

芦瑞杰,刘树林,康文平,等.结合GEE平台与机器学习算法的荒漠信息提取[J].中国沙漠,2023,43(6):60-70.

结合GEE平台与机器学习算法的 荒漠信息提取

芦瑞杰^{1,2}, 刘树林¹, 康文平¹, 冯 坤¹, 郭紫晨¹, 支 莹^{1,2}

(1.中国科学院西北生态环境资源研究院 沙漠与沙漠化重点实验室, 甘肃 兰州 730000; 2.中国科学院大学, 北京 100049)

摘要: 快速、准确获取不同荒漠类型的分布信息,对于环境保护和生态修复有着重要意义。然而,受光谱与分辨率等因素影响,目前关于不同荒漠类型信息提取研究存在明显不足。选择青海省都兰县作为典型区,基于谷歌地球引擎(Google Earth Engine, GEE)并结合多源数据对研究区荒漠进行了分类,对比评估了不同分类特征组合应用于3种机器学习方法(RF、SVM、CART)的分类性能。结果表明:(1) RF的性能要优于CART和SVM,使用RF分类器并以光谱特征、雷达特征、地形特征、纹理特征为分类依据的总体分类精度最高,整体准确率为95.68%,*Kappa*系数为0.95, F_m 得分为94.28%,获得的都兰县荒漠面积为29 039 km²。(2) 在特征重要性得分评估中,海拔与VH对荒漠分类的贡献比较突出,其他特征则差异不大。(3) 在使用光谱数据的基础上,雷达特征对识别砾质荒漠与壤土荒漠的贡献突出,而地形特征则更适用于识别其他类型的荒漠。

关键词: 荒漠分类; 机器学习; GEE平台; 都兰县

文章编号: 1000-694X(2023)06-060-11

DOI: 10.7522/j.issn.1000-694X.2023.00072

中图分类号: P931.3

文献标志码: A

0 引言

荒漠是干旱区的典型原生生态系统,是陆地生态系统的重要组成部分,具有不同于其他生态系统的结构和功能,在防风固沙、保育土壤、调控水资源、固碳释氧、维持生物多样性、文化旅游等方面发挥着重要作用^[1-2]。中国荒漠类型多样、分布广泛,对荒漠类型进行划分,研究其物质组成、分布规律和类型特征,是了解荒漠生态系统时空格局和支撑荒漠生态系统分类管理的基础,对于环境保护与生态修复等都具有重要意义^[3-5]。

目前对荒漠类型与分布的研究,主要借助各种土地覆盖产品进行,如国际地圈-生物圈计划(IGBP) DISCover 产品,马里兰大学(UMD)土地覆盖产品,基于MODIS的土地覆盖产品,基于Landsat与Sentinel影像的GlobeLand30、FROM_GLC30、ESA WorldCover10等^[6]。然而,由于每种产品的数

据源、分类方案和方法不同,它们在特定区域应用的准确性和适用性往往不理想且对荒漠类型的划分不够精细^[7-8],不能很好地揭示荒漠生态系统内部的分布特征。目前,虽然也有一些较为详细的对荒漠类型进行划分的研究,如姚爱冬等^[9]对民勤未利用地的细分和杨发相等^[10]对新疆荒漠类型的划分,但这些研究多使用人机交互的解译方式进行分类,需要大量专家知识参与决策,耗时耗力。因此,需要探索一种快速准确的荒漠信息提取方法。

由于荒漠地区光谱响应相似,仅使用光学遥感数据很难对荒漠类型进行区分,而合成孔径雷达(SAR)数据可以提供物体的表面粗糙度、纹理和介电特性等信息,两种数据类型之间存在着互补性^[11-12]。将光学数据与SAR数据协同使用,可以综合利用两种数据的优势,更好地解释、增强和分析不同地表的特征^[13]。此外,研究表明:将地形数据、纹理数据与遥感数据结合,可以提高图像分类、物

收稿日期:2023-03-23; 改回日期:2023-05-31

资助项目:第二次青藏高原综合科学考察研究项目(2019QZKK0305)

作者简介:芦瑞杰(1999—),男,山西临汾人,硕士研究生,主要研究方向为荒漠生态遥感。E-mail: luruijie@nieer.ac.cn

通信作者:刘树林(E-mail: liusl@lzb.ac.cn)

体识别的准确性^[14]。因此,将地理数据与遥感数据相结合已经成为一个新的趋势。传统方法处理大量的地理数据与遥感数据,需要预先对数据进行下载,效率较低^[15]。近年来,遥感大数据和云计算的快速发展,为多源数据的处理提供了技术支持。特别是,谷歌地球引擎(Google Earth Engine, GEE)作为一个在线处理平台,其强大的云端计算能力让用户可以从海量遥感数据中快速提取有用信息,实现了数据获取、处理、分析与应用于一体,极大提高了遥感图像处理的效率^[16];其提供的丰富遥感数据和机器学习算法在环境监测领域备受关注,已经被广泛应用于作物、湿地、草地、森林、城市土地利用分类等方面^[17-18],但目前对荒漠分类中不同分类器的分类性能和效果的比较还鲜有研究。

柴达木盆地是中国荒漠集中分布区^[19-20],既有干旱的温性荒漠,又有受低温影响的高山寒漠,是开展荒漠信息提取研究的理想区域。本文基于GEE平台,利用Sentinel-1、Sentinel-2、SRTMGL1_003等多源数据,对柴达木盆地都兰县典型荒漠区进行分类研究,评估不同机器学习算法在荒漠分类中的性能,比较不同分类特征对荒漠类型划分的影响,研究荒漠的空间分布特征,以期选取适用于都兰县荒漠类型划分的特征指标和分类方法,为研究区和类似区域荒漠生态系统管理提供借鉴。

1 材料与方法

1.1 研究区概况

都兰县位于35°20′—37°22′N、95°34′—99°14′E,地处柴达木盆地的东南部,面积约4.5万km²,海拔2 660~5 507 m(图1)。地貌类型主要为山地和

