Geomatics Science and Technology 测绘科学技术, 2023, 11(2), 88-97
Published Online April 2023 in Hans. https://www.hanspub.org/journal/gst
https://doi.org/10.12677/gst.2023.112010

Hans汉斯

基于PIE-Engine的红塔区烟草种植面积提取研究

卢成卓*，王　涛#，梁桂华

玉溪师范学院，地理与国土工程学院，云南　玉溪

收稿日期：2023年2月19日；录用日期：2023年4月5日；发布日期：2023年4月12日

摘　要

烟草是云南省的主要经济作物之一，其种植面积和产量信息是农业部门制定相关政策的重要依据。因此，实时、精准和更具成本效益地确定烟草种植面积和监测烟草生长状况的方法极为重要。本文以云南省玉溪市红塔区为研究区域，采用Sentinel-2卫星影像为数据源，利用PIE-Engine遥感云平台，对2021年该研究区的遥感影像进行解译，采用随机森林、支持向量机、神经网络和深度学习四种分类方法分别提取烟草种植面积信息，并进行对比分析获取最优分类算法，为相应的农业生产提供指导，通过研究分析得出深度学习提取结果最为准确，总体精度(OA)达到94.70%，Kappa系数为0.92，提取的烟草面积为1989.36 hm^2，最为接近红塔区2021年统计公报中的烟草种植面积，误差仅为3.12%。

关键词

遥感云平台，随机森林，支持向量机，深度学习，神经网络

A Study on the Extraction of Tobacco Planting Area in Hongta District Based on PIE-Engine

Chengzhuo Lu*, Tao Wang#, Guihua Liang

College of Geography and Land Engineering, Yuxi Normal University, Yuxi Yunnan

Received: Feb. 19th, 2023; accepted: Apr. 5th, 2023; published: Apr. 12th, 2023

Abstract

Tobacco is one of the major cash crops in Yunnan Province, and its planted area and yield infor-

*第一作者。
#通讯作者。

文章引用: 卢成卓, 王涛, 梁桂华. 基于 PIE-Engine 的红塔区烟草种植面积提取研究[J]. 测绘科学技术, 2023, 11(2): 88-97. DOI: 10.12677/gst.2023.112010

mation are an important basis for the agricultural sector to formulate relevant policies. Therefore, a real-time, accurate, and more cost-effective method of determining tobacco acreage and monitoring tobacco growth is extremely important. In this paper, we use Sentinel-2 satellite images as the data source and the PIE-Engine remote sensing cloud platform to interpret the remote sensing images of this study area in 2021 and use four classification methods: random forest, support vector machine, neural network, and deep learning to extract tobacco planting area information respectively, and compare them, and conclude that deep learning The extracted tobacco area was 1989.36 hm², which was the closest to the tobacco cultivation area in the statistical bulletin of Hongta District in 2021, with an error of only 3.12%.

Keywords

Remote Sensing Cloud Platform, Random Forest, Support Vector Machine, Deep Learning, Neural Network

1. 引言

烟草是我国重要的经济作物，在国民经济发展中发挥着重要作用。快速、实时和准确地获取烟草生长信息是推动现代烟草农业发展的基础[1]。遥感影像可根据农作物光谱特征，通过传感器记录的地表信息，识别作物，经过影像处理可获取作物其面积，从而降低人力、物力和财力的投入，降低生产成本，提高生产管理效率。

通过查阅已发表的相关文献，Sentinel-2 遥感影像在农作物的面积提取和生长监测中已有应用，如魏梦凡[2]提出了一种 V2OAE 的分类方法，利用 Sentinel-2A 遥感影像提高了冬小麦种植面积的提取精度。严欣荣等[3]采用 Sentienl-2 遥感影像，利用随机森林、反向传播神经网络、支持向量机三种机器学习分类方法对沧源县的丛生竹林空间信息进行提取，张阳等[4]采用决策树分类法利用 Sentinel-2A 遥感影像对湖南省茶陵县烤烟种植面积进行提取，其结果可满足烤烟生产管理的实际需求。薛宇飞等[5]利用 Sentinel-2 遥感影像提取了云南省德宏州芒市烟草等地物光谱特征，计算了植被指数、红边指数，提取的总体精度达到 94.38%。

