Framing the Forest: A Comparative Analysis of Google Earth Engine Classifiers for Accurate Species Extraction

Luís Monteiro[1] [0000-0001-6594-7885],

Bruna Almeida[1] [0000-0002-3349-1470],

Bruna Duarte[1],

Pedro Cabral[1,2*] [0000-0001-8622-6008] cabral@nuist.edu.cn

[1] NOVA Information Management School (NOVA IMS), Campus de Campolide, Universidade Nova de Lisboa, 1070-312 Lisboa, Portugal

[2] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information

Science and Technology, Nanjing 210044, China

* Corresponding author

This is the Author Peer Reviewed version of the following conference paper published by Springer:

Monteiro, L., Almeida, B., Duarte, B., & Cabral, P. (2024). Framing the Forest: A Comparative Analysis of Google Earth Engine Classifiers for Accurate Species Extraction. In H. Carlos-Martinez, R. Tapia-McClung, D. A. Moctezuma-Ochoa, & A. J. Alegre-Mondragón (Eds.), Recent Developments in Geospatial Information Sciences: Selected papers from IGISc 2023 (pp. 159-171). (Lecture Notes in Geoinformation and Cartography). Springer Nature Switzerland AG. Advance online publication. https://doi.org/10.1007/978-3-031-61440-8_13

Framing the Forest: A Comparative Analysis of Google Earth Engine Classifiers for Accurate Species Extraction

Luís Monteiro[1][0000-0001-6594-7885], Bruna Almeida[1][0000-0002-3349-1470], Bruna Duarte[1] and Pedro Cabral[2,1][0000-0001-8622-6008]

[1] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
[2] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
cabral@nuist.edu.cn

Abstract. Forests are critical ecosystems that provide multiple ecological services; therefore, accurate classification of their species is crucial for effective forest management and conservation. In this study, we compared and evaluated the performance of several Google Earth Engine (GEE) classifiers for extracting main forest species in Portugal. For this purpose, a time series of Sentinel-2 images was investigated for 2018 using a two-stage supervised classification workflow, which involved Random Forest, Naive Bayes, Classification and Regression Trees, Gradient Boosting Trees, and Support Vector Machine classifiers. In the first stage, Sentinel-2 images were analyzed to provide a binary land cover map from where the forest layer was extracted. The layer obtained was then used for masking the input features to the second stage. The latter stage discriminated six forest species by analyzing several input feature configurations. These included multiple spectral indices selected from the time series, as well as topographic factors. The performance of each classifier in species discrimination was assessed by comparing accuracy metrics obtained from the selected input configurations. The highest overall accuracy of the map was estimated at 68.75%, with a kappa of 0.62, for the Support Vector Machine classifier. The study's findings highlight the importance of algorithm selection using the GEE platform for accurate forest species classification in similar landscape characteristics, informing researchers, conservation practitioners, and forest managers for improved management and conservation strategies.

Keywords: Earth Observation, Machine-Learning, Forest Species, Google Earth Engine, Ecosystem Services.

1 Introduction

Forests cover about 4,06 billion ha - about 31% of the world's land surface [1]. They are critical ecosystems that provide a myriad of ecological services that contribute to societal well-being, such as carbon sequestration, biodiversity conservation, and water regulation [2, 3]. Understanding the composition and distribution of forest species is crucial for effective forest management, conservation planning, and ecosystem health monitoring [4]. Therefore, a rapid and efficient method for accurately mapping and monitoring changes in forest cover over time is crucial to support effective management and conservation actions in this ecosystem.

With the advent of remote sensing technology, researchers have leveraged satellite imagery to study forests at regional and global scales [5, 6]. The biggest challenge in classifying vegetation and forest species from satellite imagery is the complex nature of forest ecosystems, the diversity of species, and the variability in spectral signatures [7]. Recent technological developments help advance the remote sensing processing approach. Among the many remote sensing tools available, Google Earth Engine (GEE) has emerged as a powerful platform for analyzing and processing massive amounts of satellite data [8]. The platform supports the processing of cloud-based image classifications, making it possible to implement machine learning algorithms in the classification process. The machine learning-based classification system at GEE provides a wide range of algorithms that can be used as a basis for mapping various land covers, including forest coverage, through automated species extraction from satellite imagery.

Machine learning is a computer-based approach widely used for developing systems capable of learning and solving problems automatically [9]. Supervised classification employs machine learning methods as alternative algorithms for classifying large-dimensional input data [10]. These methods utilize training data to train the algorithm in recognizing objects while testing data is used to evaluate algorithm performance [11]. The GEE platform provides a diverse range of machine-learning algorithms suitable for the supervised classification of image data[8]. These algorithms utilize various machine learning techniques, including decision trees, support vector machines, random forests, and deep learning models, to classify pixels into different vegetation or species classes based on their spectral characteristics [8, 12]. Despite the availability of multiple classification algorithms in GEE, there is a need for systematic evaluations and comparisons to determine their effectiveness in accurately extracting forest species. Each algorithm has its strengths and limitations, and the choice of algorithm can significantly impact the accuracy of species classification [13].

The main goal of this study is to systematically compare and evaluate various GEE classifiers for extracting forest species, and provide insights into their strengths, weaknesses, and applicability in forest species classification using GEE. Additionally, this research aims to contribute to the enhancement of accurate mapping of forest ecosystems, which will lead to improved conservation strategies and the promotion of sustainable forest management practices.

Assessing and comparing different GEE classifiers is crucial to identifying the most suitable algorithm for precise species extraction. This understanding of classifier performance characteristics will empower researchers, conservation practitioners, and forest managers to make informed decisions regarding using satellite imagery for species mapping and monitoring purposes.

