此, 本研究根据前人的研究成果及经验, 选择了 53 种与植物生物量和生长状态相关的植被指数作为估算花生生物量的候选特征, 部分植被指数的计算见表 2。

表 2 部分植被指数及计算公式

Table 2 Partial vegetation indices and calculation formulas

植被指数	计算公式
Vegetation index (VI)	Calculation formula
修正地面叶绿素指数 MMTCI	$((R_{750}-R_{680}+0.03)(R_{750}-R_{710}))/((R_{710}-R_{680})$
Modified medium resolution transformed chlorophyll index	
双峰冠层氮素指数 DCNI	$((R_{750}-R_{670}+0.09)(R_{750}-R_{700}))/((R_{700}-R_{670})$
Double peak canopy nitrogen index	
增强植被指数 EVI	$2.5(R_{800}-R_{670})/(1+R_{800}+6R_{670}-7R_{479})$
Enhanced vegetation index	
红边叶绿素指数 RECI	$(R_{750}/R_{710})-1$
Red edge chlorophyll index	
中分辨率陆地叶绿素成像指数 MTCI	$(R_{750}-R_{710})/(R_{710}-R_{680})$
Medium resolution imaging spectrometer chlorophyll index	
绿色归一化植被指数 GNDVI	$(R_{801}-R_{550})/(R_{800}+R_{550})$
Green normalized difference vegetation index	
新植被指数 NVI	$(R_{777}-R_{747})/R_{673}$
New vegetation index	
复归一化差值植被指数 RDVI	$(R_{800}-R_{670})/(\sqrt{R_{800}+R_{670}})$
Renormalized difference vegetation index	
植被衰老反射率指数 PLSRI	$(R_{680}-R_{500})/R_{750}$
Plant leaf senescence reflectance index	
新型植被指数 LICl	$R_{735}/R_{720}-(R_{573}-R_{680})/(R_{573}+R_{680})$
Light-use efficiency chlorophyll index	

注: R_{λ} 表示波长为 λ nm 的光谱反射率。
Note: R_{λ} is spectral reflectance at a wavelength of λ nm.

除了常见的植被指数, 本文研究使用采样点 ROI 的光谱反射率均值、光谱反射率的一阶微分作为特征输入, 一阶微分的计算式如下所示:

$$D_i = \frac{R_{i+1} - R_i}{\lambda_{i+1} - \lambda_i}$$

(1)

式中 R_i 是指第 i 波段的反射率值, λ_i 是指第 i 个波段的波长值, D_i 是指第 i 波段的一阶微分特征值。采样点 ROI 区域的光谱反射率和一阶微分值如图 3 所示。

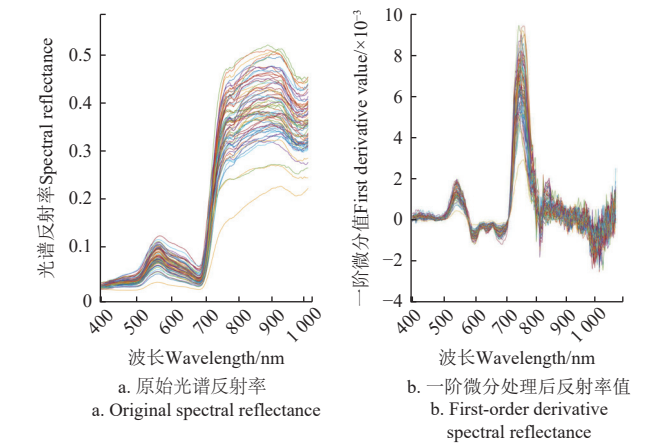


图 3 采样点光谱反射率和一阶微分值

Fig.3 Spectral reflectance and first order differential values of sampling points

1.5 特征筛选与模型构建

在遥感领域中, 高光谱数据由于其高光谱分辨率, 通常包含超过 100 个波段。这些波段在细致刻画地物反射光谱特征方面具有重要优势, 但也带来了数据冗余和维数灾难的问题^[29]。为了有效利用这些数据, 我们需要对其进行筛选和优化, 以减少影像数据的维度, 提升数据处理的效率和准确性。在这一过程中, 变量投影重要性 (variable importance in projection, VIP) 方法被用来

进行特征筛选，通过保留重要的光谱信息并剔除冗余数据，构建花生生物量的估算模型，从而实现对目标特征的精确识别和量化估算。

1.5.1 变量投影重要性

变量投影重要性（VIP）^[30] 是一种基于偏最小二乘（PLS）回归的变量筛选方法，通过衡量自变量对因变量的解释能力对自变量进行排序。其算式如下：

$$V_j = \sqrt{\frac{k}{\sum_{h=1}^m r^2(y, c_h)} \sum_{h=1}^m r^2(y, c_h) w_{hj}^2} \quad (2)$$

式中 k 为自变量个数； m 为从自变量中提取的主成分； c_h 为相关自变量提取的主成分； $r(y, c_h)$ 为花生生物量因变量和主成分的相关系数，表示主成分对 y （花生生物量）的解释能力， w_{hj} 是自变量在主成分上的权重。 V_j 代表自变量对模型拟合的重要程度， V_j 值越大，代表该自变量对于花生生物量的解释能力越强。本文通过计算光谱反射率、一阶微分光谱反射率、植被指数等特征的 VIP 值，来进行特征筛选，从而构建花生生物量的估算模型。

1.5.2 机器学习模型

随机森林（random forest, RF）是一种典型的机器学习方法，通过集成学习的思想构建多个决策树组成的森林来进行预测。每棵决策树的构建过程中，采用 bootstrap 抽样方法，从原始数据集中随机抽取样本形成训练集，以保证训练集的差异性。在节点分裂过程中，RF 随机选择部分特征进行分裂，降低决策树之间的相关性，提高模型的多样性和泛化能力^[31]。在 RF 模型构建中，决策树数量（ntree）和最小叶子节点数量（leaf）等超参数对模型性能有显著影响。为了减小优化搜索空间并加速算法收敛，本文设定 ntree 的初始范围为 10 到 500，leaf 的初始范围为 1 到 5。

