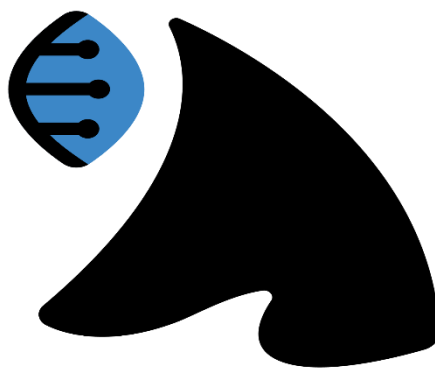


# eLasmobranc Dataset: Un conjunto de datos de imágenes de elasmobrancos



**e-Lasmobranc**



Cofinanciado por  
la Unión Europea



MINISTERIO  
DE AGRICULTURA, PESCA  
Y ALIMENTACIÓN



Fondos Europeos



MINISTERIO DE  
TRANSICIÓN ECOLÓGICA  
Y RETO DE ENERGÍA



Fundación Biodiversidad



Pleamar



Universitat d'Alacant  
Universidad de Alicante

## Autores del informe:

Ismael Beviá Ballesteros <sup>1</sup>

Nieves Aranda Garrido <sup>2, 3</sup>

Irene Antón Linares <sup>2, 3</sup>

Isabel Abel Abellán <sup>2</sup>

Marcelo Saval Calvo <sup>1</sup>

Jorge Azorín López <sup>1</sup>

Andrés Fuster Guilló <sup>1</sup>

Francisca Giménez-Casalduero <sup>2,3,4</sup>

<sup>1</sup>Departamento de Tecnología Informática y Computación (dTIC), Universidad de Alicante, Alicante, España.

<sup>2</sup>Centro de Investigación Marina de Santa Pola (CIMAR), Universidad de Alicante, Alicante, España.

<sup>3</sup>Instituto Multidisciplinario de Estudios Ambientales “Ramón Margalef”, Universidad de Alicante, Alicante, España

<sup>4</sup>Departamento de Ciencias del Mar y Biología Aplicada, Universidad de Alicante, Alicante, España.

Este documento se elabora en el proyecto e-Lasmobranc el cual se desarrolla con la colaboración de la Fundación Biodiversidad del Ministerio para la Transición Ecológica y el Reto Demográfico, a través del Programa Pleamar, y se cofinancia por la Unión Europea por el FEMPA (Fondo Europeo Marítimo, de Pesca y de Acuicultura).

## ÍNDICE

|                               |    |
|-------------------------------|----|
| Introducción .....            | 4  |
| Objetivo de la actividad..... | 5  |
| eLasmobranc Dataset .....     | 7  |
| Metodología .....             | 7  |
| eLasmobranc Dataset .....     | 10 |
| Conclusiones .....            | 14 |
| Referencias.....              | 15 |

## Índice de Figuras

|                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------|----|
| Ilustración 1. Distribución de imágenes obtenidas a partir de fuentes externas (izquierda) y de fuentes internas (derecha). ..... | 10 |
| Ilustración 2. Distribución de las especies en el conjunto de datos. ....                                                         | 10 |
| Ilustración 3. Correspondencia entre imágenes e individuos en el conjunto de datos por especie. ....                              | 11 |
| Ilustración 4. Representación de los cinco años con mayor número de imágenes obtenidas, diferenciados por especie. ....           | 11 |
| Ilustración 5. Representación de todos los meses del año con el número de imágenes obtenidas para cada especie. ....              | 12 |
| Ilustración 6. Distribución del número de imágenes según el país de procedencia. ....                                             | 12 |
| Ilustración 7. Distribución del número de imágenes en España por área, desglosada por especie según su procedencia. ....          | 12 |

## Índice de Tablas

|                                                                                                                                          |   |
|------------------------------------------------------------------------------------------------------------------------------------------|---|
| Tabla 1. Resumen de los criterios empleados para el filtrado de las imágenes, incluyendo criterios básicos y de validación experta. .... | 8 |
| Tabla 2. Los valores indican el número de imágenes tras cada fase del proceso de filtrado en los distintos datasets. ....                | 8 |

## Introducción

En el marco del proyecto e-Lasmobranc, el presente informe tiene como objetivo documentar y verificar la correcta ejecución de la actividad relacionada con el desarrollo del conjunto de datos de imágenes de elasmobranquios, denominado eLasmobranc Dataset. Para ello, se describe la metodología aplicada y los resultados obtenidos hasta la fecha.

