Data in Brief 59 (2025) 111419

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Gran canaria vegetation segmentation dataset from multi-year aerial imagery for environmental monitoring and conservation

José Salas-Cáceres[a,*], Riccardo Balia[b], Marcos Salas-Pascual[c], Javier Lorenzo-Navarro[a], Modesto Castrillón-Santana[a]

[a] *Universidad de Las Palmas de Gran Canaria, Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (SIANI), Las Palmas de Gran Canaria 35017, Spain*
[b] *University of Cagliari, Department of Mathematics and Computer Science, Cagliari, Via Ospedale 72, 09124, Italy*
[c] *Universidad de Las Palmas de Gran Canaria, Instituto Universitario de Estudios Ambientales y Recursos Naturales (IUNAT), Campus de Tafira, Las Palmas de Gran Canaria 35017, Spain*

A R T I C L E I N F O

Article history:
Received 10 December 2024
Revised 6 February 2025
Accepted 17 February 2025
Available online 21 February 2025

Dataset link: Vegetation mapping dataset of Gran Canaria, Canary Islands, Spain. (Original data)

Keywords:
Vegetation mapping
Ecology
Remote sensing
Semantic segmentation
Computer vision

A B S T R A C T

Vegetation maps are an essential tool for territorial planning, enabling the identification of areas requiring protection and facilitating the study of key ecosystem dynamics such as their evolution over time and the threats they face. These aspects are especially critical in island territories, where their fragmented nature and isolation from the mainland pose significant challenges to the development of such documentation. Traditionally, these maps have relied on local experts, requiring extensive fieldwork, significant time and financial resources. To address these challenges, a novel dataset focused on Gran Canaria (Canary Islands, Spain) is presented, designed to allow researchers to develop and test deep learning models that automatically generate vegetation maps using computer vision techniques.

This dataset is unique in the field of aerial image-based semantic segmentation, as it provides detailed annotations for 20 well-defined vegetation communities, going beyond the broad classifications commonly found in existing datasets (e.g., forests or grasslands). Additionally, an alternative version of the dataset includes five non-vegetal classes, such as

* Corresponding author.
E-mail address: jose.salas@ulpgc.es (J. Salas-Cáceres).
Social media: @JavierLorenzoN (J. Lorenzo-Navarro), @otsedom1 (M. Castrillón-Santana)

https://doi.org/10.1016/j.dib.2025.111419

water bodies, roads, or buildings to support more visually comprehensive segmentation tasks.

Specifications Table

Subject	Vegetation Mapping, Ecology, Computer Vision and Pattern Recognition, Applied machine learning.
Specific subject area	Aerial images annotated with existing vegetation describing the mapping of the vegetation communities in the analysed area for deep segmentation models training.
Type of data	Image and Table *(.csv format)*
Data collection	Data were collected through five steps: (1) Image Acquisition: Aerial images of the island (2016, scale 1:1000, resolution 1964×982 pixels) were processed in QGIS to extract 2000+ RGB images. (2) Selection of Representative Images: 185 images highlighting key vegetation communities were chosen. (3) Temporal Information: Additional images from 2018 and 2021 were captured, totalling 555 images. (4) Mask Creation: Masks with 20–25 classes were drawn by a Canary Islands vegetation expert, then refined for temporal changes. (5) Cleaning and Review: Three reviewers ensured accuracy. Supplementary altitude and infrared data (2018–2021) were added.
Data source location	Gran Canaria, Canary Islands, Spain. 27°57′31″ N, 15°35′33″ W.
Data accessibility	Repository name: Zenodo Data identification number: https://doi.org/10.5281/zenodo.14215453 Direct URL to data: https://zenodo.org/records/14215453 Instructions for accessing these data: The data is open for download in the linked repository [1].
Related research article	None

1. Value of the Data

The significance of this dataset lies in several key aspects:

- This is the first collection of images created to map and relate the diverse types of vegetation present on one of the Canary Islands. Prior to this dataset, the vegetation cartography of the Canary Archipelago has primarily relied on the expertise manual annotation of the authors in both the terrain [2] and its vegetation [3,4]. The mentioned manual annotation made with field work limits the number of individuals who can contribute to such cartographic efforts and poses challenges for independent validation by other researchers. With the creation of this database and its posterior use, another methodology not so dependent on individuals is defined.
- Researchers working on nearby regions, such as other islands within the Canary Islands or other Macaronesian archipelagos (e.g., Maderia, Azores, Cabo Verde) [5], can leverage this dataset for similar studies in their respective zones, as these island groups share significant vegetation similarities. Moreover, researchers from more distant regions with different botanical contexts can adopt the methodology and objectives underpinning this dataset to conduct similar studies.
- Another notable feature of this dataset is its potential for long-term comparison. Since the dataset includes accurately georeferenced images, it enables researchers to study vegetation dynamics by repeating the same analysis over different time periods. This characteristic facilitates future revisions of the maps generated, providing an invaluable tool for monitoring ecological changes.

- In addition to its biological relevance, this dataset is also a valuable resource for evaluating deep learning segmentation models. Unlike other datasets, such as those published under the Spacenet name [6] or LoveDA [7,8], which typically focus on specific urban elements (e.g., buildings, roads, water bodies) and either consolidate all vegetation into a single class, or ignore it entirely, this dataset presents a unique challenge due to its detailed classification of diverse vegetation communities. Even in datasets designed with similar purposes, such as UOPNOA or UOS2 [9], the vegetation-related classes are limited to only five, encompassing broad categories such as “Forest” or “Grass with Trees.” The complexity of the dataset introduced in this work makes it a valuable benchmark for testing segmentation models in more intricate and biologically nuanced contexts.

2. Background

Vegetation cartography is an essential tool for territorial management, particularly in overseeing natural resources and land use. In oceanic island regions, its significance is even greater due to the rarity of the flora, which predominantly consists of endemic species, unique to these regions [10]. Often, these islands, rich in botanical diversity, are also popular tourist destinations, attracting visitors from around the world and experiencing some of the highest levels of human pressure [11], Consequently, these regions also tend to have the highest number of endangered species [12].

Managing such territories is a complex challenge that requires a thorough understanding of the values and resources that must be preserved. This circumstance necessitates identifying the vegetation communities, their habitats, and their locations. Traditionally, creating a vegetation map required extensive fieldwork and a large team of specialized personnel [13]. For instance, the current vegetation map of the Canary Islands [2] took over 15 years to complete, requiring several partial updates since its publication.

The database presented here, along with the methodology used to generate automatically vegetation maps from these images, offers a means to produce such maps at a significantly lower cost and in considerably less time. This innovation has the potential to transform vegetation mapping and improve the management of these ecologically sensitive territories.

The dataset described is part of a broader study aimed at reducing the time and costs associated with conducting plant species inventories. This data article not only provides an in-depth exploration of the dataset’s characteristics but also details the process followed to construct it, enhancing the reproducibility of the work and facilitating its application to similar territories. This work provides a comprehensive description of the various vegetation communities depicted in the images, offering detailed explanations in the Data Description section that precisely define each community for its members.

3. Data Description

This dataset contains 555 aerial images in PNG format, taken on the island of Gran Canaria, located in the northwest of Africa. These images are distributed across 185 distinct locations over three years: 2016, 2018 and 2021. The pictures have a scale of 1:1000, see Fig. 1a. For each image, two segmentations masks, also in PNG format, are provided:

- **Base Masks or Vegetation-Focused Masks**: The base version emphasizes the vegetation communities and was developed by a botanist with expertise in endemic flora. It comprises a total of 20 classes, grouping all non-vegetation pixels into a single class called ‘No Vegetation’. The remaining 19 classes each correspond to a specific vegetation community. This version has the sole purpose of building vegetation masks, not adding any extra information. See Fig. 1b
- **Detailed Masks**: The Detailed annotation version builds on the initial Vegetation-Focused version by refining mask boundaries to achieve finer detail, enabling sharper separation of

