



**COMILLAS**  
UNIVERSIDAD PONTIFICIA

ICAI

GRADO EN INGENIERÍA MATEMÁTICA E  
INTELIGENCIA ARTIFICIAL

TRABAJO FIN DE GRADO

IDENTIFICACIÓN DE ESPECIES VEGETALES EN  
PINTURA: CREACIÓN DE UN DATASET Y  
MODELADO EXPLORATORIO EN VISIÓN POR  
COMPUTADOR

Autor: Claudia García Birichinaga

Director: Dido Carrero Muñiz

Madrid

Junio de 2026

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título  
Identificación de especies vegetales en pintura: Creación de un *dataset* y modelado  
exploratorio en visión por computador

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el  
curso académico 2025/26 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido  
tomada de otros documentos está debidamente referenciada.

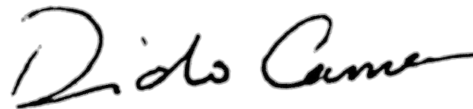


Fdo.: Claudia García Birichinaga

Fecha: 14/6/2026

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Dido Carrero Muñiz

Fecha: 15/6/2026



**COMILLAS**

UNIVERSIDAD PONTIFICIA

ICAI

# GRADO EN INGENIERÍA MATEMÁTICA E INTELIGENCIA ARTIFICIAL

TRABAJO FIN DE GRADO

## IDENTIFICACIÓN AUTOMÁTICA DE ESPECIES VEGETALES EN PINTURA: CREACIÓN DE UN DATASET Y MODELADO EXPLORATORIO EN VISIÓN POR ORDENADOR

Autor: Claudia García Birichinaga

Director: Dido Carrero Muñiz

Madrid

Junio de 2026

# IDENTIFICACIÓN AUTOMÁTICA DE ESPECIES VEGETALES EN PINTURA: CREACIÓN DE UN DATASET Y MODELADO EXPLORATORIO EN VISIÓN POR ORDENADOR

**Autor:** Claudia García Birichinaga

Director: Dido Carrero Muñiz

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

## RESUMEN DEL PROYECTO

Este Trabajo Fin de Grado aborda la identificación automática de motivos florales representados en pintura e ilustración botánica histórica mediante modelos de visión por ordenador. Para ello, se construyó un *dataset* propio de imágenes artísticas con etiquetas botánicas normalizadas y se compararon distintas arquitecturas preentrenadas mediante técnicas de *transfer learning* y *fine-tuning*. El mejor rendimiento se obtuvo con PlantCLEF-DINOv2, que alcanzó una *accuracy* Top-1 de 0,754, una *balanced accuracy* de 0,704 y un F1 macro de 0,644 en test.

**Palabras clave:** visión por ordenador, clasificación botánica, *transfer learning*, *fine-tuning*, CNN, *Vision Transformers*, DINOv2, patrimonio artístico.

## 1. Introducción

La representación de flores y plantas ha tenido un papel relevante en la historia del arte, tanto por su valor ornamental como por su función simbólica, iconográfica y compositiva.

La identificación de estos motivos vegetales puede aportar información útil para interpretar obras, estudiar su contexto cultural y analizar la presencia de determinados géneros botánicos en distintos periodos, escuelas y estilos artísticos. Sin embargo, este análisis se ha realizado tradicionalmente de forma manual, mediante la observación experta de historiadores del arte, botánicos o especialistas en iconografía vegetal.

Desde el punto de vista de la visión por ordenador, esta tarea presenta una dificultad específica. Los modelos actuales de clasificación botánica suelen entrenarse sobre fotografías naturalistas, donde las plantas aparecen de forma directa y con rasgos morfológicos claros. En cambio, en pintura las flores pueden aparecer estilizadas, parcialmente ocultas, idealizadas o integradas en composiciones complejas. Por tanto, existe un problema de cambio de dominio entre la fotografía botánica y la representación artística.

Este trabajo estudia si los modelos preentrenados pueden adaptarse a este contexto mediante técnicas de *transfer learning* y *fine-tuning*. El objetivo no es sustituir el criterio experto, sino explorar el uso de inteligencia artificial como herramienta de apoyo para analizar colecciones digitales con contenido vegetal.

## 2. Objetivos

El objetivo principal del trabajo es evaluar la viabilidad de aplicar modelos de visión por ordenador preentrenados a la identificación automática de motivos florales en obras pictóricas e ilustraciones botánicas históricas. La pregunta principal es si un modelo con

conocimiento visual previo puede adaptarse al reconocimiento de flores representadas artísticamente.

Para ello, se plantearon tres objetivos específicos. En primer lugar, construir un dataset propio de recortes florales procedentes de fuentes como el Museo del Prado, Rijksmuseum, The Metropolitan Museum of Art y *Curtis's Botanical Magazine*. En segundo lugar, definir un vocabulario botánico normalizado, organizado en familia, género y especie. En tercer lugar, comparar distintas arquitecturas de clasificación de imágenes y analizar su comportamiento mediante métricas globales, métricas equilibradas y análisis por clase, fuente y contexto artístico.

### 3. Descripción del modelo/sistema/herramienta

La solución desarrollada se organiza como un pipeline completo de construcción de datos, entrenamiento, evaluación y análisis, como se muestra en la Figura 1. La unidad básica del dataset es el recorte manual de una flor o motivo floral identificable dentro de una obra artística. Esta decisión permite asociar cada imagen a una única etiqueta botánica y evita la ambigüedad de las obras que contienen varias especies vegetales.

El conjunto final empleado en los experimentos contiene 619 imágenes, organizadas en una estructura jerárquica de familia, género y especie. La clasificación principal se realizó a nivel de género botánico, con 31 géneros considerados. Cada imagen se asoció a un archivo de metadatos que recoge su nombre, especie, género, familia, fuente, identificador original y partición de entrenamiento, validación o test.

Para reducir el riesgo de *data leakage*, la división del *dataset* se realizó agrupando los recortes relacionados con una misma obra, fuente o etiqueta botánica. Durante el entrenamiento se aplicó *data augmentation* mediante transformaciones moderadas de aumento de datos, como recortes aleatorios, volteo horizontal, pequeñas rotaciones y variaciones controladas de color. En validación y test se utilizaron transformaciones deterministas para garantizar una evaluación comparable.

Los modelos evaluados fueron ResNet-50, EfficientNet-B3, ConvNeXt-Small, ViT-Base y PlantCLEF-DINOv2. Todos ellos se adaptaron sustituyendo su cabeza de clasificación final por una nueva capa ajustada al número de géneros botánicos. Además, se compararon dos estrategias: *linear probing*, en la que el *backbone* permanece congelado, y *full fine-tuning*, en la que se actualizan todos los pesos del modelo.

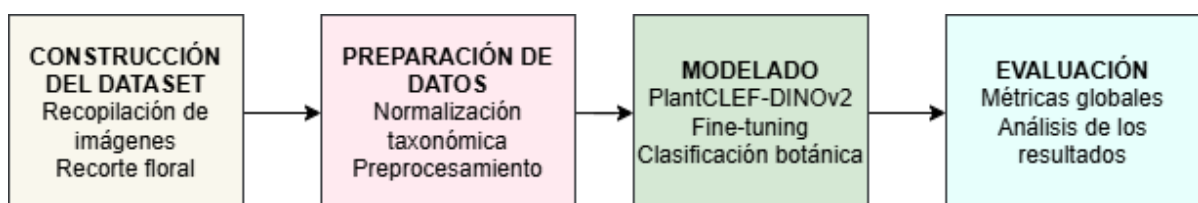


Figura 1. Flujo general del sistema desarrollado: construcción del dataset, preparación de datos, ajuste fino del modelo y evaluación de resultados.

## 4. Resultados

La evaluación se realizó sobre el conjunto de test mediante métricas de clasificación multiclase: *accuracy* Top-1, Top-3 y Top-5, *balanced accuracy*, F1 macro, *F1 weighted*, *Cohen's kappa* y *Matthews correlation coefficient*. Las métricas equilibradas son especialmente importantes porque el *dataset* presenta desbalance entre clases.

El mejor modelo fue PlantCLEF-DINOv2 con *full fine-tuning*, que obtuvo una *accuracy* Top-1 de 0,754, una *balanced accuracy* de 0,704, un F1 macro de 0,644 y un MCC de 0,744. ConvNeXt-Small fue el mejor modelo generalista, con una *accuracy* Top-1 de 0,719 y los mejores resultados Top-3 y Top-5, aunque quedó por debajo de PlantCLEF-DINOv2 en *balanced accuracy* y F1 macro. Esto indica que el preentrenamiento especializado en plantas aporta una ventaja relevante para la clasificación botánica en imágenes artísticas.

La comparación entre *linear probing* y *full fine-tuning* mostró que adaptar el modelo completo mejora claramente el rendimiento. En PlantCLEF-DINOv2, el *full fine-tuning* aumentó la *balanced accuracy* en 0,1300 y el F1 macro en 0,1101 respecto al entrenamiento únicamente de la cabeza final.

El análisis por clase mostró que el modelo reconoce mejor géneros visualmente distintivos, como *rhododendron*, *narcissus*, *tulipa* e *iris*. En cambio, presenta más dificultades en géneros como *rosa*, *anemone* y *dianthus*, probablemente por su mayor variabilidad pictórica y similitud con otros géneros ornamentales. Además, el rendimiento fue sólido tanto en ilustraciones botánicas como en pintura barroca y bodegones, donde la flor suele aparecer con mayor protagonismo visual.

## 5. Conclusiones

Los resultados muestran que la identificación automática de motivos florales en imágenes artísticas es viable mediante modelos preentrenados y técnicas de *fine-tuning*. PlantCLEF-DINOv2 obtuvo el mejor rendimiento global, especialmente en métricas equilibradas, lo que sugiere que el preentrenamiento botánico facilita la adaptación al dominio pictórico.

La principal aportación del trabajo es doble. Por un lado, se ha construido un *dataset* específico que combina imágenes artísticas, recortes florales, etiquetas botánicas normalizadas y metadatos de procedencia. Por otro lado, se ha desarrollado un pipeline completo para entrenar, evaluar y analizar modelos de clasificación botánica en pintura.

El sistema no sustituye al análisis experto, pero puede servir como herramienta de apoyo para explorar colecciones digitales, organizar imágenes con contenido vegetal y generar hipótesis preliminares sobre la presencia de determinados géneros en obras de arte.

Como trabajo futuro, sería conveniente ampliar el *dataset*, reforzar las clases más problemáticas e incorporar una etapa previa de detección automática de flores dentro de la obra completa.

# **AUTOMATIC IDENTIFICATION OF PLANT SPECIES IN PAINTING: DATASET CREATION AND EXPLORATORY MODELLING IN COMPUTER VISION**

**Author: García Birichinaga, Claudia.**

Supervisor: Carrero Muñoz, Dido

Collaborating Entity: ICAI – Universidad Pontificia Comillas

## **ABSTRACT**

This Bachelor's Thesis addresses the automatic identification of floral motifs represented in paintings and historical botanical illustrations using computer vision models. To this end, a dedicated artistic-botanical image dataset was created, with normalized botanical labels and associated metadata. Several pretrained image classification architectures were compared through transfer learning and fine-tuning. The best performance was obtained by PlantCLEF-DINOv2, which achieved a Top-1 accuracy of 0.754, a balanced accuracy of 0.704 and a macro F1-score of 0.644 on the test set.