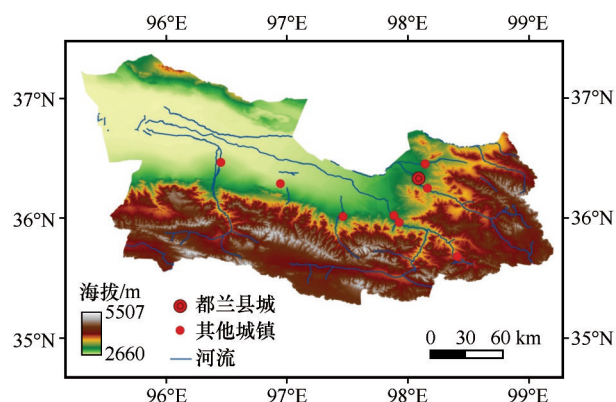


图1 研究区位置和地理概况

Fig.1 Location and geographical overview of the study area

平原,地势由东南向西北倾斜。属高原干旱大陆性气候,年均气温2.7~4.4℃,年蒸发量2 088~2 716 mm,年降水量37~296 mm且集中在5—9月,雨热同期,有利于农作物的生长。

都兰县荒漠分布广泛、类型多样^[21-22],包括高山寒漠、裸岩、戈壁、沙漠、盐碱地等多种地表类型,属于防风固沙生态功能区,其生态保护和建设是构建国家生态安全屏障的重要组成部分,已被纳入“两屏三带”的生态安全格局中^[23-24]。

1.2 数据源及预处理

本研究使用的地理坐标系为CGS_WGS_1984,在参与分类时将Sentinel-1、Sentinel-2、SRTMGL1_003数据的空间分辨率重采样至10 m。

1.2.1 Sentinel-1 影像

根据研究区范围,选取处于2019年至2021年7月、8月的Sentinel-1数据在VV、VH两种极化方式下生成的影像,作为研究使用SAR数据(GEE平台已对该数据进行了热噪声去除、辐射校正、地形校正等预处理^[25])。筛选出的影像通过中值合成,生成研究区的SAR影像。

1.2.2 Sentinel-2 影像

根据研究范围、云量小于20%,筛选与Sentinel-1数据同期的Sentinel-2A产品。共选取Sentinel-2影像328景,通过QA波段去云、中值合成等处理,镶嵌成一幅研究区的高质量影像。

1.2.3 地形数据

研究使用GEE平台提供的SRTMGL1_003地形数据,该数据的空间分辨率为30 m,由美国国家航空航天局(NASA)提供^[26]。

1.3 研究方法

在GEE云平台上,基于上述处理好的Sentinel-1、Sentinel-2、SRTMGL1_003等数据,采用随机森林(RF)、支持向量机(SVM)、分类和回归树(CART)3种分类算法,对柴达木盆地都兰县荒漠进行划分。总体工作流程如图2所示。

1.3.1 样本选择方法

本文参考中国科学院和农业部建立的中国土地利用分类系统^[27]和刘亚群等^[28]对中国陆地生态系统的分类方案,根据荒漠的遥感光谱特征,结合海拔、地面物质组成等因子,构建适合都兰县荒漠

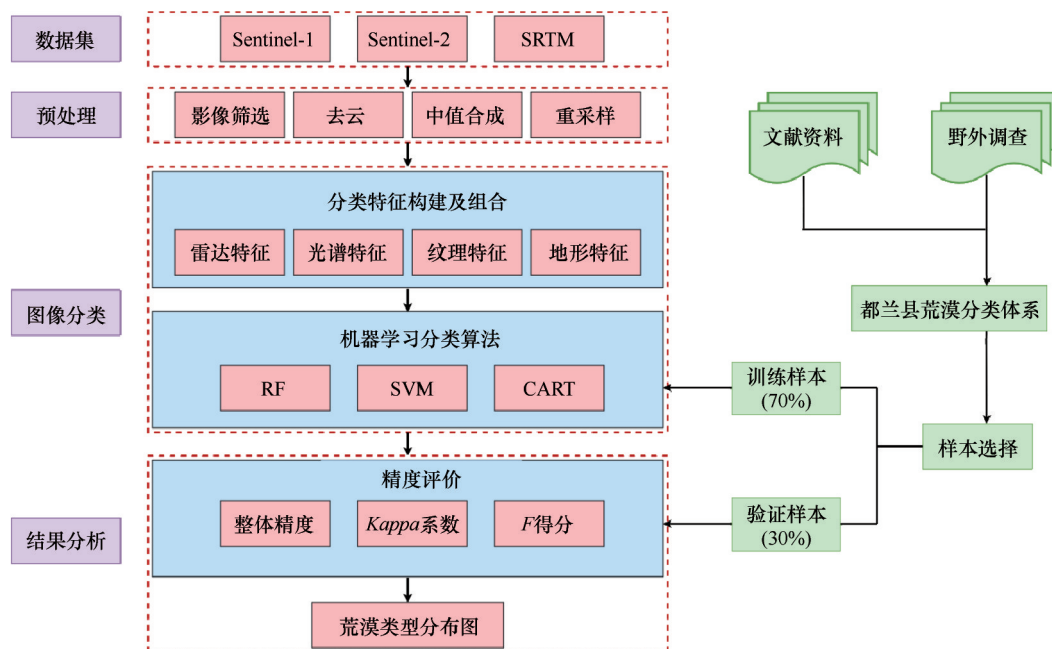


图2 荒漠分类技术流程

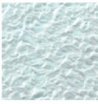
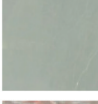
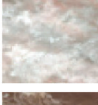
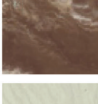

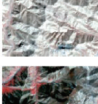

Fig.2 Technical flow chart of desert classification

类型的分类体系,并结合 Sentinel-2 假彩色合成影像建立了都兰县荒漠类型图像识别标志(表 1)。分类方案具体如下:首先,以植被覆盖度 20% 为指标,将研究区划分为荒漠区与非荒漠区两部分;然后依据

地形、气候、地面物质组成等特征将荒漠区分为 6 个类型,分别是沙质荒漠、砾质荒漠、盐碱荒漠、壤土荒漠、岩石荒漠、高山寒漠;考虑到研究区内盐碱荒漠具有不同的光谱特征,将其分为轻度与重度两种

