ADVANCES IN FOREST FIRE RESEARCH

2022

Edited by
DOMINGOS XAVIER VIEGAS
LUÍS MÁRIO RIBEIRO

Vegetation Mapping with Random Forest using Sentinel 2- A case study for Lousã region, Portugal

Pegah Mohammadpour*[1,2]; Crismeire Isbaex [1]; Emilio Chuvieco[2]; Xavier Viegas[1]; Carlos Viegas[1]

[1] *Univ of Coimbra, ADAI, Department of Mechanical Engineering, Rua Luís Reis Santos, Pólo II, 3030-788 Coimbra, Portugal. {Pegish2014@gmail.com or pegah@adai.pt}*
[2] *University of Alcala, Department of Geology, Geography and the Environment, Alcalá de Henares, Spain*

**Corresponding author*

Keywords

Vegetation classification, Sentinel 2, Random Forest, Accuracy Assessment, Independent Variables, GLCM texture feature

Abstract

Vegetation mapping requires accurate information to allow its use in applications such as sustainable forest management against the effects of climate change and the threat of wildfires. Remote sensing provides a powerful resource of fundamental data at different spatial resolutions and spectral regions, making it an essential tool for vegetation mapping and biomass management. Due to the ever-increasing availability of free data and software, satellites have been predominantly used to map, analyze, and monitor natural resources for conservation purposes. This study aimed to map vegetation from Sentinel-2 (S2) data in a complex and mixed vegetation cover of the Lousã district in Portugal. We used ten multispectral bands with a spatial resolution of 10 m, and four vegetation indices, including Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Enhanced Vegetation Index (EVI), and Soil Adjusted Vegetation Index (SAVI). After applying Principal Component Analysis (PCA) on the 10 S2A bands, four texture features, including Mean (ME), Homogeneity (HO), Correlation (CO), and Entropy (EN), were derived for the first three Principal Components. Textures were obtained using the Gray-Level Co-Occurrence Matrix (GLCM). As a result, 26 independent variables were extracted from S2. After defining the land use classes by object-based, the Random Forest (RF) classifier was applied. The map accuracy was evaluated by the confusion matrix, using the metrics of Overall Accuracy (OA), Producer Accuracy (PA), User Accuracy (UA), and Kappa Coefficient (Kappa). The described classification methodology showed a high OA of 90.5% and Kappa of 89% for vegetation mapping. Using GLCM texture features increased the accuracy up to 2%. The ME and CO had the highest contribution to the accuracy of the classification among the GLCM textures. GNDVI had outperformed other vegetation indices in variable importance. Also, using only S2A spectral bands, especially bands 11, 12, and 2, showed a high potential to classify the map with an OA of 88%. This study showed that adding at least one GLCM texture feature and at least one vegetation index into the S2A spectral bands may effectively increase the accuracy metrics and tree species discrimination.

Notation

B_2 S2A band2
B_3 S2A band3
B_4 S2A band4
B_8 S2A band8
N Number of grey levels
P Normalized symmetric GLCM of dimension $N \times N$
$P(i, j)$ Normalized grey level value in the cell i, j of the co-occurrence matrix
d Distance
θ Direction
$\bar{x}$ Mean of P_x
$\bar{y}$ Mean of P_y
σ_x Standard deviation of P_x
σ_y Standard deviation of P_y

1. Introduction

Wildfires have wide-ranging effects on people's lives, wildlife, the natural environment, and the economy worldwide. Wildfire is a key component of the Mediterranean landscapes and ecosystems due to climate change, changes in the Land Cover and Land Use (LCLU), agriculture abandonment and expansion of forests and shrublands, expansion of Wildland-Urban Interfaces (WUI), and reduction in forest management (Xavier et al. 2006). Southern European nations such as Spain and Portugal with Mediterranean Basin (MB) forests would face the greatest fire dangers, given they are already prone to severe and frequent wildfires (Investment in Disaster Risk Management in Europe 2021, Costa H et al. 2020). Portugal experienced an unprecedented fire season in 2017, with a record of 540,000 ha burnt, 119 deaths, and millions of euros in losses and damage (Benali and Fernandes 2021, Monterio-Henriques and Fernandes 2018). Severe drought and the emergence of meteorological conditions conducive to large wildfires amplified the Portuguese 2017 fire season (Benali and Fernandes 2021). According to the Portuguese Institute of the Sea and the Atmosphere (IPMA), May 2022 was the warmest May in the past 92 years because of severe drought conditions, unusually high temperatures, and low average precipitation (IPMA 2022). Consequently, MB forests' protection from wildfire and climate change will be critical, necessitating large-scale fuel mapping and management (Kolström et al. 2011).

Conventionally, vegetation mapping has been performed using field surveys, image interpretation, and ancillary data analysis. Nowadays, remote sensing data with active and passive sensors are the main source of information for earth observation and producing up-to-date LCLU classification (Aragoneses and Chuvieco 2021, Xie et al. 2008, Chaves et al. 2020). Forest mapping and monitoring can be significantly improved by using freely available middle-resolution remote sensing data such as the Sentinel constellation. Sentinels' data provide high temporal frequency, different spatial coverage, and characterize various fuel types and their conditions with several spectral bands ((Aragoneses and Chuvieco 2021, Xie et al. 2008, Zeng et al. 2020, Kaplan 2021). Sentinel-2 mission with two satellites, Sentinel 2 level A (S2A) and Sentinel 2 level B (S2B), was launched by the European Union's Copernicus Earth Observation program of the Europe Space Agency (ESA) in 2015 and 2017, respectively. Sentinel-2 mission offers spatial resolution varying between 10 m and 60 m and a revisit frequency of 5 days (Sentinel-2 User Handbook. Vol. 2). Sentinel-2 includes three red-edged vegetation and the SWIR bands that are highly susceptible to chlorophyll content and amend distinguishing different vegetation types and LCLU classification accuracy (Chaves et al. 2020). Several studies found Sentinel-2 data with high potential in different applications such as crop classification (Hernandez et al. 2020), tree species classification (Costa et al. 2022, Wessel et al. 2018, Persson et al. 2018), mapping burned area (Pacheco et al. 2021), and forest type classification (Kaplan 2021). Most of the recent studies have shown that non-parametric machine learning approaches such as Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) have a great potential to classify heterogeneous land covers (Wessel et al. 2018, Pacheco et al. 2021, Sheykhmousa et al. 2020). Ma et al. (2017) found RF as the mostly used supervised classifier and more stable object-based image analysis (OBIA) with the highest mean accuracy of 85.81%, followed by SVM through reviewing 173 publications on supervised object-based classification. For example, Wessel et al. (2018) classified tree species based on multitemporal Sentinel-2 data with the highest overall classification accuracy of 91% for the SVM object-based algorithm. Persson et al. (2018) also studied the classification of common tree species in Sweden by applying RF classifier on multitemporal Sentinel-2 data. They achieved the highest overall accuracy of 88.2% for the combination of all spectral bands for four image dates.

Several operational programs are undergoing to produce global land cover maps, such as CORINE Land Cover (CLC), Land Change Monitoring, Assessment, and Projection (LCMAP), North American Land Change Monitoring System (NALCMS), and recently WorldCover produced by ESA at 10m resolution (Costa et al. 2018). These harmonized LCLU mappings have great potential for providing valuable information on Earth's surface; however, they face several challenges. The thematic accuracy of these global or continental land cover maps is spatially dependent on a low-frequency basis (e.g., every five years or more). Usually, they cannot comply with the user-specific requirements in space and time (Costa et al. 2018). Several researchers found the Mediterranean countries, such as Portugal and Spain, with the lowest overall accuracy, below 70% out of all European countries mapped (Costa et al. 2022, Liu et al. 2021). National land cover mapping with higher accuracy and up-to-date and detailed data can use as a complementary product for larger mapping scales (Costa et al. 2022).

SMOS (Sistema de Monitorização de Ocupação do Solo) is the land cover monitoring system for mainland Portugal which is in preparation by Directorate General for Territory (DGT) based on COS (Cartografia de Uso

e Ocupação do Solo), the traditional LCLU map of Portugal. An annual land cover cartography product of mainland Portugal (COSsim) based on sentinel-2 was started to overcome the limitation of COS. Costa et al. (Costa et al. 2022) presented an approach to map COSsim for 2018 with an overall accuracy of 81.3% using Random Forest (RF) classification and Sentinel-2 multi-temporal data. Although they achieved an accuracy of over 90% for southern Portugal, the classification overall accuracy for the center and North Portugal was below 80%, especially for the Coimbra district with approximately 75%. Their methodology was developed based on the following stages such as image classification, spatial stratification, knowledge-based rules (combination of expert knowledge and auxiliary data), intra-annual change, and manual editing. They classified this map into 13 thematic classes including the predominant species such as evergreen oaks, Eucalyptus, Maritime pine, and Stone pine (Costa et al. 2022). Costa et al. (2018) also studied a methodology with the combination of supervised and rule-based classification under the scope of the Land Use and Coverage Area frame Survey (LUCAS) to produce annual land cover statistics over Portugal from 2010 to 2015. They collected temporal Landsat data and auxiliary data such as Land Use and Land Cover map (COS2010) and Land Parcel Identification System (LPIS). ANN, SVM, and RF have been used for classification in their study and final classification with 15 classes obtained through the voting process. They reached the highest overall accuracy of 87.5% for the year 2010 and the lowest accuracy of 86.4% for the years 2014 and 2015.

The main objective of this study is to apply RF classification algorithms on S2A data to map the vegetation in the Lousa region, Portugal. The specific purpose of this study is to increase the interclass separability of forest and vegetation classes. According to the literature, no up-to-date vegetation mapping with Sentinel-2 and high overall accuracy was seen for the mentioned study area. Principal component analysis (PCA) was applied to the S2A spectral bands resampled to 10m resolution. Four common Vegetation Indices (VIs) and 12 GLCM-based textural variables, extracted from three first principal components (PCs), were integrated into the S2A spectral bands to better classify various vegetation in complex land cover. A comparison has been performed between the classification with all 26 independent variables and three different combinations of S2A spectral bands, vegetation indices, and GLCM textures. The importance of input variables has been analyzed based on Mean Decrease Accuracy (MDA) and Mean Decrease Gini (Gini). The obtained high-accuracy vegetation map can be used for biomass management and wildfire risk assessment.

