

引用格式: Jiang Yufeng, Qi Jianguo, Chen Bowei, *et al.* Classification of Mangrove Species with UAV Hyperspectral Imagery and Machine Learning Methods [J]. Remote Sensing Technology and Application, 2021, 36(6): 1416-1424. [姜玉峰, 齐建国, 陈博伟, 等. 基于无人机高光谱影像和机器学习的红树林树种精细分类[J]. 遥感技术与应用, 2021, 36(6): 1416-1424.]

DOI: 10.11873/j.issn.1004-0323.2021.6.1416

基于无人机高光谱影像和机器学习的红树林 树种精细分类

姜玉峰^{1,2}, 齐建国¹, 陈博伟², 闫敏², 黄龙吉³, 张丽²

(1. 山东农业大学 信息科学与工程学院测绘系, 山东 泰安 271018;

2. 中国科学院空天信息创新研究院, 数字地球重点实验室, 北京 100090;

3. 海南东寨港国家级自然保护区管理局, 海南 海口 571129)

摘要: 利用海南省文昌市清澜港红树林保护区的无人机高光谱影像, 采用递归特征消除的随机森林算法 (Recursive Feature Elimination-Random Forest, RFE-RF) 优选植被光谱特征和纹理特征, 通过机器学习中的随机森林 (Random Forest, RF) 和支持向量机 (Support Vector Machine, SVM) 算法对研究区内的红树林树种进行精细分类, 并对对比分析和评价分类模型参数设置对总体精度的影响。结果表明: RF 分类方法的总体精度为 92.70%、Kappa 系数为 0.91, 与传统的 SVM 分类方法相比, RF 算法均提高了 5 类树种的生产者精度和使用精度, 能够有效地对红树林树种进行精细分类, 可为种植资源规划和生态环境保护等方面提供技术支持。

关键词: 机器学习; 随机森林; 高光谱; 特征提取; 精细分类

中图分类号: P23; TP79 **文献标志码:** A **文章编号:** 1004-0323(2021)06-1416-09

1 引言

红树林是生长在热带、亚热带海岸潮间带的以红树植物为主体的常绿灌木或乔木组成的潮滩湿地木本生物群落, 具有抗洪抗潮、防止海岸侵蚀、过滤污染物、净化海水、改善生态环境等多重功能^[1-3], 被誉为“天然海岸卫士”、“消浪先锋”和“生物净化筛”。2000 年, 全球红树林总面积估算为 1 377 万 hm^2 , 占全世界热带森林面积的 0.7%, 分布于 118 个国家和地区^[4]。红树林作为敏感的生态环境指示物, 对其进行监测分析有助于理解海岸带沿线生态环境变化进程, 为红树林的资源规划调查和生态环境改善发展部署提供理论依据。

传统的植被分类多采用监督分类和非监督分类的方法, 由于植被“同物异谱”和“同谱异物”的难

题, 导致植被分类的层次仅停留在林地、草地和耕地之分。随着遥感影像质量、多源遥感技术和信息技术的不断发展, 遥感分类的瓶颈才得以突破, 分类精度逐步提高, 分类的精细化也渐渐成为研究的重点方向之一。近年来, 很多学者将机器学习算法和无人机技术应用到树种精细分类研究中。刘怀鹏等^[5-6]在对种间分类方法对比研究中发现传统的最大似然分类方法在处理高维数据时容易发生休斯现象, 相比于机器学习算法分类精度较低。肖伟山等^[7]利用 ALOS PLASAR 数据 L 波段的特性提取漳江口的红树林。李想等^[8]利用资源三号卫星影像对北部湾东部地区的红树林精细分类, 星载影像信息匮乏带来的局限性限制了种间分类的精细化程度以及分类精度的提高。冯家莉等^[9]基于无人机

收稿日期: 2020-10-27; 修订日期: 2021-11-20

基金项目: 中国科学院战略性先导科技专项(A类)(XDA13020506), 国家自然科学基金项目(41771392)资助。

作者简介: 姜玉峰(1995-), 男, 山东威海人, 硕士研究生, 主要从事植被遥感研究。E-mail: jyfl1098060049@163.com

通讯作者: 张丽(1975-), 女, 新疆伊犁人, 研究员, 主要从事植被生态与海岸带遥感研究。E-mail: zhangli@aircas.ac.cn

可见光影像,采用面向对象的最近邻法识别红树林树种类型,总体精度高达91.9%。崔小芳等^[10]基于机载高光谱数据对大丰市沿海滩涂湿地区的盐蒿等植被进行精细分类,采用随机森林的分类方法精度高达90.3%。研究发现,无人机遥感凭借携带方便、成本低、受环境影响小等优势广泛应用于树种识别的研究中,机器学习作为遥感界研究的热点^[11],凭借着能对空间结构等细节信息有效挖掘与利用的优势,成为现阶段树种分类领域的主流算法^[12]。总体来说,高分辨率多光谱数据与机器学习分类算法的结合应用在种间分类方面取得较好进展,但将信息量丰富的高光谱数据应用到红树林种间分类的研究却很少。而高光谱无人机遥感影像分辨率高,波段窄、地物光谱信息丰富的优势,能够为植被精细分类提供更好的数据基础^[13]。实验以无人机高光谱影像为数据源,对比分析多个机器学习算法,对海南省文昌市清澜港八门湾头苑村红树林树种进行植被精细分类,探讨机器学习在红树林种间分类的潜力。