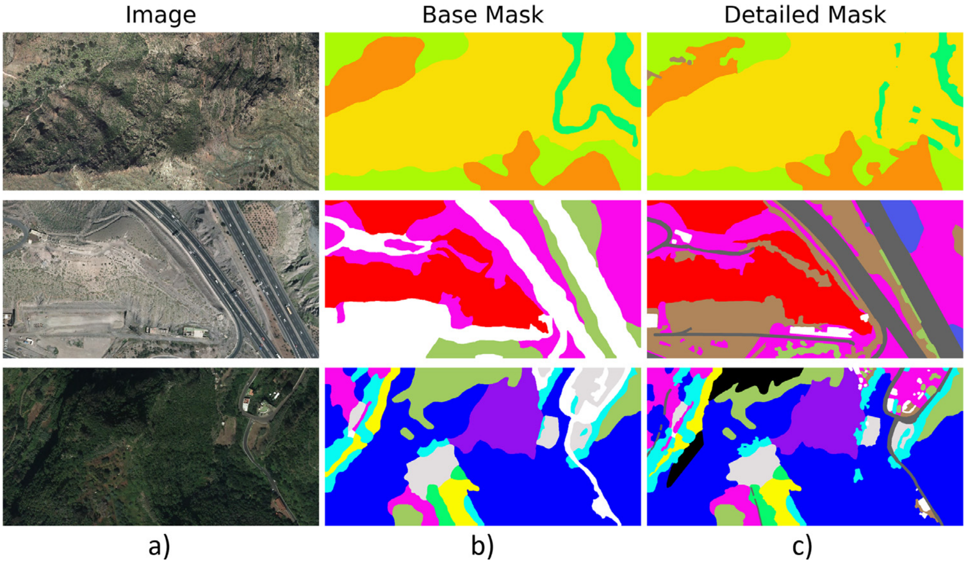


Fig. 1. Examples of some images with both of their corresponding masks.

plant communities. See Fig. 1c. This refinement process also includes annotating individual instances of trees, palms, and reeds that were previously omitted from the original masks due to their small size, making them too small to be considered a community, or because they were artificially planted. This additional detail allows the models to recognize these elements independently of their semantics. Furthermore, the non-vegetation class has been subdivided to improve class balance, incorporating additional categories such as "Water," "Roads," and "Shadows." This Detailed version comprises 25 distinct classes. This version is not meant to generate vegetation masks, but to have the most information about the elements in the zone.

Additionally, altitude data for each location is included, stored in .npy files. Infrared images in PNG format are also available for the years 2018 and 2021. This supplementary data is intended to serve as complementary information to better characterize the region.

Both versions of the masks account for 20 distinct vegetation communities, which are detailed in Table 1. These communities are the most characteristic of the island. Note that these associations are not evenly distributed, either on the island or within the dataset. The class distribution in both datasets can be visualized in Fig. 2.

Lastly, several CSV files are provided with detailed metadata, including the classes present in each image, the color associated with each vegetation community in the masks, the coordinates of each location and a translation of the local community names into English.

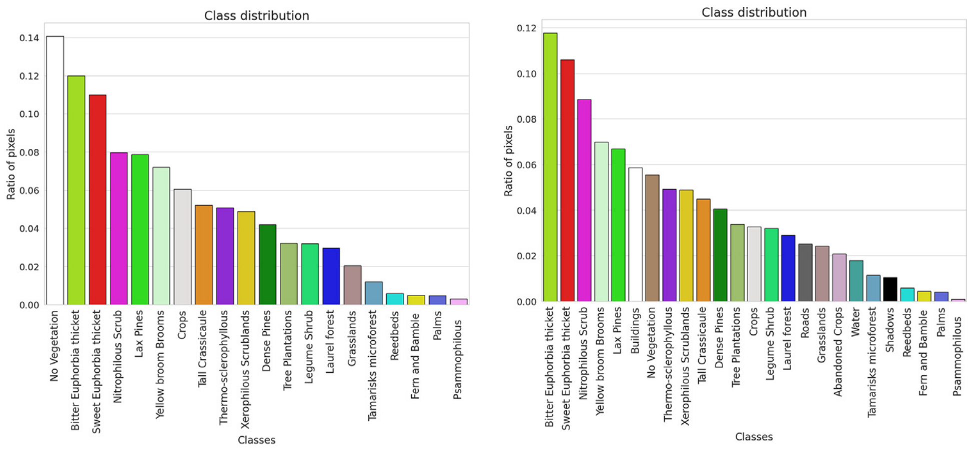
Both mask types are in the folder named "Masks," the original images are in the folder "Sat_images." The infrared information and altitude data are in the folder "Additional_data." Finally, the *csv* files are in the folder named "CSV_files."