2. Materials and Methodology

2.1. Study area

The study area is in the central region of Portugal. The study area is 138.4 km^2 (13840 ha) in size, latitudes 40º and 40º 3' N and longitudes 8º 09' and 8º 19' W, and is located in the municipality of Lousa in the Coimbra district of Portugal, as shown in **Figure 1**. The Lousa district has an elevation ranging from 57 m near the River Ceira in the north to 1205 m at Trevim in the south, which is the highest point of Serra da Lousa. Over 68% of the territory is below 400 meters of altitude. Slope values vary from the flat region in the northwest, with slopes typically less than 10°, to mountainous areas in the southeast, with slopes larger than 20° (Plano Municipal de Defesa da Floresta Contra Incêndios 2020).

The forest area (about 10,385 ha) is the municipality's dominant land use, accounting for approximately 75% of the municipality's total area. Agriculture land occupies 1399.27 ha. As a result, forestry and agricultural areas comprise most of the land, covering approximately 85% of the total area (Incêndios 2020). The predominant vegetation cover includes *Pinus pinaster* and *Eucalyptus globulus* species represent about 50% and 25% of the total forest species in the municipality, respectively (Incêndios 2020).

The climate of Lousa is the Mediterranean, with moderate, rainy winters and hot, dry summers. The months of autumn, winter, and early spring see the most precipitation. Their values are highly dependent on altitude and range, on average, between 1000 and 1800 mm annually. In 2017, 28 distinct wildfires burned about 4560 ha of forest in the municipality. The largest fires detected in Lousa municipality happened during the summer when temperatures average above 30 C°, and relative humidity falls below 30%, except for October 2017, which was affected by Storm Ophelia (Incêndios 2020). The severity of the Portuguese 2017 wildfires was exacerbated by a severe drought and the occurrence of meteorological conditions conducive to large wildfires (Viegas et al. 2017).

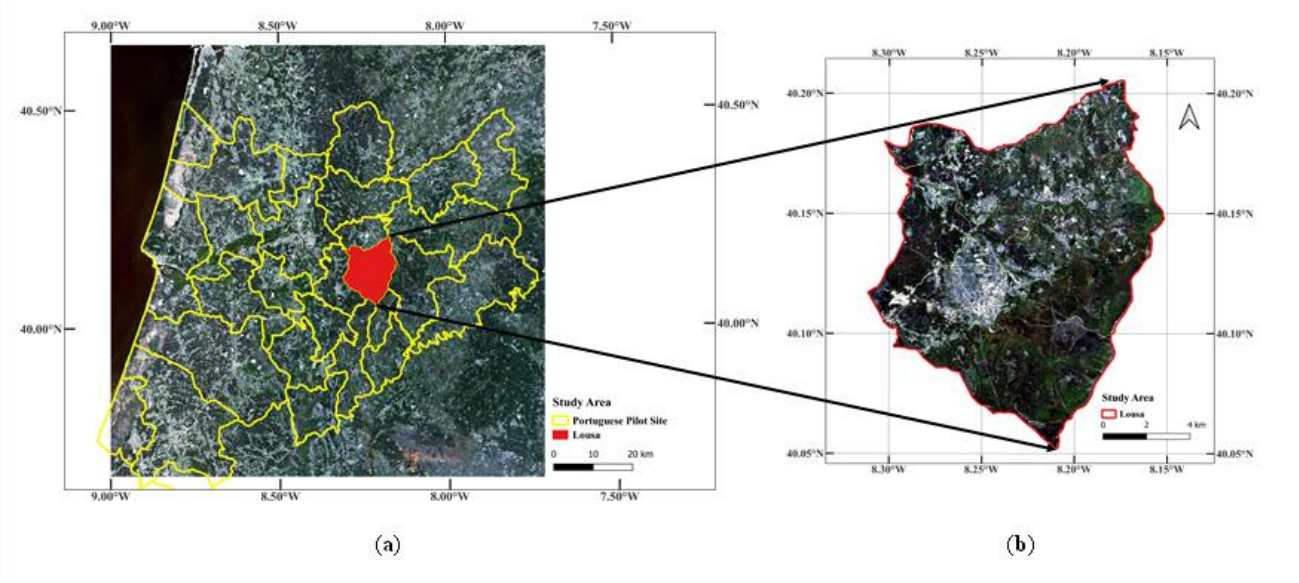


Figure 1- Geographical framework of the location and limits of the study area within a firEUrisk project and mainland Portugal: a) Sentinel-2 false-color composite image and b) study area- Lousa.

2.2. Materials, Data, and Model Algorithm

The S2A image processing procedures are synthetically represented in the workflow (**Figure 2**). Several processing stages were needed to achieve the vegetation mapping.

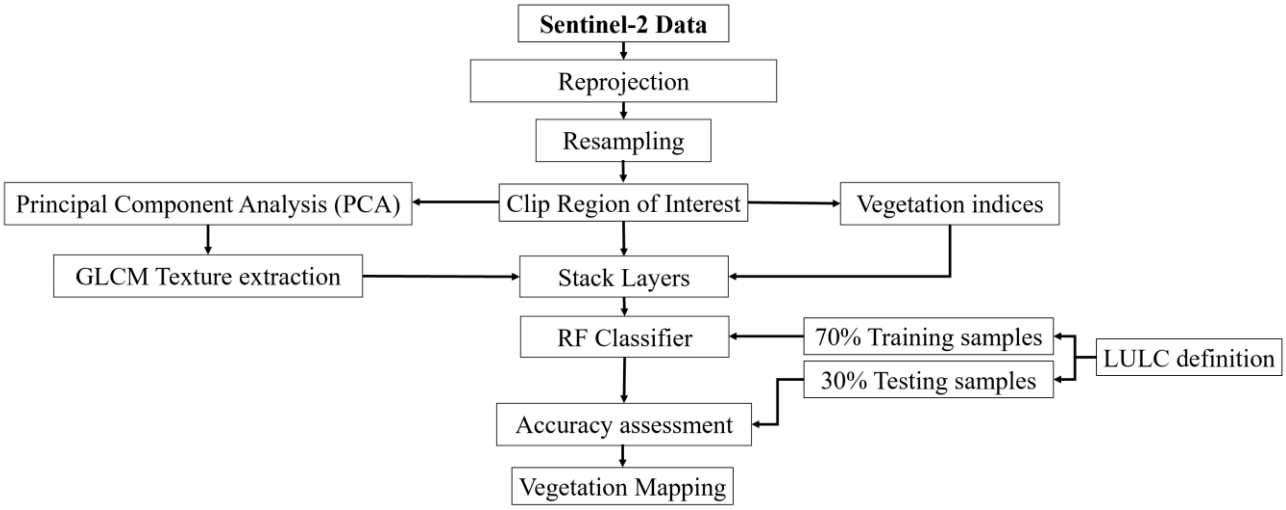


Figure 2- Scheme of vegetation mapping with Sentinel-2 data.

2.3. Dataset and preprocessing

Sentinel-2A MSI Level-2A images taken on July 18th, 2020, were derived from the Sentinels Scientific Data Hub for the study area (https://scihub.copernicus.eu/). The image was taken during the dry season to have less cloud coverage percentage (0.25%). The image was projected from WGS84 Geographic latitude/longitude coordinates to EPSG:3763 - ETRS89 / Portugal TM06, which preserves the area measure.

Sentinel-2 is equipped with MultiSpectral Instrument (MSI), which has four 10 m bands (visible and infrared – NIR), six 20 m bands (red-edge vegetation and short-wave infrared – SWRI), and three 60 m bands (Kaplan 2021). It is possible to increase the spatial resolution of all bands to 10 m using the resampling method and downscale the coarse resolution to fine resolution (Kaplan 2021, Sentinel-2 User Handbook. Vol. 2 2015, Zheng et al. 2017). Pixels of 20 m bands were resampled to the pixel size of 10 m with the Nearest Neighbor technique using ArcGIS Pro (version 2.8).

The high spectral resolution of Sentinel-2 imagery enables the extraction of different features. Based on the literature review and the sensitivity of optical features to biomass, four common spectral vegetation indices (VIs) were derived, including Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI) using ArcGIS Pro(Noi Phan et al. 2020, Xia et al. 2017). The complete list of formulas can be found in Table 1 (B2, B3, B4, and B8 are the correspondence S2A bands). NDVI is strongly related to vegetation content, and NDVI high values indicate denser and healthier vegetation. GNDVI is a modified form of NDVI that can detect varying chlorophyll concentration rates by using green reflectance instead of red reflectance (Gitelson et al. 1996). EVI was developed to increase the sensitivity of vegetation signals in high biomass locations (Bhatnagar et al. 2021). SAVI also is a vegetation index that aims to reduce the impact of soil luminance, particularly in areas with low vegetation cover (Somvanshi and Kumari 2020).

Table 1- Vegetation indices (VIs) and applications from S2A.

Vegetation Indices	Formulation	Application
NDVI	$\frac{B_8 - B_4}{B_8 + B_4}$	Detection of vegetation communities in various seasons (29) Estimating changes in vegetation state (28) Determining the density of greenness (31)
GNDVI	$\frac{B_8 - B_3}{B_8 + B_3}$	Determining water and nitrogen uptake in the crop canopy (32)(28)
EVI	$\frac{B_8 - B_4}{B_8 + (6 \times B_4) - (7.5 \times B_2) + 1} \times 2.5$	Detection of vegetation communities in various seasons (29) Land cover classification (26)
SAVI	$\frac{B_8 - B_4}{B_8 + B_4 + 0.5} \times 1.5$	Minimizing soil brightness influences (33) (30) Land cover classification (26)

In addition to VIs, the previous research proves that using textural features significantly improves classification accuracy, especially for the species with similar spectral characteristics but different spatial patterns (Gini et al. 2018). PCA is commonly used in data mining to explore data and reduce data dimensionality (Guo et al. 2002, Xiao 2010). It reduces dimensionality by identifying a new set of variables that is smaller than the original set and captures big principal variability in data while disregarding minor variability (Xiao 2010). The three first PCs with cumulative eigenvalues of more than 99% were extracted from the spectral bands in ArcGIS Pro, as shown in **Appendix A (Table A1)**.