上述学者在作物信息提取中，对于作物的研究集中在单一方法或单一影像方面，而在多种方法和多景影像融合提取上的研究较少。本文以红塔区为研究区域，该区域为红塔集团驻地所在，利用遥感云平台对

Sentinel-2 遥感影像数据进行处理，采用随机森林(RF) [6]、支持向量机(SVM) [7]、神经网络(NNC) [8]、深度学习(DL) [9]四种分类算法分别提取烟草种植区域信息，并对四种提取算法进行对比分析，获得最佳的提取方法。

2. 材料与方法

2.1. 研究区概况

红塔区位于云南省中部，在 24˚08′30″~24˚32′18″N、102˚17′32″~102˚41′37"W，如图 1 所示，气候类型为中亚热带半湿润冷冬高原季风气候，海拔 1500~2614 m。土地面积 100,400 hm^2，其坝区面积 17,300 hm^2，

南北最长 20 km，东西宽约 8 km，坝区平均高程 1650 m。降水期主要集中在 5~10 月份，年均降雨量 800~1000 mm，降水有效性高，能有效满足烤烟的生长；气温相对适宜，年平均气温 16.5℃；年均日照时数 2103 h，能充分满足烤烟生长期所需的积温，昼夜之间温差较大，有利于烟株的糖分积累与分解以及合成出更多的芳香物质[10]。