2 研究区与数据

2.1 研究区概况

研究区位于海南省文昌市清澜港八门湾头苑村红树林保护区南部,地理位置为 $110^{\circ}47'22.70''\sim$

$35.65''\text{E}$, $19^{\circ}37'33.46''\sim 39.65''\text{N}$ 。该区域属热带季风岛屿型气候,年平均气温 23.9°C ,雨量丰富,土壤干湿分明,地势平坦。八门湾红树林保护区属于省级自然保护区,有“海上森林公园”之称,是世界上海拔最低的森林,也是我国红树林品种较为丰富的地区。

实验区内的主要红树树种有正红树、海莲、黄槿和海桑,它们和椰树、鱼藤等植物杂交生长,符合海南红树林群落树种分布的典型特征,研究范围如图1所示。

2.2 数据与预处理

实验选取的高光谱影像数据,成像时间是2020年1月4日,数据获取当天无(少)云,植被生长处于茂盛时期。大疆无人机M600Pro飞行器搭载Rikola光谱仪CMOS传感器以推扫式获取光谱数据,航高为120 m,飞行速度为4.5 m/s,共45个波段,其中1~31波段是可见光波段(Visible),32~45波段是近红外波段(Near-Infrared),主要参数如表1所示。

根据Rikola高光谱遥感数据的特点,在图像质量检查的基础上对高光谱影像进行了预处理(图2),主要包括:辐射定标、暗电流校正、镜头渐晕校正、波段配准、大气校正和波谱降噪处理,获得了较好质量的图像,明显地改善了部分优势树种光谱特征曲线高度相似的问题,为后期样本选择和分类处理提供了有力保障。

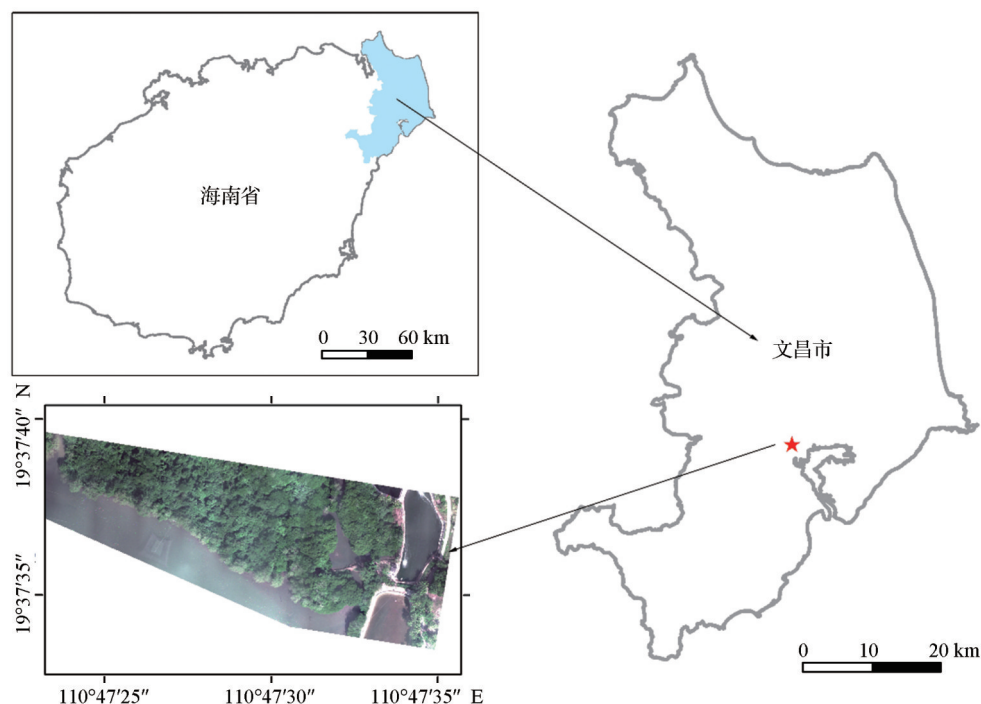


图1 研究区位置及无人机高光谱影像

Fig.1 Study area and the UAV hyperspectral image

审图号:GS(2021)5448

表1 Rikola 高光谱数据主要参数

Table 1 Key parameters of Rikola Hyperspectral data

参数名	参数值
光谱范围	500~900 nm
空间分辨率	0.075 m
波段	45 个
光谱分辨率	10 nm
光谱采样间隔	9 nm
视场角	36.5°
焦距	9 mm
量化值	12 bit