Este conjunto de datos constituye una herramienta fundamental dentro del proyecto. No solo resulta clave para el desarrollo de herramientas basadas en inteligencia artificial, sino que también presenta un elevado valor para la realización de análisis estadísticos, estudios poblacionales y tareas de apoyo a la investigación científica. En este sentido, la incorporación de información adicional asociada a cada imagen, como la fecha de captura y la localización geográfica, enriquece significativamente el valor del recurso. Asimismo, el eLasmobranc Dataset actúa como un registro de avistamientos, lo que lo convierte en un recurso especialmente relevante para el estudio y seguimiento de áreas de importancia para tiburones y rayas (ISRAs) [1,2], aportando información de gran utilidad para la evaluación de la presencia y distribución de estas especies.

El conjunto final se concibe como un recurso abierto, con el objetivo de cubrir una carencia actualmente existente, dado que no se tiene constancia de la disponibilidad de un recurso similar que recopile de manera sistemática imágenes y metadatos de estas especies concretas. De este modo, se pretende que el dataset alcance un gran valor, con aplicaciones tanto en el ámbito de las ciencias del mar como en el de la visión por computador. El resultado del trabajo se plantea como un conjunto de datos inicial, consolidado y en continua evolución, que va a constar de ampliaciones y actualizaciones progresivas que permitan incorporar nuevos registros y aumentar el número de especies sometidas a estudio.

Para la exposición y publicación de los resultados, se ha desarrollado un repositorio general del proyecto disponible en <https://github.com/Tech4DLab/eLasmobranc>, donde se centraliza el código, la documentación y los recursos asociados. Además, se ha habilitado un apartado específico dedicado al desarrollo y explicación del eLasmobranc Dataset, accesible en <https://github.com/Tech4DLab/eLasmobranc-Dataset>. Finalmente, la versión pública del conjunto de datos, junto con sus metadatos y

material complementario, se encuentra disponible para su descarga a través de Zenodo en <https://zenodo.org/records/18549737>.

## Objetivo de la actividad

En el marco de esta tarea, el desarrollo del eLasmobranc Dataset se ha llevado a cabo atendiendo a diversos subobjetivos y necesidades que han ido surgiendo durante el avance del proyecto. De forma general, la actividad principal puede resumirse como el **diseño y desarrollo de un dataset de imágenes de las especies objetivo de tiburones y rayas, estructurado y validado para su posterior explotación científica y tecnológica.**

Cabe destacar que, en línea con los objetivos y el contexto del proyecto, el dataset se centra en las especies objetivo presentes en la Región de Murcia y se compone de imágenes en las que los ejemplares presentan sus rasgos característicos completamente visibles y cumplen un umbral mínimo de calidad. Estas imágenes han sido capturadas mayoritariamente fuera del agua, con el fin de ajustarse al contexto del estudio y minimizar problemas de visibilidad y distorsión visual, salvo en casos puntuales debidamente justificados. Asimismo, las imágenes son revisadas y validadas por expertos, garantizando la fiabilidad de la identificación taxonómica.

Para alcanzar este objetivo, se han definido y ejecutado las siguientes acciones:

1. **Revisión de datasets previos**, tanto del propio grupo de investigación como de repositorios públicos, con el fin de analizar el panorama actual, identificar limitaciones y aprovechar recursos ya disponibles posibles.
2. **Catalogación de fuentes de imágenes**, incluyendo propias en distintos contextos de captura, colaboraciones y trabajos previos, así como fuentes externas.
3. **Definición de pautas de trabajo específicas** para la captura y tratamiento de imágenes de tiburones y rayas.
4. **Obtención de un nuevo conjunto de imágenes**, estimado entre 800 y 1.000 imágenes.

5. **Estandarización de metadatos clave**, considerados desde un punto de vista científico y técnico, asegurando coherencia y reutilización.
6. **Diseño y definición de la estructura del dataset**, incluyendo la organización de imágenes, etiquetas y metadatos.

## eLasmobranc Dataset

En este apartado se describen los tres últimos objetivos de la actividad, así como el conjunto de datos final, que se analiza en detalle. Esta estructura se debe a que los objetivos previos ya han sido abordados y documentados en informes independientes.