A detailed comparison of the occurrence of each community in both types of masks can be observed in Table 2 and Fig. 3. Table 2 presents the raw data on the occurrence of each vegetation community in both versions of the dataset, including the number of images in which each community appears, the total number of occurrences, and the mean size of these occurrences. These data serve as the basis for the bar graphs shown in Fig. 3. Examining the table reveals a general trend of increased occurrences for each community in the detailed version, both in total count and in number of images in that the community appears. This is due to the inclusion of smaller occurrences that are not large enough to be considered a community in the base ver-

Table 1
Description of the vegetation that constitutes each one of the classes used for both versions with common and scientific names.

Index	Vegetation description	Scientific name (syntax)	Common Name used
1	Thermo-sclerophyllous woodlands	*Pistacio lentisci-Oleetum cerasiformis, Rhamno crenulatae-Hypericetum canariensis, Echio decaisnei-Retametum rhodorhizoidis*	Thermo-sclerophyllous
2	Psammophilous shrub community dominated by *Traganum moquinii*	*Traganetum moquinii*	Psammophilous
3	Reedbed	*Arundo donax community*	Reedbed
4	Tall crassicaule scrub characterized by *Euphorbia canariensis*	*Aeonio percarnei-Euphorbietum canariensis*	Tall Crassicaule scrub
5	Crops	None	Crops
6	Shrub legume community dominated by *Chamaecytisus proliferus*	*Chamaecytiso canariae-Adenocarpetum villosi*	Legume Shrub
7	Laurel forest, heather thicket, and willow grove	*Pruno hixae-Lauretalia novocanariensis, Morello fayae-Ericetum arboreae, Rubo ulmifolii- Salicetum canariensis*	Laurel Forest
8	Nitrophilous and halo-nitrophilous scrublands and shrublands	*Carthametalia lanati, Chenoleoidetalia tomentosae* and *Forsskaoleo angustifoliae-Rumicetalia lunariae*	Nitrophilous Scrub
9	*Phoenix canariensis* palm grove	*Periploco laevigatae-Phoenicetum canariensis*	Palms
10	Grasslands and herbaceous lands	*Poetea bulbosae* and *Stellarietea mediae*	Grasslands
11	Dense pine forest of *Pinus canariensis*	*Micromerio pineolentis-Pinetum canariensis* and *Pinetum canariensis*	Dense Pine
12	Lax pine forest *Pinus canariensis*	*Pinetum canariensis subass. juniperetosum canariensis*	Lax Pines
13	Without vegetation	*None*	No Vegetation
14	Dendroid *Euphorbia regis-jubae* shrubby communities	*Euphorbietum balsamiferae*	Bitter Euphorbia thicket
15	Crassicaule shrublands characterized by *Euphorbia balsamífera*	*Euphorbietum regis-jubae, Euphorbietum lamarckii, Plocametum pendulae*	Sweet Euphorbia thicket
16	Microforest of tamarisk, *Tamarix canariensis*	*Atriplici ifniensis-Tamaricetum canariensis, Suaedo verae-Tamaricetum canariensis*	Tamarisks microforest
17	Community of ferns and bramble bushes	*Rubio periclymeni-Rubetum ulmifolii*	Ferns and Bramble
18	Xerophilous scrublands dominated by *Cistus monspeliensis*	*Euphorbio regis-jubae-Cistetum canariensis*	Xerophilous Scrublands
19	Tree plantations (eucalyptus, chestnut, etc.)	*None*	Tree Plantations
20	Broom shrub community dominated by *Teline microphylla*	*Micromerio benthami-Telinetum microphyllae*	Yellow broom Brooms

sion but are classified as individual species in the detailed version. Notably, the occurrence of “Crops” decreases in the detailed version, as the class was separated in two “Crops” which refers to active cultivation and is categorized as a vegetation class, and “Abandoned Crops” which is not. On the other hand, analysing both, graphs 3.a and 3.b, it can be stated that the detailed masks exhibit a higher number of occurrences compared to the base masks, although these occurrences are smaller in area than those in the base masks. This difference arises because, as previously mentioned, while the base masks segment by communities, focusing on the collective rather than individual instances, the detailed masks employ a finer segmentation approach. This method prioritizes visual perception over community structure, subdividing large communities into smaller ones when they are widely dispersed across the territory and isolating individual specimens that are separated from the collective.



a) Class distribution in the Vegetation-Focused version.

b) Class distribution in the Detailed version.