BP 神经网络（backpropagation neural network, BPNN）是一种常用的多层前馈神经网络，主要通过逐层调整网络中的权重和偏置来最小化预测误差，从而提高模型的准确性。其训练过程包括前向传播和反向传播两个阶段。在前向传播过程中，输入数据经过各层的线性变换和非线性激活函数得到输出结果；在反向传播过程中，通过计算预测误差的梯度，逐层更新权重和偏置，使模型逐步逼近最优状态^[32]。BPNN 在处理复杂非线性关系、分类和回归问题时表现出色，其中隐含层中的神经元数量对模型性能影响显著。为优化模型性能，本文设定隐含层中神经元数量的范围为 1 到 30。

支持向量机（support vector machine, SVM）作为一种广义线性分类器，主要用于解决二元分类问题。其目标是寻找一个最优超平面，以最大化分类间隔，将不同类别的样本分开，从而提高模型的泛化能力和鲁棒性。SVM 在处理高维数据和小样本数据时表现尤为优异^[33]。在 SVM 中，惩罚系数 C 和径向基函数参数 γ 对模型性能有显著影响。 C 用于控制优化过程中对训练数据中误分类样本的惩罚程度，而 γ 则决定了单个训练样本对决策

边界的影响范围。本文将惩罚系数 C 的初始范围设定为 1 到 500，径向基函数参数 γ 的初始范围设定为 0.1 到 5，以优化模型性能。

1.5.3 粒子群优化算法

粒子群优化算法（particle swarm optimization, PSO）结合了社会学习和个体探索的机制，使得成员（粒子）在解空间中进行全局搜索和局部搜索，从而找到最优解^[34]。得益于其收敛速度快且易于实现等优点，PSO 算法被广泛应用于多种优化问题，常被用于机器学习和深度学习中，寻找最优参数和对权重的优化^[35-36]。在本试验中，使用三种机器学习方法：随机森林（RF）、支持向量机（SVM）和反向传播神经网络（BPNN），每种方法均涉及多个参数（如树的数量、正则项系数、隐藏层节点数等），这些参数的组合对模型的性能和泛化能力有显著影响。为寻找最佳参数组合，我们采用粒子群优化算法（PSO）。具体来说，PSO 算法设定学习因子 c_1 和 c_2 分别为 1.5 和 1.7，惯性权重为 0.7，使用均方根误差（RMSE）作为适应度函数。在初始化种群和模型超参数后，将其应用于 RF、SVM 和 BPNN 模型中。每次迭代时，记录适应度值并判断是否达到终止条件。如果未达到，则继续更新粒子的位置和速度，直至满足终止条件。

1.6 精度验证

本试验旨在利用无人机高光谱影像和地面花生生物量实测值等数据，通过机器学习方法构建花生生物量估测模型。详细试验流程如图 4 所示，特征输入包括三类数据来源：原始波段反射率、光谱反射率一阶微分和植被指数。

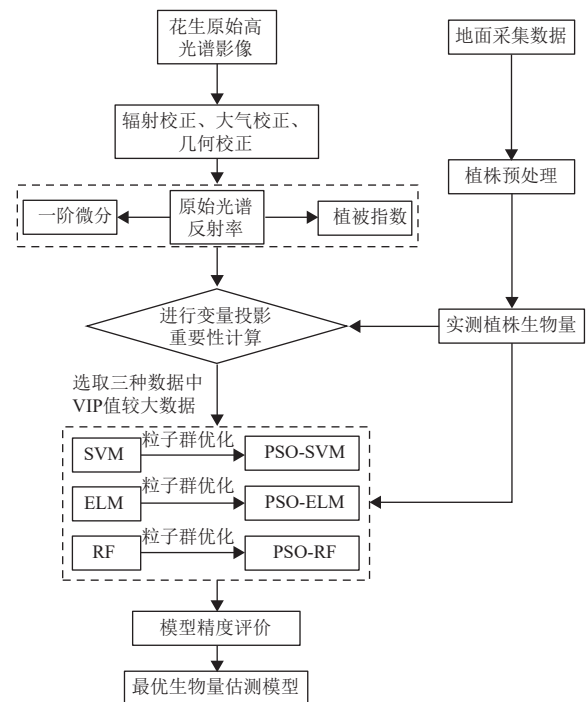


图 4 试验流程图

Fig.4 Experimental flowchart

为筛选出关键特征变量，采用变量投影重要性（VIP）

方法对 3 种数据特征进行相关性评估和 VIP 值排序。使用支持向量机 (SVM)、反向传播神经网络 (BPNN) 和随机森林 (RF) 3 种机器学习算法构建花生生物量预测模型, 并通过粒子群优化 (PSO) 算法对这三种模型进行优化, 利用相关性最高的特征组合及模型, 实现对试验区域内花生生物量的估测。本试验中涉及多种模型的对比和精度评价, 使用决定系数 (R^2)、均方根误差 (RMSE)、均方误差 (MSE)、平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE) 等指标来评价模型, 具体计算公式如下所示, R^2 越趋近于 1、RMSE、MSE、MAE 和 MAPE 越小, 则表明模型精度越高。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (5)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (6)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

式中 y_i 表示第 i 个实际观测值, kg/m^2 ; $\hat{y}_i$ 表示第 i 个模型预测值, kg/m^2 ; $\bar{y}$ 表示实际观测值的均值, kg/m^2 ; n 表示观测值个数。