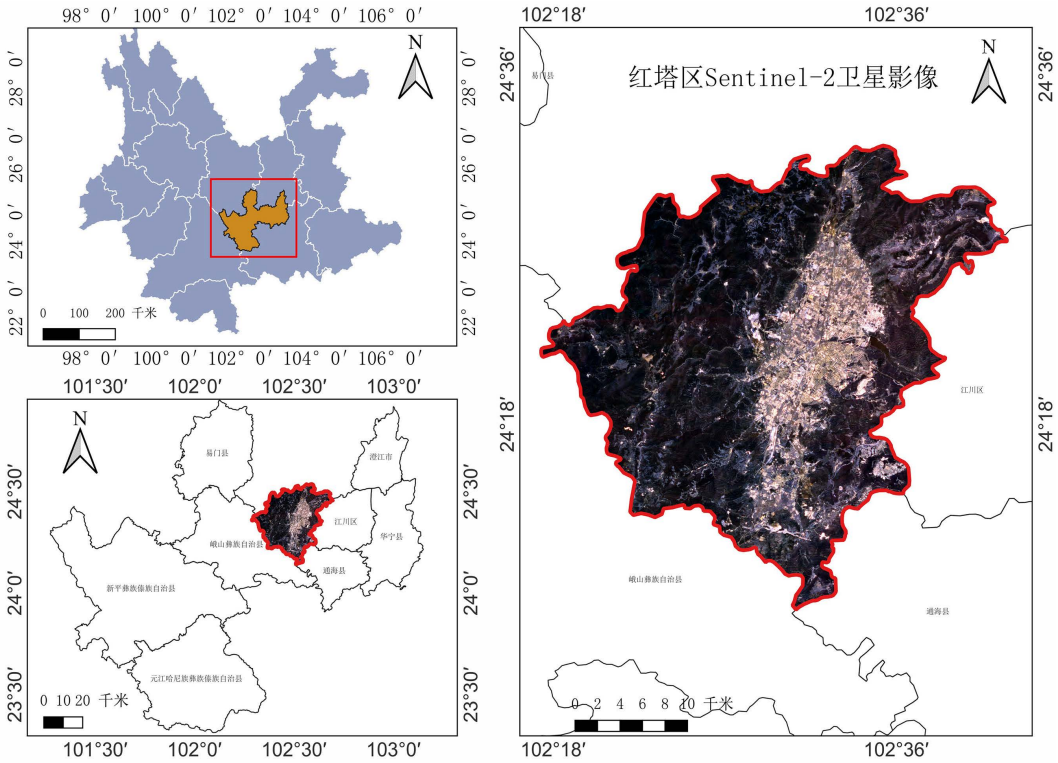

Figure 1. Location of the study area and satellite imagery
图 1. 研究区位置及卫星影像图

2.2. 数据源

本文采用的 Sentinel-2 遥感影像来源于欧洲航空局[11]。Sentinel-2 卫星影像拥有可见光到短波红外 13 个波段，与其他光学卫星相比，具有更丰富的光谱信息、更高的时间和空间分辨率。辅助数据包括红塔区边界矢量数据。经过对研究区进行实地踏勘调查，该区域主要作物类型为烤烟、水稻、玉米、小香葱等。根据烟草在该地的物候信息，1 月中旬为烟草育苗期，主要在苗床完成，4 月至 5 月初为烟草的移栽期，之后烟草进入生长期，7 月出开始进行采收，9 月中旬完成烟草初步生产。因此，影像数据的获取集中在 4 月至 7 月。

2.3. 遥感云计算平台

本文采用的遥感计算云服务平台为 PIE-Engine，该平台是基于云计算、物联网、大数据和人工智能技术自主研发的地球科学大数据实时计算平台，可处理多源遥感数据，极大减少了遥感科学技术人员与遥感工程技术人员的时间和资料成本，极大的推动了中国遥感技术生态圈的发展[12] [13]。它是目前国内最接近 GEE 的产品，弥补国内缺失 GEE 的局面，推动中国遥感技术生态圈的发展。包含大量遥感图像和矢量数据，还涵盖了气象数据、土地利用数据和地形地貌数据等，同时也可以上传自己的矢量、影像、

表格等数据到个人空间，方便引用。相比较于传统的 ENVI、ArcGIS 等传统遥感影像处理工具，该平台可以快速批量的处理影像。

2.4. 训练样本选取

在遥感影像的解译过程中，分类效果的好坏与样本点的种类、质量和数量有关[14]。通过实地样本选择，结合 Google Earth Pro，依据准确性、代表性、独立性、统计性原则[15]，共选择 1700 个样本，随机选取其中各类的 20%作为验证样本，其余的均为训练样本。如表 1 为研究区训练样本集的可分离性及详细信息。

Table 1. Study area training sample set separability
表 1. 研究区训练样本集可分离性

			样本集				训练样本可分离性		
类别	训练样本(个)	验证样本(个)	林地	水域	蔬菜大棚	烟草	其他作物	城市	其他
林地	280	70	2.0000	1.9999	2.0000	1.9930	1.9970	1.9990	1.9970
水域	100	25	1.9999	2.0000	1.9999	2.0000	2.0000	2.0000	1.9950
蔬菜大棚	128	32	2.0000	1.9999	2.0000	1.9880	1.9030	1.8390	2.0000
烟草	310	77	1.9930	2.0000	1.9880	2.0000	1.6780	1.9860	2.0000
其他作物	360	90	1.9970	2.0000	1.9030	1.6780	2.0000	1.8310	1.9990
建设用地	146	36	1.9990	2.0000	1.8390	1.9860	1.8310	2.0000	1.9970
其他	36	10	1.9970	1.9950	2.0000	2.0000	1.9990	1.9970	2.0000

2.5. 分类方法

2.5.1. 随机森林法

随机森林(RF)是指利用多棵决策树对样本数据进行训练、分类并预测的一种方法，它在对数据进行分类的同时，还可以给出各个变量的重要性评分，评估各个变量在分类中所起的作用[6]。RF 是以单一决策树为基础的集成分类算法，{h(x, Θk), k = 1, ⋯,}，其中 x 为输入的特征向量，{Θk}为独立同分布的随机向量，k 为决策树的数量，最终由所有决策树投票决定输入向量 x 的最终输出结果[6]。它通过对大量不同的训练样本进行有放回的抽样，如 Bagging 或者 Bootstrap，使得树不停生长，从而增加树的多样。因此具备复杂地物分类的能力，对于噪声和存在缺损值的数据具有良好的鲁棒性，兼具较快的学习速度，对多维特征空间数据重要性进行度量，依据特征贡献率进行最优特征筛选从而达到对高维特征空间进行降维的目的，相较当前流行的分类算法具有较高的准确性和稳健性。