## Metodología

El conjunto de imágenes se desarrolló tras un análisis exhaustivo de diversas fuentes de datos, las cuales han sido descritas en detalle en su informe correspondiente. De forma resumida, las fuentes externas incluyen datasets públicos de gran escala, imágenes de plataformas web, trabajos previos del grupo y el material generado u obtenido directamente por el grupo de investigación. En todos los casos, se verificó que los recursos contaran con licencias Creative Commons [3] que permitieran explícitamente su reutilización.

Las especies objetivo consideradas finalmente en esta versión conjunto de datos fueron seleccionadas en función de su disponibilidad de datos y presencia en la zona de estudio. En concreto, el dataset incluye las siguientes especies de elasmobranquios: *Galeorhinus galeus*, *Galeus melastomus*, *Leucoraja naevus*, *Mustelus mustelus*, *Oxynotus centrina*, *Scyliorhinus canicula* y *Torpedo marmorata*.

La selección final de imágenes se realizó mediante la aplicación de un conjunto de criterios de filtrados claramente definidos, una serie de criterios básicos fácilmente aplicables y otros que son necesariamente aplicados por expertos.

| Criterios Básicos                                                                            | Criterios expertos                                                                                           |
|----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| Debe corresponder a una de las especies de elasmobranquios definidas como objetivo.          | Se descartan aquella donde no sean visibles las características morfológicas mínimas para la identificación. |
| Debe cumplir un umbral mínimo de calidad visual.                                             | Se excluyen imágenes catalogadas erróneamente.                                                               |
| Debe haber sido capturada fuera del agua, admitiéndose excepciones debidamente justificadas. | Se eliminan duplicados y redundancias no detectadas.                                                         |
| Debe ser visible de forma completa o suficiente para su identificación.                      |                                                                                                              |
| Debe contener una única especie de elasmobranquio.                                           |                                                                                                              |

Se excluyen imágenes duplicadas o redundantes.

*Tabla 1. Resumen de los criterios empleados para el filtrado de las imágenes, incluyendo criterios básicos y de validación experta.*

En este sentido, el proceso de filtrado de las fuentes externas se estructuró en dos fases: una primera etapa basada en la aplicación de criterios simples y una segunda etapa de validación por parte de un experto. En el caso concreto de los datasets genéricos, se partió de subconjuntos correspondientes a las especies objetivo, seleccionados a partir de las etiquetas originales. Por otro lado, las fuentes propias, incluyendo el dataset DeepFish, ya cumplen los criterios simples establecidos por lo que solo son sometidos a la identificación y criterios expertos por revisión experta.

| Dataset         | Inicial | Criterios básicos | Criterios expertos |
|-----------------|---------|-------------------|--------------------|
| iNaturalist [4] | 7 892   | 835               | 766                |
| GBIF [5]        | 1325    | 86                | 74                 |
| DeepFish [6]    | 14599   | -                 | 4                  |

*Tabla 2. Los valores indican el número de imágenes tras cada fase del proceso de filtrado en los distintos datasets.*

Con respecto a los metadatos, para todas las fuentes se revisó, o se anotó en el caso de material propio, el país y área de captura, así como la fecha. La fecha se normalizó siguiendo un formato estándar DD-MM-YYYY. Para la normalización de las ubicaciones geográficas, se utilizó como estandar el servicio de geocodificación Nominatim de OpenStreetMap. A partir de esta información se completaron los campos de país (country) y área (area), este último seleccionando el primer valor disponible según el siguiente orden de preferencia: state, archipelago, county, region; y asignando un valor vacío en caso de no disponer de información adicional. Por otro lado, en los casos donde no fue posible atribuir a un país concreto, cuando la captura se produjo en zonas oceánicas abiertas, se asignó como valor el nombre de la zona u océano correspondiente.

El dataset organiza su estructura en un directorio principal que contiene: subcarpetas independientes para cada especie, un CSV de atribuciones, un archivo de texto citations y un documento Excel explicativo de las especies.

Cada imagen se identifica mediante un identificador único que codifica la especie, el individuo y el número de imagen asociada a dicho individuo. Por ejemplo, el ID

GG\_001\_01 corresponde a la primera imagen del individuo 001 de la especie *Galeorhinus galeus*. Asimismo, cada subcarpeta de especie incluye, además, un archivo CSV propio que recoge los metadatos asociados a cada imagen, incluyendo su identificador, la fecha de observación, el país y el área.