表 1 都兰县荒漠分类体系

Table 1 Desert classification system of Dulan County

类型	缩写	地表特征描述	Sentinel-2 影像特征
沙质荒漠	SD	地表被沙子覆盖,主要包括平缓积沙地和沙丘区	
砾质荒漠	GD	地表被砾石覆盖,地势平坦,主要为各种砾石含量较高的冲洪积平原和风蚀戈壁	
轻度盐碱荒漠	MS	表土盐碱含量较高,颜色偏白	
重度盐碱荒漠	SS	地表盐碱聚集,土壤板结现象严重,颜色偏黑	
壤土荒漠	LD	地表土质覆盖,颜色偏黄	
岩石荒漠	RD	地表主要为岩石,土壤发育较差,地势崎岖,分布在山区	
高山寒漠	AC	相较于岩石荒漠海拔更高,主要分布于雪线以下、植被线以上,通常颜色为黑色或黑褐色	

类型;此外,冰雪、水体、农田和高覆盖率植被等地区被视为非荒漠地区。

按照此分类体系,参考 Google Earth 高分辨率实时影像,选取研究区内的分类样本。然后,根据 GEE 中提供的 random Column 算法将样本点的 70% 作为分类的训练样本,剩余的 30% 作为分类结果的验证样本。

1.3.2 分类特征

光谱指数特征。根据遥感影像的光谱波段进行数学组合形成的光谱指数已广泛应用于各种类型的地物区分^[29]。本文计算了归一化植被差异指数(*NDVI*)、盐度指数(*SI*)、归一化差异水体指数(*NDWI*)、归一化差异雪指数(*NDSI*)、增强型植被指数(*EVI*)、裸土指数(*BSI*)、修正土壤调节植被指数(*MSAVI*)、表土粒径指数(*TGSI*)作为研究使用的分类特征。

$$NDVI = (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED}) \quad (1)$$

$$SI = \sqrt{\rho_{BLUE} \times \rho_{RED}} \quad (2)$$

$$NDWI = (\rho_{GREEN} - \rho_{NIR}) / (\rho_{GREEN} + \rho_{NIR}) \quad (3)$$

$$NDSI = (\rho_{GREEN} - \rho_{SWIR}) / (\rho_{GREEN} + \rho_{SWIR}) \quad (4)$$

$$EVI = 2.5 \times (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + 6 \times \rho_{RED} - 7.5 \times \rho_{BLUE} + 1) \quad (5)$$

$$BSI = (\rho_{RED} + \rho_{SWIR} - \rho_{NIR} - \rho_{BLUE}) / (\rho_{RED} + \rho_{SWIR} + \rho_{NIR} + \rho_{BLUE}) \quad (6)$$

$$MSAVI = \left[2 \times \rho_{NIR} + 1 - \sqrt{(2 \times \rho_{NIR} + 1)^2 - 8 \times (\rho_{NIR} - \rho_{RED})} \right] / 2 \quad (7)$$

$$TGSI = (\rho_{RED} - \rho_{BLUE}) / (\rho_{RED} + \rho_{GREEN} + \rho_{BLUE}) \quad (8)$$

式中: ρ_{NIR} 、 ρ_{RED} 、 ρ_{GREEN} 、 ρ_{SWIR} 、 ρ_{BLUE} 分别代表地物在近红外、红色、绿色、短波红外、蓝色波段的表面反射率。

雷达特征。将 Sentinel-1 在 VV、VH 两种极化方式下产生的数据,作为特征变量构建中的雷达特征。

地形特征。在 GEE 平台提供的 SRTMGL1_003 地形数据基础上,提取海拔与坡度两个因子,作为参与分类的地形特征。

纹理特征。纹理是图像的空间特征,相较于其他空间特征(形状、大小等),纹理特征不需要事先对图像进行分割,使用较方便^[30]。灰度共生矩阵(GLCM)是一种常用的纹理统计分析方法,它通过研究灰度的空间相关特性来描述纹理,该方法 1973 年由 Haralick 等^[31]提出。本文对 Sentinel-2 影像的近红外、红色、绿色波段进行线性组合生成 GLCM 需要的灰度图像^[32]:

$$Gray = 0.3 \times \rho_{NIR} + 0.59 \times \rho_{RED} + 0.11 \times \rho_{GREEN} \quad (9)$$

基于 GEE 云平台的“glcmTexture”函数,为 gray 图像选取 7 个常用的 GLCM 度量:角二阶矩(asm)、对比度(con)、相关性(corr)、方差(var)、逆差距(idm)、总和平均值(savg)、熵(ent)。之后,对生成的纹理特征进行主成分分析(PCA),选取第一主成分作为参与分类的纹理特征。

为了比较不同特征对荒漠分类的效果和性能,将分类特征(CF)组合成 5 种情景(表 2):①CF1 旨在探讨仅使用 Sentinel-2 影像光谱特征对荒漠分类的效果。②CF2 在 CF1 的基础上添加 Sentinel-1 雷达数据,探讨雷达数据是否可以提高荒漠的分类精度。③CF3 在 CF1 的基础上考虑了研究区的地形因素,探讨地形对于荒漠分类的影响。④CF4 在 CF1 的基础上添加了纹理特征作为分类特征之一。⑤CF5 则将所有分类特征作为一组特征变量。

表 2 不同特征变量组合所包含的分类特征

Table 2 Classification features contained in different combinations of feature variables

CF 组合	光谱波段	光谱指数	雷达特征	地形特征	纹理特征
CF1	√	√			
CF2	√	√	√		
CF3	√	√		√	
CF4	√	√			√
CF5	√	√	√	√	√

1.3.3 机器学习分类算法

随机森林(RF)。随机森林是由 Breiman^[33]提出的一种非参数集成机器学习算法,它通过随机选择训练样本和变量子集生成多个决策树,最终分类类别由所有决策树的多数投票决定。同时,它还可以通过回溯采样和随机改变不同决策树分类过程中预测变量的组合来提高单个决策树的分类性能,进而提高分类器性能^[34]。

支持向量机(SVM)。支持向量机是一种理论上优越的机器学习算法,它的基本原理是创建一个最优超平面,也称决策边界,最大化最近样本与平面之间的距离,有效地分离类。其分类基础是:只有位于类边界上的训练样本才能有效地对类进行区分^[35-36]。因此,支持向量机虽然对样本的需求量较小,但对样本质量的要求较高。

分类和回归树(CART)。CART 是一种基于分层决策树框架的二元分类器,其结构简单,计算速

度快,输入-输出关系易于理解和解释,被广泛应用于遥感领域的各种研究中^[37]。GEE提供的CART算法可以调整两个参数——最大叶节点数和最小叶数。本文中CART算法参数设置为默认值。