2.5.2. 支持向量机

支持向量机分类(SVM)是一种建立在统计学习理论基础上的机器学习方法[7]。其基本原理是：假设训练样本为 $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，其中 $x_i \in R^d$ 表示输入模式， $y_i \in \{\pm 1\}$ ，表示目标输出。设最优决策面方程为

$$w^{\mathrm{T}} + b = 0$$

则权值向量 w 和 b 偏置须满足约束： $y_i\left(w^{\mathrm{T}} + b\right) \geq 1 - \varepsilon_i$ ， ε_i 表示线性不可分条件下的松弛变量，是用来表示模式对理想线性情况下的偏离程度。其目的为找寻一个决策面让它在训练数据上的平均分类误差

最低，可以推导出下面的优化公式

$$\varphi(\omega,\varepsilon)=\frac{1}{2}\omega^{\mathrm{T}}\omega+c_i\sum_{i=1}^{n}\varepsilon_i$$

c 表示用户指定正参数，用于支持向量机对错分样本惩罚程度，是算法复杂度与错分样本比例间平衡的一个参数[7]。

2.5.3. 神经网络分析法

神经网络分类(NNC)是将人类大脑神经元的模式简化为人工神经网络中的处理单元，通过计算机去仿照人脑的结构，用一系列小的处理单元去模拟生物大脑的神经元，再通过算法来实现人脑的识别、记忆、思考过程，并最终用在影像分类[9]。公式如下：

$$h_j=f\left(\sum_{i=1}^{m}\left(\omega_{ij}x_i-\theta_j\right)\right)$$

$$y_k=f\left(\sum_{i=1}^{n}\left(k_{ij}x_j-\theta_k\right)\right)$$

上式中分别为隐含节点和输出点的阈值，表示输入层 i 节点与隐含层 j 节点之间的连接权值，为隐含层 j 节点与输出层 k 节点之间的连接权值，为输入层 i 节点输入的样本信息。

2.5.4. 深度学习法

深度神经网络训练原理就是通过误差反向传播尽快实现与目标函数的拟合。其本质是在参数空间中，依靠梯度寻找损失函数的下凸点，并求解损失函数极小值。为了强化边缘部分的学习效果、突出边缘形态，在损失函数中加入了对边缘强化的函数计算部分[9]。具体的损失函数为

$$L=-\lambda\sum_{n=1}^{n}\left(y_i\ln a_i\right)$$

上式中：损失函数 L 中 y_i 为第 i 个分类的真值；a_i 为第 i 个分类的预测值；λ 为像素权重调节系数，用于提高神经网络对边缘附近的像素分类误差，强化图斑边缘附近语义分割的学习效果，使边缘较为模糊的图斑可以取得更好的分类精度。定义像素权重调节系数 λ 计算公式为

$$\lambda=\omega_0\left(1+\exp\left(-\frac{\left\|p_i-p_j\right\|^2}{\sigma^2}\right)\right)$$

[9]上式中：基础权重 ω_0 与边缘计算范围 σ 为超参数，$\left\|p_i-p_j\right\|^2$ 计算像素 p_i 到最近的边缘点 p_j 的距离。

3. 结果与分析

3.1. 种植面积提取结果

通过遥感云平台 PIE-Engine 筛选出研究区 4 月~8 月云量较少的 Sentinel-2 影像，并进行去云、融合处理，再分别采用四种分类方法进行烟草种植面积提取，分类结果如图 2 所示。

3.2. 分类结果精度评价

结果精度评价是遥感影像分类或信息提取的最后一项工作，通过精度分析能判断分类方法是否有效，是否需要改善分类方法，提高分类精度[16]。分类结果精度评价通常选取用户精度(CA)、生产者精度(PA)、

总体精度(OA)和 Kappa 系数作为评价因子。本文中深度学习分类总体精度和 Kappa 系数最高，分别为94.70%和 0.92，神经网络次之，其总体分类精度达 93.40%，Kappa 为 0.89，与其他两种分类方法相比，分类结果精度较高，但分类过程需要多次迭代，耗时较长，如表 2 所示。四种分类方法中生产者精度最大是 87.88%、最小为 80.66%，用户精度最大、最小分别为 92.99%、80.52%，如图 3 所示。