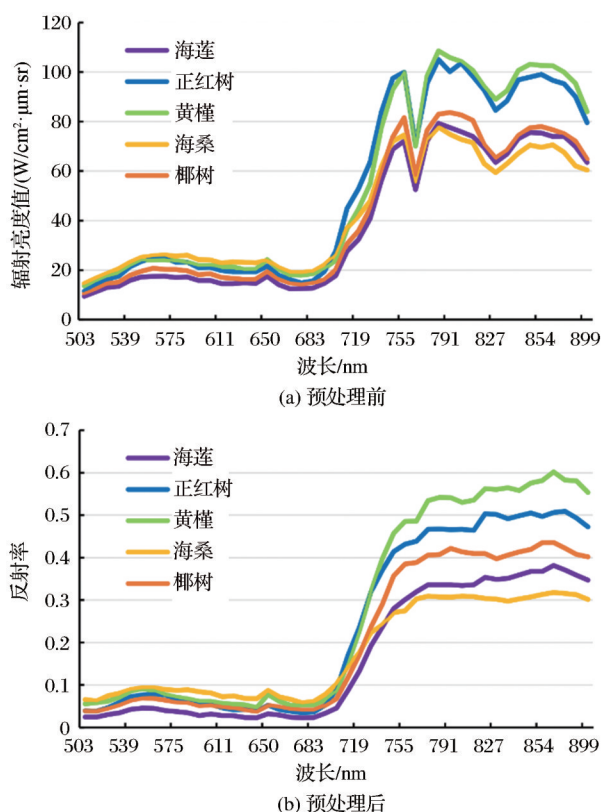


图2 优势树种预处理前后光谱曲线

Fig.2 Spectral curves before and after pretreatment of dominant tree species

同期开展了现场红树林树种调查实验,首先对前期2018年采集的无人机影像进行目视解译,标注并勾画疑似树种分布区块,并以此为基础校验图,在2020年1月3日通过现场调查方式了解红树林冠层植被类型信息,辅助以GPS定位技术,标注研究区域内的树种分布情况,确定最终树种分布信息,为分类时选取训练样本和验证样本提供参考。

3 研究方法

3.1 特征变量提取

高光谱数据的维数灾难给图像分类方面带来

了严峻挑战,维数约简是实现数据降维的关键技术,分为特征提取和特征选择。图像特征提取是图像分类的重要前提,特征提取的表现决定着图像分类的效果^[14]。特征选择是在原始特征集和提取特征集中选择出最能解释类别变量的子集,从而降低特征空间的维数^[15]。

3.1.1 光谱特征变量

近年来,随着图像空间分辨率高而光谱分辨率低的壁垒被击破,光谱特征提取成为计算机视觉领域研究热点。光谱的曲线特征、运算与变换特征、相似性度量特征等逐渐被开发出来,人们常常把光谱特征中的植被指数作为遥感图像分类的主要特征。

植被指数是通过不同波段数值间的数学变化得来,既能增强植被信息,又能放大植被类间差异,是高光谱植被精细分类所需的必要特征变量。根据影像45个波段数据计算了MNF变换后的归一化植被指数(Normalized Difference Vegetation Index—NDVI)、比值植被指数(Simple Ratio Index—SR)、绿度总和指数(Sum Green Index—SG)、红边归一化植被指数(Red Edge Normalized Difference Vegetation Index—NDVI705)、Vogelmann红边指数(Vogelmann Red Edge Index—VOG)、红边位置指数(Red Edge Position Index—REP)、光化学植被指数(Photochemical Reflectance Index—PRI)、红绿比值指数(Red Green Ratio Index—RG)、植被衰减指数(Plant Senescence Reflectance Index—PSRI)、类胡萝卜素反射指数(Carotenoid Reflectance Index—CRI)和花青素反射指数(Anthocyanin Reflectance Index—ARI)11种经典植被指数^[16-18]。

3.1.2 纹理特征变量

传统分类中仅利用影像的光谱信息,造成了空间信息的冗余浪费,纹理信息的加入能够充分体现无人机高分辨率影像的优势,极大地提高分类精度。

灰度共生矩阵法是一种常用的纹理分析方法,它是图像中2个像素灰度级联合分布的统计形式,能较好地反映纹理灰度级相关性的规律。基于影像灰度共生矩阵获取了中值(Mean)、协方差(Variance)、同质性/逆差距(Homogeneity)、二阶矩(Angular second moment)、反差/对比度(Contrast)、差异性(Dissimilarity)、熵(Entropy)和自相关(Correlation)8个纹理算子,其中移动窗口为3×3和5×5,

移动步长为1,移动方向为0°。另外,实验尝试将纹理算子45个波段的均值作为纹理特征变量参与到分类中,取得了较好的效果。

3.2 特征变量选择与降维

特征选择技术(变量选择或子集选择),可以去掉不相关的变量以达到分类模型精度提高和规避过拟合的作用。变量选择时面对的一个经典问题就是如何克服不稳定性,避免小样本扰动而完全改变优选的特征子集。实验采用的递归特征消除的随机森林算法(RFE-RF)把需要的特征集合初始化为整个数据集,从全集出发进行后向序列选择,每次消去一个排序准则得分最低的特征,直到优选出对分类结果最重要的特征子集,可以有效地提高变量选择的稳定性^[19]。

3.3 分类方法与验证

基于RFE-RF进行特征选取确定的10个变量,采用随机森林和支持向量机的算法进行分类建模,利用R/RStudio语言提供的Caret包进行十折交叉验证和参数调优,随机选取样本数据中的70%作为训练数据,30%作为验证数据,并采用总体精度、Kappa系数和混淆矩阵对分类结果进行精度评价分析。

3.3.1 随机森林算法

随机森林(Random Forest, RF)算法是由Breiman等于2001年提出的,是由多棵独立决策树组成,对随机抽取的样本信息进行类别预测的无参数分类器^[20],其本质是一种由机器学习演变的集成算法。在构建决策树时,通过Bootstrap抽样方法从训练样本集中提取数据,每颗决策树是众多二叉树组成,二叉树从根节点自上至下以递归分裂的方式对训练样本集按照节点纯度最小的原则进行二分类,直到满足分裂停止的规则为止^[21]。每棵决策树都是一个分类器,众多独立决策树集成随机森林算法模型,随机森林对每颗决策树的分类结果进行投票,决定最终分类结果。随机森林算法具有对参数不敏感、不易过拟合、训练速度快的优点,在种间分类问题上表现良好。

随机森林算法在分类建模过程中需要调整的参数有节点值mtry和决策树数目ntree,实验将ntree设置为1 000,通过调用R/RStudio语言中randomForest包中的tuneRF函数得到最优的mtry为6。