2 Material and Methods

2.1 Study area

For the present research, the study area corresponds to the landscape units of Montes entre Larouco e Marão, and Trás-os-Montes, merged into a larger mapping zone (see Figure 1). This grouping of landscape units into a larger zone allows for more effective and efficient image classification processes, as the original landscape units are relatively small. The mapping zone, situated in the Northeast of Portugal bordering Spain, showcases a diverse landscape encompassing high plateaus, deep valleys, and mountain ranges with altitudes varying from 1000 to 1200 meters, along with the lower areas of the Tâmega River valley at approximately 200 meters. The region exhibits a wide-ranging agricultural system, characterized by rounded relief and valleys of Douro River tributaries. The predominant vegetation includes cork oak along the main river tributaries and the Douro International section, while black oak dominates the remaining area. Towards the western part, forest systems featuring eucalyptus and pine trees, as well as shrubland, prevail [15].

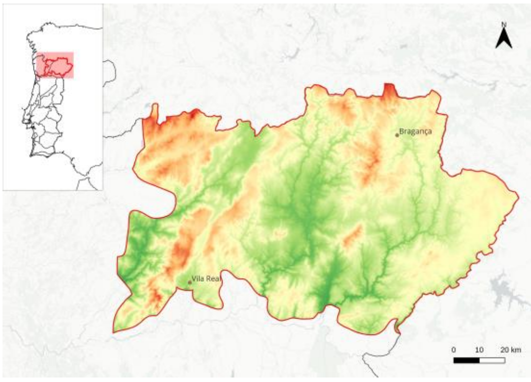


Fig. 1. Location of the study area: mapping zone containing the landscape units of Montes entre Larouco e Marão, and Trás-os-Montes.

2.2 Materials

Satellite Data. Available Sentinel-2 imagery from the entire year of 2018 was accessed using the GEE platform, and then filtered by cloud cover and composited into a single cloud-free, in order to capture the typical annual cycle of natural vegetation and agriculture observed. The multitemporal composites were generated by selecting only pixels from images with a reported cloud cover of less than 10%, which are not flagged as cloud or cirrus in the quality assessment band and calculating a median value of all valid observations. All spectral bands from Sentinel-2 imagery were utilized, except bands 1, 9, and 10, which are related to atmospheric effects. Clouds or cloud shadows were masked from the images to enhance the quality of the data, and spectral indices, including the Normalized Difference Vegetation Index (NDVI) [16], Enhanced Vegetation Index (EVI) [17], and Soil Adjusted Vegetation Index (SAVI) [18], were computed, to improve the efficiency of image classification tasks.

The selection of these spectral indices for classifying forest species was considered highly appropriate due to their established effectiveness in characterizing vegetation in various other studies. Additionally, topographic characteristics such as slope and elevation were derived from the ALOS World 3D – 30m dataset and added to the composites as data stacks, along with the bands and spectral indices. This data stack, composed of 18 bands, was used as input features in the supervised image classification process outlined in Part 2.3.

The composite NDVI was also used as ground truth data and was separately utilized to delineate the training data on the map for use as samples in the classification process.

Auxiliary Data. In this study, auxiliary data was utilized as a set of reference polygons about vegetation classes examined to aid in the classification process. Specifically, the Portuguese land use and land cover map for 2018 (COS2018) was utilized as ground truth data for a stratified sampling approach for each class. COS2018 is available in vector format with a minimum mapping unit (MMU) of 1 ha, and a minimum width of linear elements (MMW) of 20 m, and its nomenclature includes more than 80 classes. Therefore, to ensure the compatibility of the map and achieve the study's specific goal of focusing on the most representative forest species, it was necessary to reclassify into six classes, encompassing both individual species and groups of species, since the datasets had different numbers of classes (see Table 1).

Table 1. Name and description of forest classes scheme.

Class name	Font size and style
Eucalyptus	Forests where the dominant species is eucalyptus (*Eucalyptus spp.*).
Other broadleaves	Areas comprising broadleaf tree species, including willows, poplars, and various oak species, excluding cork and holm oak.
Cork oak and holm oak	Cork and holm oak forest, including trees in agroforestry systems and other non-forest areas.
Other coniferous	Forests dominated by a single species of resinous trees that do not belong to the maritime pine class and stone pine.
Maritime Pine	Forests where the dominant species is maritime pine (*Pinus pinaster*).
Stone Pine	Forests where the dominant species is stone pine (*Pinus pinea*).

GEE algorithms. In this study, four machine learning algorithms (Classification and Regression Trees, Gradient Boosting Trees, Random Forest, and Support Vector Machine) were employed to assess their suitability and accuracy in mapping forest species, using the default GEE parameters.

- Classification and Regression Trees (CART) is a classification algorithm that utilizes entropy structure, represented by a decision tree, to classify objects based on their inherent characteristics within a given class [19]. It can be seen as an IF-THEN function for object classification. However, one weakness of CART is its high sensitivity to training data, meaning that even slight changes in the training dataset can lead to different classification results [11].

- Gradient Boosting Trees (GBT) is an ensemble learning algorithm that combines multiple weak prediction models, typically decision trees, to create a strong predictive model. It works by sequentially adding new models that correct the mistakes made by previous models, leading to a highly accurate final prediction. Gradient Boosting Trees effectively capture complex relationships between features and the target variable. However, one limitation of this algorithm is its potential to overfit the training data if not properly regularized, which may result in poor generalization to unseen data [20].