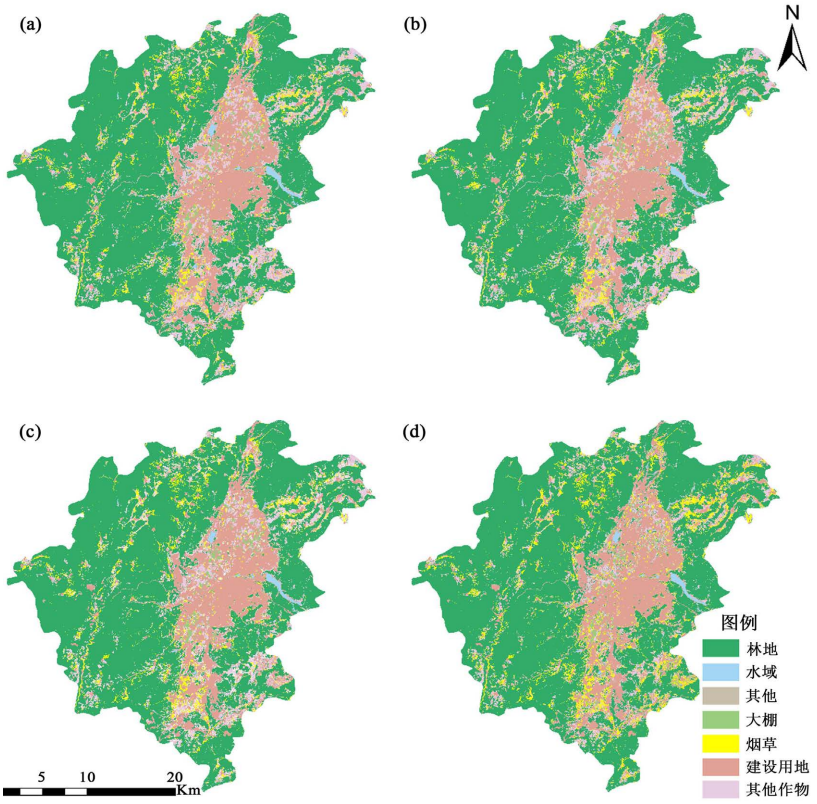

Figure 2. The tobacco planting area extraction result plot in the study area. (a) Is the deep learning extraction result; (b) The neural network extraction result; (c) The random forest extraction result; (d) The support vector machine extraction result
图 2. 研究区烟草种植面积提取结果图。(a) 为深度学习提取结果；(b) 为神经网络提取结果；(c) 为随机森林提取结果；(d) 为支持向量机提取结果