3.3.2 支持向量机算法

支持向量机(Support Vector Machine, SVM)是

一种基于统计学习理论的二分类模型,早在1964年由Vapnik等提出,到20世纪90年代受到了广泛关注,其核心内容得以迅速地发展^[22]。SVM算法的学习策略就是在特征空间中寻找最优超平面使得类内间隔最小化和类间间隔最大化,可以形式化理解为求解线性约束的凸二次规划问题^[23]。随着用户需求的不断升级,SVM从只适用于二分类问题,逐步推广到多分类方面。其中较为常用的是通过某种方式构造多个二分类器并将它们组合到一起,从而实现多类分类^[24-25]。由于该算法具有较强的鲁棒性和在高维数据处理方面出众的泛化能力,在高光谱领域得到了广泛的应用^[26]。在进行分类实验时,将光谱维和空间维特征进行叠加后优选,能明显地提高SVM算法的分类精度。

支持向量机算法在分类建模过程中需要调整的参数主要有核函数、及其系数gamma和惩罚因子cost。实验采用了支持向量机算法中常用的3种核函数:线性核(Linear Kernel)、多项式核(Polynomial Kernel)和径向基核函数(Radial Basis Function, RBF)。通过调用R/RStudio语言中e1071包中的tune函数,采用网格搜索法平行高效地得到多项式核的最优cost和gamma分别为100和0.063,RBF的最优cost和gamma分别为100和11。

3.3.3 精度验证

验证样本以2020年实地调查的树种分布区块分布图作为参考进行选取,为均衡各地类的精度分量对总体精度的影响,每种地类各选取了1 500个左右的样本点,随机选取30%用作精度验证。囿于研究区东北部地形复杂,无法通达的限制因素,导致样本分布均一性欠佳^[27],但已有样本涵盖了研究区内所有优势树种。通过计算红树林树种精细分类结果的总体精度、Kappa系数以及混淆矩阵对分类结果进行描述与分析并行式精度评价,公式^[28]如下:

$$P_o = \frac{\sum_{i=1}^n P_{ii}}{N} \quad (1)$$

$$k = \frac{P_o - P_e}{(1 - P_e)} \quad (2)$$

其中: P_o 代表总体精度; n 为分类的类型数量; N 为样本像元总数; P_{ii} 为第*i*类型的被正确分类的样本数目; k 代表Kappa系数; P_e 代表由于偶然机会造成的分类结果与地面调查数据类型相一致的概率。

由于传统精度验证方法样本的点选取是基于目视解译判读的方式,会存在单木边缘和特征相似树种混交区域边缘误判的现象。为了保证分类精

度结果更加真实、可靠,实验还针对在红树林种间分类方面表现最优的 RF 算法的分类结果,以 2020 年 1 月 3 日至 7 日采集的 70 个 GPS 野外点位为验证数据进行独立的精度验证,采用混淆矩阵的方法进行精度评价。

4 结果分析与讨论

4.1 特征优选结果

实验采用 RFE-RF 的方法对 792 个特征变量组成的数据集进行特征优选,集合包括高光谱影像的 45 个光谱波段、11 个植被指数、各波段的 8 种纹理指数和每种纹理算子的波段均值。

通过多次迭代,每次剔除最不重要的特征变量,并重复这个过程,最终 22 个变量的表现情况如图 3 所示。可以看出当变量数为 16 时,总体精度为 86.02%,Kappa 系数为 0.84,分类效果达到最佳,当变量个数为 10 时,总体精度为 84.90%,Kappa 系数为 0.82,较变量数为 9 时,分类精度明显提升,且较变量数为 16 时,分类效果差距较小,考虑到特征变量冗余会造成精度降低和过拟合的问题,故选出集合中最重要的 10 个特征变量组成分类所需的特征模型,如表 2 所示。

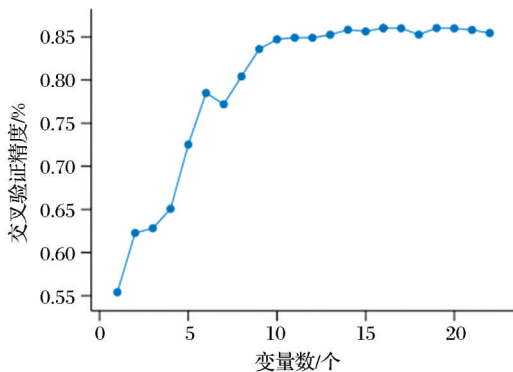


图 3 随机森林分类器精度与纳入变量数的关系

Fig.3 Relationship between random forest classifier precision and number of included variables

由表 2 可以看出,光谱特征中的植被指数特征对红树林树种精细分类的贡献要大于纹理特征,其中红边归一化植被指数 NDVI 705 和比值植被指数 SR 表现尤为突出。研究中提取的纹理特征较多,但贡献度相对较小,造成了严重的冗余现象,降低特征优选效率。8 种纹理特征算子中,Mean 值和 Entropy 值能有效提升红树林树种精细分类的精度,其他 6 种算子作用较小。

表 2 高光谱影像提取特征

Table 2 Hyperspectral image extraction features

特征类型	特征名称	描述或公式	重要度排序
光谱特征	NDVI ₇₀₅	$\frac{\rho_{750} - \rho_{705}}{\rho_{750} + \rho_{705}}$	1
	SR	$\frac{\rho_{NIR}}{\rho_{RED}}$	2
	NDVI	$\frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$	3
	VOG ₁	$\frac{\rho_{740}}{\rho_{720}}$	4
	CRI ₂	$\frac{1}{\rho_{510}} - \frac{1}{\rho_{700}}$	5
	CRI ₁	$\frac{1}{\rho_{510}} - \frac{1}{\rho_{550}}$	7
	RG	$\frac{\rho_{RED}}{\rho_{GREEN}}$	8
	PRI	$\frac{\rho_{531} - \rho_{570}}{\rho_{531} + \rho_{570}}$	9
	Mean_3*3_B29	$\frac{\sum_{i,j=0}^{N-1} P_{ij}}{N^2}$	6
纹理特征	Entropy_5*5_mean	$\frac{\sum_{B=1}^{45} \sum_{k=0}^{N-1} V_k (-\ln V_k)}{45}$	10