**Keywords:** computer vision, botanical classification, transfer learning, fine-tuning, CNN, Vision Transformers, DINOv2, artistic heritage.

## **1. Introduction**

The representation of flowers and plants has played an important role throughout the history of art, both for its ornamental value and for its symbolic, iconographic and compositional functions. Identifying these botanical motifs can provide useful information for interpreting artworks, studying their cultural context and analyzing the presence of specific botanical genera across different periods, schools and artistic styles. Traditionally, however, this type of analysis has been carried out manually, relying on the expertise of art historians, botanists or specialists in vegetal iconography.

From a computer vision perspective, this task presents a specific challenge. Current botanical classification models are usually trained on naturalistic photographs, where plants are shown directly and their morphological traits are clearly visible. In contrast, flowers in paintings may appear stylized, partially occluded, idealized or embedded within complex compositions. As a result, there is a domain shift between botanical photography and artistic representation.

This project explores whether pretrained models can be adapted to this context through transfer learning and fine-tuning. The aim is not to replace expert judgement, but to assess the potential of artificial intelligence as a support tool for analyzing digital collections that contain botanical visual material.

## **2. Objectives**

The main objective of this work is to evaluate the feasibility of applying pretrained computer vision models to the automatic identification of floral motifs in paintings and historical botanical illustrations. In the main experiments, the classification task is defined at genus

level, since this taxonomic level provides a suitable balance between visual reliability, taxonomic precision and the availability of examples.

Three specific objectives were defined. The first was to build a dedicated dataset of floral crops obtained from sources such as the Museo Nacional del Prado, the Rijksmuseum, The Metropolitan Museum of Art and Curtis’s Botanical Magazine. The second was to define a normalized botanical vocabulary, organized into family, genus and species levels. The third was to compare different image classification architectures and analyze their behavior using global metrics, class-balanced metrics and analyses by botanical class, image source and artistic context.

### 3. Description of the proposed model/system/tool

The proposed solution is organized as a complete pipeline for data construction, training, evaluation and result analysis, as shown in Figure 1. The basic unit of the dataset is a manually cropped flower or identifiable floral motif within an artwork. This decision makes it possible to associate each image with a single botanical label and avoids the ambiguity that would arise when a complete artwork contains several plant species.

The final dataset used in the experiments contains 619 images, organized in a hierarchical structure of family, genus and species. The main classification task was performed at genus level, with 31 botanical genera considered. Each image was associated with a metadata file containing its filename, species, genus, family, source, original identifier and train, validation or test split.

To reduce the risk of data leakage, the dataset split was designed by grouping related crops according to artwork, source and botanical label. During training, moderate data augmentation was applied, including random crops, horizontal flipping, small rotations and controlled colour variations. In validation and test, deterministic transformations were used to ensure stable and comparable evaluation.

The evaluated models were ResNet-50, EfficientNet-B3, ConvNeXt-Small, ViT-Base and PlantCLEF-DINOv2. In all cases, the original classification head was replaced by a new layer adapted to the number of botanical genera in the dataset. Two training strategies were also compared: linear probing, where the backbone remains frozen and only the final classifier is trained, and full fine-tuning, where all model parameters are updated.

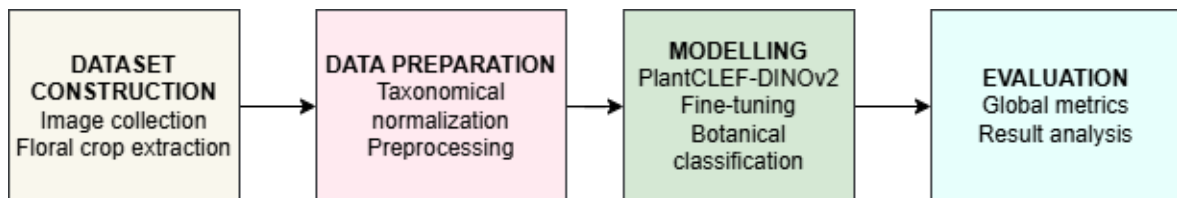


Figure 1. General workflow of the proposed system: dataset construction, data preparation, model fine-tuning and evaluation.

### 4. Results

The evaluation was carried out on the test set using multiclass classification metrics: Top-1, Top-3 and Top-5 accuracy, balanced accuracy, macro F1-score, weighted F1-score, Cohen’s kappa and Matthews correlation coefficient. Balanced metrics are particularly relevant in this project because the dataset presents class imbalance.

The best-performing model was PlantCLEF-DINOv2 with full fine-tuning, which achieved a Top-1 accuracy of 0.754, a balanced accuracy of 0.704, a macro F1-score of 0.644 and an MCC of 0.744. ConvNeXt-Small was the strongest general-purpose baseline, with a Top-1 accuracy of 0.719 and the best Top-3 and Top-5 results, although it remained below PlantCLEF-DINOv2 in balanced accuracy and macro F1-score. These results suggest that botanical pretraining provides a relevant advantage for botanical classification in artistic images.

The comparison between linear probing and full fine-tuning showed that adapting the entire model clearly improves performance. For PlantCLEF-DINOv2, full fine-tuning increased balanced accuracy by 0.1300 and macro F1-score by 0.1101 compared with training only the final classification head.

The class-level analysis showed that the model recognizes visually distinctive genera more effectively, such as *rhododendron*, *narcissus*, *tulipa* and *iris*. In contrast, it presents greater difficulty with genera such as *rosa*, *anemone* and *dianthus*, probably due to their higher pictorial variability and visual similarity to other ornamental genera. The analysis by source and artistic style also showed that the model performs well both in historical botanical illustration and in Baroque painting, especially when the flower has a clear visual presence within the composition.

## 5. Conclusions

The results indicate that automatic identification of floral motifs in artistic images is feasible, in an exploratory setting, using pretrained models and fine-tuning techniques. PlantCLEF-DINOv2 achieved the best overall performance, especially in class-balanced metrics, which suggests that botanical pretraining facilitates adaptation to the pictorial domain.

The main contribution of this work is twofold. First, a dedicated dataset was created, combining artistic images, floral crops, normalized botanical labels and source metadata. Second, a complete pipeline was developed to train, evaluate and analyse botanical classification models in paintings and historical botanical illustrations.

The system is not intended to replace expert analysis. However, it may serve as a support tool for exploring digital collections, organizing images with botanical content and generating preliminary hypotheses about the presence of specific plant genera in artworks. Future work should focus on expanding the dataset, strengthening the most problematic classes and incorporating a prior detection stage to automatically locate flowers within complete artworks.

## Índice de la memoria

<b>Capítulo 1. Introducción</b>	<b>1</b>
1.1 Contexto y Motivación	1
1.2 Objetivos	3
1.3 Alineación con los ODS	5
1.4 Estructura del Trabajo	5
<b>Capítulo 2. Estado del Arte</b>	<b>6</b>
2.1 Conceptos fundamentales	6
2.2 Técnicas y modelos relevantes	7
2.3 Trabajos relacionados	8
2.4 Limitaciones de las soluciones actuales	9
<b>Capítulo 3. Sistema/Modelo Desarrollado</b>	<b>11</b>
3.1 Planteamiento del problema	11
3.1.1 Definición formal del problema	11
3.1.2 Pruebas y métricas de evaluación para la validación del sistema	12
3.2 Construcción del conjunto de datos	13
3.2.1 Fuentes de imágenes	15
3.2.2 Selección y etiquetado de imágenes	18
3.2.3 Vocabulario botánico y normalización taxonómica	20
3.2.4 Organización del dataset	22
3.2.5 Descripción final del dataset	24
3.3 Diseño de la solución	25
3.3.1 Enfoque general	25
3.3.2 Arquitectura del pipeline	26
3.3.3 Modelos evaluados	29
3.3.4 Estrategias de entrenamiento	31
3.4 Implementación	36
3.4.1 Tecnologías y recursos HW/SW empleados	36
3.4.2 Estructura del código	37
3.4.3 Pipeline de preprocesado	38
3.4.4 Pipeline de entrenamiento	40

---

3.4.5 Pipeline de evaluación .....	42
3.4.6 Reproducibilidad .....	43
<b>Capítulo 4. Resultados.....</b>	<b>47</b>
4.1 Configuración experimental.....	47
4.2 Comparación global de modelos .....	49
4.3 Comparación entre linear probing y full fine-tuning.....	52
4.4 Análisis por clase botánica.....	53
4.5 Análisis por fuente y estilo artístico .....	55
4.6 Análisis de errores .....	58
<b>Capítulo 5. Conclusiones y Trabajos Futuros.....</b>	<b>60</b>
<b>Capítulo 6. Bibliografía.....</b>	<b>64</b>
<b>Anexo I. Vocabulario botánico y normalización taxonómica .....</b>	<b>67</b>
<b>Anexo II. Arquitectura de los modelos evaluados .....</b>	<b>81</b>
<b>Anexo III. Análisis de errores del modelo.....</b>	<b>89</b>

## *Índice de figuras*

Figura 1. Flujo general del sistema desarrollado: construcción del dataset, preparación de datos, ajuste fino del modelo y evaluación de resultados. ....	5
Figura 2. Ejemplo de generación de muestras a partir de una obra con varios motivos florales. A partir de la imagen original se extraen recortes independientes asociados a etiquetas botánicas distintas. Fuente de la obra original: Museo del Prado, elaboración propia.....	19
Figura 3. Ejemplo de data augmentation en entrenamiento para una imagen.....	27
Figura 4. Esquema de la arquitectura general del pipeline .....	29
Figura 5. Esquema de la estrategia de linear probing .....	32
Figura 6. Esquema de full fine-tuning .....	33
Figura 7. Distribución de imágenes por género botánico antes de aplicar el criterio mínimo de representación .....	35
Figura 8. Distribución de imágenes por género botánico tras aplicar el criterio mínimo de representación.....	35
Figura 9. Esquema de la estructura del código desarrollado.....	38
Figura 10. Curva ROC-AUC para los modelos evaluados .....	51
Figura 11 Arquitectura general de ResNet-50. Fuente: adaptado de Towards Data Science [23].....	81
Figura 12. Esquema general de la arquitectura ConvNeXt, organizada en etapas jerárquicas que reducen progresivamente la resolución espacial y aumentan el número de canales. Fuente: adaptado de Medium [24].....	84
Figura 13 Esquema general de la arquitectura Vision Transformer, basada en la división de la imagen en parches y el uso de mecanismos de atención para modelar relaciones globales entre regiones de la imagen. Fuente: adaptado de GeeksforGeeks [25] .....	86

## *Índice de tablas*

Tabla 1. Ejemplo de la normalización taxonómica.....	22
Tabla 2. Campos del archivo de metadatos .....	23
Tabla 3. Distribución de las imágenes en el dataset desarrollado .....	25
Tabla 4. Arquitectura de cada modelo evaluado .....	30
Tabla 5. Configuración utilizada en los modelos evaluados .....	48
Tabla 6. Configuración utilizada por modelo evaluado .....	49
Tabla 7. Resultados en el conjunto de test para los diferentes modelos entrenados .....	50
Tabla 8. Comparación de resultados entre full fine-tuning y linear probing para el modelo PlantCLEF-DINOv2 .....	53
Tabla 9. Resultados por género botánico para las clases con mayor soporte.....	55
Tabla 10. Resultados del modelo por fuente .....	56
Tabla 11. Resultados del modelo por estilo artístico .....	58
Tabla 12. Vocabulario botánico utilizado para la organización taxonómica del conjunto de datos .....	80
Tabla 13. Número de errores Top-1 por clase real en el conjunto de test.....	90
Tabla 14. Clases predichas incorrectamente con mayor frecuencia .....	91