A total of four textural variables, including Mean (ME), Entropy (EN), Homogeneity (HO or also called Inverse Difference Moment- IDM), and Correlation (CO), were derived using the first three PCs (Kattenborn et al. 2019). These textures were obtained through the Gray-Level Co-Occurrence Matrix (GLCM) statistical approach with Probabilistic Quantizer (Quantization levels of 32), the inter-pixel distance of 1, and window size of 5×5 using SNAP (version 8.0.0). The window size was optimized based on this particular study area since the texture measures are dependent on the window size (Chatziantoniou et al. 2017). The formulas of textures and their application are listed in **Table 2**. The GLCM method examines the distribution of gray levels across adjacent pixels by considering the pixels' spatial position in an image (Humeau-Heurtier 2019, Xu and Gowen 2020, Zhang et al. 2017). Humeau-Heurtier (2019) defined the GLCM $P(i,j \mid d,\theta)$ as a relative frequency of the occurrence of the same intensity value i (reference pixel) adjacent to the different intensity value j (neighbor pixel) in a specific spatial relation at the distance d and direction of θ. Based on these parameters, ME used the frequency of occurrence of a certain neighboring pixel to weight the pixel value (Held 1998). EN assesses the degree of disorder and complexity of the texture distribution in GLCM (Humeau-Heurtier 2019, Zhang et al. 2017). HO shows the level of homogeneity (Humeau-Heurtier 2019), and CO measures the linear relationship between pixel values for the grayscale in the horizontal or vertical direction (Humeau-Heurtier 2019,Xhang et al. 2017).

Table 2- Texture metrics Formula

Texture Metrics	Formulation	Application
ME	$\sum_{i}^{N}\sum_{j}^{N} i \times P(i,j\|d,\theta)$	Weighting pixel value based on the frequency of its occurrence in conjunction with a specific neighbor pixel value (42,43) Calculating the mean processing window value (44)
EN	$-\sum_{i}^{N}\sum_{j}^{N} P(i,j\|d,\theta)\log_2 P(i,j\|d,\theta)$	Assessing the disorder of the GLCM (39) Reflecting the complexity of the texture distribution (41)
HO	$-\sum_{i}^{N}\sum_{j}^{N} \frac{P(i,j\|d,\theta)}{1+(i-j)^2}$	Measuring the level of homogeneity in pairs of pixels (39)
CO	$\frac{\sum_{i}^{N}\sum_{j}^{N}(i-\bar{x})(j-\bar{y})P(i,j\|d,\theta)}{\sigma_x\sigma_y}$	Measuring grey level linear relation between pixels (39,41,45)

2.4. Reference Data

Training samples for each ground object must be selected separately since RF is a supervised classifier. The basis for the selection of training and validation samples was acquired visually from Portugal Directorate-General for the Territory (DGT; Direção-Geral do Território 2018), Portugal Institute for Nature Conservation and Forests (ICNF; Instituto da Conservação da Natureza e das Florestas 2015), and Google Earth (orthophoto maps). Vegetation category was selected and verified based on dominant species, as presented in **Table 3**, such as *Pinus pinaster*, *Eucalyptus*, *Castanea*, *Pinus pinea*, *Quercus (*including *Quercus robur*, *Quercus suber*, and *Quercus ilex)*, *Acacia*, and other land use including cropland and agriculture land, shrubland and grassland, Water and barren (area without vegetation including roads and urban area) (Incêndios C 2020). Therefore, samples' polygons were manually created based on visual interpretation and reference sources.

Table 3- Dominant species represented in ha and percentage adapted from (Incêndios C 2020).

Species	Area (ha)	Percentage of the Study Area (%)
Pinus pinaster	5190.17	37.5
Eucalyptus	2611.95	18.9
Quercus	1273.36	9.2
Castanea	510.93	3.7
Acacia	506.26	3.6
Pinus Pinea	11.18	0.1

2.5. Image Classification

A total of 26 independent variables (including 10 spectral bands, four vegetation indices, and 12 texture measures) were used as an input to supervised classification to improve the classification performance (Aragoneses and Chuvieco 2021). RF algorithm was applied as a machine learning algorithm for vegetation classification in RStudio (version 2022.02.3+492). RF algorithm, an ensemble classifier that produces multiple decision trees, is commonly used in LULC due to its' high accuracy results achieved, solving highly non-linear problems on relatively small-size databases, and handling a large number of input features (Kluczek and Zagajewski 2022, Wójtowicz et al. 2021, Adeli et al. 2022). Furthermore, RF enables input feature ranking through random permutation, which has been studied in this paper using the RandomForest package in R (Zheng et al. 2017). The number of trees (ntree) and the number of random samples per node (mtry) were 500 and five, respectively. As shown in **Appendix A (Figure A1**), by increasing the ntree, out-of-bag (OOB) Error decreases and becomes stable at around 200 (Feng et al. 2015, Hanes et al. 2022). Researchers showed satisfactory results with the default parameter for the RF classifier which is ntree of 500 trees (Immitzer et al. 2016,Wang et al.

2015). In addition, a mtry that is excessively small or large reduces the individual tree's prediction ability. The optimum value for mtry calculates through 1/3 or the square root of the number of input variables (43,53). In this paper, the optimum mtry was calculated as five through the tuneRF function in R, as it is shown in **Appendix A (Figure A2)**. Samples were randomly subdivided into 70% for the training of the classification algorithm and 30% to validate the model. Finally, an accuracy assessment of vegetation classification has been performed by calculating the confusion matrix.

2.6. Accuracy Assessment

Accuracy assessments have traditionally relied on a confusion matrix (Stehman and Foody 2019). The confusion matrix measures OA, PA, UA, and Kappa to describe the fitness between the generated classes and the reference data (Xie et al. 2008, Sheykhmousa et al. 2020). OA is directly correlated to the percentage of the study area that is correctly classified, and the Kappa is also measuring the performance of RF (Stehman and Foody 2019, Landis and Koch 1977). OA is a relatively coarse measurement since it does not provide information on class-specific accuracy, while PA and UA respectively provide class-specific accuracy on the reference and classified area per class (Stehman and Foody 2019). The AO, PA, and UA may consider high precision with accuracy above 79% (56), and the Kappa greater than 80% represents a high degree of agreement (Landis and Koch 1977).

3. Results and Discussion

A total of 26,100 pixels were selected for the training and validation samples by visual interpretation using reference sources, including DGT, ICNF, and Google Earth. The effective sample size is suggested to be between 0.2% to 3% of the total dataset pixels (Blatchford et al. 2021). The sampling size of this study is approximately 1.3% of the total pixels. Increasing the training sample size generally boosts classification accuracy. Many researchers have found a positive correlation between classification accuracy and sample size (Ma et al. 2017). According to Moraes et al.(2021), there is no recommended minimum sample size, and the sufficient sample size depends on several parameters such as the classifier, predictor variables, class definition, and size and spatial features of the study area. They analyzed the influence of sample size on the LCLU in the North of Portugal using S2 data and the RF classifier. They compared OA for the sample size of 50 to 6000 per class. They found a similar OA value with a variation of 2%, and the highest value of 73.7% indicated the low sensitivity of the RF classifier to the sample size (Moraes et al. 2021). According to the actual occupied area in (Incêndios C 2020), Pinus pinaster and Eucalyptus have the greatest number of pixels, and other classes have a similar proportion of actual occupied. Samples were divided into 70% (17394 pixels) as training samples and 30% (8706 pixels) as validation samples using Stratified random sampling in R. **Appendix B (Table B1)** shows the total of 10 classes for this study.

Accuracy assessment of the classification was evaluated based on the confusion matrix and OA, PA, UA, and Kappa metrics. The rows and columns of the confusion matrix shown in **Table 4** indicate the map classification and the reference classification, respectively. Diagonal cells of the matrix show the correct classifications, and off-diagonal cells indicate misclassifications. The repetition of 10 times was conducted for an unbiased evaluation, and the mean accuracy metrics are presented in **Table 5**. OA 90.9% (±1) and Kappa of 89% (±1) have achieved for this classification, indicating a high classification accuracy. The lowest and highest PA was obtained for *Acacia* and Water, respectively. The other classes have a high accuracy of over 90%, except for *Quercus,* with a PA of 79.5%. In the case of UA, barren has the highest accuracy of 97.1%, while *Acacia* has the lowest with a value of 70.1%. Predominant species of *Pinus pinaster* and *Eucalyptus* has a high UA and PA of over 90%.

Table 4- Confusion Matrix and statistical measures for RF vegetation classification.

		Reference										**Total**	**PA**
Classification	**Classes**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	N° Pixels	%
	1	**2136**	40	36	18	63	0	0	5	4	0	2302	92.8
	2	101	**2052**	13	11	4	11	26	51	1	7	2277	90.1
	3	32	10	**580**	64	7	0	17	20	0	0	730	79.5
	4	24	2	41	**333**	1	0	0	1	0	0	402	82.5
	5	30	2	6	9	**176**	0	0	0	0	0	223	78.8
	6	0	0	0	0	0	**43**	0	3	0	1	47	91.5
	7	10	5	5	6	0	2	**638**	1	0	17	684	93.3
	8	19	21	13	1	0	3	11	**1083**	0	0	1151	94.1
	9	0	0	0	0	0	0	0	0	**39**	0	39	100
	10	0	2	1	0	0	0	9	0	0	**831**	843	98.6
Total	N° Pixels	2352	2134	695	442	251	59	701	1164	44	856	**8706**	
UA	%	90.8	96.2	83.5	75.3	70.1	72.9	91	93	88.6	97.1		
OA	**90.9%**												
K	**89%**												

Figure 3 shows the classified vegetation map based on the proposed methodology. **Figure 4** represents the area and percentage of each class in the classified map and reference data. According to the result, the largest area is occupied by *Pinus pinaster* followed by *Eucalyptus* mostly in the south and north of the study area, respectively. However, the Pinus pinaster has the majority in the reference data and occupied approximately 37% of the study area (Incêndios C 2020,DGT 2018,ICNF 2016). Since the reference data were based on COS2018 (DGT 2018), they were not up-to-date and accurate. It is necessary to perform a field measurement to detect the changes and assess the final classification of this paper. Land cover changes may happen due to wildfire or cutting trees either for commercial purposes by private land owners or forest management. In addition, the study area is very heterogeneous and has several areas with mixed species that it is also hard to distinguish visually using Google Earth. Also, Cropland/Agriculture land class has a greater area than the reference data mentioned in **Section2.1**. Some of these misclassifications were due to the new plantation field of tree species and confusion with the Shrubland/Grassland class. Both of these fields have a similar pattern and spectral features of Cropland/Agriculture land class. The other six classes' areas predicted similar results compared to the reference data with an error of less than 3%, referring to **Appendix B (Table B1)** for more information. According to Costa et al. (2022), there is always expected to have spectral confusion between shrubland and tree species or grassland and agricultural land because of similar spectral features of some classes. They also mentioned the broad conceptual definition of the classes as another factor of misclassification (Costa et al. 2022). Shrubland/Grassland class may have a broad conceptual definition since it includes natural, semi-natural, and agricultural grassland and also shrubs. Costa et al. (2022) also found several classification contradictions between their LCLU classified map and COS2018 (DGT 2018) for the land cover of mainland Portugal. They presented a significant increase in the occupied area of shrubland and grassland as one of the major land covers of Portugal in comparison with COS (DGT 2018). Problematic classes may also result from the fragmented species distribution, either rare or mixed with other classes, especially for tree stands (Immitzer et al. 2016). The major reasons for different classification statistics may relate to the higher resolution imagery of S2A with 10 m compared with traditional maps, including COS, class definition, and impact of the 2017 wildfire and burnt area on land cover changes (DGT 2018). Immitzer et al. (2016) mentioned the effect of a parameter such as a stand age and density, crown coverage, and understory on the spectral behavior of the tree species and their intra-class separability. Several Pine and Eucalyptus stands have been seen in this study area with different ages due to the plantation or cutting. All these effective factors have been considered in this study while comparing the result with the reference data.