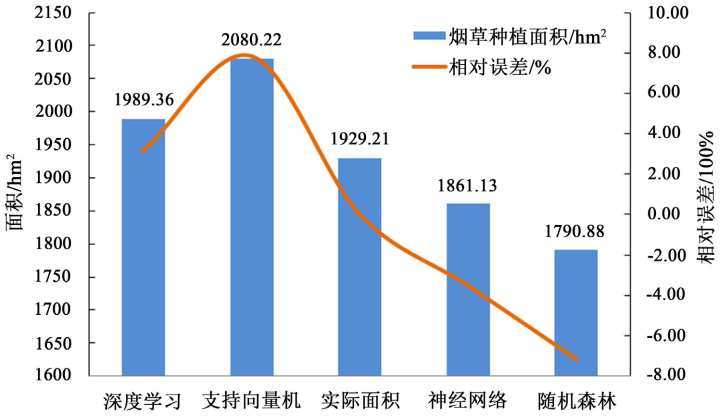

Figure 3. Area extraction and relative error
图 3. 面积提取及相对误差

对红塔区同时相的遥感影像进行分类提取烟草种植面积，利用四种不同的分类算法进行分类，其总体分类精度和卡方系数的不同主要是因为算法的结构有所差异，其中深度学习的精度最高，深度学习算法学习能力强，覆盖范围广，适用性好，可以映射到任意函数，模拟效果更好，能解决很复杂的问题，但其耗时较长。随机森林分类算法的分类结果精度较差，原因是分类算法中决策树的数量不够多，加大决策树的数量则会导致模型训练更慢，所需要的时间和空间更大，因此随机森林要达到更好的结果，需要更好的硬件条件和云计算服务。其余两种分类算法介于上两种之间，神经网络是基于人脑神经系统的分类算法，具有自学习功能和高速寻找最优化解的能力，但受数据的限制，当数据不充分时则无法进行，也无法解释自己的推理过程和依据，只有结果；支持向量机是一种传统机器学习方法，分类的复杂度和结过精度主要受支持向量的数目影像，对参数和函数的选择较为敏感。

Table 2. Overall classification accuracy and Kappa coefficient
表 2. 总体分类精度和 Kappa 系数

分类方法	总体分类精度(OA)	Kappa 系数
支持向量机(SVM)	91.7049%	0.8669
随机森林(RF)	83.2481%	0.8113
深度学习(DL)	94.6960%	0.9194
神经网络(NNC)	93.4047%	0.8805

3.3. 提取结果误差分析

通过查阅《红塔区 2021 年国民经济和社会发展统计公报》，红塔区 2021 年烟草种植面积为 1929.21 hm^2。经过统计分析计算，如图 4 所示，其中相对误差值大于 0，表示分类过程中存在错分现象；相对误差值小于 0 则表示存在漏分现象。深度学习算法提取面积为 1989.36 hm^2，精度为 96.88%，误差值约为 3.12%，部分地区存在漏分现象；而随机森林法提取面积为 1790.88 hm^2，面积精度为 92.83%，误差值最大，为 7.17%，错分现象较为明显，其余两种分类方法介于以上两者之间。

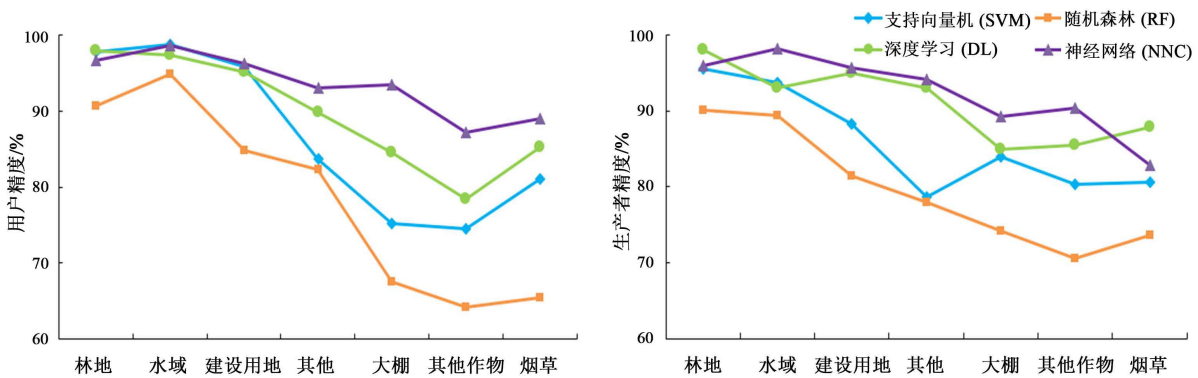

Figure 4. User accuracy and producer accuracy
图 4. 用户精度和生产者精度

通过混淆矩阵计算出各方法的错分误差和漏分误差，如图 5 所示，神经网络方法提取烟草的错分误差最小为 7.01%，而漏分误差为 17.17%，支持向量机的漏分误差最小 9.34%，而错分误差为 18.99%，随