1.3.4 精度评价方法

选择基于混淆矩阵生成的整体精度(OA)、生产者精度(PA)、用户精度(UA)、Kappa系数、F得分^[38],评估都兰县荒漠分类的结果。

$$OA = \left(\sum_{i=1}^n P_{ii} / N \right) \times 100\% \quad (10)$$

$$PA = (P_{ii} / P_{i+}) \times 100\% \quad (11)$$

$$UA = (P_{ii} / P_{+i}) \times 100\% \quad (12)$$

$$Kappa = \frac{N \times \sum_{i=1}^n P_{ii} - \sum_{i=1}^n (P_{i+} \times P_{+i})}{N^2 - \sum_{i=1}^n (P_{i+} \times P_{+i})} \quad (13)$$

$$F_i = 2 \times (PA \times UA) / (PA + UA) \quad (14)$$

$$F_M = \sum_{i=1}^n F_i / n \quad (15)$$

式中: n 是混淆矩阵中的总列数即总类别数; N 代表用于精度评估的样本总数; P_{ii} 代表混淆矩阵中第

行第*i*列的样本数; P_{+i} 代表第*i*列的样本总数; P_{i+} 代表第*i*行的样本总数; F_i 代表第*i*列所代表类别的F得分,它是基于生产者精度与用户精度的调和平均值; F_M 代表所有类别*F_i*得分的平均。

在本研究中, OA 、 $Kappa$ 、 F_M 用来评估分类器的整体性能, F_i 用来评估不同荒漠类型的分类精度。

2 结果

2.1 3种机器学习分类算法整体性能差异

在仅使用影像光谱特征对都兰县荒漠进行分类时,RF、SVM、CART等3种分类器的分类精度均达不到理想效果。但当这3种分类器在应用CF1分类的基础上分别单独加入雷达特征、地形特征、纹理特征参与分类时都会不同程度地提升分类器的整体性能(图3)。具体而言:对于RF和CART,地形特征的加入对于分类精度的提升帮助最大,其次是雷达特征、纹理特征。对于SVM,雷达特征的加入则对分类精度的提升起到最大帮助,其次是纹理特征,而地形特征反而对分类器整体精度的提升起到的作用最小。

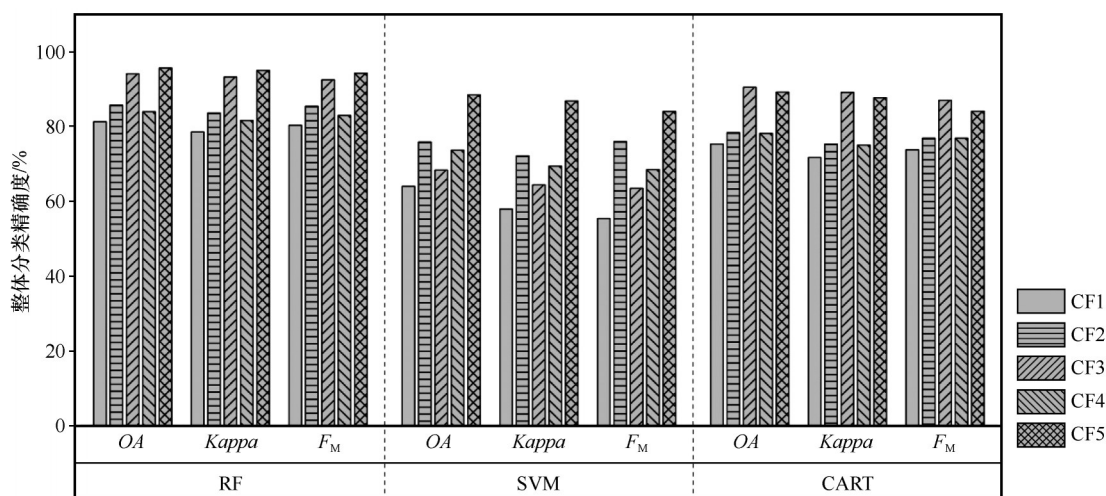


图3 RF、SVM和CART在不同分类特征组合下的整体精确度

Fig.3 Overall accuracy of RF, SVM, and CART with different classification feature combinations

无论使用哪种分类特征组合,RF分类器的性能都要优于CART和SVM。RF与SVM分类器的整体分类效果在同时应用光谱波段、光谱指数、雷达特征、地形特征、纹理特征参与分类时达到最佳。其中,RF的最高OA为95.68%,Kappa系数为0.95、 F_M 得分为94.28%;SVM的最高OA为88.50%、Kappa系数为0.87、 F_M 得分为84.05%。对于CART分类器,当

所有的分类特征同时参与分类时呈现的整体分类精度却没有表现出最佳效果,其最高精度在应用光谱波段、光谱指数、地形特征参与分类时获得:OA为90.52%、Kappa系数为0.89、 F_M 得分为87.03%。

2.2 不同类型荒漠分类的精确性评估

在应用CF1参与分类的基础上,单独加入雷达

特征、地形特征、纹理特征对不同类型荒漠的分类精度影响有所不同。总体而言,地形特征对沙质荒漠、轻度盐碱荒漠、重度盐碱荒漠、岩石荒漠、高山寒漠的分类精度提升作用最大,雷达特征则能更好地提高砾质荒漠与壤土荒漠的分类精度,而纹理特征对于各荒漠类型的区分表现最差(图4)。具体来讲:对于沙质荒漠,地形特征、雷达特征的加入分别提升了9.73%、4.77%的分类精度,但纹理特征加入时却降低了其分类精度。对于砾质荒漠,雷达特征、地形特征、纹理特征的加入都提高了其分类精度,但提升幅度都较小,分别为1.89%、1.84%、0.65%。对于轻度盐碱荒漠,地形特征、雷达特征、纹理特征分别提升了14.83%、5.96%、1.84%的分类精度。对于重度盐碱荒漠,雷达特征对其分类精度的提升效果不佳,仅0.4%,地形特征和纹理特征则分别提升了8.79%、1.91%。对于壤土荒漠,雷达特征表现出对分类精度提升最佳的效果,达到15.7%,地形特征则为7.97%,纹理特征对提高壤土荒漠的分类精度帮助最小,为4.94%。对于岩石荒