- Random Forest (RF) is a classifier derived from a decision tree that enables multi classification or regression using an ensemble of multiple decision tree models and is nowadays one of the most used algorithms [12]. It addresses the limitations of the decision tree algorithm, particularly its sensitivity to training data, by constructing a statistical data ensemble [21]. The model generates many decision trees using a random selection of training datasets and factors, and its most important input criteria are the size of the training dataset, and the number of trees generated [22].

- Support Vector Machine (SVM) is considered an effective algorithm that overcomes limitations in remote sensing data classification [23]. By optimizing a hyperplane based on training data, SVM formulates a classification function that successfully separates classes. The hyperplane is determined by measuring the margin between the hyperplane and the nearest instances of each class, known as support vectors. Critical parameters such as kernel functions, cost parameters, and gamma play a crucial role in determining the optimal hyperplane for accurate classification [24].

2.3 Methods

The methodology utilized in this study used a spatial and methodological division to tailor the classification mapping to the unique local conditions. A spatial stratification procedure was implemented as a primary step, to account for the diverse landscape of Portugal. This method involved reclassifying the original landscape units into a larger mapping zone, ensuring that the resulting zone was created based on environmental factors that influence the distribution of the target forest species. Also, the division into a mapping zone concentrated the sampling effort in areas with a higher likelihood of the forest species, leading to a more efficient and accurate national image classification [25]. Afterward, a two-phase classification approach for forest species mapping was developed for each classifier, employing multiple stages of data analysis that iteratively improved the results of the previous stage, as depicted in Figure 2.

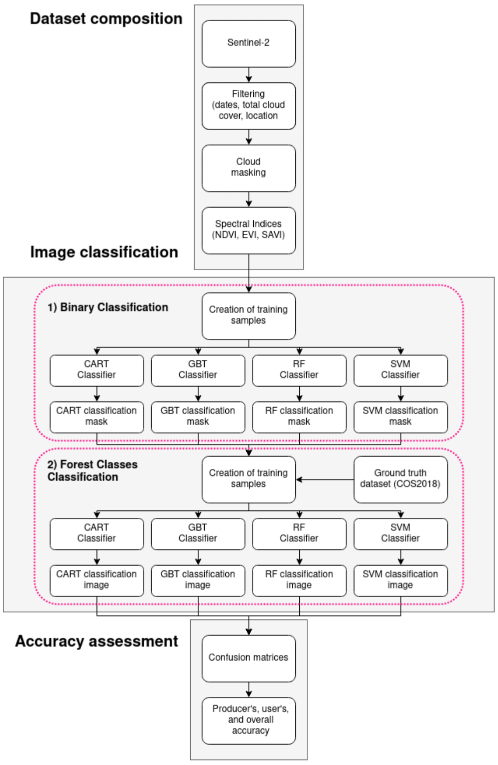


Fig. 2. Methodological approach for forest species mapping.

In GEE, the initial phase involved a binary classification to distinguish the forest from non-forest pixels, which served as an input filter for the second phase. A separate classification was performed using the data stack, which included bands, indices, and topographic factors. This two-phase approach is intended to improve the accuracy of forest species mapping by filtering out non-forest areas and enabling a more focused analysis of the relevant features.

The second stage of the classification process focused on a multi-class problem, by discriminating between forest species. These species were identified within the forest's pixels recognized in the first phase of the classification exercise. In both classification phases, all selected algorithms were utilized for image classification in the GEE.

The resulting land cover map was then post-processed to eliminate small, isolated pixels, and noise. The necessary accuracy assessments were conducted for each classifier using a stratified random sampling protocol. Validation samples were composed of randomly located points or pixels as sampling units and assessed in GEE. The classification result was then subjected to an accuracy assessment procedure, encompassing the calculation of various metrics such as overall accuracy, confusion matrix, kappa coefficient, and producer and user accuracies for each class.

Binary Forest Classification. The first phase employed a binary-class, index-driven, pixel-based classification, to detect pixels corresponding to the forest class. Areas identified with substantial vegetation cover, the NDVI were used as a basis for photo interpretation, with pixel values of NDVI greater than 0,6 were selected as representative of vegetated areas, and training points were manually marked on the map [26]. A total of 100 reference points were considered for each land cover class, and they were distributed randomly and evenly throughout the map. The drawn points were divided into training and validation datasets randomly for the classifiers' training and subsequent map validation. For each class, 70% of the points were allocated to the training dataset, while the remaining 30% were used for validation.

The algorithms utilized the training points by sampling them across the spectral bands and indices, to generate the final map. The pixels were then classified into either forest or non-forest classes using a random sampling scheme, and the output forest layer was then extracted from the final map, to be used in the second phase as a mask of the input features.

During this phase, the classifiers were utilized with default parameters, as an exhaustive analysis of hyperparameters and feature importance was not applicable due to the simplicity of the binary classification task.

Forest Classes Classification. During the second phase, a data-driven, pixel-based classification was employed to detect six forest species classes. The classification was performed only on the pixels of the forest layer identified in the first stage. The mapped forest species reference points were randomly divided into 70% and 30%, respectively, to train and validate the classifiers for each class. These points were marked on the map, over areas corresponding to these classes from the COS2018 dataset. A total of 30 reference points per class were considered, resulting in 180 reference points for each mapping zone.

The classifiers were fine-tuned to optimize their performance, where applicable, by exploring the influence of different features and adjusting hyperparameters. This meticulous optimization process significantly improved the accuracy of species classification, ensuring a more precise delineation of distinct forest types within the study area.

It is important to note that while some classifiers allowed for comprehensive analyses of feature importance and hyperparameter tuning, specific models had limitations. Despite these limitations, efforts were made to optimize the classifiers to the extent possible, given the available tools and functionalities. Incorporating these optimization techniques played a pivotal role in refining the classification results, enabling a more accurate and nuanced representation of the forest ecosystem in the final land cover map.