机森林错分误差为 19.5%，漏分误差为 18.32%。深度学习在提取上的错分误差和漏分误差相差不大，接近 10%，相对其他三种为最优提取算法。

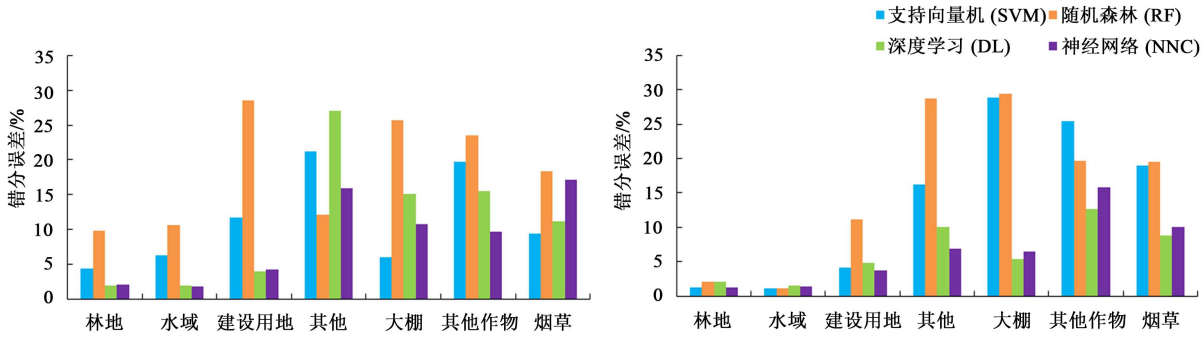

Figure 5. Missed errors and mis-split errors
图 5. 漏分误差和错分误差

3.4. 烟草空间分布

从遥感影像分类结果图可以得出研究区内烟草种植主要集中分布在小石桥乡、洛河乡、北城街道、研和街道和春和街道北部，该区域海拔相对较高，其分布特征主要沿村庄周围和道路两侧，交通便利，便于劳作，在北城街道南部、春和东部、大营街道则呈零星状分布，该地区地势相对平坦，主要经济作物为小香葱和大棚蔬菜等。而在玉带路街道、玉兴路街道、凤凰路街道、高仓街道和李琪街道的西北部分布少，该地区主要为红塔区的建成区。烟草分布如图 6 所示。

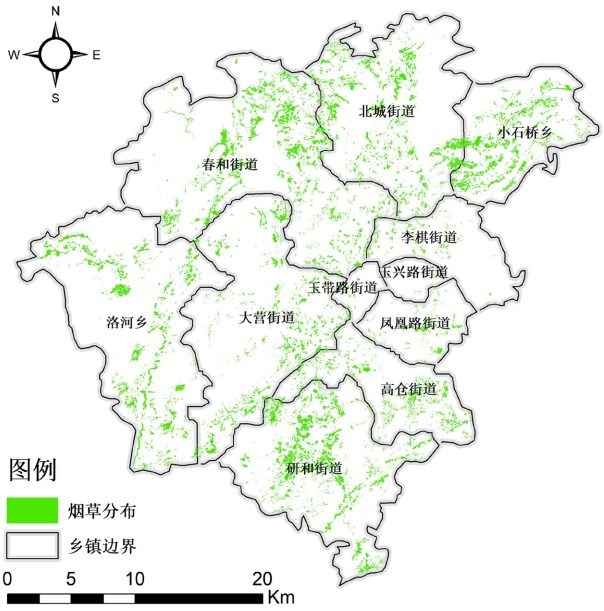

Figure 6. Spatial distribution of tobacco
图 6. 烟草空间分布图

4. 结论及讨论

本文利用遥感云计算平台的超大算力，实时批量处理遥感影像数据，以研究区内的多景 Sentinel-2

卫星影像为数据源，采用随机森林、支持向量机、神经网络和深度学习四种分类方法，对红塔区烟草种植面积信息进行提取，对不同方法提取出的面积结果进行对比分析，得出深度学习的分类结果最为接近真实值，误差仅为 7.61%，提取精度较高，可满足农业部门快速统计需求。