## Capítulo 1. INTRODUCCIÓN

La identificación de especies vegetales representadas en obras pictóricas constituye una tarea de interés tanto para la Historia del Arte como para la Botánica, ya que permite profundizar en el análisis iconográfico, simbólico y contextual de las obras. Sin embargo, este proceso ha dependido tradicionalmente del análisis manual realizado por especialistas, lo que limita su aplicación a gran escala y dificulta el estudio sistemático de grandes colecciones de imágenes artísticas. Desde el punto de vista de la visión por ordenador el reconocimiento de motivos vegetales en pintura plantea un reto técnico relevante debido al cambio de dominio entre imágenes fotográficas naturalistas y representaciones artísticas, caracterizadas por diferentes grados de estilización. La adaptación de modelos preentrenados al dominio artístico permite evaluar la viabilidad de aplicar técnicas de *transfer learning* y *fine-tuning* al análisis botánico en obras de arte. Este enfoque no pretende sustituir el juicio experto, sino proporcionar una herramienta de apoyo que facilite el análisis exploratorio y la organización de grandes volúmenes de datos visuales, contribuyendo al desarrollo de nuevas metodologías en el ámbito de las humanidades digitales

### 1.1 CONTEXTO Y MOTIVACIÓN

La representación de la naturaleza y, en particular, de elementos florales y vegetales, ha ocupado un lugar relevante en la historia del arte occidental. Las plantas no aparecen únicamente como elementos decorativos, sino que pueden desempeñar funciones simbólicas, iconográficas, narrativas y compositivas dentro de la obra. Su identificación permite profundizar en la interpretación de las escenas representadas, en la relación entre arte y conocimiento científico, y en el contexto cultural en el que fueron producidas las obras. Desde esta perspectiva, el estudio botánico aplicado al arte constituye una línea de análisis interdisciplinar de interés tanto para la Historia del Arte como para la Botánica, ya que permite reconocer especies, estudiar su presencia en distintos periodos y comprender los significados asociados a su representación.

En la actualidad, el interés por el estudio botánico en el arte se refleja en publicaciones especializadas, exposiciones, itinerarios y conferencias organizadas por instituciones museísticas. En este contexto, destacan las publicaciones de Eduardo Barba Gómez, como *El jardín del Prado: un paseo botánico por las obras de los grandes maestros* [1] y *El paraíso a pinceladas: jardines en las obras de arte* [2], así como iniciativas del Museo Nacional del Prado como *Un paseo botánico por el Prado* y *La botánica en el arte. Las plantas en las colecciones del Museo del Prado* [3]. Estos trabajos muestran que los motivos vegetales pueden aportar información relevante para la lectura de las obras, más allá de su valor puramente ornamental.

No obstante, el análisis sistemático de estos elementos continúa siendo, en gran medida, una tarea manual que requiere conocimiento experto y una revisión individual de las obras. Esta limitación dificulta el estudio a gran escala de grandes colecciones digitales, especialmente cuando las plantas aparecen integradas en composiciones complejas, parcialmente ocultas o representadas con distintos grados de estilización. Por tanto, las herramientas de visión por ordenador pueden ofrecer nuevas posibilidades para apoyar la exploración, organización y análisis de imágenes artísticas con contenido vegetal.

La aplicación de técnicas de inteligencia artificial al análisis de imágenes ha experimentado un avance significativo en los últimos años, especialmente en tareas de clasificación, detección y reconocimiento visual. En el ámbito de la botánica, los modelos de aprendizaje profundo han demostrado una elevada capacidad para identificar especies vegetales a partir de fotografías naturalistas, donde las plantas suelen aparecer representadas de forma directa y con rasgos morfológicos claramente visibles.

Sin embargo, la aplicación de estos modelos al dominio artístico constituye todavía un campo incipiente y presenta importantes lagunas en cuanto a *datasets* especializados, protocolos de anotación y estrategias de adaptación de dominio para entornos artísticos. Además, las obras pictóricas presentan una serie de limitaciones intrínsecas para la identificación automática, ya que no reproducen necesariamente la realidad de forma objetiva, sino que están condicionadas por el estilo del autor, la técnica utilizada, el periodo

histórico y la función simbólica o decorativa de los motivos representados. En consecuencia, una flor puede aparecer idealizada, simplificada, parcialmente oculta o integrada dentro de una composición compleja, lo que dificulta su identificación automática.

Esta diferencia entre imágenes fotográficas y representaciones artísticas constituye un problema de cambio de dominio. Un modelo entrenado sobre fotografías de plantas puede aprender patrones visuales muy útiles en contextos naturalistas, pero estos patrones no son necesariamente generalizables a obras artísticas. Por ello, resulta necesario estudiar hasta qué punto los modelos preentrenados pueden adaptarse a este nuevo contexto mediante técnicas de *transfer learning* y *fine-tuning*.

La motivación principal de este trabajo surge, por tanto, de la posibilidad de combinar herramientas de visión por ordenador con el estudio del patrimonio artístico. La identificación automática de especies vegetales en pintura podría facilitar el análisis de grandes colecciones digitales, apoyar tareas de catalogación y abrir nuevas vías de investigación en humanidades digitales. Además, el estudio de los errores del modelo puede aportar información relevante sobre las limitaciones de la clasificación automática cuando las especies vegetales aparecen sometidas a distintos grados de estilización artística.

En este sentido, el proyecto se plantea como un análisis exploratorio de la relación entre representación artística, morfología vegetal y aprendizaje automático. A partir de un conjunto de datos específico, formado por imágenes procedentes de distintas fuentes, se comparan varias arquitecturas de visión por ordenador y se estudia su rendimiento en función de las clases botánicas, las fuentes visuales y las características estilísticas de las obras.

## **1.2 OBJETIVOS**

El objetivo principal de este Trabajo Fin de Grado es evaluar la viabilidad de aplicar modelos de visión por ordenador preentrenados al reconocimiento automático de motivos vegetales representados en obras pictóricas. En los experimentos principales, la clasificación se plantea a nivel de género botánico, por ser el nivel taxonómico que ofrece un mejor equilibrio entre

fiabilidad visual, precisión taxonómica y disponibilidad de ejemplos. El trabajo busca estudiar la capacidad de estos modelos para adaptarse al dominio artístico mediante técnicas de transferencia de aprendizaje y ajuste fino, así como analizar su comportamiento ante diferentes clases botánicas, fuentes visuales y estilos de representación.

Para alcanzar este objetivo general, se plantean los siguientes objetivos específicos:

- Diseñar y construir un conjunto de datos de imágenes artísticas con motivos florales, procedentes de repositorios digitales de museos, colecciones abiertas e ilustraciones botánicas históricas.
- Definir un vocabulario botánico adecuado para la tarea de clasificación, estableciendo el nivel taxonómico más apropiado en función de la disponibilidad de datos y de la fiabilidad de las anotaciones.
- Documentar el proceso de recopilación, selección, normalización y etiquetado de las imágenes, incluyendo la organización del *dataset* y la generación de metadatos asociados.
- Aplicar técnicas de *transfer learning* y *fine-tuning* sobre modelos de visión por ordenador preentrenados, adaptando sus capas finales al número de clases del conjunto de datos.
- Comparar el rendimiento de diferentes arquitecturas de clasificación de imágenes, incluyendo modelos convolucionales, modelos basados en *transformers* visuales y modelos con preentrenamiento específico en el dominio botánico.
- Evaluar los modelos mediante métricas adecuadas para problemas multiclase y conjuntos de datos desbalanceados, como *accuracy*, *balanced accuracy*, *precision*, *recall* y *F1-score*.
- Analizar los resultados obtenidos por clase botánica, identificando qué flores o géneros presentan un mejor rendimiento y cuáles generan mayores confusiones.

- Estudiar la influencia de las características visuales y estilísticas de las obras en el rendimiento del modelo, diferenciando entre ilustraciones botánicas, pinturas históricas y representaciones ornamentales o estilizadas.
- Identificar las principales limitaciones del enfoque propuesto y plantear posibles líneas de mejora.

### ***1.3 ALINEACIÓN CON LOS ODS***

El proyecto se alinea principalmente con el ODS 4, Educación de calidad, al aplicar herramientas de inteligencia artificial al análisis interdisciplinar del patrimonio cultural, y con el ODS 9, Industria, innovación e infraestructura, al explorar el uso de modelos de visión por ordenador para desarrollar metodologías innovadoras de análisis de imágenes artísticas.

### ***1.4 ESTRUCTURA DEL TRABAJO***

El presente trabajo se organiza en cinco capítulos. El Capítulo 1 introduce la motivación del proyecto, sus objetivos y la relación con los Objetivos de Desarrollo Sostenible. El Capítulo 2 presenta el estado del arte relacionado con la visión por ordenador aplicada al arte, la identificación automática de plantas y las limitaciones de las soluciones existentes. El Capítulo 3 describe la construcción del conjunto de datos, el diseño del sistema, los modelos evaluados y la implementación desarrollada. El Capítulo 4 recoge los resultados experimentales y su análisis por modelo, clase botánica, fuente y estilo artístico. Finalmente, el Capítulo 5 presenta las conclusiones, limitaciones y posibles líneas de trabajo futuro.

## **Capítulo 2. ESTADO DEL ARTE**

En este capítulo se revisan los principales conceptos, técnicas y trabajos relacionados con la identificación automática de flores en obras pictóricas. El objetivo es situar el proyecto dentro del contexto actual de la visión por ordenador, la clasificación botánica y el análisis computacional del arte, identificando las aportaciones previas y las limitaciones que justifican el desarrollo de este trabajo.

### ***2.1 CONCEPTOS FUNDAMENTALES***

La identificación de especies vegetales representadas en obras de arte se sitúa en la intersección entre la Historia del Arte, la Botánica y la visión por ordenador. Desde el punto de vista histórico-artístico, los motivos florales pueden desempeñar funciones simbólicas, iconográficas, narrativas o compositivas, además de funciones decorativas.

El análisis botánico aplicado al arte ha adquirido un interés creciente en los últimos años, como muestran distintas publicaciones, exposiciones e itinerarios museísticos centrados en la identificación de plantas en obras artísticas [1]-[4]. Sin embargo, este análisis continúa realizándose principalmente de forma manual, mediante la observación experta de rasgos morfológicos como la forma de los pétalos, la disposición de las hojas, el color o la estructura floral.

Desde el punto de vista computacional, la tarea presenta un problema de cambio de dominio: los modelos entrenados sobre fotografías naturalistas no tienen por qué generalizar correctamente a pinturas, dibujos o ilustraciones históricas, donde las flores pueden aparecer idealizadas, simplificadas, parcialmente ocultas o integradas en composiciones complejas.