3 Results

Among the compared classification results, the SVM classifier exhibited the highest overall classification accuracy at 69%, outperforming RF at 63%, GBT at 60%, and CART at 56% (see Table 2). Similarly, when considering the kappa coefficient, SVM attained the highest value of 0,62, while RF and GBT achieved 0,54 and 0,52 kappa, respectively followed by CART with 0,43. Additionally, when examining producer and user accuracy, SVM consistently achieved the highest values, with producer accuracy (PA) and user accuracy (UA) both with 71%.

Table 2. Overall accuracy (OA), producer accuracy (PA), user accuracy (UA), and kappa coefficient for CART, GBT, RF, and SVM classifiers.

Classifier	OA	PA	UA	Kappa
CART	0,56	0,59	0,59	0,47
GBT	0,60	0,63	0,62	0,53
RF	0,63	0,65	0,63	0,55
SVM	0,69	0,71	0,71	0,62

Analysis of producer and user accuracies highlighted the varying performance of different classes across different classifiers, with certain classes exhibiting high accuracy rates. In contrast, others show variations depending on the classifier used. Among all the classified classes, the Other Broadleaved class consistently achieved the highest user accuracy across most classifiers. Specifically, CART yielded a user accuracy of 73%, GBT achieved 69%, and SVM achieved 91%. However, the RF classifier showed a slightly different pattern, with the highest user accuracy of 80% observed for the Cork Holm Oak class. On the other hand, the Eucalyptus class performed consistently poorly across all classification algorithms, with the lowest user accuracy (CART – 48%, GBT – 53%, RF – 53%, SVM – 56%).

When considering the producer accuracy, the distribution of high and low accuracy rates across classes showed more dispersion. Specifically, for both CART and SVM classifiers, the Other coniferous class showcased the highest producer accuracy, reaching 83%. In comparison, the GBT classifier achieved a slightly lower producer accuracy of 79% for the other broadleaves class. In the case of RF, the Cork and Holm Oak class demonstrated a producer accuracy of 75%. The classes with the lowest producer accuracy were stone pine for GBT (37%) and SVM (53%), and maritime pine for CART (33%) and RF SVM (47%).

When compared with other outputs, the CART classifier displayed a tendency to overestimate the number of pixels classified as maritime pine in the northwest region of the image, which predominantly consisted of eucalyptus and other broadleaves with scattered beech stands. The classification images also exhibited a minor salt-and-pepper effect attributed to the shadows of trees within the forest (see Figure 3).

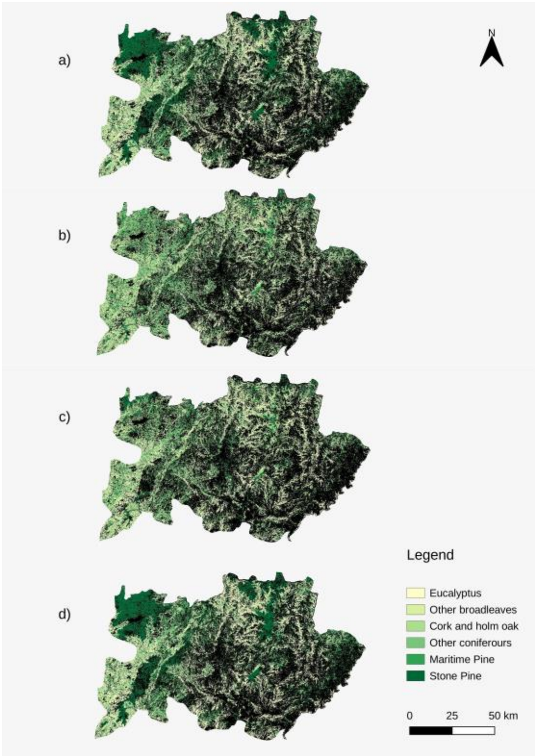
Table 3 showcases the estimated abundances of classified forest species in the study area derived from the outputs of the evaluated classification algorithms.

Table 3. Extent of forest species classes according for each algorithm classification (hectares).

Class	CART	GBT	RF	SVM
Eucalyptus	45,03	66,76	68,12	72,64
Other broadleaves	7,63	7,53	8,40	3,79

Cork oak and holm oak	18,95	8,58	8,02	9,36
Other coniferous	6,32	3,23	3,35	3,10
Maritime Pine	16,64	9,75	9,42	8,45
Stone Pine	5,42	4,16	2,70	2,67

Fig. 3. Classification of output images using CART (a), GBT (b), RF (c), and SVM (d) classifiers. (The color black symbolizes masked areas – non-forest vegetation)



In analyzing the classification results across various classes, notably, the SVM demonstrated the highest accuracy for the "Eucalyptus" class, covering 72,64 hectares, surpassing GBT (66,76 hectares) and RF (68,12 hectares). In contrast, CART's classification underestimated the extent of the Eucalyptus forest, covering only 45,03 hectares. For the "Other broadleaves" class, RF achieved the highest coverage at 8,40 hectares, while the other classifiers achieved similar extents, around 7,5 hectares. A contrast emerges in the "Cork oak and Holm oak" class, where SVM excelled, covering

9,36 hectares, while CART's classification significantly overestimated this class with 18,95 hectares. In the "Other coniferous" class, SVM produced the most accurate result (3,10 hectares), showcasing a notable difference from the other classifiers. For "Maritime Pine," GBT provided the highest extent (9,75 hectares), while SVM (8,45 hectares) and RF (9,42 hectares) were also competitive. Lastly, "Stone Pine" forests were most accurately represented by SVM (2,67 hectares), with RF and GBT also performing reasonably well.