本研究利用 sentinel-2 数据提取烟草及其他地物信息，利用不同的分类算法进行分类，因为算法本身结构具有较大差异，分类结果也有明显差异，而研究本人对算法结构的研究还不够深入，还需进一步的提升；本研究虽然较传统的统计方法来说效率有了提升，但遥感影像空间分辨率有限，分类结果仍然存在一定的误差，下一步深入的研究可选择更高分辨率的影像，如高分二号等。在不同分类方法上，深度学习和神经网络的提取精度优于其他两种，但神经网络和随机森林分类算法的漏分误差较大，支持向量机和随机森林的错分误差较大，可能是存在同谱异物的现象，进一步可根据多时相、多源卫星遥感数据提取更精确的信息。深度学习方法在红塔区烟草信息提取研究中效果良好，提取的烟草信息更符合实际情况，但就总体分类精度而言，深度学习方法的耗时约为神经网络方法耗时的 2 至 3 倍，支持向量机方法耗时的 3 至 4 倍，得到 1.29%和 2.99%的提升，投入与产出失衡，下一步可提高和优化分类算法，以提高生产效率。基于多时段合成对烟草种植区的提取精度还有待进一步提高下一步可根据多时相、多源卫星遥感数据对烤烟种植区域进行提取，也可加入多时相雷达孔径数据来提取种植数据，以避免因天气影响而导致的精度降低。

基金项目

大学生创新创业训练计划项目(202111390019)；云南省地方本科高校基础研究联合专项(202001BA070001-109)。

参考文献

[1] 李朋彦. 基于无人机高光谱遥感的烤烟生长监测[D]: [硕士学位论文]. 郑州: 河南农业大学, 2019.

[2] 魏梦凡. 基于 Sentinel-2A 卫星遥感影像的开封市冬小麦种植面积提取技术研究[D]: [硕士学位论文]. 郑州: 河南大学, 2019.

[3] 严欣荣, 张美曼, 郑亚雄, 等. 基于 Sentinel-2 的丛生竹林信息提取方法比较及分布特征[J]. 生态学杂志, 2020, 39(3): 1056-1066.

[4] 张阳、屠乃美、陈舜尧, 等. 基于 Sentinel-2A 数据的县域烤烟种植面积提取分析[J]. 烟草科技, 2020, 53(11): 15-22.

[5] 薛宇飞, 张军, 张萍, 等. 基于 Sentinel-2 遥感影像的烟草种植信息精准提取[J]. 中国烟草科学, 2022, 43(1): 96-106.

[6] 赵志国. 基于随机森林方法的遥感影像分类方法[J]. 北京测绘, 2021, 35(9): 1173-1176.

[7] 张伐伐, 李卫忠, 卢柳叶, 等. SVM 多窗口纹理土地利用信息提取技术[J]. 遥感学报, 2012, 16(1): 67-78.

[8] 张雷, 刘昌华, 石林峰, 等. 基于 U-net 神经网络的烟草种植信息提取[J]. 农业与技术, 2021, 41(22): 44-47.

[9] 朱明, 李景文, 吴博, 等. 基于深度学习的高分辨率卫星影像地表覆盖分类方法[J]. 桂林理工大学学报, 2022, 42(1): 115-121.

[10] 刘秀玲, 武仕强, 王忠兴, 等. 玉溪市红塔区现代烟草农业种植比较优势评价[J]. 农家之友(理论版), 2011(2): 52-54.

[11] 常文涛, 王浩, 宁晓刚, 等. 融合 Sentinel-2 红边波段和 Sentinel-1 雷达波段影像的扎龙湿地信息提取[J]. 湿地科学, 2020, 18(1): 12-21.

[12] 卞晓东, 禹定峰, 刘东升, 等. 基于 PIE-Engine Studio 的黄河口及其邻近海域水质遥感监测[J]. 山东轻工业学院学报, 2022, 36(2): 53-58.

[13] 程伟, 钱晓明, 李世卫, 等. 时空遥感云计算平台 PIE-Engine Studio 的研究与应用[J]. 遥感学报, 2022, 26(2): 335-347.

[14] 李杰, 张军, 李宇宸. Sentinel-2A 与 GF-1 数据在油菜种植提取中的差异性分析及提取方法对比研究[J]. 云南大学学报, 2019, 41(4): 678-688.

[15] 孙逸飞, 柳平增, 张艳, 等. 基于 Sentinel-2A 遥感影像的潍坊市冬小麦种植面积提取研究[J]. 中国农机化学报, 2022, 43(7): 98-105.

[16] 屈旭洲, 施冬. 基于高光谱遥感卫星时序序列玉米种植面积的提取[J]. 绿色科技, 2021, 23(12): 236-237.