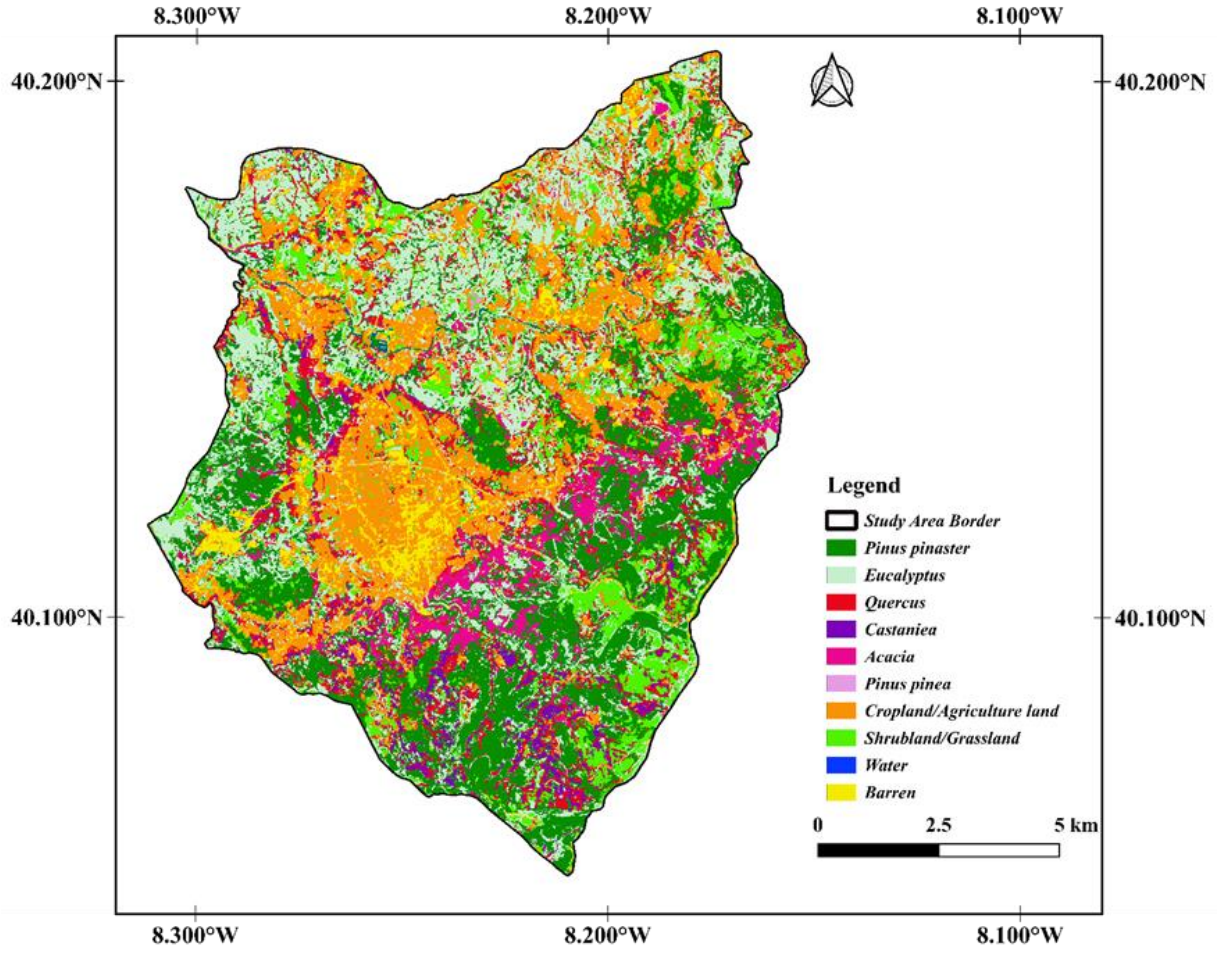


Figure 3- Vegetation map generated based on 26 independent variables of S2A imagery using RF classification for Lousa, Portugal.

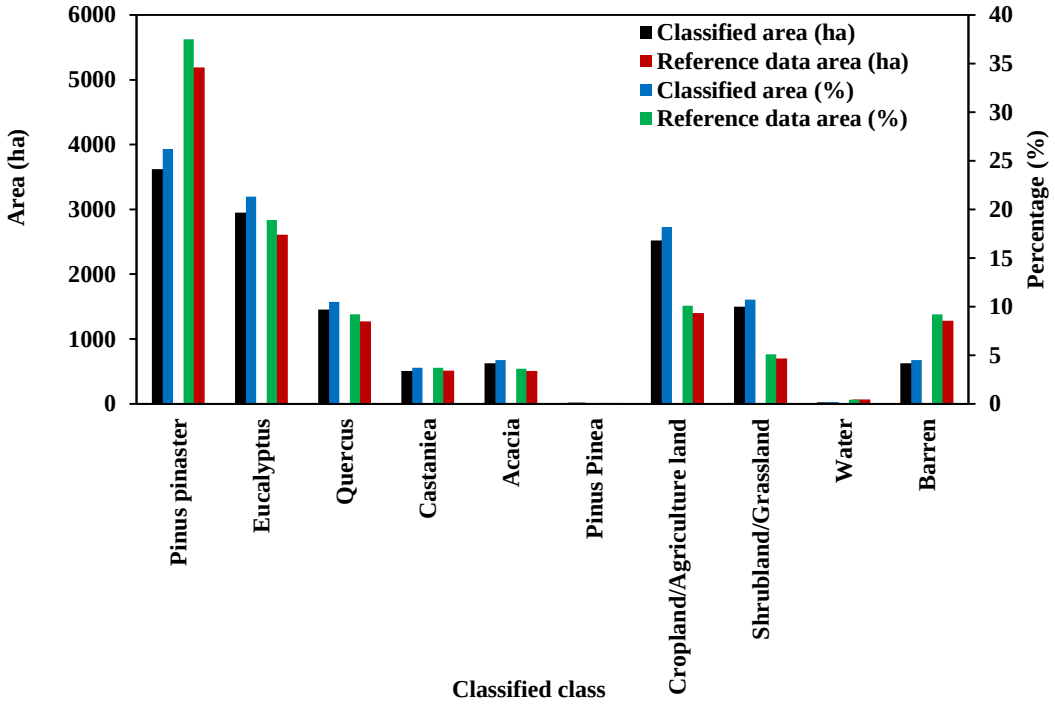


Figure 4- Comparison of classification result and reference data represented in the area (ha) and percentage.

3.1. Importance of Independent Variables

In order to assess the importance of each variable in RF classification accuracy, MeanDecreseAccuracy (MDA) and MeanDeacreseGini (Gini) have been used through random permutation. **Figure 5 (a) and (b)** show the ranking of the variables based on their contribution to distinguishing the classes. This ranking was almost the same for all the repetitions. Band 11 and 12 (SWIR) and band 2 (Blue) have the highest MDA and Gini among the S2A spectral bands, although the NIR band of 8 achieved low scores. In the case of the GLCM textures, the ME and CO of the PC3 have the highest rank among all 26 variables in MDA. Generally, the ME and CO for the PC1 and PC3 were ranked among the top five variables in both MDA and Gini. Additionally, GNDVI and NDVI have the highest contribution among VIs.

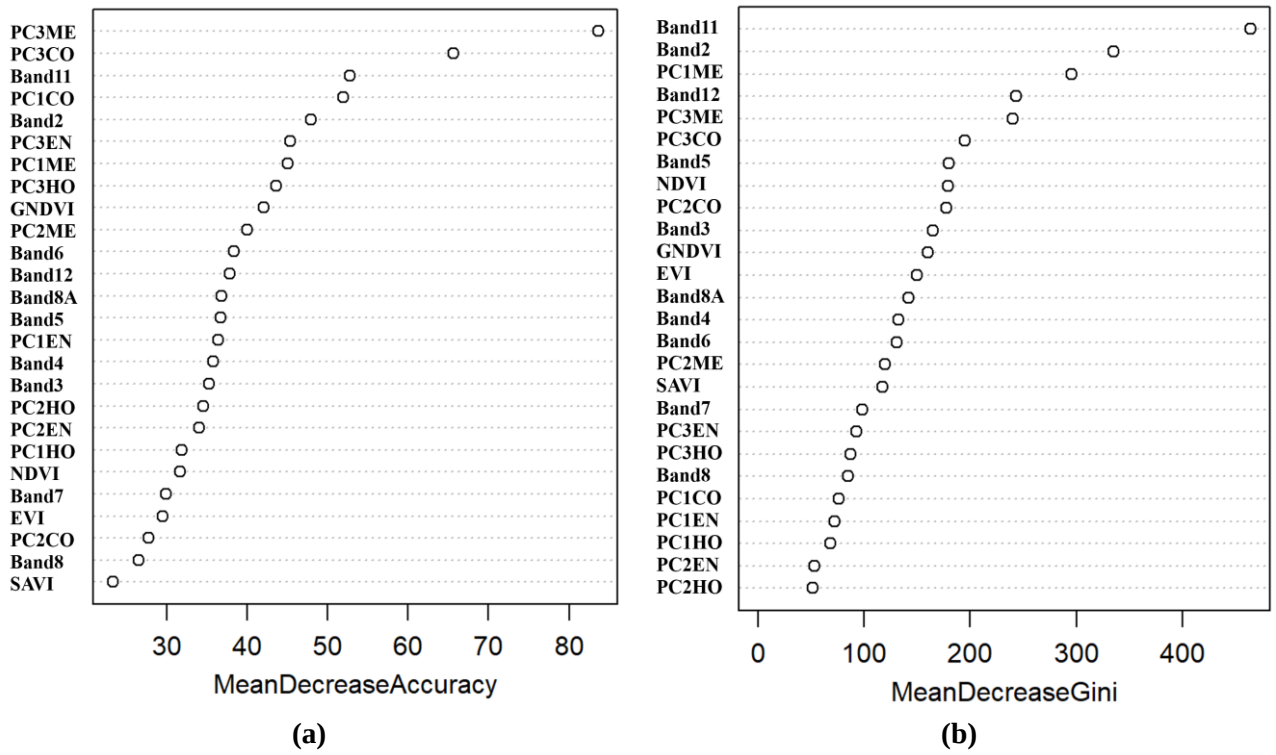


Figure 6 - Band Importance ranking based on mean decrease accuracy and mean decrease Gini.