注:光谱特征公式中 ρ_{NIR} = 800 nm, ρ_{RED} = 680 nm, ρ_{RED} 代表红光波段均值, ρ_{GREEN} 代表绿光波段均值。纹理特征公式中的 i 是行号; j 是列号; P_{ij} 是单元格 i, j 中的归一化值; N 是行数或者列数; B 是波段数; V_k 是单元格 i, j 矩阵中的值, $k=1, 2, \dots, n$ 。

4.2 分类结果及精度评价

实验针对正红树、海莲、黄槿、海桑、椰子 5 种植被类型和不透水层、水体共 7 种地类,采用机器学习中随机森林和支持向量机的方法进行分类,并通过聚类处理得到了研究区的分类结果如图 4 所示。

从图 4 的分类效果对比中可以看出, SVM 方法的分类结果出现“椒盐”现象,各树种分布较散,不符合树木生长的空间分布规律,主要是因为 SVM 方法对 RFE-RF 优选的特征不适用造成的。SVM (Polynomial) 分类方法对海莲的可分性较差,主要错分为正红树和海莲,与它们的冠层形态相似有很大的关系。

由红树林分类结果精度评价(表 3)可以看出,在红树林树种精细分类方面, RF 算法的总体分类精度为 92.70%, 相比 SVM 最优(RBF)算法提高 20.21%, Kappa 系数提高 0.23, SVM 算法中基于 RBF 的分类精度最高, Linear Kernel 精度最低。与周在明等^[29]在漳江口开展的红树林种间分类研究结果相比,选用特征数更少,精度更高,更适合应用于多树种混交林区域。