2 结果与分析

2.1 特征排序

对高光谱影像中 176 个原始反射率波段、一阶微分处理后的 175 个特征波段以及 53 个植被指数分别与实测花生生物量进行 VIP 值计算, 3 种数据按照 VIP 值由大到小的顺序排列, 结果如图 5 所示。

在原始波段数据中, 748.8 和 745.3 nm 两个波段的 VIP 值最高, 分别为 1.42 和 1.40。在一阶微分数据中, df99 和 df100 两个一阶微分特征的 VIP 值最高, 分别为 2.57 和 2.56; 其中, DF99 和 DF100 分别是根据式 (1) 由原始反射率数据中相邻波段 734.8、731.3、731.3、727.8 nm 计算得到的。植被指数 DCNII (双峰冠层氮素指数) [37] 和 MMTCI (修正地面叶绿素指数) [38] VIP 值最高, 分别达到了 1.59 和 1.52。通过相关性分析可以发现, 与花生生物量相关的特征波段主要集中在波长 700~750 nm 的光谱反射率范围。

2.2 基于单一类型特征的生物量估测模型构建

本文根据 VIP 排序结果, 选取前 1% 的特征 (植被指数选取 2 个) 作为输入因子, 分别构建基于 SVM、BPNN 和 RF 的花生生物量估算模型, 模型在训练集和测试集的试验结果如表 3 所示。结果表明, 相较于原始波段反射率特征, 由一阶微分和植被指数分别构建的机

器学习模型表现更佳, 这与张凡凡等 [39] 的研究结果一致。其中, 由植被指数的三种机器学习模型在训练集上的决定系数 R^2 均超过 0.59; 由一阶微分的三种机器学习模型在训练集上的决定系数 R^2 均超过 0.48。

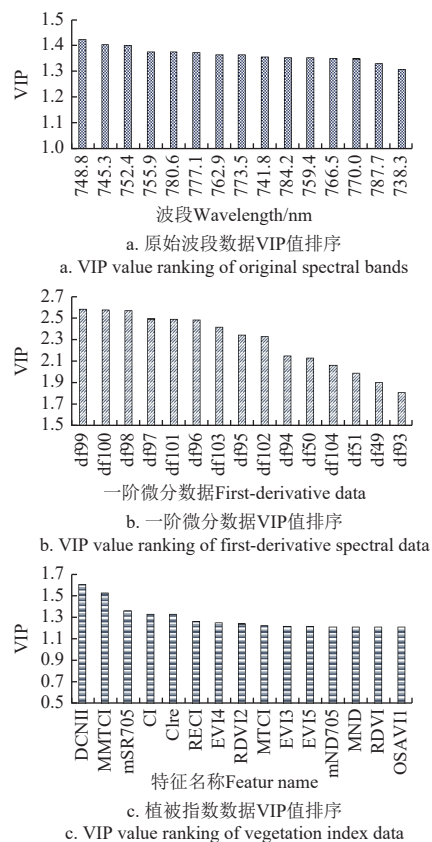


图 5 VIP(variable importance in projection) 排序结果

Fig.5 Sort result of VIP analysis

表 3 基于三种机器学习方法构建花生生物量预测模型精度对比

Table 3 Comparison of peanut biomass prediction models constructed using three machine learning methods

数据 Data	建模方法 Modeling method	变量个数 Number of variables	训练集 R^2 Training set R^2	训练集 RMSE Training set RMSE/ ($\text{kg} \cdot \text{m}^{-2}$)	测试集 R^2 Testing set R^2	测试集 RMSE Testing set RMSE/ ($\text{kg} \cdot \text{m}^{-2}$)
原始波段反 射率 Raw band data	SVM	2	0.197	0.167	0.147	0.158
	BPNN	2	0.278	0.159	0.104	0.162
	RF	2	0.541	0.076	0.126	0.157
植被指数 Vegetation index data	SVM	2	0.590	0.119	0.364	0.137
	BPNN	2	0.622	0.115	0.342	0.139
	RF	2	0.851	0.072	0.564	0.113
一阶微分 First derivative data	SVM	2	0.482	0.134	0.302	0.143
	BPNN	2	0.487	0.134	0.330	0.140
	RF	2	0.846	0.073	0.686	0.096

特别是使用一阶微分作为 RF 方法的特征输入效果最佳, 该模型在训练集上的决定系数 R^2 为 0.846, 在测试集上的决定系数 R^2 为 0.686, RMSE 为 0.096 kg/m^2 。结果表明, 单一特征类型在花生生物量估算中的整体精度较低, 难以全面反映植株生物量的复杂变化。因此需要探索不同特征组合对提高对花生生物量估算精度的影响。

2.3 基于特征组合的花生生物量估算模型构建

由以上结果可知, 经过 VIP 相关性筛选后的特征与

花生生物量具有较好的相关性，因此本试验将筛选后的原始波段、一阶微分、植被指数进行特征组合，使用机器学习方法构建花生生物量估算模型，对每种类型特征中 VIP 值最高的两个数据进行组合作为模型输入，并讨论多种组合情况下，不同机器学习方法对花生生物量估测精度的影响。具体组合及试验结果表 4 所示。

结果表明，相较于基于单一类型特征的生物量估算模型，使用组合特征构建的三种机器学习模型性能出现了不同程度的变化。由一阶微分和植被指数特征组合后构建的 RF 模型表现最佳，在训练集和测试集上的决定系数 R^2 分别为 0.734 和 0.754，RMSE 分别为 0.096 和 0.085 kg/m²。相比之下，由一阶微分和原始波段特征组合后构建的机器学习模型性能略有下降，可能是由于原始波段和一阶微分特征之间存在一定冗余。由植被指数

和原始波段特征组合后构建的 RF 模型，性能相比单纯使用植被指数构建的 RF 模型下降较大（训练集 R^2 下降 34%、测试集 R^2 下降 16%），这可能是因为筛选后的植被指数 DCNII 和 MMTCI 都是由筛选后的原始反射率波段 B750 组成的。