The methodology has been applied to different independent variables input to analyze the importance of the GLCM textures. The classification was repeated for three additional scenarios of input bands: (I) Only S2A spectral bands, (II) S2A spectral bands and VIs, and (III) S2A spectral bands and GLCM textures. The Rf classifier based on S2A spectral band and GLCM texture achieved the maximum OA and Kappa, as presented in **Table 7**. Using all 26 variables has ranked second with a difference of 1.2%, lower than the combination of spectral bands and GLCM textures. Wang *et al. (2015)* also achieved high accuracy of 97.1% by using only GLCM feature on IKONOS multispectral bands, which is higher than the accuracy of the combination of all spectral bands and GLCM and GI features (96.9%). They believed that lower accuracy of all bands is associated with increasing omission and commission errors of some Gi features. In this paper, the lower accuracy of classification with all bands may cause by vegetation indices. However, the combination of spectral bands and four vegetation indices yielded higher accuracy than only spectral bands and improved the accuracy. The minimum OA and Kappa also were obtained for employing only S2A spectral bands. Adding GLCM texture to S2A spectral bands has increased the accuracy metrics by 4%.

Table 6- OA and Kappa for RF classification with different combinations of independent variables.

	All bands	**Spectral bands**	**Spectralbands+ VIs**	**Spectralbands+ GLCM texture**
OA (%)	90.8	88	88.6	92
Kappa (%)	89	86	86.1	90.2

Appendix B (Figures B1) shows the band importance for these three additional classifications. In all of the four scenarios of input variables, band11 and band 2, ME and CO of PC1 and PC3 outperformed other variables with the high value of MDA and Gini. It can be concluded that adding at least one GLCM texture may improve the classification accuracy. In this paper, the ME and CO have enhanced the RF classification with higher accuracy. Several studies have indicated the importance of texture feature extraction in increasing the accuracy of the classified map (Zheng et al. 2017, Zhang et al. 2017,Adeli et al. 2022). Wang et al. (2015) found the ME as the most important variable in mapping forest health conditions using IKONOS imagery. Zhang et al. (2017) emphasized the importance of GLCM texture feature extraction on classification accuracy improvement. These textures provide information about different objects with the same spectral, while spectral bands provide the data for the same objects in the different spectrums (Zheng et al. 2017). Zheng et al. (2017) have suggested using Extended Attribute Profiles (EAPs) in addition to GLCM textures, which can improve shrubland, agricultural land, and barren discrimination. Furthermore, S2A spectral bands have shown a high potential in the LCLU classification of the regional study area of this paper without using additional variables, which were also mentioned previously in several research studies (Chaves et al. 2020,Zheng et al. 2017, Immitzer et al. 2016). In this paper, bands 11 (SWIR) and band 2 (blue) were almost among the five first important variables in all combinations. Immitzer et al. (2016) found bands 2 and 11 as the highest score while NIR band 8A was the lowest score for forest species classification. Feng et al.(2015) declared that spectral features contribute the most in their study with the UAV for flood mapping. In their study, the red and blue bands followed by ME were the most important variables.

3.2. Effect of Bootstrap Sample Size

Rf classifier uses bootstrap aggregation, randomly creating different training subsets of the original training dataset to grow the individual regression trees (Immitzer et al. 2016, Wang et al. 2015). Wang et al.(2015) suggested using two-thirds of the validation dataset per subset for bootstrap sample size. The minimum value of the bootstrap sample size in this paper was equal to the minimum class pixels (Water), and the largest sample size was approximately equal to the extraction of highest and lowest class pixels (*Pinus pinaster* and Water). As shown in **Figure 7**, the lowest and highest OA and Kappa were calculated for the sampling size of 137 and 7000, respectively. The highest OA was increased by around 15.7%, and Kappa was increased by 19.3%. According to UA and PA, the lowest UA and PA has appeared in the sample size of 137 with zero percentage of water and *Pinus pinea* detection; more information was presented in **Appendix B (Table b2)**. The minimum PA remained below 70% until the sample size of 5000. Thomlinson et al.(1999) proposed the minimum OA and per-class accuracy of 85% and 70% for high precision classification, respectively. Although the bootstrap sample size of 6000 and 7000 has increased OA and Kappa by approximately 1%, the sample size of 5000 was used for this research to avoid overfitting and high computational process based on previous studies (Zheng et al. 2017, Thomlinson et al. 1999). 5000 bootstrap sample size is also equal to two-thirds of the validation dataset in this paper. Furthermore, the lowest and highest OOB error were obtained for the samples of 5000 and 137, respectively.

This paper has achieved high value for the accuracy metrics based on a single image of medium resolution S2A multispectral bands compared with previous studies that used multi-temporal or high-resolution imagery (Costa et al. 2022, Immitzer et al. 2016). The main purpose of this study was to increase the interclass separability of forest and vegetation classes and to provide up-to-date data based on S2A for the Lousa region, Portugal. However, the major limitation is the lack of recent reference data or fieldwork in Portugal. To better analyze the accuracy of the classified map, field measurement data and reliable and up-to-date ground truth data as reference data are needed. For future study, adding Light Detection and Ranging equipment (LiDAR) data, UAV imagery with finer resolution, Sentinel-1 SAR data, and knowledge of local experts are expected to increase the discrimination of the LCLU classes and provide more information on vertical features of the species. Recently, Portugal has started capturing LiDAR data for the whole country. Nevertheless, it is not yet covered in the study area of this paper.

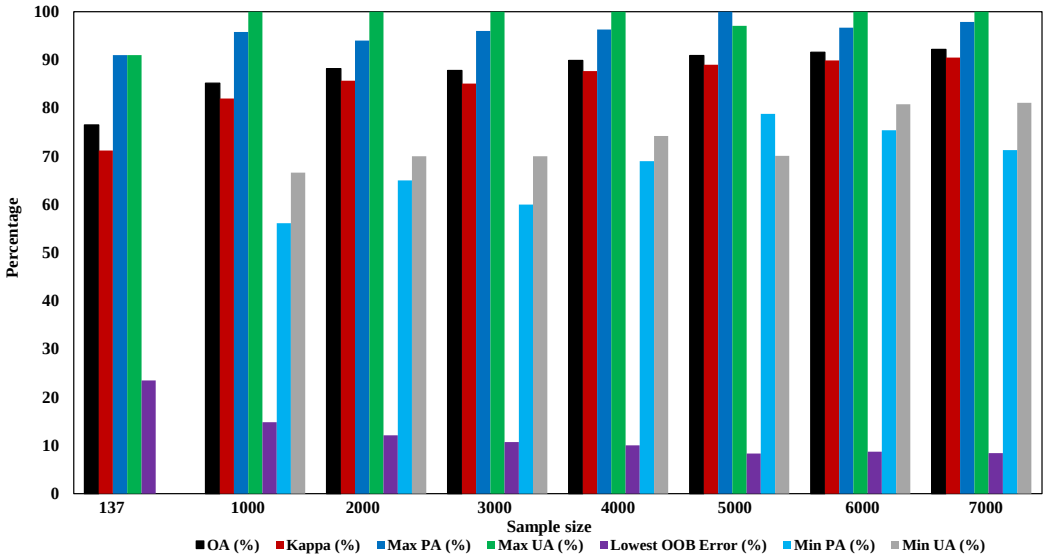


Figure 7- Effect of sample size on OA, UA, PA, Kappa, and OOB Error in RF classification.

4. Conclusion and Future Works

This paper proposed a methodology for vegetation mapping by applying a Random Forest classifier on the Sentinel2-A image. Twenty-six independent variables were used by combining ten S2A multispectral bands resampled to 10 m pixel size with four vegetation indices and 12 GLCM texture feature variables. These GLCM texture features were calculated from the three first principal components of downscaled S2A spectral bands using PCA. Training and validation data was acquired through forestry data of ICNF and DGT and visual interpretation using Google Earth. RF classifier with 500 *ntree* and 5 *mtry* was used to classify the study area into ten classes, including the predominant tree species. A high precision classification was performed with an overall accuracy of 90.5% and a Kappa coefficient of 89%. Nevertheless, employing all 26 variables did not achieve the highest accuracy compared to the combination spectral bands and GLCM textures with an OA of 92%. The comparison of classification accuracy for four different combinations of input variables shows the importance of GLCM textures especially ME and CO in improving the performance of the RF classifier. Spectral bands 11 and 2 also have high scores in MDA and GINI for all the scenarios. Generally, S2A 10m spectral bands have a strong potential for precise and rapid vegetation classification. GNDVI contributes the most to the accuracy of classification amongst vegetation indices. The results of this study indicate that incorporating at least one GLCM texture feature and at least one vegetation index into the S2A spectral bands can effectively improve the accuracy of assessment and tree species classification.

Few studies explored vegetation mapping with high accuracy in Portugal, mainly in the Lousa region, using Sentinel 2 images. Exploring the multispectral band and texture feature analysis can contribute to the spectral separability of vegetation, facilitating studies of aboveground biomass estimation and wildfire simulation. Accuracy assessment of the implemented methodology has shown a high potential for generating high-accuracy vegetation mapping over forestry areas. It is foreseen to expand the methodology for the data fusion of Sentinel 2 and Sentinel 1 to include data regarding the vertical structure of the vegetation and canopy. This data will enable the creation of a fuel type tridimensional semantic map to be used by fire behavior models and allow the definition of biomass management strategies towards wildfire risk reduction or accurate fire spread simulations for the decision support during wildfire occurrences. These results are relevant for the ESA 2025 Agenda, whose challenge is integrating information from Copernicus mapping carbon to promote sustainable land management and improved resilience against natural risks and impacts of climate change. Therefore, more studies are necessary to be conducted in order to cover more datasets in Portugal.