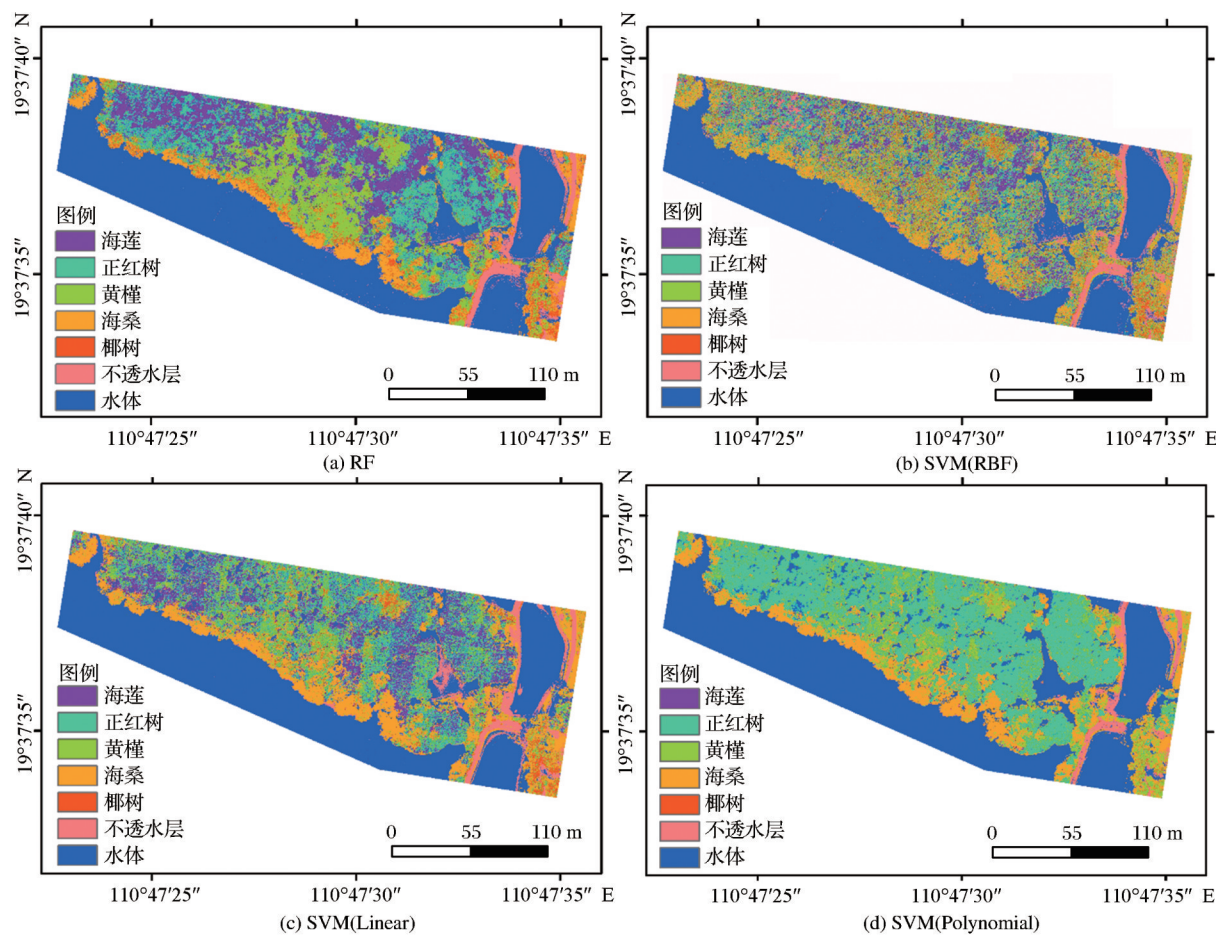


图 4 基于无人机高光谱影像的机器学习分类结果图 审图号:GS(2021)5448

Fig.4 Machine learning classification maps based on UAV Hyperspectral images

表 3 红树林分类结果精度评价

Table 3 Accuracy assessment of mangrove classification results												
分类方法	总体精度 /%	Kappa	海莲		正红树		黄槿		海桑		椰树	
			生产者	使用者	生产者	使用者	生产者	使用者	生产者	使用者	生产者	使用者
			精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%
RF	92.70	0.91	90.51	92.26	89.93	88.16	89.38	88.15	94.34	95.86	86.88	87.10
SVM(RBF)	72.49	0.68	67.30	68.75	52.80	57.01	55.89	57.35	68.33	60.89	66.58	64.51
SVM(Linear)	60.97	0.54	11.04	100	92.56	46.29	69.54	39.51	87.10	75.34	4.20	81.33
SVM(Polynomial)	63.55	0.57	0.35	100	91.79	46.10	60.99	39.47	86.76	75.13	17.47	63.34

由独立验证结果精度评价(表4)对比表3可以看出,RF算法的GPS真实点位独立验证结果与随机样本验证的结果具有高度一致性,总体精度降低了5.56%,Kappa系数低了0.06,变化幅度不大,5类树种的生产者精度和使用者精度大多有所下降,但与扩展样本验证的分类结果表现基本一致。

在单个树种分类精度方面,表现最好的仍然是RF算法,它在5类树种的生产者精度和使用者精度方面均达到86.88%以上,其中对椰树的可分性相对较差。综合来看,5类树种中海桑的可分性最佳,

呈现的空间位置分布规律符合海桑常生长于红树林与海水交界边缘的特性,椰树的可分性最差。

红树中分类精度较低的树种是正红树、黄槿和海莲,主要是因为藤蔓和三叶鱼藤会缠绕生长在海桑、海莲和正红树的树干和冠层,致使训练时不同红树树种的样本中会有少部分的藤蔓和三叶鱼藤的信息,混淆分类的训练和验证样本,干扰分类模型构建,导致分类精度降低。另外,在GPS定点调查标记红树物种信息时,局部视野有遮挡,且较难获得实时准确的全局视野信息,会导致误判遥感影

表 4 独立验证结果精度评价

Table 4 Accuracy assessment of independently verified results

类名	海莲	正红树	黄槿	海桑	椰树	水体	不透水层	行总计	使用者精度/%
海莲	13	2	0	0	0	0	0	15	86.67
正红树	1	16	0	0	0	0	0	17	94.12
黄槿	2	0	8	0	1	0	0	11	72.73
海桑	0	0	0	6	0	0	0	6	100.00
椰树	0	0	1	1	4	0	0	6	66.67
水体	0	0	0	0	0	8	0	8	100.00
不透水层	0	0	0	0	0	0	6	6	100.00
列总计	16	18	9	7	5	8	6	70	
生产者精度/%	81.25	88.89	88.89	85.71	80.00	100.00	100.00		
总体分类精度:87.14/%					Kappa 系数:0.85				

像垂直拍摄的部分点位的冠层红树林树种信息。研究区内红树林物种丰富,高度密集混交,不同树种高低错落生长分布,部分树高较高、叶子稀疏的红树下方,往往会生长着另一树种的红树,致使无人机拍摄的此类样本点的影像单个像元本身显示的是多个树种的混合信息,从而影响机器学习分类建模。红树林新叶、老叶和病叶的光谱信息差异等都会造成树种精细分类精度降低。

5 结 语

实验基于文昌市红树林保护区机载 Rikola 高光谱影像,采用递归特征消除的随机森林算法对特征变量进行优选,以优选特征为数据集,通过随机森林和支持向量机算法对其进行分类。RF 算法在红树林种间分类的总体精度为 92.70%,Kappa 系数为 0.91,相比于 SVM 算法更具有优势,分类精度高,分类速度快。利用 RFE-RF 算法对高光谱影像的光谱、空间特征进行优选,能够大幅度降低数据维度,提高分类精度,避免过拟合现象,该处理对 RF 算法精度有很大提高,但对 SVM 算法精度的提高并不明显。光谱特征中的植被指数特征对红树林树种精细分类的贡献要大于纹理特征,其中 NDVI 705 和 SR 贡献尤为突出。

实验对随机森林算法和支持向量机算法在小范围内高复杂性树种混交分布的红树林群落进行了初步研究,未来可考虑将该算法应用于更长时空尺度上,监测红树林树种空间位置分布变化;应用到更多数据源中,讨论该算法对不同分辨率数据、不同分类领域的适用性,做好从高到低下沉推广应用。红树林高度混交的生长习性给种间分类带来了不小的挑战,未来可以考虑不断积累样本数据,

建立红树林样本数据库,将深度学习分类算法应用到红树林树种精细分类的研究中。

参考文献 (References):

[1] Lin Peng. A review on the mangrove research in china [J]. Journal of Xiamen University (Natural Science Edition), 2001,40(2):592-603.[林鹏. 中国红树林研究进展[J]. 厦门大学学报(自然科学版),2001,40(2):592-603.]

[2] Wang Yan. The significance and proposal of mangrove protection[J]. Forest Engineering, 2008(4):19-20,57.[王燕. 红树林保护的意义的建议[J]. 森林工程,2008(4):19-20,57.]