此外，将一阶微分、植被指数和原始波段等三种数据来源进行特征组合后构建的机器学习模型性能并不最优，在训练集和测试集上的决定系数 R^2 、 R_1^2 别为 0.651 和 0.640，RMSE、RMSE₁ 分别为 0.11 和 0.10 kg/m²，这与赵玉清等^[40]的研究不同，三个类别特征组合模型精度高于两类别、单一类别特征组合。而本文的试验结果表明：特征类型、数量并不总是越多越好，关键在于筛选出有效的特征。有效特征的筛选对于提高花生生物量估算的准确性至关重要。

表 4 基于特征组合的三种机器学习方法模型精度对比
Table 4 Comparison of peanut biomass prediction models constructed using three machine learning methods

数据源 Data	建模方法 Modeling method	变量个数 Number of variables	训练集 R^2 Training set R^2	训练集 RMSE Training set RMSE/ (kg·m ⁻²)	测试集 R_1^2 Testing set R_1^2	测试集 RMSE ₁ Testing set RMSE ₁ / (kg·m ⁻²)
一阶微分+植被指数 First derivative data + Vegetation index data	SVM	4	0.656	0.109	0.618	0.106
	BPNN	4	0.570	0.122	0.480	0.124
	RF	4	0.734	0.096	0.754	0.085
一阶微分+原始波段 First derivative data + Raw band data	SVM	4	0.611	0.116	0.564	0.113
	BPNN	4	0.364	0.149	0.355	0.138
	RF	4	0.751	0.093	0.674	0.098
植被指数+原始波段 Vegetation index data + Raw band data	SVM	4	0.791	0.081	0.501	0.121
	BPNN	4	0.261	0.161	0.216	0.152
	RF	4	0.554	0.125	0.469	0.125
一阶微分+植被指数+原始波段 First derivative data + Vegetation index data+ Raw band dat	SVM	6	0.660	0.107	0.530	0.117
	BPNN	6	0.479	0.135	0.344	0.139
	RF	6	0.651	0.110	0.640	0.103

2.4 基于粒子群优化的生物量估测模型构建

不同机器学习方法构建的花生生物量估算模型精度存在显著差异，这可能是由于模型内部的超参数对模型性能的影响较大。本研究利用粒子群优化算法（PSO）对三种机器学习模型的超参数进行优化，使得这些模型能够在全局范围内搜索到最优解，从而提高模型的精度。本文在表 5 所示的特征组合基础上，采用优化后的机器学习模型（PSO-SVM、PSO-BPNN 和 PSO-RF）进行实验，并将训练集和测试集的性能表现列于表 5。通过对比分析优化前后模型的性能变化，评估 PSO 优化对花生生物量估测精度的提升效果。

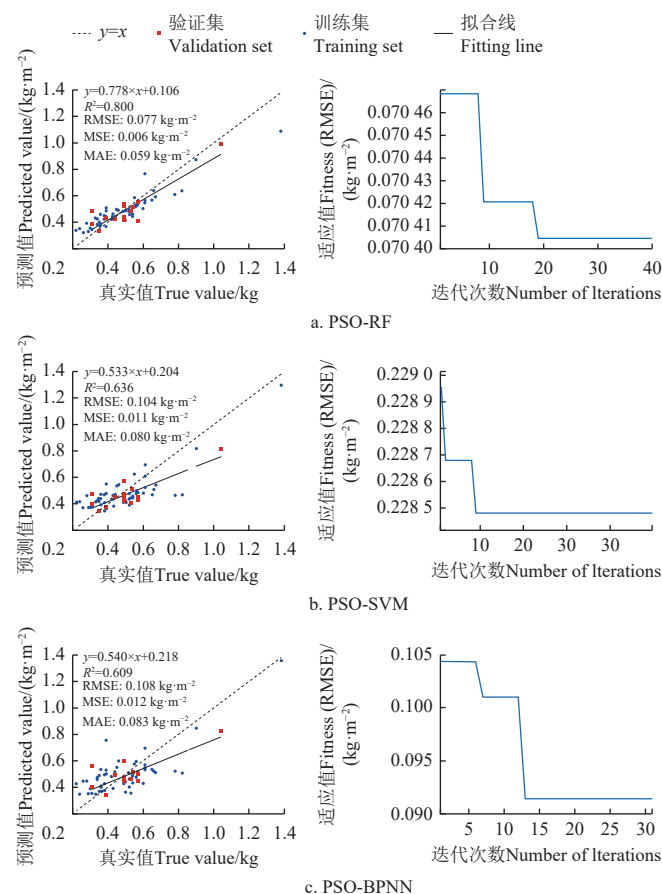
结果显示，经过粒子群算法（PSO）优化后的三种机器学习方法的建模精度均有不同程度提高，模型的训练集、测试集 R^2 分别提升了 14.1% 和 13.5%。特别是基于一阶微分、植被指数和原始波段 3 种特征组合的 PSO 优化随机森林（PSO-RF）模型，训练集和测试集的 R^2 分别提升了 27% 和 15%，3 种特征组合后的训练集 R^2 值最高达到 0.830。因此，PSO 优化有助于提高模型精度，这与王晗等^[41]的研究结果类似。在测试集中的性能表现上，以一阶微分和植被指数组合作为特征输入的 PSO-RF 模型表现最佳，其训练集和测试集的 R^2 值分别为 0.810 和 0.800，均方根误差（RMSE）分别为 0.081 和 0.076 kg/m²。

表 5 基于粒子群优化机器学习方法的花生生物量估测模型精度对比
Table 5 Comparison of peanut biomass prediction models using particle swarm optimization-based machine learning methods