## 2.2 TÉCNICAS Y MODELOS RELEVANTES

La clasificación automática de imágenes ha avanzado de forma significativa gracias al desarrollo del aprendizaje profundo. En particular, las redes neuronales convolucionales han sido ampliamente utilizadas en tareas de reconocimiento visual, ya que permiten extraer características jerárquicas de las imágenes, desde patrones simples como bordes y texturas hasta estructuras visuales más complejas. Arquitecturas como ResNet [5], EfficientNet [6] o ConvNeXt [7] constituyen referencias habituales para comparar modelos de clasificación de imágenes.

Junto a las arquitecturas convolucionales, en los últimos años han adquirido especial relevancia los modelos basados en transformadores visuales, como los *Vision Transformers* [8]. Estos modelos dividen la imagen en parches y utilizan mecanismos de atención para aprender relaciones globales entre distintas regiones de la imagen. Este enfoque ha sido especialmente importante en el desarrollo de modelos fundacionales de visión, como DINOv2 [9], que también sirve de base para variantes especializadas en botánica, como PlantCLEF-DINOv2 [10].

Cuando el conjunto de datos disponible es limitado, una estrategia habitual consiste en utilizar modelos preentrenados y adaptarlos al problema concreto mediante *transfer learning* [11]. Dentro de esta estrategia existen distintos grados de adaptación. En el *linear probing*, el extractor de características permanece congelado y solo se entrena una nueva cabeza de clasificación. En el *fine-tuning*, en cambio, se actualizan también los pesos internos del modelo, lo que permite una adaptación más profunda al dominio objetivo. Estas técnicas son relevantes para este trabajo porque permiten comparar modelos generalistas con modelos especializados en botánica bajo un escenario de cambio de dominio.

## **2.3 TRABAJOS RELACIONADOS**

Los trabajos relacionados con este proyecto pueden agruparse en tres líneas principales: el análisis computacional de obras de arte, la identificación automática de plantas y los estudios botánicos tradicionales aplicados al arte.

En el ámbito del análisis computacional del arte, Crowley y Zisserman [12] estudian la transferencia de modelos entrenados sobre imágenes fotográficas al dominio artístico para la detección de objetos en pinturas. En una línea similar, Madhu et al. [13] abordan el reconocimiento de personajes en obras artísticas. Aunque ninguno de estos trabajos se centra en la clasificación botánica, ambos muestran que los modelos de visión por ordenador pueden extraer representaciones útiles en imágenes pictóricas, incluso cuando los elementos aparecen estilizados o reinterpretados según convenciones artísticas.

En la identificación automática de plantas, iniciativas como Pl@ntNet [14] han demostrado el potencial de los sistemas de visión por ordenador para reconocer especies vegetales a partir de grandes conjuntos de fotografías anotadas. Sin embargo, estos enfoques se orientan principalmente a imágenes naturalistas, por lo que su aplicación directa a obras pictóricas introduce un problema de cambio de dominio.

Finalmente, existen trabajos de carácter humanístico y botánico centrados en el estudio de plantas representadas en obras de arte. En este sentido, destacan las publicaciones de Eduardo Barba Gómez, como *El jardín del Prado* [1] y *El paraíso a pinceladas* [2], así como iniciativas expositivas como *Un paseo botánico por el Prado* [3] y *La botánica en el arte. Las plantas en las colecciones del Museo del Prado* [4]. Estos trabajos evidencian el valor histórico-artístico de la identificación botánica en pintura y muestran cómo las plantas pueden contribuir a la interpretación simbólica, cultural y material de las obras.

## ***2.4 LIMITACIONES DE LAS SOLUCIONES ACTUALES***

Las soluciones existentes presentan varias limitaciones que justifican el desarrollo de este proyecto. En primer lugar, los modelos de identificación automática de plantas se han entrenado principalmente sobre imágenes naturalistas. Aunque estos modelos pueden reconocer rasgos morfológicos de plantas o flores, su rendimiento puede verse reducido cuando se aplican a imágenes artísticas, donde la representación de la planta no siempre conserva una correspondencia directa con la realidad.

En segundo lugar, los trabajos de visión por ordenador aplicados al arte suelen centrarse en problemas diferentes, como la detección de objetos generales o el reconocimiento de personajes en obras pictóricas [12],[13]. Estas tareas demuestran que la visión por ordenador puede aplicarse al patrimonio artístico, pero no abordan específicamente la identificación botánica con detalle taxonómico.

En tercer lugar, existe una carencia de conjuntos de datos especializados que combinen imágenes artísticas con anotaciones botánicas normalizadas. Para entrenar y evaluar modelos en este dominio es necesario recopilar imágenes de distintas fuentes, revisar la calidad visual de los motivos vegetales, normalizar los nombres científicos y decidir el nivel taxonómico adecuado. Esta fase es especialmente importante porque la identificación a nivel de especie puede ser poco fiable en algunas obras pictóricas, mientras que el género o la familia pueden ofrecer niveles de clasificación más robustos.

Además, las obras artísticas introducen dificultades propias: variabilidad estilística, diferentes técnicas pictóricas, composiciones complejas, motivos parcialmente ocultos, idealización de formas y posible ambigüedad simbólica. Estos factores hacen que el problema sea distinto de la clasificación botánica en fotografía y obligan a analizar los errores del modelo por clase, fuente visual y características artísticas, además de su rendimiento global.

El estado del arte muestra que existen avances relevantes en análisis computacional del arte y en clasificación automática de plantas, pero todavía falta una aproximación específica que combine ambos ámbitos para abordar la identificación automática de motivos botánicos en pintura. Esta laguna metodológica y experimental constituye el punto de partida del presente proyecto.

## Capítulo 3. SISTEMA/MODELO DESARROLLADO

### 3.1 PLANTEAMIENTO DEL PROBLEMA

#### 3.1.1 DEFINICIÓN FORMAL DEL PROBLEMA

El problema abordado en este trabajo consiste en una tarea de clasificación multiclase de imágenes. Dado un recorte manual de una obra pictórica o ilustración botánica histórica en el que aparece un motivo floral identificable, el sistema debe asignar una etiqueta botánica perteneciente al vocabulario definido para el proyecto.

Formalmente, sea  $X = \{x_1, x_2, \dots, x_n\}$  el conjunto de imágenes del *dataset*, donde cada imagen  $x_i$  corresponde a un recorte de una obra pictórica o ilustrada de una flor. A cada imagen se le asocia una etiqueta  $y_i$ , perteneciente a un conjunto de clases botánicas  $Y = \{c_1, c_2, \dots, c_k\}$ . Cada clase puede corresponder, según el nivel taxonómico elegido, a una familia, un género o una especie vegetal.

El objetivo consiste en aprender una función de clasificación:

$$f: X \rightarrow Y$$

A diferencia de los problemas clásicos de identificación de plantas en imágenes naturalistas, este trabajo se centra en imágenes procedentes de obras pictóricas, ilustraciones botánicas históricas y representaciones artísticas. Esto introduce una dificultad adicional, ya que las especies vegetales pueden aparecer estilizadas, parcialmente ocultas o representadas con distintos grados de realismo.

Además, el conjunto de datos presenta una distribución desigual entre clases, ya que algunas especies o géneros aparecen representados con mayor frecuencia que otros. Esta circunstancia obliga a tratar el problema como una clasificación multiclase desbalanceada, evaluando el rendimiento del sistema mediante métricas que tengan en cuenta tanto el acierto global como el comportamiento individual en cada clase.

En consecuencia, el sistema desarrollado debe resolver tres aspectos principales: identificar patrones visuales asociados a distintas especies o géneros vegetales, adaptarse al cambio de dominio entre imagen naturalista y representación artística, y mantener un rendimiento razonable en presencia de clases con distinto número de ejemplos.

### **3.1.2 PRUEBAS Y MÉTRICAS DE EVALUACIÓN PARA LA VALIDACIÓN DEL SISTEMA**

Para evaluar el rendimiento del sistema desarrollado, se emplean diferentes métricas de clasificación multiclase.

La primera métrica que se tiene en cuenta es la *accuracy* top-1 que mide el porcentaje de imágenes cuya etiqueta predicha coincide con la etiqueta real. De esta forma se puede tener una visión general del rendimiento del modelo.

También se utilizan la *accuracy* top-3 y top-5 que indican si la clase correcta para una imagen aparece entre las tres y cinco predicciones más probables, respectivamente.

Dado que la representación de las diferentes especies botánicas en el arte no es uniforme, es decir hay un desbalance de clases, se emplea también la *balanced accuracy*. Esta métrica calcula el promedio de la *accuracy* por clase, evitando que las clases mayoritarias dominen la evaluación global del modelo.

Además, se calculan *precision*, *recall* y *F1-score* en sus variantes *macro* y *weighted*. La variante *macro* otorga el mismo peso a todas las clases, independientemente de su número de ejemplos, mientras que la variante *weighted* pondera cada clase según el tamaño de la muestra.

Por último, se realiza un análisis cualitativo de errores para identificar los géneros botánicos más problemáticos y las confusiones más frecuentes entre clases.

## **3.2 CONSTRUCCIÓN DEL CONJUNTO DE DATOS**

La construcción del conjunto de datos es una fase de gran importancia en el proyecto, ya que condiciona la calidad del entrenamiento, y por tanto la fiabilidad de las predicciones y calidad del modelo. Esta fase adquiere especial relevancia en el presente trabajo, dado que no se dispone de una base de datos consolidada que combine información botánica con representaciones vegetales en obras pictóricas.

Antes de iniciar la recopilación de imágenes, fue necesario delimitar el tipo de datos que se iban a emplear. El conjunto de datos se definió a partir de imágenes artísticas en las que apareciesen flores identificables, con el objetivo de entrenar modelos capaces de reconocer motivos botánicos en contextos visuales alejados de la fotografía naturalista convencional.

En la propuesta inicial de construcción de la base de datos se planteó incluir obras comprendidas entre los siglos XIV y XIX, abarcando movimientos como el Renacimiento, Barroco, Rococó, Romanticismo, Impresionismo y Postimpresionismo. Sin embargo, la selección final estuvo condicionada por la disponibilidad de imágenes con representación floral claramente identificable en las fuentes utilizadas. Como resultado, el conjunto de datos quedó compuesto principalmente por obras realizadas entre los siglos XVI y XIX, procedentes de las colecciones del Museo del Prado, el Rijksmuseum y The Metropolitan Museum of Art, complementadas con ilustraciones botánicas históricas de los siglos XVIII y XIX procedentes de *Curtis's Botanical Magazine*.

Las obras seleccionadas presentan una amplia variedad de estilos artísticos. Aunque existe una presencia destacada de pinturas barrocas y bodegones florales, el conjunto incluye también ejemplos vinculados al Renacimiento, Manierismo, Barroco con influencia flamenca, Rococó, transición entre Rococó y Neoclasicismo, Realismo académico, Realismo académico con influencia romántica, Realismo, Impresionismo, Postimpresionismo y Movimiento Estético. Esta diversidad permite incorporar representaciones vegetales producidas bajo convenciones visuales diferentes, desde

composiciones florales de carácter naturalista hasta imágenes más decorativas, simbólicas o estilizadas.

En cuanto a las escuelas artísticas, el conjunto incluye principalmente obras de la escuela española, junto con ejemplos de escuela flamenca, neerlandesa, italiana, francesa, americana y de La Haya. También aparecen algunas obras sin escuela documentada y casos de influencia mixta, como obras españolas con influencia flamenca. Esta diversidad geográfica y estilística contribuye a ampliar la variabilidad visual del conjunto de datos, evitando que el modelo aprenda únicamente las convenciones formales de una tradición artística concreta.