[3] Marois D E, Mitsch W J. Coastal protection from tsunamis and cyclones provided by mangrove wetlands—a review[J]. International Journal of Biodiversity Science, Ecosystem Services & Management, 2015, 11(1):71-83. DOI: 10.1080/21513732.2014.997292.

[4] Giri C, Ochieng E, Tieszen L L, *et al.* Status and distribution of mangrove forests of the world using earth observation satellite data [J]. Global Ecology and Biogeography, 2011, 20(1):154-159. DOI: 10.1111/j.1466-8238.2010.00584.x.

[5] Liu Huaipeng. Typical urban greening tree species classification based on WorldView-2[D]. Huhhot: Inner Mongdio: Inner Mongolia Agricultural University, 2016.[刘怀鹏. 基于 WorldView-2 数据的城市典型绿化树种分类[D]. 呼和浩特: 内蒙古农业大学,2016.]

[6] Zhang Heng, Wang Shixin, Zhou Yi, *et al.* Comparison of different methods of mangrove extraction from multi-source remote sensing images[J]. Wetlang Science, 2015, 13(2):145-152.[章恒,王世新,周艺,等. 多源遥感影像红树林信息提取方法比较[J]. 湿地科学,2015,13(2):145-152.]

[7] Xiao Weishan, Wang Xiaoqin, Ling Feilong. The application of ALOS PALSAR data on mangrove forest extraction [J]. Remote Sensing Technology and Application, 2010, 25(1):91-96.[肖伟山,汪小钦,凌飞龙. ALOS PALSAR 数据在漳江口红树林提取中的应用[J]. 遥感技术与应用, 2010, 25(1):91-96.]

- [8] Li Xiang, Liu Kai, Zhu Yuanhui, *et al.* Study on mangrove species classification based on ZY-3 image[J]. *Remote Sensing Technology and Application*, 2018, 33(2): 360-369. [李想, 刘凯, 朱远辉, 等. 基于资源三号影像的红树林物种分类研究[J]. *遥感技术与应用*, 2018, 33(2): 360-369.]
- [9] Feng Jiali, Liu Kai, Zhu Yuanhui, *et al.* Application of unmanned aerial vehicles to mangrove resources monitoring[J]. *Tropical Geography*, 2015, 35(1): 35-42. [冯家莉, 刘凯, 朱远辉, 等. 无人机遥感在红树林资源调查中的应用[J]. *热带地理*, 2015, 35(1): 35-42.]
- [10] Cui Xiaofang, Liu Zhengjun. Wetland vegetation classification based on object-based classification method and multi-source remote sensing images[J]. *Geomatics & Spatial Information Technology*, 2018, 41(8): 113-116. [崔小芳, 刘正军. 基于随机森林分类方法和多源遥感数据的湿地植被精细分类[J]. *测绘与空间地理信息*, 2018, 41(8): 113-116.]
- [11] Zhang Run, Wang Yongbin. Research on machine learning with algorithm and development[J]. *Journal of Communication University of China (Science and Technology Edition)*, 2016, 23(2): 10-18. [张润, 王永滨. 机器学习及其算法和发展研究[J]. *中国传媒大学学报: 自然科学版*, 2016, 23(2): 10-18.]
- [12] Xing Jin. Comparative analysis of machine learning algorithms for land cover classification in high-resolution remote sensing images[J]. *Gansu Science and Technology*, 2020, 36(3): 27-34. [邢瑾. 机器学习算法在高分辨率遥感影像土地覆被分类中的对比分析[J]. *甘肃科技*, 2020, 36(3): 27-34.]
- [13] He Z, Shi Q, Liu K, *et al.* Object-oriented mangrove species classification using hyperspectral data and 3 D siamese residual network[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(12): 2150-2154. DOI:10.1109/LGRS.2019.2962723.
- [14] Zhai Junhai, Zhao Wenxiu, Wang Xizhao. Research on image feature extraction[J]. *Journal of Hebei University (Natural Science Edition)*, 2009, 29(1): 106-112. [翟俊海, 赵文秀, 王熙熙. 图像特征提取研究[J]. *河北大学学报(自然科学版)*, 2009, 29(1): 106-112.]
- [15] Yan Kesheng, Li Taifu, Wei Zhengyuan, *et al.* Feature selection for data classification based on pls supervised feature extraction and false nearest neighbors[J]. *Computer and Applied Chemistry*, 2012, 29(7): 817-821. [颜克胜, 李太福, 魏正元, 等. 融合PLS监督特征提取和虚假最近邻点的数据分类特征选择[J]. *计算机与应用化学*, 2012, 29(7): 817-821.]
- [16] Pan Chen, Du Peijun, Luo Yan, *et al.* Decision tree classification of remote sensing images based on vegetation indices[J]. *Computer Applications*, 2009, 29(3): 777-780. [潘琛, 杜培军, 罗艳, 等. 一种基于植被指数的遥感影像决策树分类方法[J]. *计算机应用*, 2009, 29(3): 777-780.]
- [17] Dezhi W, Bo W, Penghua Q, *et al.* Artificial mangrove species mapping using pléiades-1: an evaluation of pixel-based and object-based classifications with selected machine learning algorithms[J]. *Remote Sensing*, 2018, 10(2): 294. DOI: 10.3390/rs10020294.
- [18] Dong Di, Zeng Jisheng, Wei Zheng, *et al.* Integrating space-borne optical and SAR imagery for monitoring mangroves and *Spartina alterniflora* in Zhangjiang Estuary[J]. *Journal of Tropical Oceanography*, 2020, 39(2): 107-117. [董迪, 曾纪胜, 魏征, 等. 联合星载光学和SAR影像的漳江口红树林与互花米草遥感监测[J]. *热带海洋学报*, 2020, 39(2): 107-117.]
- [19] Wu Chenwen, Liang Jinghan, Wang Wei, *et al.* Random forest algorithm based on recursive feature elimination[J]. *Statistics and Decision Making*, 2017(21): 60-63. [吴辰文, 梁靖涵, 王伟, 等. 基于递归特征消除方法的随机森林算法[J]. *统计与决策*, 2017(21): 60-63.]
- [20] Liu Shu, Jiang Qigang, Ma Yue, *et al.* Object-oriented wetland classification based on hybrid feature selection method combining with relief F, multi-objective genetic algorithm and random forest[J]. *Transactions of The Chinese Society of Agricultural Machinery*, 2017, 48(1): 119-127. [刘舒, 姜琦刚, 马玥, 等. 基于多目标遗传随机森林特征选择的面向对象湿地分类[J]. *农业机械学报*, 2017, 48(1): 119-127.]
- [21] HeYuanhui, Wang Changlin, Jia Huicong, *et al.* Research on extraction of winter wheat based on random forest[J]. *Remote Sensing Technology and Application*, 2018, 33(6): 1132-1140. [贺原惠子, 王长林, 贾慧聪, 等. 基于随机森林算法的冬小麦提取研究[J]. *遥感技术与应用*, 2018, 33(6): 1132-1140.]
- [22] Zhang Xuegong. Interoduction to statistical learning theory and support vector machines[J]. *Acta Automatica Sinica*, 2000, 26(1): 36-46. [张学工. 关于统计学习理论与支持向量机[J]. *自动化学报*, 2000, 26(1): 36-46.]
- [23] Feng Guohe. Parameter optimizing for Support Vector Machines classification[J]. *Computer Engineering and Applications*, 2011, 47(3): 123-124, 128. [奉国和. SVM分类核函数及参数选择比较[J]. *计算机工程与应用*, 2011, 47(3): 123-124, 128.]
- [24] 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. [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.]
- [25] Gu Xiaotian, Gao Xiaohong, Ma Huijuan, *et al.* Comparison of machine learning methods for land use/land cover classification in the complicated terrain regions[J]. *Remote Sensing Technology and Applications*, 2019, 34(1): 57-67. [谷晓天, 高小红, 马慧娟, 等. 复杂地形区土地利用/土地覆被分类机器学习方法比较研究[J]. *遥感技术与应用*, 2019, 34(1): 57-67.]
- [26] Miao Yuhong, Yang Min, Wu Guojun. Sophisticated vegetation classification based on multi-dimensional features of hyperspectral image[J]. *Journal of Atmospheric and Environmental Optics*, 2020, 15(2): 117-124. [苗宇宏, 杨敏, 吴国俊. 基于高光谱影像多维特征的植被精细分类[J]. *大气与环境光学学报*, 2020, 15(2): 117-124.]
- [27] Ren Chong. Forest types precise classification and forest resources change monitoring based on medium and high spatial