数据源 Data	建模方法 Modeling method	变量个数 Number of variables	训练集 R^2 Training set R^2	训练集 RMSE Training set RMSE/ (kg·m ⁻²)	测试集 R_1^2 Testing set R_1^2	测试集 RMSE ₁ Testing set RMSE ₁ / (kg·m ⁻²)
一阶微分+植被指数 First derivative data + Vegetation index data	PSO-SVM	4	0.637	0.112	0.636	0.103
	PSO-BPNN	4	0.590	0.119	0.536	0.117
	PSO-RF	4	0.810	0.081	0.800	0.076
一阶微分+原始波段 First derivative data + Raw band data	PSO-SVM	4	0.621	0.115	0.601	0.108
	PSO-BPNN	4	0.540	0.127	0.414	0.131
	PSO-RF	4	0.815	0.080	0.709	0.092
植被指数+原始波段 Vegetation index data + Raw band data	PSO-SVM	4	0.792	0.085	0.568	0.113
	PSO-BPNN	4	0.285	0.158	0.270	0.146
	PSO-RF	4	0.818	0.079	0.499	0.121
一阶微分+植被指数+原始波段 First derivative data + Vegetation index data+ Raw band dat	PSO-SVM	6	0.662	0.108	0.564	0.113
	PSO-BPNN	6	0.556	0.124	0.504	0.121
	PSO-RF	6	0.830	0.077	0.742	0.087

通过比较三种机器学习模型的性能可以发现, 随机森林模型在 PSO 优化前后均表现最优。PSO 在提升机器学习模型精度方面的有效性, 尤其是在处理复杂数据特征时具有明显优势。

图 6 展示了优化后的 3 种模型在测试集上的回归拟合表现, 其中, PSO-RF 模型的预测精度最高, 在 RMSE、MSE 和 MAE 等指标上均为最低。具体而言, PSO-RF 模型的 R^2 、MSE、MAE 分别为 0.800、0.006 kg/m² 和 0.059 kg/m², 迭代优化后的决策树数量 (ntree) 为 105, 最小叶子节点数量 (leaf) 为 2 时, 适应度指标值 RMSE 达到最低, 为 0.077 kg/m²。PSO-SVM 模型, 其 R^2 、MSE、MAE 分别为 0.636、0.011 kg/m² 和 0.081 kg/m², 迭代优化后的惩罚系数值为 350, 径向基函数参数值为 1.0 时, 适应度指标值 RMSE 最低, 为 0.106。PSO 对三种算法进行超参数优化时, 最大迭代次数不超过 20, 表明算法具有较高的时效性和收敛效率。未经过 PSO 优化的模型在各项指标上表现相对较差, 经过 PSO 优化后可以显著提升模型的精度和稳定性。



注: a、b、c 中左图为回归拟合图, 右图是 PSO 优化模型的寻优过程
Note: In Fig. a, b, and c, the left panels show the regression fitting plots, while the right panels depict the optimization process of the PSO model.

图 6 一阶微分数据与植被指数数据组合基于 PSO-RF、PSO-SVM、PSO-BPNN 模型对花生生物量预测模型
Fig.6 Peanut biomass prediction models based on the combination of first-order derivative data and vegetation indices using PSO-RF, PSO-SVM, and PSO-BPNN models

本试验利用 PSO-RF 模型估算花生生物量的精度最

高, 在训练集和测试集上的 R^2 最高分别为 0.830 和 0.800, 这与周敏姑等^[42]的研究结果精度相当。尽管如此, 利用无人机高光谱影像估算花生植株生物量仍面临环境因素变化、影像分辨率低、模型特征冗余等问题, 仍需进一步探索不同条件、特征组合对生物量估算模型精度的影响。

3 结论

本文通过利用无人机高光谱遥感技术, 结合特征筛选和优化算法, 构建了河南花生试验田生物量估测模型。具体研究方法和结论如下:

1) 提取高光谱影像中的光谱反射率数据, 构建反射率一阶微分和植被指数作为特征输入, 并利用变量投影重要性 (VIP) 方法筛选特征波段, 构建了基于原始波段、一阶微分、植被指数和组合特征的 RF、SVM 和 BPNN 模型。试验发现, 有效特征的筛选与特征组合对于提高花生生物量估算的准确性至关重要。

2) 使用粒子群优化 (PSO) 算法对三种模型进行参数优化, 设定均方根误差 RMSE 为优化适应度指标; 相比其他两种模型, RF 模型在决策树数量 (ntree) 为 105、最小叶子节点数量 (leaf) 为 2 时的 RMSE 值为 0.077 kg/m²。

3) 使用 PSO 优化的 RF 模型在花生生物量估算中表现最佳, R^2 、RMSE、MSE、MAE 分别为 0.800、0.076 kg/m²、0.005 kg/m²、0.059 kg/m², 这一模型的应用为基于无人机高光谱遥感的花生生物量估算提供了理论支持和技术参考。

本研究仍存在一定的局限。仅使用了单个年份的数据进行建模, 无法完全涵盖不同气候条件、病虫害影响下的生长特征, 未来研究将进一步结合多源遥感数据, 与作物产量、长势参数及病虫害等级等关键农业指标进行综合分析。进一步扩展样本数据的覆盖范围, 探索建模在多气候条件和不同农作物生长阶段的适用性。

[参 考 文 献]

- [1] 廖伯寿. 我国花生生产发展现状与潜力分析[J]. 中国油料作物学报, 2020, 42(2): 161.
 - [2] 闫琰, 王秀东, 王济民, 等. “双循环”背景下国家粮食安全战略研究[J]. 中国工程科学, 2023, 25(4).
 - [3] QI H, WU Z, ZHANG L, et al. Monitoring of peanut leaves chlorophyll content based on drone-based multispectral image feature extraction[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106292.
 - [4] 曹中盛, 李艳大, 黄俊宝, 等. 监测花生叶面积指数和地上部生物量的最优植被指数及适宜波段带宽[J]. 中国油料作物学报, 2022, 44(6): 1320-1328.
- CAO Zhongsheng, LI Yanda, HUANG Junbao, et al. Sensitive