El archivo CSV de atribuciones recoge, para cada imagen del conjunto de datos, información relativa a la procedencia de la fuente (interna o externa), la licencia original, así como un texto de atribución que identifica al propietario original del recurso y la licencia asociada. Asimismo, se especifica la fuente de procedencia y, en el caso de fuentes externas, el identificador original de la imagen. Para las imágenes procedentes de datasets generales se incluye el identificador correspondiente, mientras que para aquellas obtenidas a partir de fuentes web se almacena la URL exacta de origen. En el caso de los datos propios, el material ha sido distribuido bajo la licencia CC-BY-NC.

Por último, el dataset incorpora un archivo de texto citations, en el que se proporciona la cita identificativa correspondiente de cada fuente y un documento Excel que explica las especies de elasmobranquios incluidas y sus rasgos característicos.

## eLasmobranc Dataset

El dataset final se compone de un total de 1117 imágenes, de las cuales 902 proceden de fuentes externas y 215 de fuentes internas. En concreto, el conjunto de datos incluye las siguientes especies de elasmobrancios: *Galeorhinus galeus*, *Galeus melastomus*, *Leucoraja naevus*, *Mustelus mustelus*, *Oxynotus centrina*, *Scyliorhinus canicula* y *Torpedo marmorata*. Es importante señalar que el conjunto de datos no presenta una correspondencia uno a uno entre imágenes e individuos. De las 1117 imágenes, 807 corresponden a individuos distintos.

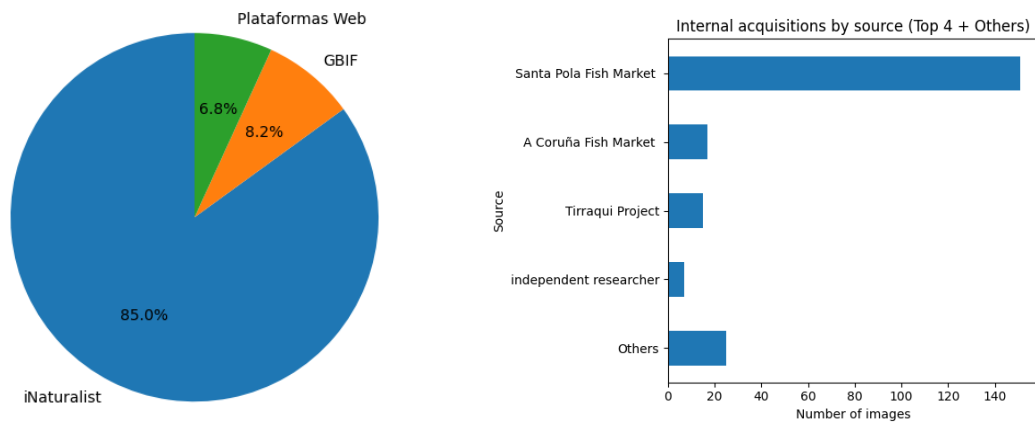


Ilustración 1. Distribución de imágenes obtenidas a partir de fuentes externas (izquierda) y de fuentes internas (derecha).

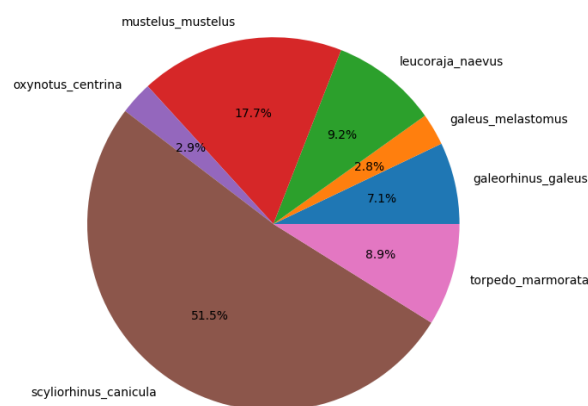


Ilustración 2. Distribución de las especies en el conjunto de datos.