漠,仅用光谱指数与光谱波段很难有效地对岩石荒漠进行识别,精度仅50.40%,虽然加入纹理特征和雷达特征分别提升了16.01%和12.59%的分类精度,但分类效果依然不理想,而地形特征的加入则明显改善了对岩石荒漠的分类精度,且可以达到较为理想的分类效果。对于高山寒漠,地形特征参与分类时其分类精度提升了20.51%,雷达特征则为4.39%,纹理特征对分类精度提升的贡献最小,仅为0.09%。

对于所有类型的沙漠,当所有分类特征都参与分类时,将获得最高的 F_1 得分和最佳的分类效果。沙质荒漠的最高 F_1 为93.79%、砾质荒漠的最高 F_1 为96.12%、轻度盐碱荒漠的最高 F_1 为93.85%、重度盐碱荒漠的最高 F_1 为93.84%、壤土荒漠的最高 F_1 为89.55%、岩石荒漠的最高 F_1 为95.94%、高山寒漠的最高 F_1 为98.96%。而重度盐碱荒漠与高山寒漠在应用光谱特征与地形特征参与分类时、壤土荒漠在应用光谱特征与雷达特征参与分类时也可达到最高的分类精度。

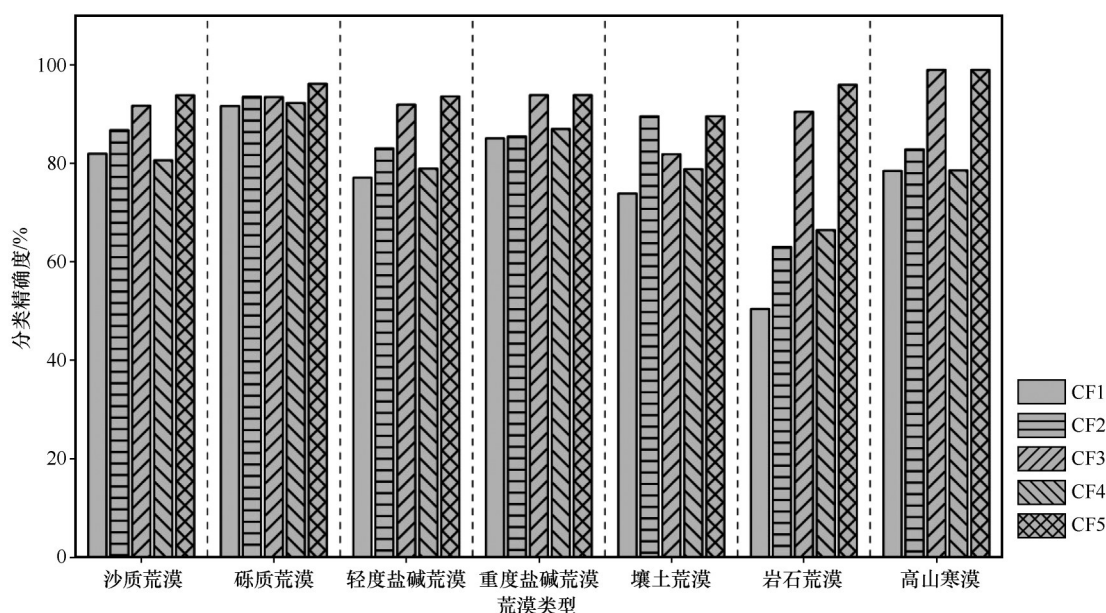


图4 不同的分类特征组合下各荒漠类型的分类精度

Fig.4 Classification accuracy of each desert type under different combinations of classification features

2.3 随机森林分类算法中各变量重要性大小

GEE平台为随机森林分类算法提供了“解释”功能,可以对参与分类的特征变量进行重要性得分评估。从图5可以看到,地形特征中的海拔特征在分类中重要性得分最高(429.63),而坡度特征则对分类的贡献较小。雷达特征中VH在分类中的重要

性要比VV更好,具有较高的重要性得分(257.31)。光谱指数特征中,TGSI、BSI的重要性得分要高于其他的光谱指数,这表明这几个指数对提取荒漠类型的帮助更大,而EVI、MSAVI、NDVI这几个常用的植被指数的则对于荒漠类型的提取帮助较小。Sentinel-2影像的光谱波段中,短波红外的重要性要高

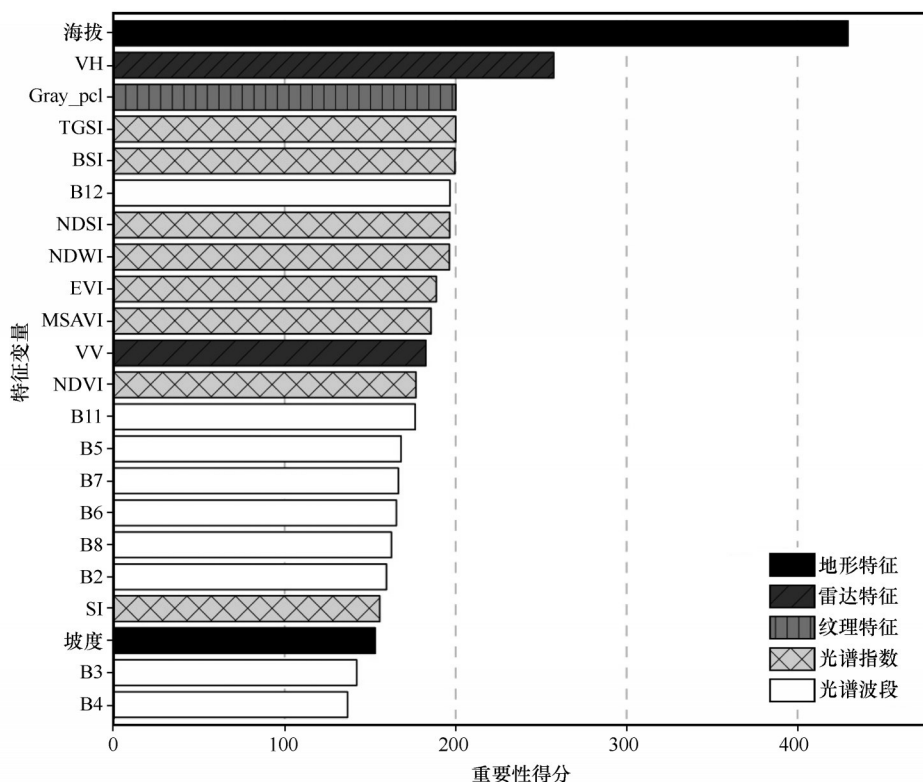


图5 特征变量重要性得分

Fig.5 Importance scores of feature variables

于近红外波段、植被红边和可见光波段。基于光谱波段的Gray纹理特征的重要性要高于光谱波段,其重要性得分为200.18。但总体而言,海拔与VH极化方式对荒漠分类的贡献较为突出,其他的分类特征重要性得分较低且差异不大。