- vegetation indices and optimal bandwidths for monitoring peanut LAI and AGB[J]. *Journal of the Chinese Oil Crops*, 2022, 44(6): 1320-1328. (in Chinese with English abstract)
- [5] SHU M, LI Q, GHAFOR A, et al. Using the plant height and canopy coverage to estimation maize aboveground biomass with UAV digital images[J]. *European Journal of Agronomy*, 2023, 151: 126957.
- [6] 曾斌, 唐敏, 唐伟, 等. 花生粕和花生秸秆在动物饲用中的研究进展[J]. *中国饲料*, 2023(5): 152-162.
ZENG Bin, TANG Min, TANG Wei, et al. Research progress of peanut meal and peanut straw in animal feeding[J]. *China Feed*, 2023(5): 152-162. (in Chinese with English abstract)
- [7] ZHAO F T, WANG F, ZHANG F, et al. Catalytic hydrolysis of crop straws with different biochemical composition[J]. *International Journal of Hydrogen Energy*, 2023, 48(19): 6927-6936.
- [8] 冯惠芬, 李映雪, 吴芳, 等. 机器学习结合高光谱植被指数与 SPAD 值估算冬小麦氮含量[J]. *农业工程学报*, 2024, 40(1): 227-237.
FENG Huifen, LI Yingxue, WU Fang, et al. Estimating winter wheat nitrogen content using SPAD and hyperspectral vegetation indices with machine learning[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(1): 227-237. (in Chinese with English abstract)
- [9] 刘一博, 裴杰, 方华军, 等. 利用无人机影像反演水稻 SPAD 值的最优空间窗口确定[J]. *农业工程学报*, 2023, 39(19): 165-174.
LIU Yibo, PEI Jie, FANG Huajun, et al. Optimizing spatial window selection for rice SPAD value retrieval using multispectral UAV images[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(19): 165-174. (in Chinese with English abstract).
- [10] 刘修宇, 张锦水, 吴俊旭, 等. 时-空-谱融合自动纯化生成样本的农作物遥感分类[J]. *农业工程学报*, 2024, 40(10): 157-167.
LIU Xiuyu, ZHANG Jinshui, WU Junxu, et al. Automatic purifying and generating graded samples for crop remote sensing classification using temporal-spatial-spectral fusion[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(10): 157-167. (in Chinese with English abstract)
- [11] 兰仕浩, 李映雪, 吴芳, 等. 基于卫星光谱尺度反射率的冬小麦生物量估算[J]. *农业工程学报*, 2022, 38(24): 118-128.
LAN Shihao, LI Yingxue, WU Fang, et al. Winter wheat biomass estimation based on satellite spectral-scale reflectance[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(24): 118-128. (in Chinese with English abstract)
- [12] 赵越, 徐大伟, 范凯凯, 等. Landsat 8 和机器学习估算蒙古高原草地地上生物量[J]. *农业工程学报*, 2022, 38(24): 138-144.
ZHAO Yue, XU Dawei, FAN Kaikai, et al. Estimating above-ground biomass in grassland using Landsat 8 and machine learning in Mongolian plateau[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(24): 138-144. (in Chinese with English abstract)
- [13] 刘涛, 张寰, 王志业, 等. 利用无人机多光谱估算小麦叶面积指数和叶绿素含量[J]. *农业工程学报*, 2021, 37(19): 65-72.
LIU Tao, ZHANG Huan, WANG Zhiye, et al. Estimation of the leaf area index and chlorophyll content of wheat using UAV multi-spectrum images[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(19): 65-72. (in Chinese with English abstract)
- [14] 王志业, 刘涛, 张寰, 等. 基于多光谱影像特征的小麦锈病反演模型构建[J]. *河南农业大学学报*, 2023, 57(5): 826-834.
WANG Zhiye, LIU Tao, ZHANG Huan, et al. Construction of wheat rust disease inversion model based on multi-spectral image features[J]. *Journal of Henan Agricultural University*, 2023, 57(5): 826-834. (in Chinese with English abstract)
- [15] 苏宝峰, 刘砥柱, 陈启帆, 等. 基于时间序列植被指数的小麦条锈病抗性等级鉴定方法[J]. *农业工程学报*, 2024, 40(4): 155-165.
SU Baofeng, LIU Dizhu, CHEN Qifan, et al. Method for the identification of wheat stripe rust resistance grade using time series vegetation index[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(4): 155-165. (in Chinese with English abstract).
- [16] 许鑫, 张力, 岳继博, 等. 农田环境下无人机图像并行拼接识别算法[J]. *农业工程学报*, 2024, 40(9): 154-163.
XU Xin, ZHANG Li, YUE Jibo, et al. Parallel mosaic recognition algorithm for UAV images in farmland environment[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(9): 154-163. (in Chinese with English abstract)
- [17] ZHANG Z, ZHU L. A review on unmanned aerial vehicle remote sensing: Platforms, sensors, data processing methods, and applications[J]. *Drones*, 2023, 7(6): 398.
- [18] LI W, WEISS M, JAY S, et al. Daily monitoring of effective green area index and vegetation Chlorophyll content from continuous acquisitions of a multi-band spectrometer over winter wheat[J]. *Remote Sensing of Environment*, 2024, 300: 113883.
- [19] 帅爽, 张志, 张天, 等. 特征优化结合随机森林算法的干旱区植被高光谱遥感分类方法[J]. *农业工程学报*, 2023, 39(9): 287-293.
SHUAI Shuang, ZHANG Zhi, ZHANG Tian, et al. Hyperspectral image classification method for dryland vegetation by combining feature optimization and random forest algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(9): 287-293. (in Chinese with English abstract)
- [20] ABURAED N, ALKHATIB M Q, MARSHALL S, et al. A review of spatial enhancement of hyperspectral remote sensing imaging techniques[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 2275-2300.
- [21] 唐子竣, 张威, 黄向阳, 等. 基于地面高光谱遥感的大豆产量估算模型研究[J]. *农业机械学报*, 2024, 55(1): 145-153, 240.
TANG Zijun, ZHANG Wei, HUANG Xiangyang, et al. Soybean Seed Yield Estimation Model Based on Ground