- resolution remote sensing images[D]. Beijing: Chinese Academy of Forestry, 2016.[任冲.中高分辨率遥感影像森林类型精细分类与森林资源变化监测技术研究[D].北京:中国林业科学研究院,2016.]
- [28] Xu Wenning, Wang Pengxin, Han Ping, *et al.* Application of kappa coefficients to accuracy assessments of drought forecasting models: a case study of Guanzhong Plain[J]. Journal of Natural Disasters, 2011, 20(6): 81-86.[许文宁,王鹏新,韩萍,等.Kappa系数在干旱预测模型精度评价中的应用——以
- 关中原的干旱预测为例[J].自然灾害学报,2011,20(6): 81-86.]
- [29] Zhou Zaiming, Chen Benqing, Xu Ran, *et al.* Identification of the mangrove species using UAV hyperspectral images: a case study of Zhangjiangkou mangrove national nature reserve [J]. Haiyang Xuebao, 2021, 43(9): 137-145.[周在明,陈本清,徐冉,方维.基于无人机高光谱特征的红树林种群识别研究——以漳江口红树林国家级自然保护区为例[J].海洋学报,2021,43(9):137-145.]

Classification of Mangrove Species with UAV Hyperspectral Imagery and Machine Learning Methods

Jiang Yufeng^{1,2}, Qi Jianguo¹, Chen Bowei², Yan Min²,
Huang Longji³, Zhang Li²

(1.Department of Surveying and Mapping, School of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China;

2.Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

3.Hai Nan Dong Zhai Gang National Nature Reserve Authority, Haikou 571129, China)

Abstract: In this paper, we used the UAV hyperspectral images of the mangrove reserve at Qinglan Harbor, Wenchang, Hainan Province, and then preferentially selected vegetation spectral features and texture feature variables using Recursive Feature Elimination-Random Forest (RFE-RF). We further used the Random Forest (RF) and Support Vector Machine (SVM) algorithms to classify the mangrove tree species in the study area, and further the results of the classification model parameters on the overall accuracy were analyzed and evaluated. The results showed that the overall accuracy of RF classification was 92.70% and the Kappa coefficient was 0.91. Compared with the traditional SVM classification method, RF improved the producer accuracy and user accuracy of five types of tree species, which could effectively classify mangrove tree species and provide technical support for germplasm resource planning and ecological environmental protection.

Key words: Machine learning; Random forest; Hyperspectral; Feature extraction; Species classification