2.4 都兰县荒漠类型识别结果

都兰县各种类型荒漠的总面积为29 039.43 km²,占整个都兰县面积的64.15%(图6)。沙质荒漠面积为2 455.38 km²(5.42%),砾质荒漠面积为4 341.50 km²(9.59%),轻度盐碱荒漠面积为1 664.78 km²(3.68%),重度盐碱荒漠面积为5 360.53 km²(11.84%),壤土荒漠面积为1 424.70 km²(3.15%),岩石荒漠面积为8 956.25 km²(19.79%),高山寒漠面积为4 836.29 km²(10.68%)。

都兰县的荒漠在空间分布上随海拔梯度大致呈现出条带状分布的特点。高山寒漠和岩石荒漠主要分布在都兰县东部和南部的山区,壤土荒漠则多分布在山脚的冲洪积扇或河流两侧,砾质荒漠大面积分布在都兰县南北两侧的山前冲洪积平原上,沙质荒漠则集中分布在研究区东部,盐碱荒漠则广泛分布在地势低洼的中西部。

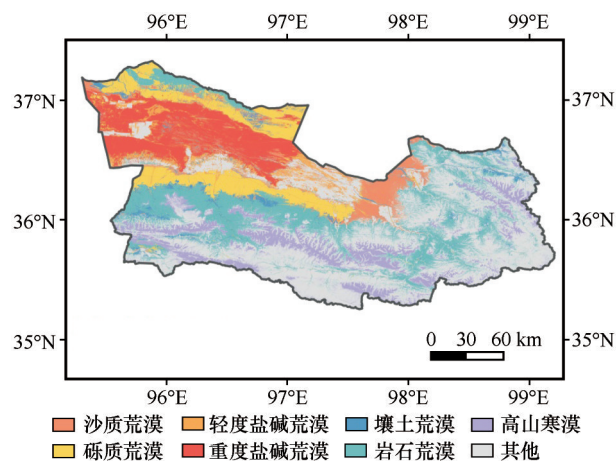


图6 都兰县荒漠类型的空间分布

Fig.6 Spatial distribution of desert types in Dulan County

3 讨论

3.1 不同机器学习方法对都兰县荒漠提取的整体性能

机器学习算法已被证明在处理高维、复杂的数据上也能产生良好的分类结果^[39]。但以往的研究表明:研究对象的差异对不同分类器的影响较大。如:Abida等^[40]比较了RF、CART、SVM、RVM等4种机器学习方法应用于半干旱地区的土地利用分类

的性能,结果显示RF是最好的方法。而Mustak等^[41]指出,与RF和CART相比,SVM分类器在作物识别方面的准确性更高。然而,不同的机器学习分类算法在荒漠类型提取中的性能却鲜有研究。在本研究中,我们使用基于像素的分类程序评估和比较了3种机器学习分类方法(RF、CART、SVM)对荒漠分类时的性能。结果显示:在使用相同的特征变量组合时,RF分类方法取得的整体精度要高于SVM和CART,在荒漠分类中具有明显的优势。

本研究结果显示,特征变量的类型与数量,会对不同机器学习分类算法的性能产生影响。地形特征对RF与CART分类精度提升的贡献最大,而雷达特征则对SVM分类器表现出更好的效果。这种结果可能是由于分类器模型结构的差异造成的,RF与CART分类器都是基于决策树框架构建的,而SVM分类器则与之不同^[39,42]。同时,尽管RF与CART采用了相似的分类策略,但添加过多的分类特征会降低CART分类器的准确性,而RF是一种集成机器学习算法,具有更好的鲁棒性^[43]。

3.2 分类特征对荒漠类型识别的影响

已有大量的光谱指数被用于监测和评估荒漠化^[44],但我们选取的3种常用植被指数(EVI、MSAVI、NDVI)在荒漠类型提取中的作用不大。这可能是由于植被指数对低植被覆盖度的敏感性较低所致^[45]。Xiao等^[46]根据表层土壤的颗粒组成与荒漠化程度密切关系,提出了用TGSI来监测荒漠化。我们也认识到,物质运移遵循由远及近、由粗到细的分异规律,对不同类型荒漠分布具有一定影响^[4]。BSI的分类性能会随着干旱度的增加而提高,是区分裸露土壤与其他土壤的有效指标^[40]。本研究中国TGS和BSI用于荒漠分类时,比其他指数表现更优。然而,由于荒漠地区的低光谱差异性,仅使用光谱波段和光谱指数对荒漠进行分类很难达到理想的分类效果。

GLCM纹理信息已被证明可以提高图像的分类精度^[47]。在本研究中,我们将基于近红外波段和可见光波段建立的纹理特征与光学数据结合用于荒漠分类,分类精度却没有明显改善。而在对影像光谱波段的重要性得分评估中,短波红外波段对荒漠分类的重要性要高于其他波段。因此,使用基于短波红外的纹理信息,可能会提高纹理特征在荒漠分类中的作用。

张志伟等^[48]在对高寒地区荒漠分类体系的探讨中也指出,海拔是影响水热分布的重要因子,与荒漠的分布关系密切。本研究也表明,海拔因子在所有参与分类的特征变量中取得了最高的重要性得分,对提高研究区内各种荒漠类型的分类精度具有良好的效果。此外,有研究指出表层土壤粒径的差异会引起微波后向散射系数的变化,在植被稀疏的干旱地区,这一响应会更加明显^[49]。本研究区由于壤土荒漠与砾质荒漠大部分处于同一海拔梯度上,然而其土壤表层粒径大小却存在明显差异。本研究将Sentinel-1雷达数据与光学数据结合后发现,相较于地形特征,雷达特征可以更好地地区分壤土荒漠与砾质荒漠。