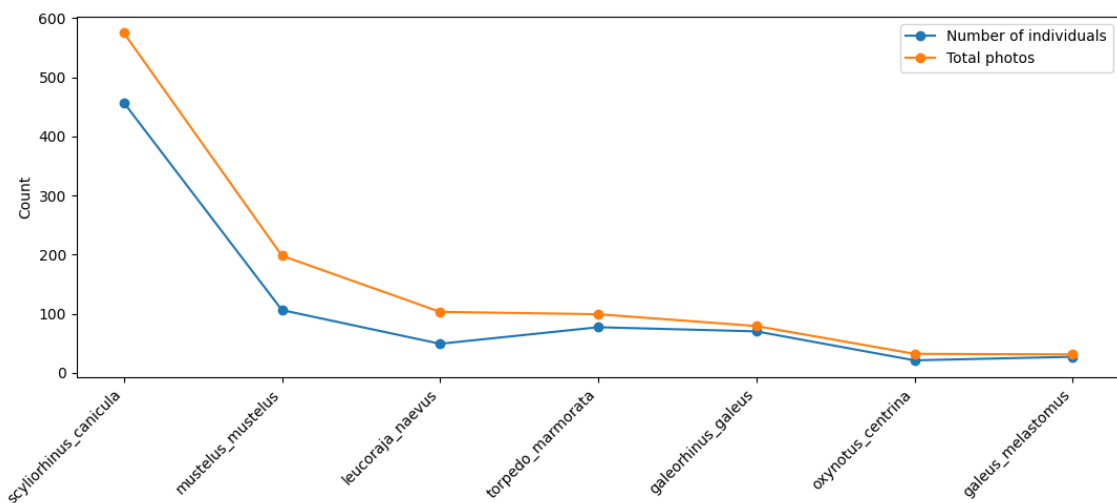


Ilustración 3. Correspondencia entre imágenes e individuos en el conjunto de datos por especie.

Con respecto a los metadatos asociados, se ha realizado un análisis temporal y geográfico considerando el año, el mes de captura, así como el país y el área concreta de procedencia. Del total de imágenes, en 7 casos no se dispone de información temporal a ningún nivel, 27 imágenes carecen de información sobre el país y en 164 imágenes no se pudo determinar el área específica. Estas imágenes fueron excluidas del análisis correspondiente.

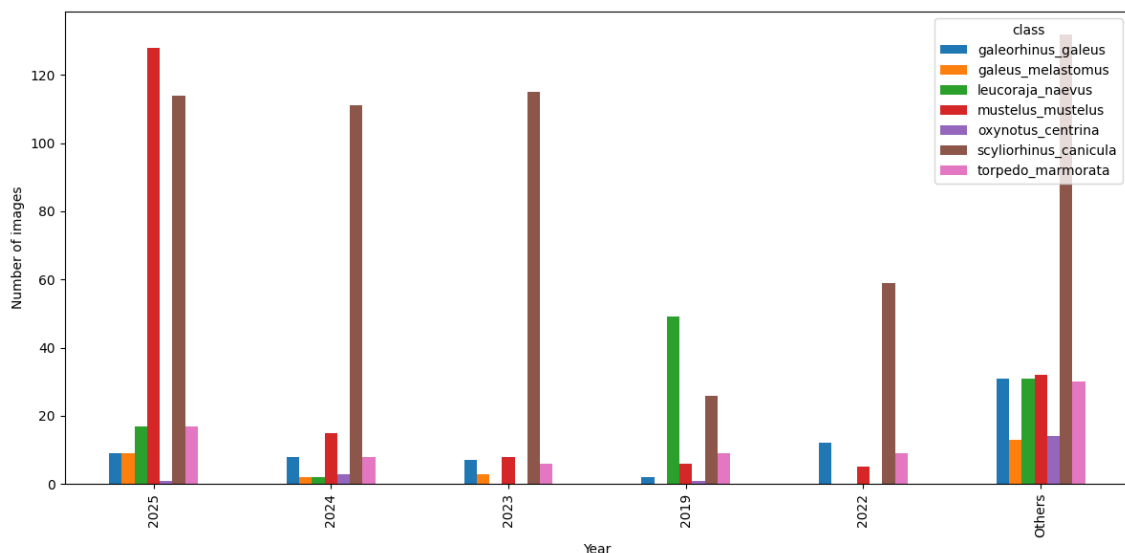


Ilustración 4. Representación de los cinco años con mayor número de imágenes obtenidas, diferenciados por especie.