- Hyperspectral Remote Sensing Technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 145-153,240. (in Chinese with English abstract)
- [22] 齐浩, 吕亮杰, 孙海芳, 等. 基于无人机高光谱与机器学习的小麦品系产量估测研究[J/OL]. 农业机械学报, [2024-06-03]. <http://kns.cnki.net/kcms/detail/11.1964.S.20240520.1707.002.html>.
- QI Hao, LV Liangjie, SUN Haifang, et al. Yield estimation of wheat lines based on UAV hyperspectral and machine learning[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, [2024-06-03]. <http://kns.cnki.net/kcms/detail/11.1964.S.20240520.1707.002.html>. (in Chinese with English abstract)
- [23] FENG H, TAO H, FAN Y, et al. Comparison of winter wheat yield estimation based on near-surface hyperspectral and UAV hyperspectral remote sensing data[J]. *Remote Sensing*, 2022, 14(17): 4158.
- [24] CHEN X, LI F, SHI B, et al. Estimation of winter wheat plant nitrogen concentration from UAV hyperspectral remote sensing combined with machine learning methods[J]. *Remote Sensing*, 2023, 15(11): 2831.
- [25] XU T, WANG F, SHI Z, et al. Dynamic estimation of rice aboveground biomass based on spectral and spatial information extracted from hyperspectral remote sensing images at different combinations of growth stages[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 202: 169-183.
- [26] SAHOO R N, REJITH R G, GAKHAR S, et al. Drone remote sensing of wheat N using hyperspectral sensor and machine learning[J]. *Precision Agriculture*, 2024, 25(2): 704-728.
- [27] 江凯伦, 安吉庆, 赵雨薇, 等. 采用 RNCA-PSO-ELM 的水稻叶绿素光谱特征分析与反演[J]. 农业工程学报, 2022, 38(8): 178-186.
- JIANG Kailun, AN Jiqing, ZHAO Yuwei, et al. Analysis and inversion of rice chlorophyll spectral characteristics using RNCA-PSO-ELM[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(8): 178-186. (in Chinese with English abstract)
- [28] 梁文广, 吴勇锋, 石一凡, 等. 基于无人机高光谱的太浦河水质反演[J]. 测绘通报, 2024(4): 29-34.
- LIANG Wenguang, WU Yongfeng, SHI Yifan, et al. Inversion of Taipu River water quality parameters by UAV hyperspectral imaging technology[J]. Bulletin of Surveying and Mapping, 2024(4): 29-34. (in Chinese with English abstract)
- [29] 彭俊杰, 汪泓, 王宇, 等. 基于 GA-PLS-SPA 的辣椒叶片叶绿素含量高光谱估测[J]. 江苏农业科学, 2024, 52(7): 184-192.
- PENG Junjie, WANG Hong, WANG Yu, et al. Hyperspectral estimation of chlorophyll content in pepper leaves based on GA-PLS-SPA[J]. Jiangsu Agricultural Sciences, 2024, 52(7): 184-192. (in Chinese with English abstract)
- [30] MAHIEU B, QANNARI E M, JAILLAIS B. Extension and significance testing of Variable Importance in Projection (VIP) indices in Partial Least Squares regression and Principal Components Analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 2023, 242: 104986.
- [31] BELGIU M, DRĂGUT L. Random forest in remote sensing: A review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24-31.
- [32] WANG L, ZENG Y, CHEN T. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting[J]. *Expert Systems with Applications*, 2015, 42(2): 855-863.
- [33] MAVROFORAKIS M E, THEODORIDIS S. A geometric approach to support vector machine (SVM) classification[J]. *IEEE Transactions on Neural Networks*, 2006, 17(3): 671-682.
- [34] WANG D, TAN D, LIU L. Particle swarm optimization algorithm: an overview[J]. *Soft Computing*, 2018, 22: 387-408.
- [35] 师翊, 王应宽, 赵龙, 等. 基于 CART 重要度排序和混合 ELM 模型的蒸散预测[J]. 农业工程学报, 2023, 39(14): 89-96.
- SHI Yi, WANG Yingkuan, ZHAO Long, et al. Evapotranspiration prediction using CART importance ranking and hybrid ELM models[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(14): 89-96. (in Chinese with English abstract)
- [36] 严格齐, 赵婉莹, 于镇伟, 等. 基于超参数优化算法的随机森林模型预测奶牛呼吸频率[J]. 农业工程学报, 2024, 40(11): 195-203.
- YAN Geqi, ZHAO Wanying, YU Zhenwei, et al. Predicting respiratory rate of dairy cows using hyperparameter-optimized random forest models[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(11): 195-203. (in Chinese with English abstract)
- [37] CHEN P, HABOUDANE D, TREMBLAY N, et al. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat[J]. *Remote Sensing of Environment*, 2010, 114(9): 1987-1997.
- [38] 董恒, 孟庆野, 王金梁, 等. 一种改进的叶绿素提取植被指数[J]. 红外与毫米波学报, 2012, 31(4): 336-341.
- DONG Heng, MENG Qingye, WANG Jinliang, et al. A modified vegetation index for crop canopy chlorophyll content retrieval[J]. *Journal of Infrared and Millimeter Waves*, 2012, 31(4): 336-341. (in Chinese with English abstract)
- [39] 张凡凡, 贾子玉, 庄前友, 等. 基于高光谱遥感的针茅群落生物量估算[J/OL]. 草业科学, [2024-07-30]. <http://kns.cnki.net/kcms/detail/62.1069.s.20240723.1508.002.html>.
- ZHANG Fanfan, JIA Ziyu, ZHUANG Qianyou, et al. Estimation of Stipa biomass based on hyperspectral remote sensing[J]. *Pratacultural Science*, [2024-07-30]. <http://kns.cnki.net/kcms/detail/62.1069.s.20240723.1508.002.html>. (in Chinese with English abstract)
- [40] 赵玉清, 杨慧丽, 张悦, 等. 基于特征组合与 SVM 的小粒种咖啡缺陷生豆检测[J]. 农业工程学报, 2022, 38(14): 295-302.
- ZHAO Yuqing, YANG Huili, ZHANG Yue, et al. Detection of defective arabica green coffee beans based on feature combination and SVM[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(14): 295-302. (in Chinese with English abstract)
- [41] 王晗, 向友珍, 李汪洋, 等. 基于无人机多光谱遥感的冬油菜地上部生物量估算[J]. 农业机械学报, 2023, 54(8): 218-229.
- WANG Han, XIANG Youzhen, LI Wangyang, et al. Estimation of winter rapeseed above-ground biomass based on UAV multi-spectral remote sensing[J]. Transactions of the