4 Discussion and Conclusion

Forests play a vital role in providing multiple ecological services, highlighting the significance of accurate classification for effective forest management and conservation. Remote sensing technology has emerged as a crucial tool in precisely classifying forest species. GEE has proven to be an influential cloud-based geospatial analysis and modeling platform, and this research aimed to demonstrate the effectiveness of various classifiers in accurately mapping and classifying forest species for Portugal.

Four classification algorithms were compared using Sentinel-2 bands, specific vegetation indices, and topographic characteristics as input features. All classifiers utilized the same training data and were tested on the same test data to ensure fair comparisons. Among the SVM achieved the highest accuracies, which aligns with previous literature findings and reinforces its effectiveness in addressing complex tree species classification problems [26–28].

Several studies have reported comparable performance between SVM and RF classifiers [4, 29]. While both classifiers demonstrated high accuracy, SVM showcased superior performance in this study. One possible explanation for this difference could be that RF classifiers generally benefit from a larger number of instances to fully leverage their randomization concept and achieve robust generalization to new data [22]. Considering the focus on multispectral data and a relatively small sample size per class in this research, the SVM classifier exhibited better performance.

The lower accuracies observed for CART and GBT classifications can be attributed to their sensitivity to variations or outliers in the data, leading to decreased accuracy compared to SVM and RF classifiers [21]. In contrast, SVM and RF classifiers are renowned for their robustness in tackling such challenges and exhibit lower sensitivity than other classifiers. RF demonstrates resilience to its parameters, while SVM is more sensitive to hyperparameters [30]. Additionally, the specific distribution and characteristics of the dataset might have favored SVM and RF classifiers, as they excel in scenarios with non-linear or complex decision boundaries [9].

Lastly, it is essential to consider the relevance of selected features, as they can significantly influence the classifier's results. Using the 18-band data stack might not be the optimal number of input data for CART and GBT algorithms, as it may have failed to capture the spectral information that provides the best differentiation between classes. This notion is supported by the work of Zhang and others [31], emphasizing the importance of carefully selecting relevant bands for classification purposes.

While GEE is a robust platform, its distinct advantages over other software or platforms could be more precise, especially concerning algorithm comparison and training time. Future research should delve deeper into these aspects, exploring the impact of training sample sizes, variations in training samples, and reference data characteristics. By addressing these factors comprehensively, researchers can enhance the accuracy of forest species mapping and classification, shedding light on the specific advantages GEE offers over alternative platforms. Such investigations will pave the way for more informed decisions in choosing the optimal platform and algorithm combinations, ensuring precise and efficient forest species classification, crucial for sustainable forest management and conservation efforts.

Acknowledgments

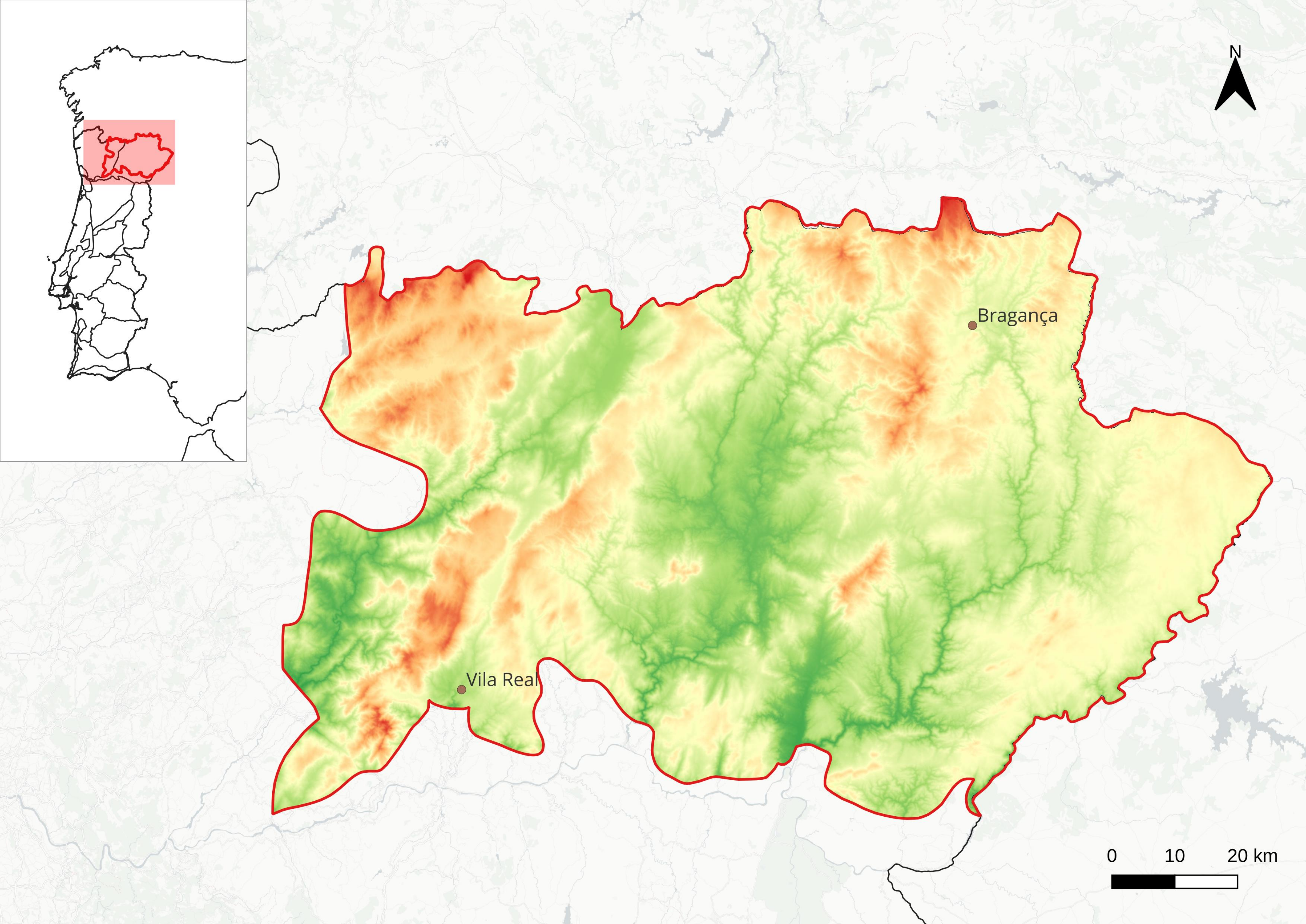
This study was supported by the research project MaSOT – Mapping Ecosystem Services from Earth Observations, funded by the Portuguese Science Foundation – FCT (EXPL/CTA-AMB/0165/2021), The authors gratefully acknowledge the financial support of the FCT, through the MagIC Research (Centro de Investigação em Gestão de Informação - UIDB/04152/2020).