Desde el punto de vista del género artístico, las obras corresponden mayoritariamente a bodegones florales, pero también se incorporan retratos, guirnaldas, pinturas alegóricas, bodegones con figuras alegóricas, bodegones de banquete, estudios florales y composiciones con elementos narrativos. Esta diversidad es relevante para el proyecto porque los motivos vegetales no siempre aparecen con el mismo grado de protagonismo: en algunos casos constituyen el elemento central de la obra, mientras que en otros forman parte de una escena más compleja.

Asimismo, se decidió incluir obras realizadas mediante distintos soportes y técnicas. El conjunto está formado principalmente por pinturas al óleo sobre lienzo y óleo sobre tabla, aunque también incorpora ejemplos de óleo sobre papel, óleo sobre cobre, óleo sobre madera, gouache y acuarela sobre papel y dibujos realizados con tiza negra y lápiz sobre papel verjurado gris. Por tanto, aunque el óleo constituye la técnica dominante, la presencia de diferentes técnicas introduce variaciones en textura, color, línea y definición de los motivos vegetales, enriqueciendo la diversidad visual del conjunto.

La diversidad técnica y visual del conjunto se amplió mediante la incorporación de ilustraciones botánicas históricas procedentes de *Curtis's Botanical Magazine*. Estas imágenes, centradas habitualmente en una única especie y con mayor detalle morfológico, complementan las pinturas de museo al aportar representaciones más claras desde el punto de vista botánico.

A partir de esta delimitación, se diseñó un proceso propio de recopilación, selección, etiquetado y organización de imágenes para construir un *dataset* adecuado para el entrenamiento de modelos de visión por ordenador. Siempre que fue posible, cada imagen se acompañó de metadatos como autor, fecha, escuela, técnica artística, movimiento, género y clasificación botánica.

En los siguientes apartados se describen las fuentes utilizadas, los criterios de selección y etiquetado, la normalización taxonómica y la organización final del *dataset* empleado en los experimentos.

### **3.2.1 FUENTES DE IMÁGENES**

Las imágenes utilizadas en el conjunto de datos se obtuvieron a partir de distintas fuentes digitales, seleccionadas con el objetivo de obtener imágenes artísticas con presencia de motivos botánicos y, al mismo tiempo, garantizar una cierta diversidad visual, temporal y estilística. Dado que el objetivo del trabajo no se limita a reconocer plantas en imágenes naturalistas, sino a estudiar su identificación en obras pictóricas, se consideró necesario combinar colecciones museísticas con fuentes de ilustración botánica histórica.

Las fuentes empleadas fueron seleccionadas atendiendo a tres criterios principales. En primer lugar, la disponibilidad de imágenes digitales con calidad suficiente para su uso en tareas de visión por ordenador. En segundo lugar, la presencia explícita de flores identificables. Y en tercer lugar, la posibilidad de asociar cada imagen a una etiqueta botánica, ya fuera a partir del nombre de la obra o ilustración, o mediante los metadatos disponibles.

En el conjunto de datos final se incorporaron imágenes procedentes del Museo del Prado, Rijksmuseum, The Metropolitan Museum of Art y *Curtis's Botanical Magazine*.

### ***3.2.1.1 Museo del Prado***

El Museo del Prado constituye una de las fuentes principales del conjunto de datos, ya que aporta obras pictóricas históricas de alta calidad con presencia de motivos vegetales. La recopilación se realizó de forma manual a partir de la colección digital del museo, revisando las obras una a una y seleccionando aquellas en las que las flores eran suficientemente visibles para la tarea de clasificación.

Una de las ventajas de esta fuente es que gran parte de las obras cuentan con información descriptiva o estudios especializados sobre los elementos vegetales representados [1]-[4], lo que facilitó la identificación inicial de determinados motivos florales y permitió seleccionar ejemplos con mayor fiabilidad. Además, las obras del Prado presentan una notable variedad visual y compositiva: las flores aparecen en bodegones, escenas religiosas, retratos y composiciones ornamentales. Por ello, esta fuente resulta especialmente útil para evaluar la dificultad de identificar motivos vegetales integrados en escenas pictóricas complejas.

### ***3.2.1.2 Rijksmuseum***

La colección del Rijksmuseum se incorporó por la presencia de bodegones, composiciones florales y obras de tradición neerlandesa en las que las flores tienen un papel destacado. Aunque inicialmente se planteó una recopilación automatizada mediante API, las limitaciones técnicas encontradas durante el proyecto llevaron finalmente a una selección manual desde la colección digital del museo [20].

El etiquetado de estas imágenes requirió revisión individual, ya que los metadatos disponibles no siempre proporcionaban una identificación botánica directa. Por ello, las muestras seleccionadas se asociaron al vocabulario botánico del proyecto a partir de la información disponible, el análisis visual de la obra y la coherencia con el resto del conjunto de datos.

### ***3.2.1.3 Metropolitan Museum of Art***

The Metropolitan Museum of Art se utilizó como fuente museística internacional para ampliar la diversidad visual y geográfica del conjunto de datos. Esta colección ofrece imágenes en acceso abierto y metadatos asociados a las obras, como título, autor, fecha, técnica y enlace a la obra original. Además, su API pública facilitó la obtención inicial de imágenes y metadatos generales [19].

Sin embargo, la principal dificultad de esta fuente fue la ausencia frecuente de información botánica específica. En muchos casos, los metadatos no indicaban qué especies vegetales aparecían representadas, más allá de referencias generales en el título. Por este motivo, las imágenes procedentes del Met requirieron una revisión manual más cuidadosa y se descartaron aquellas en las que no era posible asignar una etiqueta botánica fiable. Como consecuencia, el número final de imágenes incorporadas desde esta fuente fue inferior al volumen inicialmente recopilado.

### ***3.2.1.4 Curtis's Botanical Magazine***

*Curtis's Botanical Magazine* se incorporó como fuente de ilustración botánica histórica, utilizando imágenes disponibles a través de repositorios digitales como Biodiversity Heritage Library y Wikimedia Commons [21],[22]. A diferencia de muchas pinturas museísticas, estas imágenes suelen presentar las plantas de forma más clara y descriptiva, con mayor atención a los rasgos morfológicos. Por ello, resultan útiles para reforzar el conjunto de datos con ejemplos visualmente más próximos al dominio botánico.

Su inclusión permitió aumentar el número de muestras disponibles para determinados géneros y, al mismo tiempo, introducir una tipología visual distinta a la pintura artística. Esta diferencia resulta relevante para analizar posteriormente si los modelos obtienen mejores resultados en ilustraciones botánicas, donde la flor suele aparecer con mayor protagonismo y precisión morfológica, que en pinturas de composición más compleja.

### 3.2.2 SELECCIÓN Y ETIQUETADO DE IMÁGENES

Una vez definidas las fuentes de datos, se llevó a cabo un proceso de selección orientado a construir un conjunto útil para una tarea de clasificación supervisada. El trabajo se centró en plantas con flores, ya que los motivos florales son frecuentes en las obras analizadas y presentan rasgos visuales relativamente discriminativos, como la forma de los pétalos, la disposición floral, el color, la simetría o la presencia de hojas asociadas. Aunque en algunos casos se utilizaron hojas o tallos como información complementaria, la unidad principal de análisis fue la flor o motivo floral reconocible, no la planta completa.

El criterio principal de inclusión fue que el motivo floral fuese suficientemente reconocible desde el punto de vista visual. Para ello, debía apreciarse al menos una parte relevante de la estructura floral, como la forma general de la corola, el contorno, la disposición aproximada de los pétalos, el color predominante o su relación con hojas y tallos cercanos. Por el contrario, se descartaron motivos excesivamente pequeños, borrosos, oscurecidos, parcialmente ocultos o tan estilizados que no conservasen rasgos morfológicos mínimos. Este criterio permitió evitar que el modelo aprendiese a partir de muestras visualmente pobres o ambiguas, aunque la etiqueta procediese de una fuente fiable.

La unidad básica del *dataset* es el recorte manual de una flor o motivo floral identificable dentro de una obra, no la obra completa. Esta decisión fue necesaria porque muchas pinturas contienen varias especies vegetales en una misma composición. Utilizar la obra completa como una única muestra habría introducido ambigüedad, ya que una misma imagen podría estar asociada a varias etiquetas botánicas. Por ello, cuando una pintura contenía más de una flor o planta con flores, se realizaron recortes individuales de cada motivo de interés, de modo que cada muestra quedase asociada a una única clase.

Este proceso permitió transformar composiciones complejas en muestras independientes, cada una asociada a una única clase. En la Figura 2 se muestra un ejemplo de este procedimiento a partir de la obra *Ramillete de flores con lazo azul* de Luis Paret y Alcázar. A partir de la imagen original se extraen distintos motivos florales, cada uno de los cuales

se convierte en una muestra independiente del conjunto de datos y se asocia a su correspondiente etiqueta botánica. En este caso, una misma obra permite generar recortes correspondientes a géneros o especies como *Narcissus*, *Anemone* y *Rosa centifolia*.

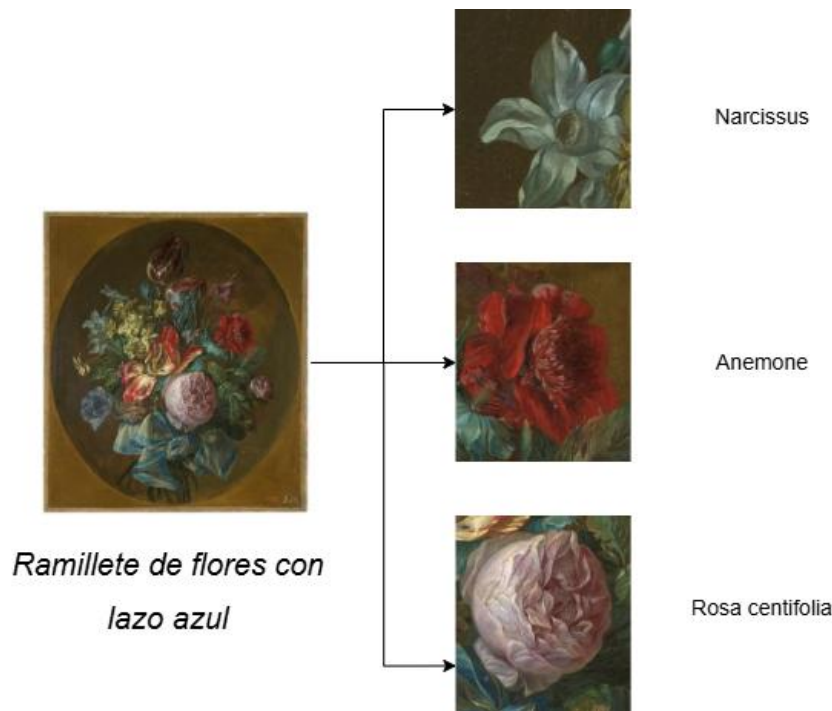


Figura 2. Ejemplo de generación de muestras a partir de una obra con varios motivos florales. A partir de la imagen original se extraen recortes independientes asociados a etiquetas botánicas distintas. Fuente de la obra original: Museo del Prado, elaboración propia.