Fig. 2. Class distribution in both versions of the dataset.

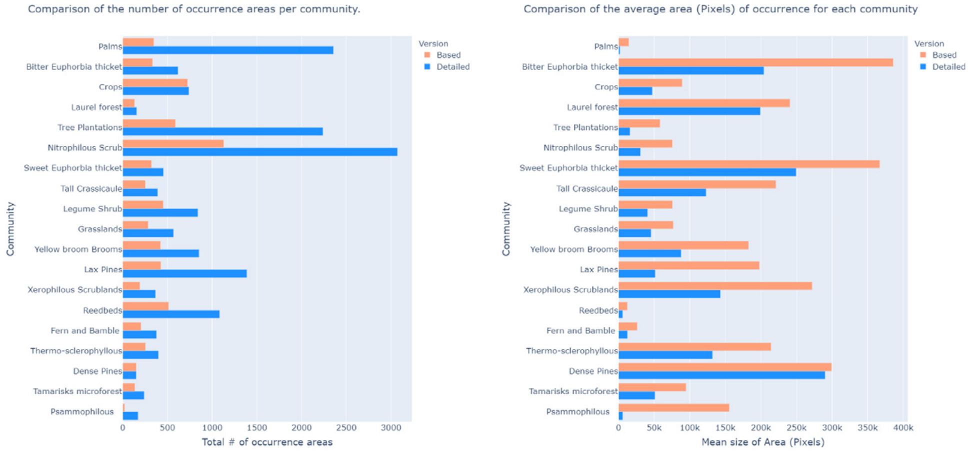
Table 2
A comparative table of each vegetal community across both types of masks. “Number of images” refers to the total number of images in which each community appears, “Total occurrences” indicates the number of individual appearances of each community, and “Mean size” represents the average size (in pixels) of these occurrences.

Vegetal Communities	Base Version			Detailed Version		
	Number of Images	Total Occurrences	Mean Size (In pixels)	Number of Images	Total Occurrences	Mean Size (In pixels)
Palms	93	348	14,403	171	2356	1,866
Bitter Euphorbia thicket	150	333	385,890	158	618	203,994
Crops	223	725	89,460	173	739	47,469
Laurel forest	39	132	240,791	39	156	199,171
Tree Plantations	144	589	58,351	235	2238	16,152
Nitrophilous Scrub	285	1130	75,452	324	3072	30,848
Sweet Euphorbia thicket	108	321	366,881	108	455	249,507
Tall Crassicaule	105	252	221,023	108	391	122,976
Legume Shrub	129	452	75,633	135	840	40,767
Grasslands	82	284	76,896	97	567	45,715
Yellow broom Brooms	110	422	182,560	114	853	87,839
Lax Pines	123	426	197,741	127	1388	51,536
Xerophilous Scrublands	84	192	271,900	93	366	143,025
Reedbeds	123	513	12,411	127	1084	5,855
Fern and Bamble	51	204	26,232	59	378	12,511
Thermo-sclerophyllous	75	254	214,190	81	399	131,952
Dense Pines	54	150	299,398	54	150	289,941
Tamarisks microforest	42	135	94,689	40	240	51,152
Psammophilous	15	21	155,402	15	171	5,917

4. Experimental Design, Materials and Methods

As illustrated in Fig. 4, the process that leads to the creation of the dataset’s masks was divided into five steps:

- **Image acquisition:** The initial step in constructing the database involved acquiring aerial images of the island. Open data provided by the Regional Government through the company GRAFCAN [14] was utilized. The aerial image of the island, downloaded in *.jp2* format, was



a) Comparison of the number of occurrence areas per community.

b) Comparison of the average area of occurrence (pixels) for each community.