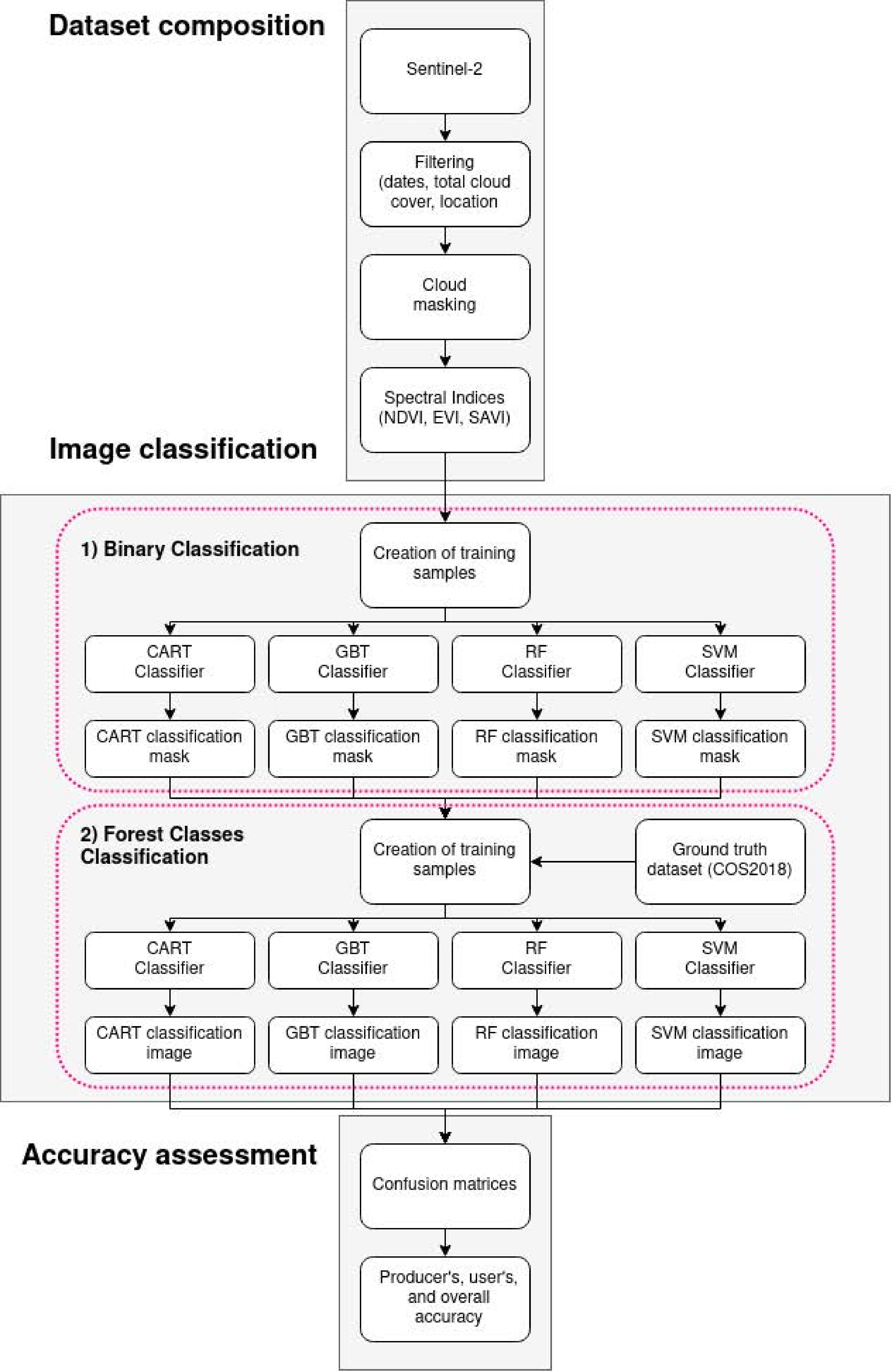
References

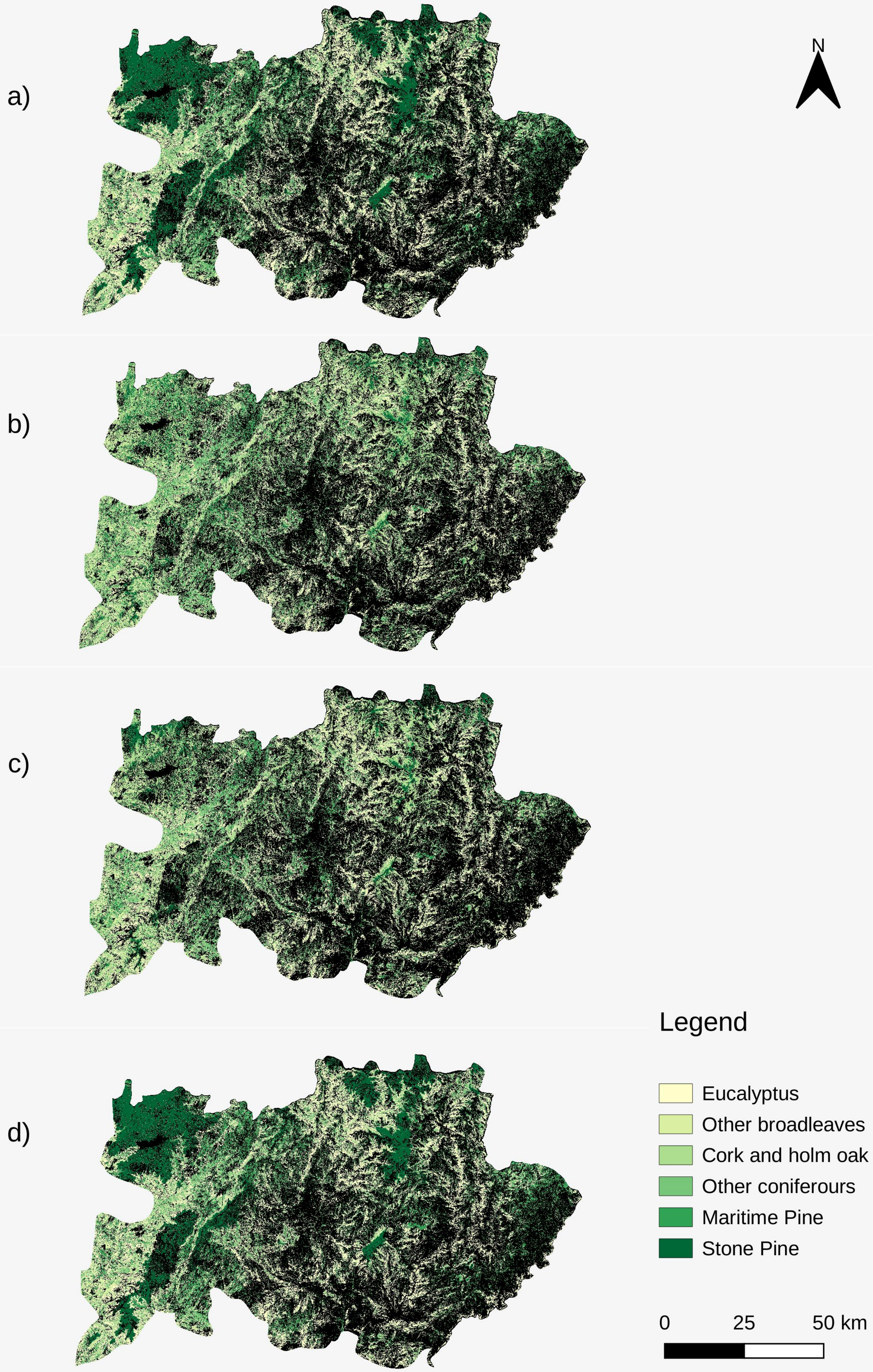
1. The State of the World’s Forests 2022. FAO (2022). https://doi.org/10.4060/cb9360en.
2. Friedlingstein, P., Jones, M.W., O’Sullivan, M., Andrew, R.M., Bakker, D.C.E., Hauck, J., Le Quéré, C., Peters, G.P., Peters, W., Pongratz, J., Sitch, S., Canadell, J.G., Ciais, P., Jackson, R.B., Alin, S.R., Anthoni, P., Bates, N.R., Becker, M., Bellouin, N., Bopp, L., Chau, T.T.T., Chevallier, F., Chini, L.P., Cronin, M., Currie, K.I., Decharme, B., Djeutchouang, L.M., Dou, X., Evans, W., Feely, R.A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R.A., Hurtt, G.C., Iida, Y., Ilyina, T., Luijkx, I.T., Jain, A., Jones, S.D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J.I., Körtzinger, A., Landschützer, P., Lauvset, S.K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P.C., Melton, J.R., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T.M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A.J., Sweeney, C., Tanhua, T., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F., Van Der Werf, G.R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A.J., Willis, D., Wiltshire, A.J., Yuan, W., Yue, C., Yue, X., Zaehle, S., Zeng, J.: Global Carbon Budget 2021. Earth Syst. Sci. Data. 14, 1917–2005 (2022). https://doi.org/10.5194/essd-14-1917-2022.
3. Millennium Ecosystem Assessment (Program) ed: Ecosystems and human well-being: synthesis. Island Press, Washington, DC (2005).