3.3 本研究的优势与局限性

本研究基于GEE平台,整合多源遥感数据,实现了对荒漠类型的高精度分类,得到了都兰县不同类型荒漠的空间分布状况。通过应用不同的分类特征组合对比了3种常用的机器学习算法在荒漠分类中的性能,从中选出了最优分类方法,并探讨了各种分类特征提取不同荒漠类型的能力,为后续针对不同荒漠类型特点,选择合适的分类指标与分类方法提供了参考。

本研究是基于像素的分类方法,该方法存在像元空间分辨率与目标对象大小不匹配、局部出现“椒盐现象”等问题^[50]。而基于对象的分类方法可以通过不同尺度的图像分割与对象进行建模^[51]。因此,在未来的研究中可以考虑使用基于对象的方法对荒漠类型进行划分。需要注意的是,欠分割和过度分割可能会影响分类结果的准确性,选择合适的分割尺度十分重要^[52]。

4 结论

地形特征、雷达特征、纹理特征的加入,均比仅使用光谱数据情况下的分类精度高。其中地形特征的加入对RF和CART分类器整体精度的提升最明显,而雷达特征的加入则对SVM分类器整体精度的提升表现最好。

在提取荒漠信息时,RF分类器的性能要优于SVM和CART。使用RF分类器,并以光谱特征、雷达特征、地形特征、纹理特征为分类依据的总体分类精度最高,整体准确率为95.68%,Kappa系数为0.95, F_M 得分为94.28%,获得的都兰县荒漠面积为

29 039 km²。

本研究区分类特征的重要程度从大到小依次为地形特征、雷达特征、纹理特征、光谱指数、光谱波段。其中,海拔与VH对荒漠分类的贡献比较突出,其他特征则差异不大。

本研究区中,地形特征对沙质荒漠、轻度盐碱荒漠、重度盐碱荒漠、岩石荒漠、高山寒漠的分类精度提升作用最大,雷达特征则能更好地识别砾质荒漠与壤土荒漠。

参考文献:

- [1] Li C, Fu B, Wang S, et al. Drivers and impacts of changes in China's drylands [J]. *Nature Reviews Earth & Environment*, 2021, 2(12): 858–873.
- [2] 程磊磊, 郭浩, 卢琦. 荒漠生态系统服务价值评估研究进展 [J]. *中国沙漠*, 2013, 33(1): 281–287.
- [3] Huang J, Zhang G, Zhang Y, et al. Global desertification vulnerability to climate change and human activities [J]. *Land Degradation & Development*, 2020, 31(11): 1380–1391.
- [4] 杨发相, 桂东伟, 岳健, 等. 干旱区荒漠分类系统探讨: 以新疆为例 [J]. *干旱区资源与环境*, 2015, 29(11): 145–151.
- [5] 王涛, 朱震达. 我国沙漠化研究的若干问题: 1. 沙漠化的概念及其内涵 [J]. *中国沙漠*, 2003, 23(3): 3–8.
- [6] Liu L, Zhang X, Gao Y, et al. Finer-resolution mapping of global land cover: recent developments, consistency analysis, and prospects [J]. *Journal of Remote Sensing*, 2021(1): 1–38.
- [7] Zhang M, Ma M, De Maeyer P, et al. Uncertainties in classification system conversion and an analysis of inconsistencies in global land cover products [J]. *ISPRS International Journal of Geo-Information*, 2017, 6(4): 112.
- [8] 许晓聪, 李冰洁, 刘小平, 等. 全球2000–2015年30 m分辨率逐年土地覆盖制图 [J]. *遥感学报*, 2021, 25(9): 1896–1916.
- [9] 姚爱冬, 车腾腾, 姜丽娜, 等. 甘肃民勤县荒漠化区未利用地的遥感分类研究 [J]. *林业科学研究*, 2014, 27(2): 195–200.
- [10] 杨发相, 李生字, 岳健, 等. 新疆荒漠类型特征及其保护利用 [J]. *干旱区地理*, 2019, 42(1): 12–19.
- [11] 肖森天, 依力亚斯江·努尔麦提, 努尔比耶·穆合塔尔, 等. 基于光学和雷达多源遥感的于田绿洲土壤盐渍化时空分析 [J]. *干旱区研究*, 2023, 40(1): 59–68.
- [12] Zhang J. Multi-source remote sensing data fusion: status and trends [J]. *International Journal of Image and Data Fusion*, 2010, 1(1): 5–24.
- [13] Joshi N, Baumann M, Ehammer A, et al. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring [J]. *Remote Sensing*, 2016, 8(1): 70.
- [14] Halmy M W A, Gessler P E. The application of ensemble techniques for land-cover classification in arid lands [J]. *International Journal of Remote Sensing*, 2015, 36(22): 5613–5636.
- [15] Phan T N, Kuch V, Lehnert L W. Land cover classification using Google Earth Engine and random forest classifier: the role of image composition [J]. *Remote Sensing*, 2020, 12: 2411.
- [16] Gorelick N, Hancher M, Dixon M, et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone [J]. *Remote Sensing of Environment*, 2017, 202: 18–27.
- [17] 张华, 安慧敏. 基于GEE的1987–2019年民勤绿洲NDVI变化特征及趋势分析 [J]. *中国沙漠*, 2021, 41(1): 28–36.
- [18] Wang L, Diao C, Xian G, et al. A summary of the special issue on remote sensing of land change science with Google Earth Engine [J]. *Remote Sensing of Environment*, 2020, 248: 112002.
- [19] Yang X, Zhang K, Jia B, et al. Desertification assessment in China: an overview [J]. *Journal of Arid Environments*, 2005, 63(2): 517–531.
- [20] 鲍锋, 董治宝, 张正偲. 柴达木盆地风沙地貌区风况特征 [J]. *中国沙漠*, 2015, 35(3): 549–554.
- [21] Han J, Wang J, Chen L, et al. Driving factors of desertification in Qaidam Basin, China: an 18-year analysis using the geographic detector model [J]. *Ecological Indicators*, 2021, 124: 107404.
- [22] Fu Y, Chen H, Niu H, et al. Spatial and temporal variation of vegetation phenology and its response to climate changes in Qaidam Basin from 2000 to 2015 [J]. *Journal of Geographical Sciences*, 2018, 28: 400–414.
- [23] 俞孔坚, 李海龙, 李迪华, 等. 国土尺度生态安全格局 [J]. *生态学报*, 2009, 29(10): 5163–5175.
- [24] 王晓峰, 勒斯木初, 张明明. “两屏三带”生态系统格局变化及其影响因素 [J]. *生态学杂志*, 2019, 38(7): 2138–2148.
- [25] Filippini F. Sentinel-1 GRD preprocessing workflow [C]// *Proceedings of the International Electronic Conference on Remote Sensing*, 2019: 11.
- [26] Van Zyl J J. The Shuttle Radar Topography Mission (SRTM): a breakthrough in remote sensing of topography [J]. *Acta Astronautica*, 2001, 48(5): 559–565.
- [27] 刘纪远. 中国资源环境遥感宏观调查与动态 [M]. 北京: 中国科学技术出版社, 1996.
- [28] 刘亚群, 吕昌河, 傅伯杰, 等. 中国陆地生态系统分类识别及其近20年的时空变化 [J]. *生态学报*, 2021, 41(10): 3975–3987.
- [29] 赵欣悦, 席海洋, 赵静, 等. 阿拉善地区土壤盐渍化的遥感反演及分布特征 [J]. *中国沙漠*, 2023, 43(1): 27–36.
- [30] Kupidura P. The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery [J]. *Remote Sensing*, 2019, 11(10): 1233.
- [31] Haralick R M, Shanmugam K, Dinstein I H. Textural features for image classification [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC(6): 610–621.
- [32] Tassi A, Vizzari M. Object-oriented lulc classification in Google Earth Engine combining SNIC, GLCM, and machine learning algorithms [J]. *Remote Sensing*, 2020, 12(22): 3776.
- [33] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45:

- 5–32.
- [34] Miao X, Heaton J S. A comparison of random forest and Ada-boost tree in ecosystem classification in east Mojave Desert [C]//Proceedings of the 2010 18th International Conference on Geoinformatics. Beijing, China, 2010: 1–6.
- [35] Huang C, Davis L S, Townshend J R G. An assessment of support vector machines for land cover classification [J]. *International Journal of Remote Sensing*, 2002, 23(4): 725–749.
- [36] Foody G M, Mathur A. A relative evaluation of multiclass image classification by support vector machines [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42(6): 1335–1343.
- [37] Shetty S. Analysis of machine learning classifiers for LULC classification on Google Earth Engine [D]. Enschede, Netherlands: University of Twente, 2019.
- [38] Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation [C]//Proceedings of the AI 2006: Advances in Artificial Intelligence; 19th Australian Joint Conference on Artificial Intelligence. Hobart, Australia, 2006: 1015–1021.
- [39] Maxwell A E, Warner T A, Fang F. Implementation of machine-learning classification in remote sensing: an applied review [J]. *International Journal of Remote Sensing*, 2018, 39(9): 2784–2817.
- [40] Abida K, Barbouchi M, Boudabbous K, et al. Sentinel-2 data for land use mapping: comparing different supervised classifications in semi-arid areas [J]. *Agriculture*, 2022, 12(9): 1429.
- [41] Mustak S, Uday G, Ramesh B, et al. Evaluation of the performance of SAR and SAR-optical fused dataset for crop discrimination [C]//Workshop on "Earth Observations for Agricultural Monitoring", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. New Delhi, India, 2019: 563–571.
- [42] Shih H, Stow D A, Tsai Y H. Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping [J]. *International Journal of Remote Sensing*, 2019, 40(4): 1248–1274.
- [43] 张亦然, 刘廷玺, 童新, 等. 基于多源遥感和机器学习方法的科尔沁沙地植被覆盖度反演 [J]. *中国沙漠*, 2022, 42(3): 187–195.
- [44] 任雨, 张勃, 陈曦东. 科尔沁沙地土地荒漠化敏感性评估 [J]. *中国沙漠*, 2023, 43(2): 159–169.
- [45] Wu W. The generalized difference vegetation index (GDVI) for dryland characterization [J]. *Remote Sensing*, 2014, 6(2): 1211–1233.
- [46] Xiao J, Shen Y, Tateishi R, et al. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing [J]. *International Journal of Remote Sensing*, 2006, 27(12): 2411–2422.
- [47] Zhou H, Fu L, Sharma R P, et al. A hybrid approach of combining random forest with texture analysis and VDVI for desert vegetation mapping based on UAV RGB Data [J]. *Remote Sensing*, 2021, 13(10): 1891.
- [48] 张志伟, 尹惠妍, 钱登峰, 等. 高寒地区荒漠分类系统初探: 以堆龙德庆区为例 [J]. *高原农业*, 2019, 3(1): 47–53.
- [49] Song Y, Zheng H, Chen X, et al. Desertification extraction based on a microwave backscattering contribution decomposition model at the dry bottom of the Aral Sea [J]. *Remote Sensing*, 2021, 13(23): 4850.
- [50] Blaschke T. Object based image analysis for remote sensing [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010, 65(1): 2–16.
- [51] Chen G, Weng Q, Hay G J, et al. Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities [J]. *GIScience & Remote Sensing*, 2018, 55(2): 159–182.
- [52] Liu D, Xia F. Assessing object-based classification: advantages and limitations [J]. *Remote Sensing Letters*, 2010, 1(4): 187–194.

Combining the GEE platform and machine learning algorithm for desert information extraction

Lu Ruijie^{1,2}, Liu Shulin¹, Kang Wenping¹, Feng Kun¹, Guo Zichen¹, Zhi Ying^{1,2}

(1. Key Laboratory of Desert and Desertification, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The rapid and accurate mapping of desert type distribution is of great significance for environmental protection and ecological restoration. However, due to the influence of spectral and resolution factors, the current research on the extraction of information about different desert types is obviously insufficient. In this study, Dulan County of Qinghai Province was selected as a typical area. The desert was classified based on GEE platform and multi-source data, and the classification performance of different classification features combined with three machine learning methods (RF, SVM and CART) was compared and evaluated. The results show that (1) RF outperforms CART and SVM, and the overall classification accuracy using the RF classifier and spectral features, radar features, terrain features and texture features is the highest, with an overall accuracy of 95.68%, a *Kappa* coefficient of 0.95, an F_M score of 94.28%, and an obtained desert area of Dulan County is 29 039 km². (2) In the assessment of feature importance scores, elevation and VH contribute more to the desert classification, while other features do not contribute much. (3) Based on the use of spectral data, radar features are the most helpful for identifying gravel and loamy deserts, while terrain features are more suitable for identifying other types of deserts.

Key words: desert classification; machine learning; GEE platform; Dulan County