El etiquetado de las imágenes combinó la información disponible en las fuentes originales con una revisión manual posterior. En algunos casos, como en determinadas obras del Museo del Prado, los metadatos, descripciones o estudios botánicos previos indicaban las especies vegetales presentes en la obra, lo que facilitó la identificación inicial. En otras fuentes, como el Rijksmuseum o The Metropolitan Museum of Art, la información botánica era más limitada y el etiquetado requirió una revisión más cuidadosa de cada imagen.

Cuando la identificación a nivel de especie no era suficientemente fiable, se optó por utilizar una etiqueta a nivel de género. Esta decisión responde a la naturaleza del problema, ya que en pintura no siempre se conservan todos los rasgos morfológicos necesarios para distinguir

especies próximas. Por tanto, el etiquetado buscó mantener un equilibrio entre precisión taxonómica y fiabilidad visual.

El resultado de este proceso fue un conjunto de imágenes recortadas y etiquetadas, en el que cada muestra corresponde a un motivo floral concreto y tiene asociada una clase botánica única. Esta organización permite formular el problema como una tarea de clasificación multiclase y facilita el análisis posterior del rendimiento del modelo por especie, género o familia.

### 3.2.3 VOCABULARIO BOTÁNICO Y NORMALIZACIÓN TAXONÓMICA

Una vez seleccionadas las imágenes, fue necesario definir un vocabulario botánico controlado que permitiera asociar cada muestra del conjunto de datos a una clase coherente. El nivel de detalle con el que se representan las plantas varía considerablemente entre obras. Algunas imágenes muestran rasgos morfológicos suficientes para una identificación específica, mientras que otras presentan una representación más estilizada o general, por lo que se optó por utilizar etiquetas a nivel de género o familia cuando era más apropiado.

El vocabulario utilizado en el proyecto se organizó siguiendo una estructura taxonómica jerárquica basada en tres niveles: familia, género y especie. Esta organización permite mantener una relación clara entre las distintas clases y facilita tanto el entrenamiento del modelo como el análisis posterior de los resultados. Por ejemplo, una imagen etiquetada como *rosa\_centifolia* queda asociada al género *rosa* y a la familia *Rosaceae*, mientras que una imagen etiquetada como *iris\_germanica* queda asociada al género *iris* y a la familia *Iridaceae*.

Para implementar esta organización se creó el módulo `taxonomy.py`, que recoge el vocabulario taxonómico utilizado en el proyecto. Este archivo define una correspondencia entre familias, géneros y especies, y permite obtener automáticamente la familia y el género asociados a cada etiqueta botánica normalizada. De este modo, la taxonomía no se utiliza

únicamente como referencia conceptual, sino también como parte activa del proceso de construcción del *dataset*.

La definición de este vocabulario permitió tratar de forma ordenada clases con distinto grado de especificidad. Cuando la representación visual de la planta era suficientemente clara y la información disponible permitía una identificación fiable, se utilizó la especie como etiqueta. En cambio, cuando existía ambigüedad o no se conservaban rasgos morfológicos suficientes para distinguir entre especies próximas, se empleó una etiqueta a nivel de género. Esta decisión permitió reducir el riesgo de introducir errores en el etiquetado y mantener la coherencia del conjunto de datos.

Además, durante la construcción del *dataset* fue necesario realizar un proceso de normalización de los nombres botánicos. Las denominaciones procedentes de distintas fuentes presentaban diferencias de formato o variaciones lingüísticas. Para evitar que una misma clase apareciera representada con nombres distintos, se unificaron las etiquetas mediante una nomenclatura común basada en minúsculas y guiones bajos. Así, nombres equivalentes o variantes ortográficas fueron transformados a una única forma normalizada.

Este proceso incluyó, por ejemplo, la corrección de errores en nombres científicos, la eliminación de numeraciones añadidas al final de los archivos, la sustitución de guiones por guiones bajos y la agrupación de variantes visuales bajo una misma clase cuando correspondían al mismo taxón.

La Tabla 1 muestra algunos ejemplos de la normalización taxonómica utilizada en el proceso de construcción del *dataset*. El vocabulario botánico completo, junto con los conceptos taxonómicos empleados y los criterios generales de normalización, se recoge en el Anexo I. Vocabulario botánico y normalización taxonómica.

<b>Clase normalizada</b>	<b>Género</b>	<b>Familia</b>
rosa_centifolia	Rosa	Rosaceae
iris_germanica	Iris	Iridaceae

paemonia_suffruticosa	Paeonia	Paeoniaceae
narcissus_jonquilla	Narcissus	Amaryllidaceae
tulipa	Tulipa	Liliaceae

Tabla 1. Ejemplo de la normalización taxonómica

### 3.2.4 ORGANIZACIÓN DEL DATASET

Tras seleccionar, recortar y etiquetar las imágenes, el *dataset* se organizó siguiendo una estructura jerárquica basada en la clasificación taxonómica explicada en el apartado de Vocabulario botánico y normalización taxonómica. De esta forma, se mantiene una correspondencia clara entre cada imagen, su etiqueta botánica asociada y los niveles superiores de la taxonomía.

La estructura final del *dataset* se definió mediante tres niveles principales: familia, género y especie, esto permite que cada imagen tenga asociada toda su información taxonómica correspondiente. La estructura es la siguiente:

```
dataset_final/
  Familia/
    genero/
      especie/
        imagen.jpg
```

Siguiendo el ejemplo de la Tabla 1, una imagen correspondiente a la especie *rosa\_centifolia* se almacenaría de la siguiente manera:

```
dataset_final/
  Rosaceae/
    rosa/
      rosa_centifolia/
        imagen.jpg
```

La construcción de esta estructura se automatizó mediante los scripts `build_dataset.py` y `build_dataset_fun.py`. Estos módulos leen las imágenes procedentes de las distintas

fuentes, extraen o normalizan la etiqueta botánica y consultan el vocabulario taxonómico para obtener la familia y el género correspondientes. A partir de esta información, el *dataset* final queda organizado siguiendo la ruta jerárquica `dataset_final/family/genus/species/filename`, manteniendo una estructura coherente con la taxonomía definida.

Además de la organización en carpetas, se generó un archivo de metadatos en formato CSV. Este archivo recoge la información principal de cada imagen incorporada al *dataset*, incluyendo el nombre final del archivo, la etiqueta botánica asignada, el género, la familia, la fuente de procedencia, el identificador original de la obra o imagen y la partición del conjunto de datos a la que pertenece.

Los campos del archivo se muestran en Tabla 2.

<b>Campo</b>	<b>Descripción</b>
filename	Nombre final de la imagen dentro del <i>dataset</i>
species	Clase botánica asignada a la imagen
genus	Género botánico correspondiente
family	Familia botánica correspondiente
source	Fuente de procedencia de la imagen
source_id	Identificador original de la obra o ilustración
original_path	Ruta original de la imagen
split	Partición asignada (train, eval, test)

*Tabla 2. Campos del archivo de metadatos*

Para reducir el riesgo de *data leakage* entre particiones, la división del conjunto de datos se realizó agrupando las imágenes por fuente, identificador original y etiqueta botánica. De este modo, los recortes correspondientes a una misma especie dentro de una misma obra o fuente se mantienen en la misma partición, evitando que variantes prácticamente equivalentes aparezcan simultáneamente en entrenamiento y evaluación.

La organización jerárquica del *dataset* permite conservar distintos niveles de información taxonómica. Cada imagen mantiene asociada su familia, género y especie, sin embargo, las evaluaciones principales se realizaron a nivel de género. Esto mantiene un equilibrio entre detalle botánico, fiabilidad del etiquetado y número suficiente de ejemplos por clase, especialmente teniendo en cuenta que en pintura no siempre se conservan los rasgos morfológicos necesarios para distinguir especies similares visualmente.

La conservación de los niveles de familia y especie aporta información adicional para el análisis posterior de resultados. Aunque la clasificación principal se realiza a nivel de género, la estructura taxonómica permite conocer la naturaleza de los errores cometidos por el modelo. Por ejemplo, una confusión entre clases pertenecientes a una misma familia puede interpretarse de forma distinta a una confusión entre familias botánicas alejadas.

### **3.2.5 DESCRIPCIÓN FINAL DEL DATASET**

Tras aplicar los criterios de selección, recorte, etiquetado y normalización taxonómica, se obtuvo el conjunto de datos final utilizado en los experimentos. Este conjunto está formado por imágenes de motivos florales procedentes de obras pictóricas e ilustraciones botánicas históricas, organizadas mediante una estructura jerárquica de familia, género y especie.

El conjunto final empleado en los experimentos principales contiene 619 imágenes, distribuidas en las particiones de entrenamiento, validación y prueba. La clasificación se realizó a nivel de género botánico, con un total de 31 géneros considerados en los experimentos. La distribución de imágenes entre las distintas particiones se muestra en la Tabla 3.

Partición	Número de imágenes
Train	385
Validación	120
Test	114
Total	619

Tabla 3. Distribución de las imágenes en el dataset desarrollado

### 3.3 DISEÑO DE LA SOLUCIÓN

#### 3.3.1 ENFOQUE GENERAL

La solución propuesta se basa en el uso de modelos de visión por ordenador preentrenados, adaptados al dominio artístico mediante técnicas de *transfer learning* y *fine tuning*. El objetivo es aprovechar representaciones visuales aprendidas previamente en grandes conjuntos de imágenes y especializarlas posteriormente en la tarea concreta de identificación de especies vegetales representadas en obras pictóricas.

El sistema parte de imágenes correspondientes a recortes de flores o motivos vegetales previamente identificados dentro de una obra artística. A partir de cada recorte, el modelo debe predecir una etiqueta botánica perteneciente al vocabulario definido en el conjunto de datos. En los experimentos principales, esta etiqueta corresponde al género botánico, ya que este nivel taxonómico ofrece un equilibrio adecuado entre precisión, fiabilidad del etiquetado y número suficiente de ejemplos por clase.

El enfoque adoptado no consiste en entrenar una red neuronal desde cero, ya que el tamaño del conjunto de datos y la complejidad del dominio artístico harían que esta estrategia fuera poco robusta y propensa al sobreajuste. En su lugar, se emplean arquitecturas previamente entrenadas, capaces de extraer patrones visuales generales como bordes, texturas, formas, colores y relaciones espaciales. Posteriormente, estas arquitecturas se adaptan al problema mediante la sustitución de la cabeza de clasificación final y el entrenamiento sobre el *dataset* construido.

### 3.3.2 ARQUITECTURA DEL PIPELINE

La solución desarrollada se organiza como un pipeline modular que integra las fases principales del sistema: preparación del conjunto de datos, preprocesamiento de las imágenes, adaptación de modelos preentrenados, entrenamiento, evaluación y análisis de resultados. Esta arquitectura permite mantener separadas las distintas etapas del proceso y comparar diferentes modelos bajo una misma estructura de trabajo.