Fig. 3. Graphs comparing the occurrence areas number and size per community between mask types.

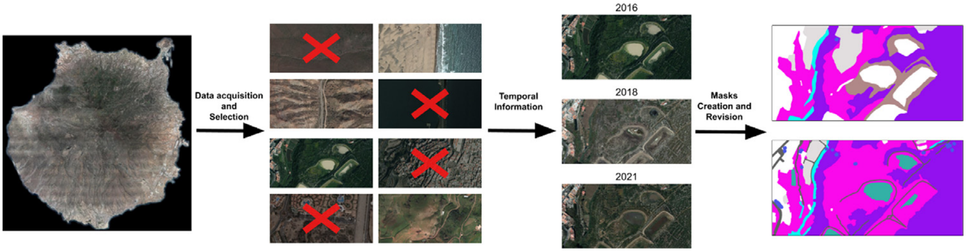


Fig. 4. Visual representation of the process of creation of the dataset.

processed using the open-source program QGIS to extract over 2000 images of the island's surface. The scale used was 1:1000 and the resulting RGB images in PNG format had a resolution of 1964×982 pixels. This scale was chosen as it is the minimum at which the details necessary to distinguish different vegetation communities can be recognized. This initial image downloaded was from 2016. Parallel to the image extractions, the coordinates where the photos were taken were saved in a CSV file. By doing this, additional data can be extracted from the same points and be used as help in the identification of vegetation.

- **Selection of representative images:** To assemble the dataset, 185 images representing the most significant vegetation communities on the island were selected.
- **Acquisition of temporal information:** Subsequently, images of the same areas represented in the original 185 images were captured in the years 2018 and 2021 to study the vegetation's evolution over this period. This approach also adds robustness to the dataset, as the images from different years are captured at separate times of the year, thereby depicting the vegetation in various states. The images from the three years add a total of 555.
- **Creation of the masks:** Masks were created from each one of the 185 original images that were going to be in the dataset. Two masks were made for each image. The first version based on 2016 images, which differentiated only between vegetation communities, grouping all non-vegetation into a single class called 'No Vegetation', resulting in a total of 20 classes. As said before, these masks were annotated by an expert in the endemic flora, using both, the visual information that each photo contains and their previous knowledge of the biome and

communities that the area depicted contains. The second version emphasized the different structures found and visual artifacts, distinguishing between 25 classes. For the rest of the images, the masks made for the 2016 photo were used and then slightly modified to cover the changes between the years. This step, as well as the selection of representative images, was performed by a renowned botanist specialist in the task of drawing vegetation maps in Canary Islands.

- **Cleaning and Reviewing of the Created masks:** Upon completing the mask creation process, three different individuals reviewed the masks from both a technical and environmental perspective, assessing the quality and fidelity of the annotations. Special attention was given to the images from 2018 and 2021, which were constructed based on the 2016 images.

Once the photographs and masks were finalized, supplementary data was collected from the same locations. Specifically, altitude data and infrared images of the captured areas were obtained. These data are intended to complement the original RGB images. For the Infrared images, only 2018 and 2021 years were covered, as the infrastructure to capture these images were not operative in 2016. This infrared information is a composition of the Infrared, Red and Green channels made from photos with 4 bands (RGBIr). Since the altitude data of the terrain is constant throughout the years, only one version of them was taken. Because the source images, visual, altitude, and infrared, are georeferenced, they could be aligned without displacement, ensuring that they provide complementary information for the exact same locations. Finally, some other CSV files were created for both versions of the dataset, gathering the information of which vegetation communities are present in each image and which of them change with the years.

Limitations

The unique limitation encountered during the dataset creation process was the unavailability of infrared maps for the year 2016, which restricted their generation for that period.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief. The current dataset does not involve human subjects, animal experiments, or data collected from social media platforms.

CRediT Author Statement

José Salas-Cáceres: Conceptualization, Methodology, Data curation, Writing – original draft, Visualization, Project administration, Investigation, Validation. **Riccardo Balia:** Data curation, Writing – review & editing, Validation. **Marcos Salas-Pascual:** Conceptualization, Data curation, Writing - Original Draft, Methodology. **Javier Lorenzo-Navarro:** Supervision, Writing – review & editing, Project administration. **Modesto Castrillón-Santana:** Supervision, Writing – review & editing, Project administration.