Chinese Society for Agricultural Machinery, 2023, 54(8): 218-229. (in Chinese with English abstract)

- [42] 周敏姑, 闫云才, 高文, 等. 基于多光谱遥感和 CNN 的玉米地上生物量估算模型[J]. 农业机械学报, 2024, 55(9): 238-248.

ZHOU Mingu, YAN Yuncai, GAO Wen, et al. Estimating aboveground biomass of maize based on multispectral remote sensing and CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(9): 238-248. (in Chinese with English abstract).

Estimation of peanut biomass based on feature selection and particle swarm optimization

LIU Tao^{1,2}, YANG Fengyuan¹, LIU Wang¹, ZHANG Huan², YIN Dongmei³,
ZHANG Quanguo², JIAO Youzhou^{2,4*}

(1. School of Urban and Rural Planning, Henan University of Economics and Law, Zhengzhou 450046, China; 2. School of Mechanical and Electrical Engineering, Henan Agricultural University, Zhengzhou 450002, China; 3. College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China; 4. Mechanical Engineering College, Henan University of Engineering, Zhengzhou 451191, China)

Abstract: Peanut is one of the most widely cultivated oil crops globally, with China leading in both production and consumption. As the demand for oil crops increases, ensuring stable peanut production and oil supply security has become a key agricultural goal. Peanut biomass, as a crucial parameter reflecting crop growth status, is essential for precision agriculture management and efficient resource utilization. The aboveground parts of peanut plants can be used not only as animal feed but also as a resource for bioenergy production. Therefore, comprehensive and accurate biomass estimation provides valuable references for yield prediction and resource management. Traditional biomass measurement methods are often labor-intensive and time-consuming, with spatial and temporal limitations. Recently, with the development of UAV remote sensing, especially the widespread application of hyperspectral imaging technology, crop biomass estimation has become more efficient. Hyperspectral imaging, known for its high resolution and rich spectral information, has been used for growth monitoring and yield estimation of crops such as soybean, rice, and wheat, demonstrating superior performance in predicting parameters like yield, chlorophyll, and nitrogen content, as well as in disease diagnosis. However, research on peanut remains limited, particularly regarding the spectral characteristics of different peanut varieties and their impact on biomass estimation accuracy. This study, using UAV hyperspectral imaging, investigated sensitive spectral bands and feature combinations for efficient and accurate field-scale peanut biomass estimation. An experimental field with 11 peanut varieties in Xingyang, Henan, was used as the study area. First, UAV hyperspectral images of the test field were collected and preprocessed with radiometric calibration and atmospheric correction to ensure data accuracy. Spectral reflectance data from ground sampling points were then extracted, and the first derivative of spectral reflectance and multiple vegetation indices were calculated to enhance the feature dimensions related to biomass. The variable importance in projection (VIP) method was used to select sensitive spectral bands and feature combinations closely related to biomass, effectively eliminating data redundancy and isolating highly relevant features. Using the selected features and ground-truth data, Support Vector Regression (SVR), back propagation neural network (BPNN), and random forest regression (RFR) models were constructed, and the estimation accuracy of different machine learning models was compared. Additionally, selected sensitive features were combined in multiple ways and input into the models to further improve estimation accuracy. The particle swarm optimization (PSO) algorithm was employed to optimize model hyperparameters, achieving the best model performance. Results showed that the sensitive features derived from the first derivative of spectral reflectance were highly correlated with peanut biomass, yielding better model performance than those derived from raw spectral reflectance and individual vegetation indices. The RF model combining the first derivative of spectral reflectance and vegetation indices achieved the highest estimation accuracy ($R^2 = 0.75$, RMSE = 0.08). Further improvement was achieved with the PSO-optimized RF model (PSO-RF), which resulted in an accuracy of $R^2 = 0.80$ and RMSE = 0.07. This study demonstrates the potential of combining UAV hyperspectral imaging with machine learning models for non-destructive peanut biomass estimation, providing essential theoretical and technical support for large-scale agricultural biomass monitoring. This study provides an effective method for accurate peanut biomass estimation and offers technical support for precision farmland management in the construction of smart villages.

Keywords: peanut; biomass; smart village; feature selection; machine learning; particle swarm optimization