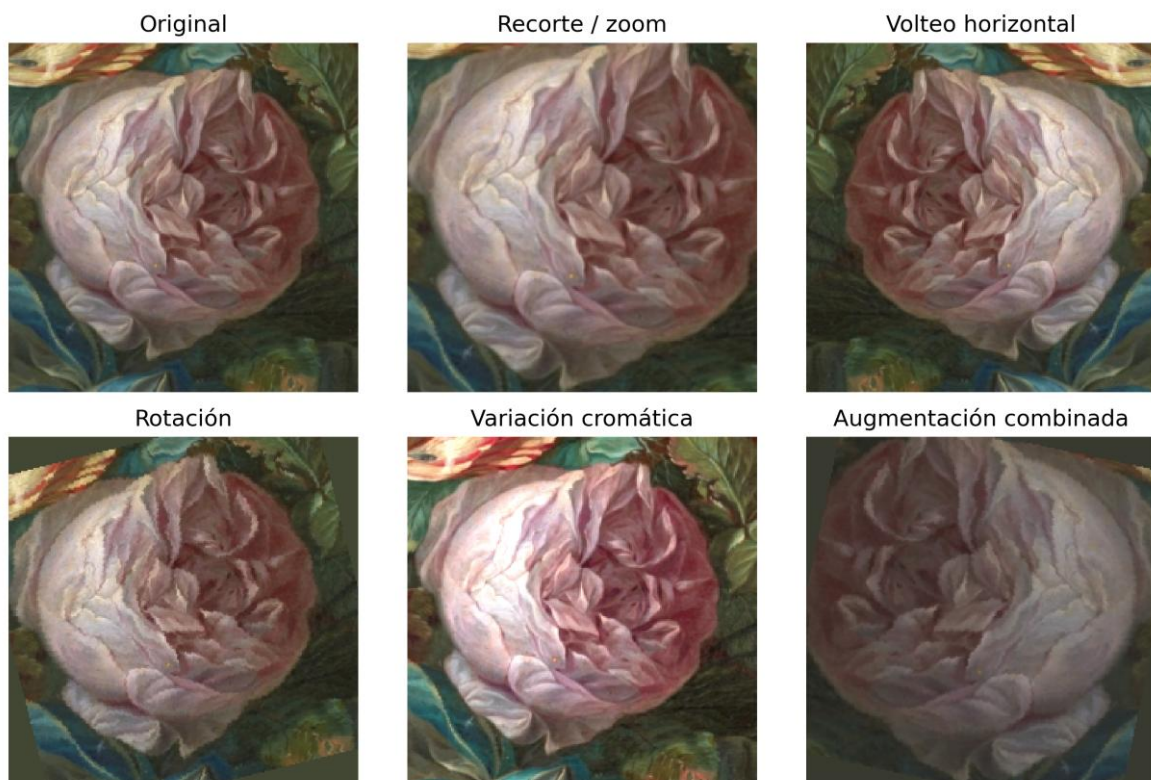
El punto de partida del pipeline son las imágenes originales procedentes de las distintas fuentes descritas anteriormente. Tras su selección, se realizan recortes manuales de los motivos vegetales identificables y se asigna a cada recorte una etiqueta botánica. Este proceso permite transformar obras pictóricas completas, que pueden contener varias flores o plantas, en muestras individuales adecuadas para una tarea de clasificación supervisada.

Una vez generados los recortes, las imágenes se organizan en el *dataset* final siguiendo una estructura jerárquica basada en la taxonomía botánica. Cada muestra queda asociada a una familia, un género y una especie, además de conservar información sobre su fuente de procedencia y la partición del conjunto de datos a la que pertenece. Esta organización constituye la base sobre la que se construyen los conjuntos de entrenamiento, validación y prueba.

A continuación, el *pipeline* aplica el preprocesamiento necesario para adaptar las imágenes al formato de entrada requerido por los modelos. Esta fase incluye el redimensionamiento de las imágenes, la conversión a tensores y la normalización de los canales RGB. Además, durante el entrenamiento se incorporan técnicas de *data augmentation*, con el objetivo de aumentar la variabilidad visual observada por el modelo y reducir el riesgo de sobreajuste, especialmente relevante dado el tamaño limitado del conjunto de datos.

En concreto, para el subconjunto de entrenamiento las imágenes se redimensionan y posteriormente, se aplica un recorte aleatorio reescalado al tamaño de entrada del modelo. También se utilizan volteo horizontal, pequeñas rotaciones y variaciones cromáticas

controladas de brillo, contraste, saturación y tono. Estas transformaciones permiten simular cambios razonables en encuadre, orientación e iluminación, sin alterar de forma agresiva la morfología floral representada. La Figura 3 muestra un ejemplo visual de estas transformaciones.



*Figura 3. Ejemplo de data augmentation en entrenamiento para una imagen*

El *data augmentation* se aplica únicamente durante el entrenamiento y de forma dinámica, es decir, no se generan nuevas imágenes almacenadas en disco. Por tanto, el número físico de imágenes del *dataset* permanece constante antes y después del aumento de datos. Sin embargo, el tamaño efectivo del conjunto de entrenamiento aumenta durante el aprendizaje, ya que una misma imagen puede presentarse al modelo con ligeras variaciones en distintas épocas. La diferencia es que, debido a las transformaciones aleatorias aplicadas en cada iteración, una misma imagen puede aparecer con versiones ligeramente distintas en épocas sucesivas. Por ejemplo, una imagen puede mostrarse en una época con un pequeño giro y en

otra con un recorte o una variación de brillo diferente. Por tanto, cuantas más épocas se entrene el modelo, mayor será la diversidad de versiones observadas por el mismo a partir de las imágenes originales.

En validación y prueba no se aplica *data augmentation*. En estos subconjuntos se utilizan transformaciones deterministas, como redimensionamiento, recorte central, conversión a tensor y normalización, con el fin de garantizar una evaluación estable, reproducible y comparable entre modelos.

La siguiente fase consiste en la carga del modelo preentrenado seleccionado. En el proyecto se contemplan distintas arquitecturas de visión por ordenador, incluyendo redes convolucionales y modelos basados en *Vision Transformers* [8]. Estos modelos se adaptan a la tarea mediante la sustitución de la cabeza de clasificación final, de forma que la salida corresponda al número de clases botánicas definidas en el conjunto de datos.

Durante el entrenamiento, el modelo recibe las imágenes preprocesadas y ajusta sus parámetros en función de las etiquetas disponibles. El proceso registra las métricas de entrenamiento y validación, y guarda el mejor modelo obtenido, lo que permite conservar el *checkpoint* más adecuado para la evaluación posterior.

Tras el entrenamiento, el modelo guardado se evalúa sobre el subconjunto correspondiente. Esta fase permite calcular métricas globales mencionadas en el apartado 3.1.2, las cuales permiten comparar el rendimiento de las distintas arquitecturas y analizar el comportamiento del sistema en un conjunto de datos con clases representadas de forma desigual.

Por último, el pipeline incorpora una fase de análisis de resultados. En esta etapa se estudian las predicciones del modelo no solo de forma global, sino también atendiendo a variables como la clase botánica, la familia, la fuente de procedencia, la escuela artística, el movimiento, la técnica o el tipo de obra. De este modo, el análisis permite relacionar el rendimiento del modelo con las características visuales y artísticas de las imágenes.

En la Figura 4 se resume la arquitectura general del pipeline desarrollado, desde las imágenes originales hasta el análisis final de los resultados.

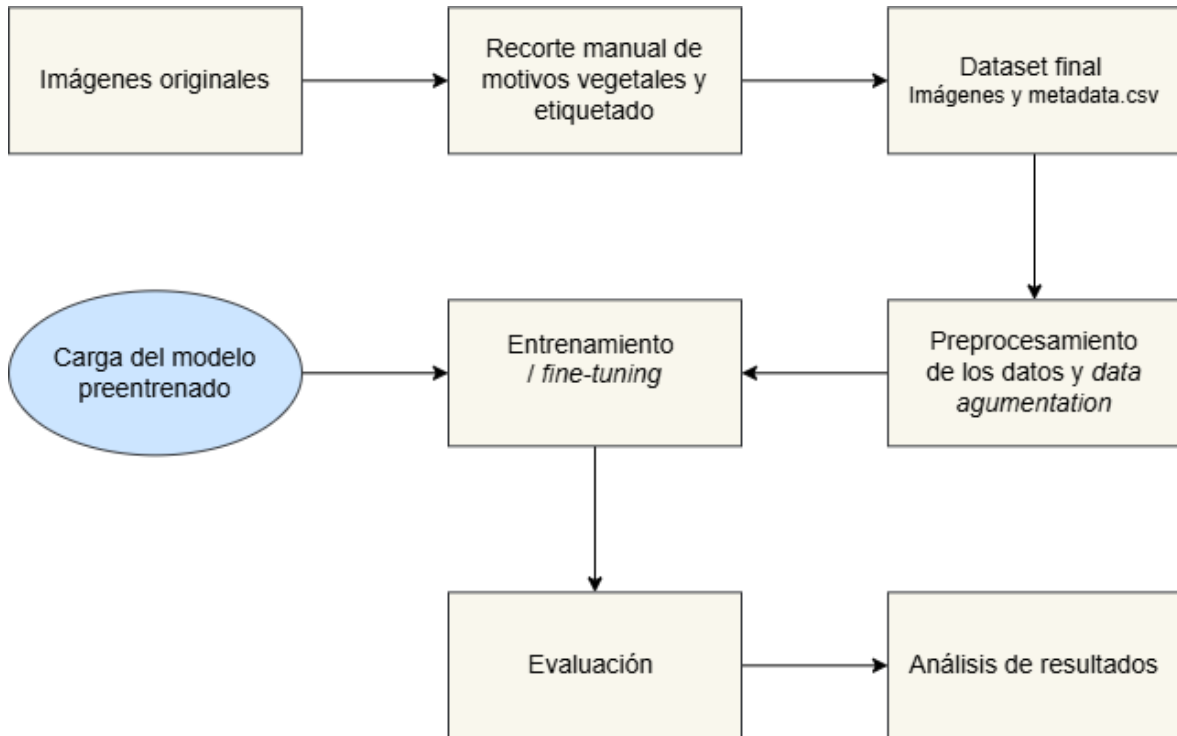


Figura 4. Esquema de la arquitectura general del pipeline

### 3.3.3 MODELOS EVALUADOS

Para evaluar qué tipo de arquitectura se adapta mejor a la identificación de flores representadas en obras pictóricas, se seleccionaron distintos modelos de visión por ordenador preentrenados. La comparación incluyó arquitecturas convolucionales clásicas, arquitecturas convolucionales modernas, modelos basados en *vision transformers* y modelos con preentrenamiento más próximo al dominio botánico.

La selección de modelos responde a la necesidad de comparar enfoques con propiedades diferentes. Las redes convolucionales han sido ampliamente utilizadas en clasificación de

imágenes y destacan por su capacidad para capturar patrones locales, texturas, contornos y formas. Los *vision transformers*, por su parte, permiten modelar relaciones más globales dentro de la imagen mediante mecanismos de atención. Finalmente, el uso de modelos con preentrenamiento botánico permite estudiar si el conocimiento previo sobre plantas facilita la identificación de motivos vegetales cuando estos aparecen representados en pintura.

Las arquitecturas evaluadas fueron ResNet-50 [5], EfficientNet-B3 [6], ConvNeXt-Small [7], ViT-Base [8] y PlantCLEF-DINOv2 [10]. La Tabla 4 resume el tipo de arquitectura de cada modelo y su motivo de inclusión en la comparación experimental.