5. Acknowledgments

This work is financed by national funds through FCT - Foundation for Science and Technology, under grant agreement 2021.05971.BD attributed to the 1st author. The authors gratefully acknowledge the Portuguese

Foundation for Science and Technology (FCT) for its support under the framework of the research projects PCIF/SSI/0096/2017 – FIREFRONT and IMFire - Intelligent Management of Wildfires, ref. PCIF/SSI/0151/2018, fully funded by national funds through the Ministry of Science, Technology, and Higher Education and project FirEURisk – Developing a Holistic, Risk-Wise Strategy for European Wildfire Management, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003890.

6. Author Contributions

Pegah Mohammadpour: Writing – review & editing, Writing – original draft, Data processing, Map generation, Visualization, Validation, Software, Methodology and study design, Formal analysis, Data curation**.** Crismeire Isbaex: Writing – review & editing**.** Emilio Chuvieco: Writing – review & editing, Supervision**.** Xavier Viegas: Writing – review & editing, Supervision**.** Carlos Viegas: Writing – review & editing, Supervision.

7. Conflicts of Interest

The authors declare no conflict of interest.

Appendix A

Table A1- Percent and accumulative eigenvalues of three first PCs.

PC Layer	Percent of Eigenvalues (%)	Accumulative of Eigenvalues (%)
PC1	92.5	92.5
PC2	6.3	98.8
PC3	0.6	99.4

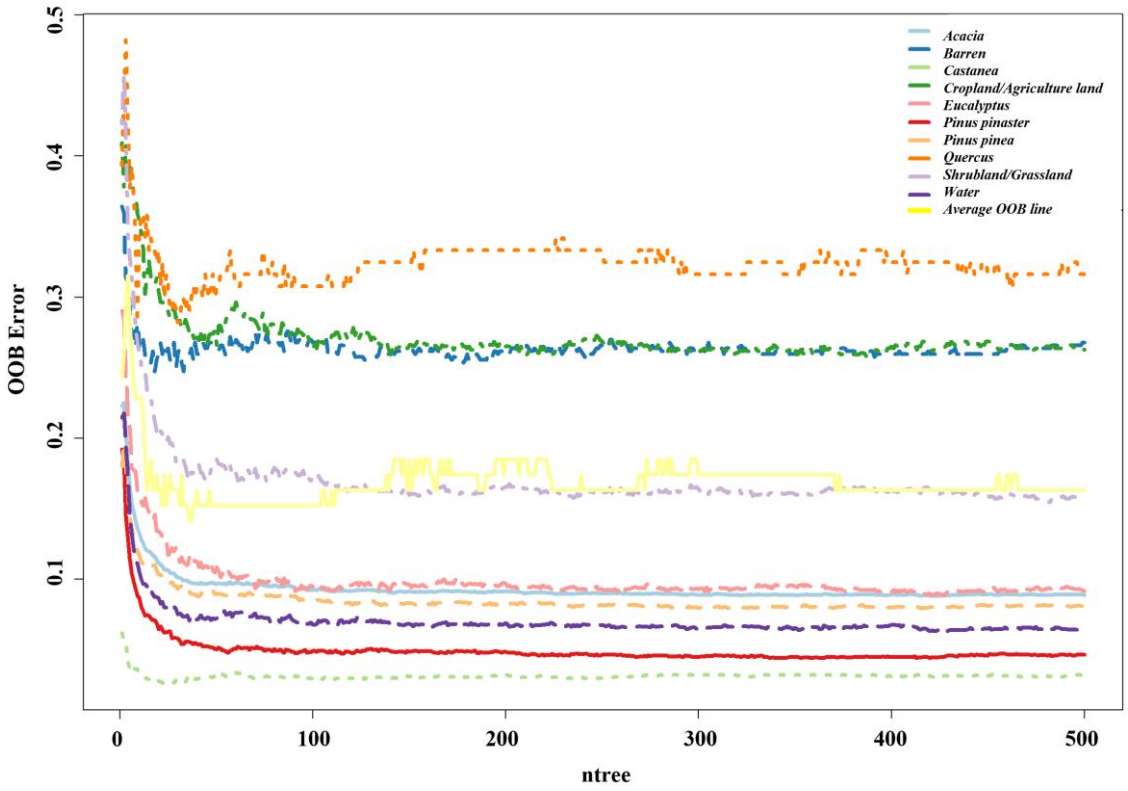


Figure A1- OOB Error of the number of RF trees, ntree.

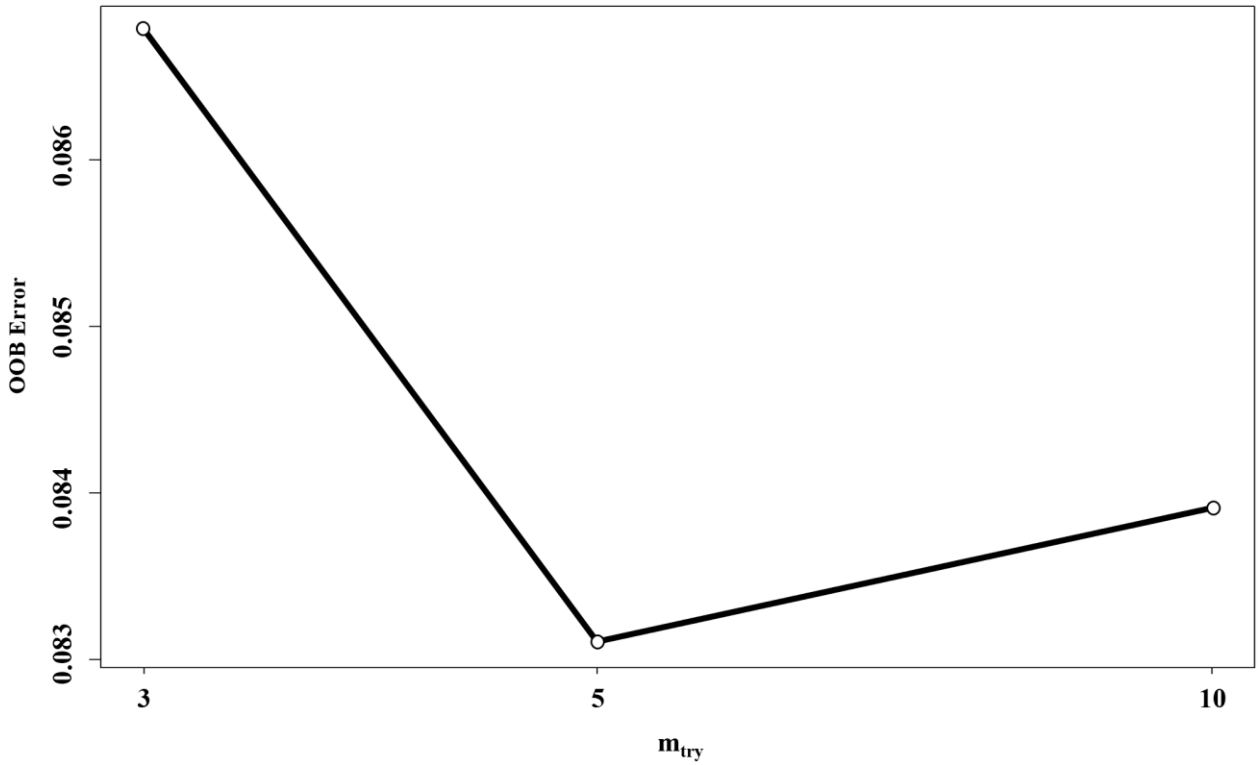
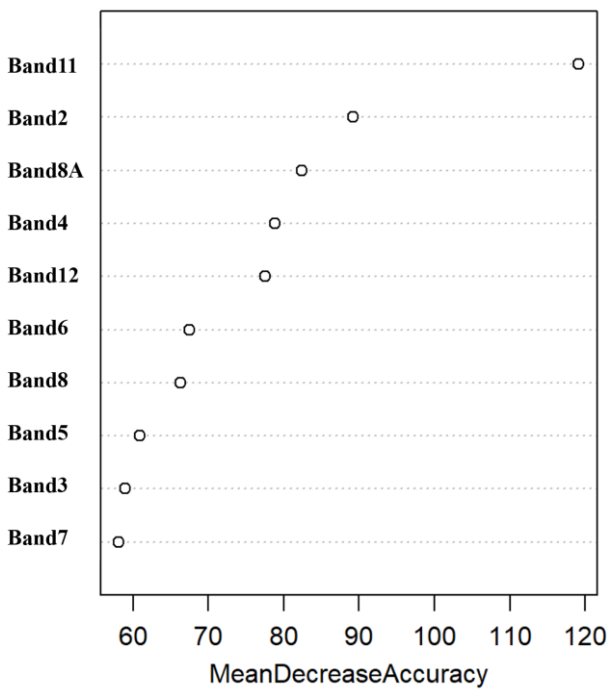


Figure A2- OOB Error of the number of RF mtry.

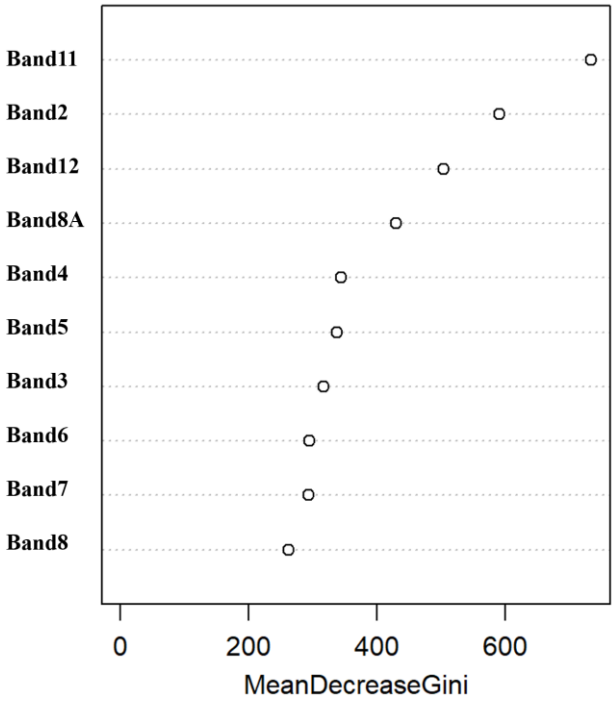
Appendix B

Table B1- Classification result represented in the area (ha) and percentage.