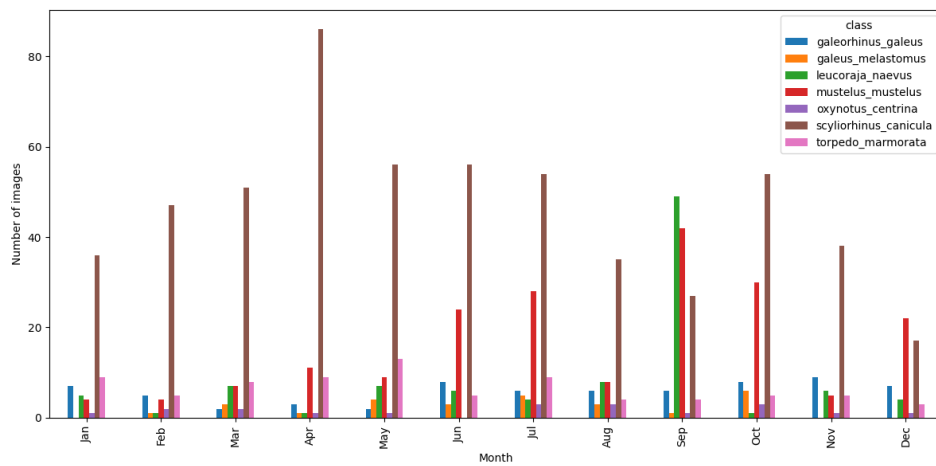


Ilustración 5. Representación de todos los meses del año con el número de imágenes obtenidas para cada especie.

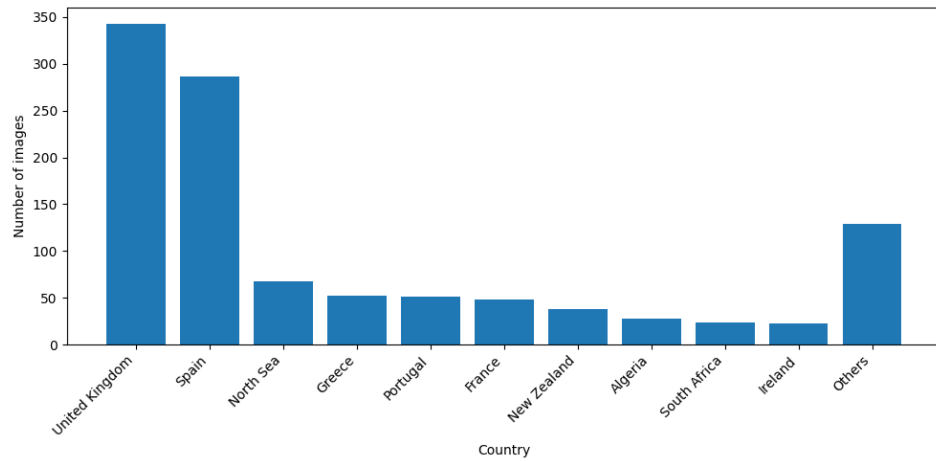


Ilustración 6. Distribución del número de imágenes según el país de procedencia.

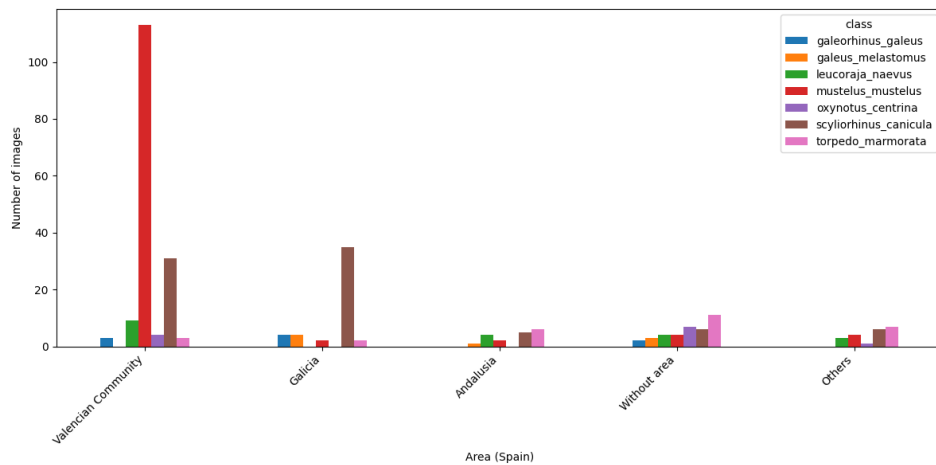


Ilustración 7. Distribución del número de imágenes en España por área, desglosada por especie según su procedencia.

Finalmente, el conjunto de datos se organiza en un directorio principal que contiene los siguientes elementos:

- **data/**

Carpeta principal que alberga los datos visuales del *dataset*.