4. White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P.: Remote Sensing Technologies for Enhancing Forest Inventories: A Review. Canadian Journal of Remote Sensing. 42, 619–641 (2016). https://doi.org/10.1080/07038992.2016.1207484.
5. Calders, K., Jonckheere, I., Nightingale, J., Vastaranta, M.: Remote Sensing Technology Applications in Forestry and REDD+. Forests. 11, 188 (2020). https://doi.org/10.3390/f11020188.
6. Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G.: High-Resolution Global Maps of 21st-Century Forest Cover Change. Science. 342, 850–853 (2013). https://doi.org/10.1126/science.1244693.
7. Lausch, A., Borg, E., Bumberger, J., Dietrich, P., Heurich, M., Huth, A., Jung, A., Klenke, R., Knapp, S., Mollenhauer, H., Paasche, H., Paulheim, H., Pause, M., Schweitzer, C., Schmulius, C., Settele, J., Skidmore, A., Wegmann, M., Zacharias, S., Kirsten, T., Schaepman, M.: Understanding Forest Health with Remote Sensing, Part III: Requirements for a Scalable Multi-Source Forest Health Monitoring Network Based on Data Science Approaches. Remote Sensing. 10, 1120 (2018). https://doi.org/10.3390/rs10071120.
8. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment. 202, 18–27 (2017). https://doi.org/10.1016/j.rse.2017.06.031.
9. Mitchell, T.M.: Machine Learning. McGraw-Hill, New York (1997).
10. Hastie, T., Tibshirani, R., Friedman, J.H.: The elements of statistical learning: data mining, inference, and prediction. Springer, New York, NY (2009).
11. Bishop, C.M.: Pattern recognition and machine learning. Springer, New York (2006).
12. Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., Xian, G.: A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sensing of Environment. 132, 159–175 (2013). https://doi.org/10.1016/j.rse.2013.01.012.
13. Adugna, T., Xu, W., Fan, J.: Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sensing. 14, 574 (2022). https://doi.org/10.3390/rs14030574.
14. Ribeiro, O.: Portugal o Mediterrâneo e o Atlântico. Livraria Sá da Costa Editora, Lisbon (1998).
15. d'Abreu, A.C.: R (Alentejo central) a V (Algarve). Direcção Geral do Ordenamento do Território [u.a.], Queluz (2004).
16. Rouse, J.W., Jr., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring Vegetation Systems in the Great Plains with Erts. NASA Special Publication. 351, 309 (1974).
17. Huete, A., Liu, H.Q., van Leeuwen, W.: A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sensing of Environment. 59, 440–451 (1997). https://doi.org/10.1016/S0034-4257(96)00112-5.
18. Huete, A.R.: A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment. 25, 295–309 (1988). https://doi.org/10.1016/0034-4257(88)90106-X.
19. Berk, R.A.: Classification and Regression Trees (CART). In: Statistical Learning from a Regression Perspective. pp. 129–186. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-44048-4_3.
20. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. Ann. Statist. 29, (2001). https://doi.org/10.1214/aos/1013203451.

21. Breiman, L.: Random Forests. Machine Learning. 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324.
22. Raczko, E., Zagajewski, B.: Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing. 50, 144–154 (2017). https://doi.org/10.1080/22797254.2017.1299557.
23. Mountrakis, G., Im, J., Ogole, C.: Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing. 66, 247–259 (2011). https://doi.org/10.1016/j.isprsjprs.2010.11.001.
24. Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., Rahman, A.: Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sensing. 12, 1135 (2020). https://doi.org/10.3390/rs12071135.
25. Cano, E., Denux, J.-P., Bisquert, M., Hubert-Moy, L., Chéret, V.: Improved forest-cover mapping based on MODIS time series and landscape stratification. International Journal of Remote Sensing. 38, 1865–1888 (2017). https://doi.org/10.1080/01431161.2017.1280635.
26. Ghosh, A., Fassnacht, F.E., Joshi, P.K., Koch, B.: A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. International Journal of Applied Earth Observation and Geoinformation. 26, 49–63 (2014). https://doi.org/10.1016/j.jag.2013.05.017.
27. Shao, Y., Lunetta, R.S.: Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing. 70, 78–87 (2012). https://doi.org/10.1016/j.isprsjprs.2012.04.001.
28. Suess, S., Van Der Linden, S., Okujeni, A., Leitão, P., Schwieder, M., Hostert, P.: Using Class Probabilities to Map Gradual Transitions in Shrub Vegetation from Simulated EnMAP Data. Remote Sensing. 7, 10668–10688 (2015). https://doi.org/10.3390/rs70810668.
29. Ballanti, L., Blesius, L., Hines, E., Kruse, B.: Tree Species Classification Using Hyperspectral Imagery: A Comparison of Two Classifiers. Remote Sensing. 8, 445 (2016). https://doi.org/10.3390/rs8060445.
30. Basheer, S., Wang, X., Farooque, A.A., Nawaz, R.A., Liu, K., Adekanmbi, T., Liu, S.: Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques. Remote Sensing. 14, 4978 (2022). https://doi.org/10.3390/rs14194978.
31. Zhang, H.K., Roy, D.P.: Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. Remote Sensing of Environment. 197, 15–34 (2017). https://doi.org/10.1016/j.rse.2017.05.024.


N
Bragança
Vila Real
0
10
20 km


Dataset composition
Sentinel-2
Filtering (dates, total cloud cover, location
Cloud masking
Spectral Indices (NDVI, EVI, SAVI)
Image classification
1) Binary Classification
Creation of training samples
CART Classifier
GBT Classifier
RF Classifier
SVM Classifier
CART classification mask
GBT classification mask
RF classification mask
SVM classification mask
2) Forest Classes Classification
Creation of training samples
Ground truth dataset (COS2018)
CART Classifier
GBT Classifier
RF Classifier
SVM Classifier
CART classification image
GBT classification image
RF classification image
SVM classification image
Accuracy assessment
Confusion matrices
Producer's, user's, and overall accuracy


a)
b)
c)
d)
N
Legend
Eucalyptus
Other broadleaves
Cork and holm oak
Other coniferours
Maritime Pine
Stone Pine
0
25
50 km