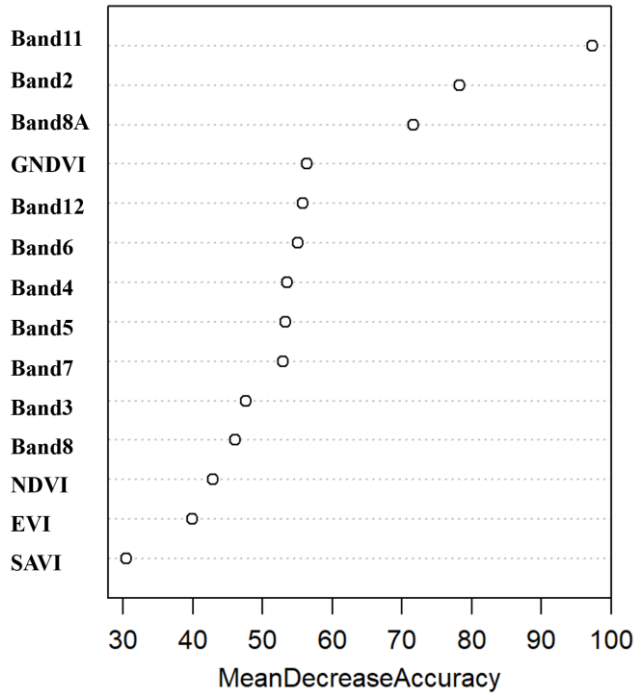
Class ID	Classified Class	Classified area (ha)	Percentage (%)	The difference with reference data area (ha)	The difference with reference data area (%)
1	*Pinus pinaster*	3619.1867	26.2	-1570.9833	-11.3
2	*Eucalyptus*	2951.4365	21.3	+339.4865	+2.4
3	*Quercus*	1453.5870	10.5	-180.227	-1.3
4	*Castanea*	506.3817	3.7	-4.5483	-0.0
5	*Acacia*	626.2897	4.5	+120.0297	+0.9
6	*Pinus Pinea*	17.1625	0.1	+5.9825	+0.0
7	Cropland/Agriculture land	2518.3886	18.2	+1119.0689	+8.1
8	Shrubland/Grassland	1499.6008	10.7	+798.0708	+5.6
9	Water	21.9260	0.2	-42.914	-0.2
10	Barren	625.9895	4.5	-655.2805	-4.7


Band11
Band2
Band8A
Band4
Band12
Band6
Band8
Band5
Band3
Band7
60
70
80
90
100
110
120
MeanDecreaseAccuracy

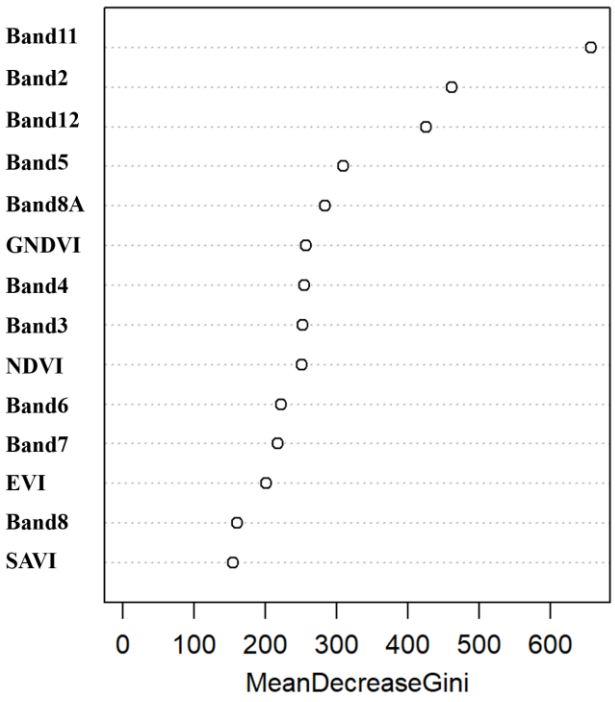
(a)


Band11
Band2
Band12
Band8A
Band4
Band5
Band3
Band6
Band7
Band8
0
200
400
600
MeanDecreaseGini

(b)


Band11
Band2
Band8A
GNDVI
Band12
Band6
Band4
Band5
Band7
Band3
Band8
NDVI
EVI
SAVI
30
40
50
60
70
80
90
100
MeanDecreaseAccuracy

(c)


Band11
Band2
Band12
Band5
Band8A
GNDVI
Band4
Band3
NDVI
Band6
Band7
EVI
Band8
SAVI
0
100
200
300
400
500
600
MeanDecreaseGini

(d)

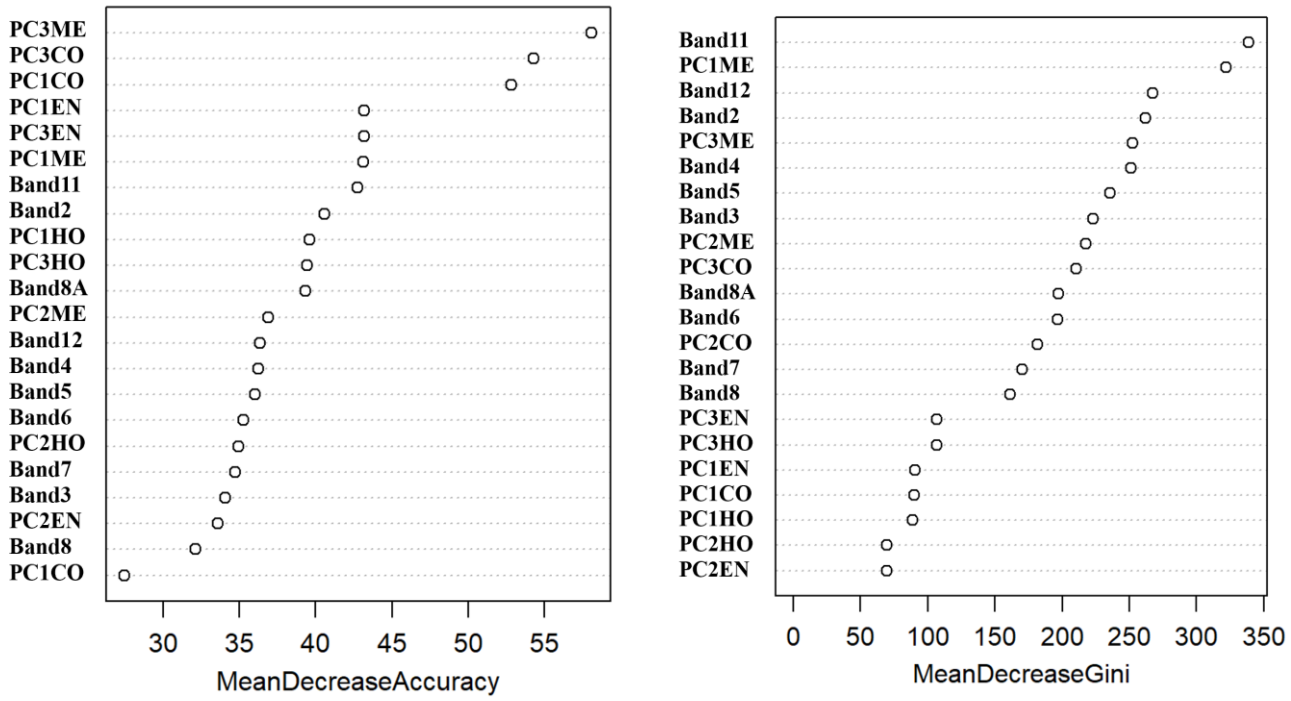


Figure B1- Mean decrease accuracy and mean decrease Gini for three different combinations of variables: (I) Spectral bands (a)-(b), (II) Spectral bands + VIs (c)-(d), and (III) Spectral bands+ GLCM texture (e)-(f).

Table B2- Effect of sample size on OA, UA, PA, and Kappa in RF classification.

Sample Size	**137**	**1000**	**2000**	**3000**	**4000**	**5000**	**6000**	**7000**
OA (%)	76.5	85.2	88.2	87.8	89.9	90.9	91.6	92.2
Kappa (%)	71.2	82	85.7	85.1	87.7	89	89.9	90.5
Max PA (%)	91 Barren	95.8 Barren	94 *Eucalyptus*	96 Barren	96.3 Barren	100 water	96.7 Barren	97.9 Barren
Min PA (%)	0 Water & *Pinus pinea*	56.1 *Pinus pinea*	65 *Acacia*	60 *Acacia*	69 *Castanea*	78.8 *Acacia*	75.4 *Castanea*	71.3 *Pinus pinea*
Max UA (%)	91 Barren	100 Water	100 Water	100 Water	100 *Pinus pinea*	97.1 Barren	100 Water	100 Water/P*inus pinea*
Min UA (%)	0 Water & *Pinus pinea*	66.6 *Quercus*	70 *Quercus*	70 *Quercus*	74.2 *Quercus*	70.1 *Acacia*	80.8 *Quercus*	81.1 *Quercus*
Lowest OOB Error (%)	23.5	14.83	12.1	10.7	10	8.3	8.7	8.4
Optimum mtry	10	10	5	10	10	5	5	5
ntree	Not stable	Stable	Stable	Stable	Stable	Stable	Stable	Stable

8. References

(ICNF) PI for NC and F. No Title [Internet]. 2015. Available from: https://geocatalogo.icnf.pt/catalogo_tema3.html

Adeli S, Quackenbush LJ, Salehi B, Mahdianpari M. THE IMPORTANCE OF SEASONAL TEXTURAL FEATURES FOR OBJECT-BASED CLASSIFICATION OF WETLANDS : NEW YORK STATE CASE STUDY. 2022;XLIII(June):6–11.

Agarwal A, Kumar S, Singh D. Development of neural network based adaptive change detection technique for land terrain monitoring with satellite and drone images. Def Sci J. 2019;69(5):474–80.