- **Subcarpetas por especie/**

Cada subcarpeta corresponde a una especie de elasmobranquio e incluye:

- Imágenes identificadas mediante un ID único (por ejemplo, *GG\_001\_01*).
- Un archivo de metadatos específico de la especie (por ejemplo, *GG\_metadata.csv*).

- **Elasmobranch taxonomic keys.xlsx**

Documento descriptivo que recoge todas las especies de elasmobranquios y sus rasgos morfológicos característicos.

- **attributions.csv**

Archivo que contiene información detallada sobre la procedencia y licencias de cada imagen.

- **citations.txt**

Archivo de texto que proporciona las citas identificativas de todas las fuentes utilizadas.

## Conclusiones

El desarrollo del eLasmobranc Dataset ha permitido consolidar un conjunto de datos estructurado y validado de imágenes de elasmobranquios. Hasta donde se conoce, no existen trabajos previos que agrupen especies de este tipo y nivel taxonómico, integrando además un enfoque orientado tanto al ámbito de las ciencias del mar como al de la visión por computador. El resultado final, compuesto por 1.117 imágenes correspondientes a 807 individuos distintos y siete especies objetivo, constituye un recurso de alto valor científico y tecnológico

La metodología aplicada para su construcción nos permite minimizar los errores presentes en datasets genéricos y asegurar que los ejemplares incluidos presentan rasgos morfológicos suficientes para su correcta identificación, reforzando así la utilidad del conjunto de datos. Asimismo, la estandarización y normalización de los metadatos asociados, incluyendo información temporal y geográfica, amplía significativamente el potencial de explotación del dataset, no solo para modelos de inteligencia artificial, sino también para análisis espaciales.

Cabe destacar que el eLasmobranc Dataset, se concibe como un recurso en continua evolución. Las futuras ampliaciones previstas, tanto en número de imágenes como en diversidad de especies y contextos de captura, permitirán incrementar progresivamente su representatividad y robustez. De este modo, el dataset sienta una base sólida para el desarrollo de herramientas avanzadas de análisis y monitorización de elasmobranquios, contribuyendo de forma directa a los objetivos del proyecto e-Lasmobranc y al conocimiento y conservación de estas especies.

Para la exposición y publicación de los resultados, se ha desarrollado un repositorio general del proyecto disponible en <https://github.com/Tech4DLab/eLasmobranc>, donde se centraliza el código, la documentación y los recursos asociados. Además, se ha habilitado un apartado específico dedicado al desarrollo y explicación del eLasmobranc Dataset, accesible en <https://github.com/Tech4DLab/eLasmobranc-Dataset>. Finalmente, la versión pública del conjunto de datos, junto con sus metadatos y material complementario, se encuentra disponible para su descarga a través de Zenodo en <https://zenodo.org/records/18549737>.

## Referencias

- [1] Pozo-Montoro, M.; Arroyo, E.; Abel, I.; Bas Gómez, A.; Clemente Navarro, P.; Cortés, E.; Esteban, A.; García-Charton, J.A.; López Castejón, F.; Ortolano, A.; et al. Important Shark and Ray Areas (ISRAs) en el SE ibérico, una declaración necesaria. In Proceedings of the XV Reunión del Foro Científico Sobre la Pesca Española en el Mediterráneo; Universitat d'Alacant: Alicante, Spain, 2025; pp. 55–64.
- [2] Jabado, R.W.; García-Rodríguez, E.; Kyne, P.M.; Charles, R.; Armstrong, A.H.; Bortoluzzi, J.; Mouton, T.L.; Gonzalez-Pestana, A.; Battle-Morera, A.; Rohner, C.; et al. Mediterranean and Black Seas: A Regional Compendium of Important Shark and Ray Areas; Technical report; IUCN SSC Shark Specialist Group: Dubai, United Arab Emirates, 2023.
- [3] Creative Commons Licenses. Information on public licenses for shared works (2026). Disponible en: <https://creativecommons.org/share-your-work/cclicenses/>
- [4] iNaturalist Licensed Observation Images was accessed on 02/12/2025 from <https://registry.opendata.aws/inaturalist-open-data>.
- [5] GBIF.org (13 December 2025) GBIF Occurrence Download <https://doi.org/10.15468/dl.39xqca>
- [6] Garcia-d'Urso, N., Galan-Cuenca, A., Pérez-Sánchez, P. et al. The DeepFish computer vision dataset for fish instance segmentation, classification, and size estimation. Sci Data 9, 287 (2022). <https://doi.org/10.1038/s41597-022-01416-0>