Data Availability

Vegetation mapping dataset of Gran Canaria, Canary Islands, Spain. (Original data) (Zenodo).

Acknowledgements

This work is part of the project PID2021-122402OB-C22, funded by MCIN/AEI/ 10.13039/501100011033/FEDER, UE. By the ACIISI-Gobierno de Canarias and European FEDER funds under project ULPGC Facilities Net and Grant EIS 2021 04, it is also supported by the Ministry of Universities, Science, Innovation and Culture of the Canary Islands and the European Social Fund Plus (FSE+) under the funding framework for doctoral research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Salas-Cáceres, J. Lorenzo-Navarro, M. Castrillón-Santana, M. Salas-Pascual, R. Balia, Vegetation mapping dataset of Gran Canaria, Zenodo, Canary Islands, Spain, 2024 [Data set], doi:10.5281/zenodo.14215453.
[2] M.J. del Arco Aguilar, W. Wildpret de la Torre, P.L. Pérez De Paz, O. Rodríguez Delgado, J.R. Acebes Ginovés, A. García Gallo, V.E. Martín Osorio, J.A. Reyes Betancort, M. Salas Pascual, M.A. Díaz, J.A. Bermejo Domínguez, R. González González, M.V. Cabrera Lacalzada, S. García Ávila, Mapa de Vegetación de Canarias, GRAFCAN, Santa Cruz de Tenerife, 2006.
[3] M.J. del Arco Aguilar, O. Rodríguez Delgado, Vegetation of the Canary Islands. Plant and vegetation, Springer, Cham, 2018.
[4] M. Salas-Pascual, Flora y vegetación de Gran Canaria, in: J. Afonso-Carrillo (Ed.), Gran Canaria: las huellas del tiempo, Instituto de Estudios Hispánicos de Canarias, Puerto de la Cruz, Tenerife, 2020, pp. 115–165.
[5] J.Y. Loidi, D. Vynokurov, The biogeographical kingdoms and regions of the world, Mediterran. Botany 45 (2) (2024) e92333, doi:10.5209/mbot.92333.
[6] SpaceNet on Amazon Web Services (AWS). “Datasets.” The SpaceNet catalog. last modified October 1st, 2018. Accessed on 20/11/2024. https://spacenet.ai/datasets/.
[7] J. Wang, Z. Zheng, A. Ma, X. Lu, Y. Zhong, LoveDA: a remote sensing land-cover dataset for domain adaptive semantic segmentation, in: J. Vanschoren, S. Yeung (Eds.), Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (Vol. 1), 2021 Retrieved from.
[8] Wang, J., Zheng, Z., Ma, A., Lu, X., & Zhong, Y. (2021). LoveDA: a remote sensing land-cover dataset for domain adaptive semantic segmentation [Data set]. doi:10.5281/zenodo.5706578.
[9] O.D. Pedrayes, D.G. Lema, D.F. García, R. Usamentiaga, Á. Alonso, Evaluation of semantic segmentation methods for land use with spectral imaging using sentinel-2 and PNOA imagery, Remote Sensing 13 (12) (2021) 2292, doi:10.3390/rs13122292.
[10] B. Tershy, K.-W. Shen, K. Newton, N. Holmes, D. Croll, The importance of islands for the protection of biological and linguistic diversity, BioScience 1 (6) (2015), doi:10.1093/biosci/biv031.
[11] B. Martín de la Rosa, Turismo en Ecosistemas insulares. Antropología en el Paraíso, ACA y PASOS, RTPC, El Sauzal (Tenerife), 2009.
[12] D. Quammen, The Song of the Dodo: Island Biogeography in an Age of Extinctions, Scribner, New York, 1996.
[13] P.L. Pérez de Paz, Panaroma actual de la cartografía de la vegetación de las Islas Canarias, Lazaroa 25 (2004) 51–62.
[14] Cartográfica de Canarias (GRAFCAN). https://www.grafcan.es, 2023 (accessed 02 December 2024).