Alam MJ, Rahman KM, Asna SM, Muazzam N, Ahmed I, Chowdhury MZ. Comparative studies on IFAT, ELISA & DAT for serodiagnosis of visceral leishmaniasis in Bangladesh. Bangladesh Med Res Counc Bull. 1996;22(1):27–32.

Aragoneses E, Chuvieco E. Generation and mapping of fuel types for fire risk assessment. Fire. 2021;4(3).

Benali A, Fernandes P. Understanding the impact of different landscape-level fuel management strategies on wildfire hazard Understanding the impact of different landscape-level fuel management strategies on wildfire hazard. 2021;(March).

Bhatnagar S, Gill L, Regan S, Waldren S, Ghosh B. A nested drone-satellite approach to monitoring the ecological conditions of wetlands. ISPRS J Photogramm Remote Sens [Internet]. 2021 Apr;174:151–65. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0924271621000125

Blatchford ML, Mannaerts CM, Zeng Y. Determining representative sample size for validation of continuous, large continental remote sensing data. Int J Appl Earth Obs Geoinf [Internet]. 2021;94(September 2020):102235. Available from: https://doi.org/10.1016/j.jag.2020.102235

Chatziantoniou A, Petropoulos GP, Psomiadis E. Co-Orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a mediterranean setting based on machine learning. Remote Sens. 2017;9(12):1–19.

Chaves MED, Picoli MCA, Sanches ID. Recent applications of Landsat 8/OLI and Sentinel-2/MSI for land use and land cover mapping: A systematic review. Remote Sens. 2020;12(18).

Costa H, Almeida D, Vala F, Marcelino F, Caetano M. Land cover mapping from remotely sensed and auxiliary data for harmonized official statistics. ISPRS Int J Geo-Information. 2018;7(4).

Costa H, Benevides P, Moreira FD, Moraes D, Caetano M. Spatially Stratified and Multi-Stage Approach for National Land Cover Mapping Based on Sentinel-2 Data and Expert Knowledge. Remote Sens. 2022;14(8).

Costa H, de Rigo D, Libertà G, Houston Durrant T, San-Miguel-Ayanz J. European wildfire danger and vulnerability in a changing climate: towards integrating risk dimensions. 2020. 59 p.

ESA. Sentinel-2 User Handbook. Vol. 2, European Space Agency ESA Standard Document. 2015. 64 p.

Feng Q, Liu J, Gong J. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier-A case of yuyao, China. Water (Switzerland). 2015;7(4):1437–55.

Gini R, Sona G, Ronchetti G, Passoni D, Pinto L. Improving tree species classification using UAS multispectral images and texture measures. ISPRS Int J Geo-Information. 2018;7(8):1–18.

Gitelson A, Merzlyak MN. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J Photochem Photobiol B Biol. 1994;

Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from EOS- MODIS. Remote Sens Environ. 1996;58(3):289–98.

Guo Q, Wu W, Massart DL, Boucon C, De Jong S. Feature selection in principal component analysis of analytical data. Chemom Intell Lab Syst. 2002;61(1–2):123–32.

Hanes CC, Wotton M, Woolford DG, Martell DL, Flannigan M. Mapping organic layer thickness and fuel load of the boreal forest in Alberta, Canada. Geoderma [Internet]. 2022;417:115827. Available from: https://doi.org/10.1016/j.geoderma.2022.115827

Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. SEG Tech Progr Expand Abstr. 1973;3:610–21.

Held M, Committee TIB. GLCM TEXTURE: A TUTORIAL. 17th Int Symp Ballist. 1998;2(March):267–74.

Hernandez I, Benevides P, Costa H, Caetano M. Exploring sentinel-2 for land cover and crop mapping in portugal. Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch. 2020;43(B3):83–9.

Humeau-Heurtier A. Texture feature extraction methods: A survey. IEEE Access. 2019;7:8975–9000.

ICNF IC da N e das F. 6.o Inventário Florestal Nacional. Relatório Final 2015. 2016.

Immitzer M, Vuolo F, Atzberger C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sens. 2016;8(3):1–27.

Incêndios C, Base I, Florestal GT. Plano Municipal de Defesa da Floresta Contra Incêndios. 2020;

Instituto Português do Mar e da Atmosfera (IPMA). May Climatological Bulletin [Internet]. 2022. Available from: https://www.ipma.pt/pt/media/noticias/news.detail.jsp?f=/pt/media/noticias/textos/Boletim_climatologico_maio.html

Investment in Disaster Risk Management in Europe Makes Economic Sense. Invest Disaster Risk Manag Eur Makes Econ Sense. 2021;

Kaplan G. Broad-Leaved and Coniferous Forest Classification in Google Earth Engine Using Sentinel Imagery. 2021;(November 2020):64.

Kattenborn T, Lopatin J, Förster M, Christian A, Ewald F. Remote Sensing of Environment UAV data as alternative to fi eld sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. Remote Sens Environ [Internet]. 2019;227(January):61–73. Available from: https://doi.org/10.1016/j.rse.2019.03.025

Kluczek M, Zagajewski B. Airborne HySpex Hyperspectral Versus Multitemporal Sentinel-2 Images for Mountain Plant Communities Mapping. 2022;

Kolström M, Vile T, Lindner M. Climate Change Impacts and Adaptation in European Forests. EFI Policy Br 6. 2011;(February):14.

Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159–74.

Liu H, Gong P, Wang J, Wang X, Ning G, Xu B. Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020 - iMap World 1.0. Remote Sens Environ [Internet]. 2021;258:112364. Available from: https://doi.org/10.1016/j.rse.2021.112364

Ma L, Li M, Ma X, Cheng L, Du P, Liu Y. A review of supervised object-based land-cover image classification. ISPRS J Photogramm Remote Sens [Internet]. 2017;130:277–93. Available from: https://doi.org/10.1016/j.isprsjprs.2017.06.001

Ma L, Li M, Ma X, Cheng L, Du P, Liu Y. A review of supervised object-based land-cover image classification. ISPRS J Photogramm Remote Sens. 2017;130:277–93.

Monteiro-Henriques T, Fernandes PM. Regeneration of native forest species in Mainland Portugal: Identifying main drivers. Forests. 2018;9(11).

Moraes, Daniel, Benevides P, Costa H, FranciscoD. MF, Caetano M. INFLUENCE OF SAMPLE SIZE IN LAND COVER CLASSIFICATION ACCURACY USING RANDOM FOREST AND SENTINEL-2 DATA IN PORTUGAL. IGARSS 2021 - 2021 IEEE Int Geosci Remote Sens Symp. 2021;

Nizalapur V, Vyas A. Texture analysis for land use land cover (LULC) classification in parts of Ahmedabad, Gujarat. Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch. 2020;43(B3):275–9.

Noi Phan T, Kuch V, Lehnert LW. Land cover classification using google earth engine and random forest classifier-the role of image composition. Remote Sens. 2020;12(15):1–22.

Pacheco ADP, Junior JADS, Ruiz-Armenteros AM, Henriques RFF. Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in central portugal using landsat-8, sentinel-2, and terra imagery. Remote Sens. 2021;13(7):1–25.

Persson M, Lindberg E, Reese H. Tree species classification with multi-temporal Sentinel-2 data. Remote Sens. 2018;10(11):1–17.

Portugal Directorate-General for the Territory (DGT). No Title [Internet]. 2018. Available from: https://www.dgterritorio.gov.pt/Carta-de-Uso-e-Ocupacao-do-Solo-para-2018

Sheykhmousa M, Mahdianpari M, Ghanbari H, Mohammadimanesh F, Ghamisi P, Homayouni S. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2020.

Somvanshi SS, Kumari M. Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data. Appl Comput Geosci [Internet]. 2020;7(June):100032. Available from: https://doi.org/10.1016/j.acags.2020.100032

Stehman S V., Foody GM. Key issues in rigorous accuracy assessment of land cover products. Remote Sens Environ [Internet]. 2019;231(May):111199. Available from: https://doi.org/10.1016/j.rse.2019.05.018

Story M, Congalton RG. Remote Sensing Brief Accuracy Assessment: A User's Perspective. Photogramm Eng Remote Sensing. 1986;52(3):397–9.

Thomlinson JR, Bolstad P V., Cohen WB. Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products. Remote Sens Environ. 1999;70(1):16–28.

Viegas DX, Figueiredo Almeida M, Ribeiro LM, Raposo J, Viegas MT, Oliveira R, et al. O Complexo de Incêndios de Pedrogão Grande E Concelhos Limítrofes, Iniciado a 17 de Junho de 2017. 2017. Iniciado a [Internet]. 2017;2017:238. Available from: https://www.portugal.gov.pt/pt/gc21/comunicacao/documento?i=o-complexo-de-incendios-de-pedrogao-grandee-%0Aconcelhos-limitrofes-iniciado-a-17-de-junho-de-2017

Wang H, Zhao Y, Pu R, Zhang Z. Mapping Robinia pseudoacacia forest health conditions by using combined spectral, spatial, and textural information extracted from IKONOS imagery and random forest classifier. Remote Sens. 2015;7(7):9020–44.

Wessel M, Brandmeier M, Tiede D. Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. Remote Sens. 2018;10(9).

Wójtowicz A, Piekarczyk J, Czernecki B, Ratajkiewicz H. A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. J Photochem Photobiol B Biol. 2021;223(February).

Xavier AC, Rudorff BFT, Shimabukuro YE, Berka LMS, Moreira MA. Multi-temporal analysis of MODIS data to classify sugarcane crop. Int J Remote Sens. 2006;

Xia H, Zhao W, Li A, Bian J, Zhang Z. Subpixel inundation mapping using landsat-8 OLI and UAV data for a wetland region on the zoige plateau, China. Remote Sens. 2017;9(1):1–22.

Xiao B. Principal component analysis for feature extraction of image sequence. 2010;

Xie Y, Sha Z, Yu M. Remote sensing imagery in vegetation mapping: a review. J Plant Ecol. 2008;

Xu JL, Gowen AA. Spatial-spectral analysis method using texture features combined with PCA for information extraction in hyperspectral images. Vol. 34, Journal of Chemometrics. 2020.

Zeng L, Wardlow BD, Xiang D, Hu S, Li D. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sens Environ. 2020;

Zhang X, Cui J, Wang W, Lin C. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. Sensors (Switzerland). 2017;17(7).

Zheng H, Du P, Chen J, Xia J, Li E, Xu Z, et al. Performance evaluation of downscaling sentinel-2 imagery for Land Use and Land Cover classification by spectral-spatial features. Remote Sens. 2017;9(12).