<b>Modelo</b>	<b>Tipo de arquitectura</b>	<b>Motivo de inclusión</b>
ResNet-50	CNN clásica	Modelo convolucional de referencia
EfficientNet-B3	CNN eficiente	Evaluar el equilibrio entre rendimiento y coste computacional
ConvNeXt-Small	CNN moderna	Comparar con un baseline generalista competitivo
ViT-Base	Vision Transformer	Evaluar una arquitectura basada en mecanismos de atención
PlantCLEF-DINOv2	DINOv2 / Vision Transformer	Analizar el efecto del preentrenamiento especializado en plantas

*Tabla 4. Arquitectura de cada modelo evaluado*

Para facilitar la comparación, todos los modelos se integraron en una misma estructura de entrenamiento y evaluación. Cada arquitectura se cargó a partir de su versión preentrenada correspondiente y se adaptó a la tarea sustituyendo su cabeza de clasificación original por una nueva capa ajustada al número de géneros botánicos del conjunto de datos. De este modo, la comparación permite analizar la capacidad de distintos *backbones* para adaptarse al dominio artístico bajo un protocolo experimental común.

La descripción ampliada de las arquitecturas y sus esquemas visuales se incluye en el Anexo II. Arquitectura de los modelos evaluados.

### 3.3.4 ESTRATEGIAS DE ENTRENAMIENTO

Una vez definidos los modelos a evaluar, se plantearon distintas estrategias de entrenamiento con el objetivo de analizar cómo influye el grado de adaptación del modelo preentrenado al dominio artístico. En todos los casos, los modelos parten de pesos previamente aprendidos y se modifican para ajustar la salida al número de clases botánicas del conjunto de datos.

La implementación del entrenamiento se apoya en PyTorch [16] y en la librería `timm` [17], que permite cargar los modelos preentrenados mencionados en 3.3.3. Para cada modelo, el código sustituye la cabeza de clasificación original por una nueva cabeza compuesta por una capa de *dropout* y una capa lineal final adaptada al número de clases del proyecto.

#### 3.3.4.1 *Linear probing*

La primera estrategia considerada fue el *linear probing*. En esta aproximación, el modelo preentrenado se utiliza como extractor fijo de características visuales. Es decir, el *backbone* permanece congelado y únicamente se entrena la nueva cabeza de clasificación, como se puede observar en el esquema de la Figura 5.

Esta estrategia permite evaluar hasta qué punto las representaciones aprendidas durante el preentrenamiento son útiles para la tarea de identificación botánica en pintura. Si el modelo obtiene buenos resultados con el *backbone* congelado, significa que las características

visuales aprendidas previamente ya contienen información suficientemente relevante para distinguir las clases botánicas del *dataset*.

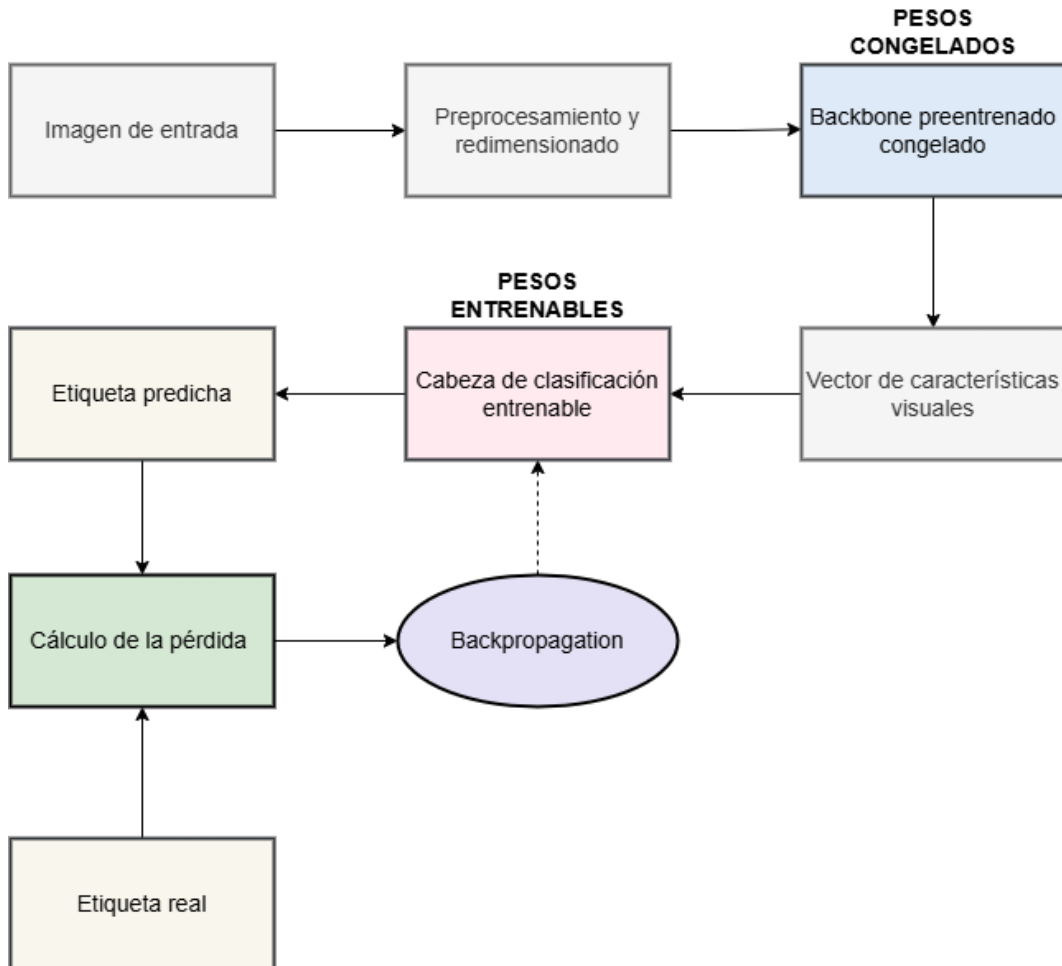


Figura 5. Esquema de la estrategia de linear probing

### 3.3.4.2 Full fine-tuning

La segunda estrategia fue el *full fine-tuning*. En este caso, además de sustituir la cabeza final de clasificación, se actualizan los pesos del modelo completo durante el entrenamiento. Por tanto, se entrena el *backbone* preentrenado, como se puede comprobar en la Figura 6. De esta manera, las representaciones internas del modelo se adaptan al dominio específico del proyecto. Esta estrategia permite que el modelo ajuste sus filtros, mapas de características o mecanismos de atención a los patrones propios de la pintura. Así, el modelo puede aprender

no solo rasgos botánicos generales, sino también cómo estos rasgos aparecen transformados en un contexto artístico.

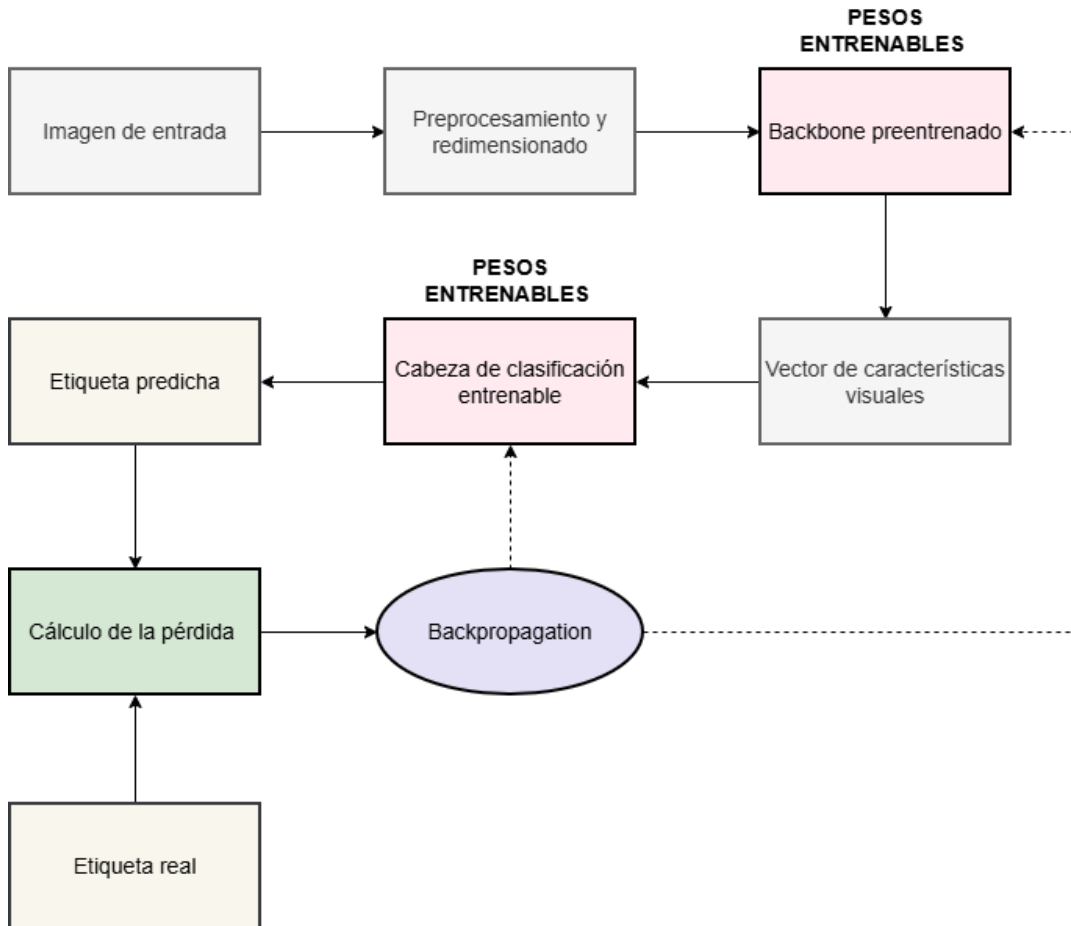


Figura 6. Esquema de full fine-tuning

### 3.3.4.3 Configuración del entrenamiento

El entrenamiento se realizó utilizando una función de pérdida de entropía cruzada. Además, el código permite aplicar *label smoothing*, que suaviza la distribución objetivo y reduce la tendencia del modelo a generar predicciones excesivamente confiadas [15]. También se emplea el optimizador AdamW, junto con un planificador de tasa de aprendizaje *CosineAnnealingLR*, que reduce progresivamente el *learning rate* durante el entrenamiento.

La selección del mejor modelo se realiza a partir de una métrica de validación configurable. En los experimentos principales se utiliza `val_balanced_acc`, es decir, la *balanced accuracy* sobre el conjunto de validación. Esta decisión es adecuada para un *dataset* con clases desbalanceadas, ya que evita seleccionar únicamente el modelo con mayor *accuracy* global y favorece modelos con mejor comportamiento medio entre clases.

Durante cada época se registran métricas de entrenamiento y validación, incluyendo pérdida, *accuracy*, *balanced accuracy* y F1 macro. Estas métricas se almacenan también en un histórico, lo que permite analizar la evolución del entrenamiento y detectar posibles problemas de sobreajuste.

#### ***3.3.4.4 Tratamiento del desbalance de clases***

El conjunto de datos presenta una distribución desigual entre clases botánicas, ya que algunos géneros aparecen con mayor frecuencia que otros en las fuentes utilizadas. Este desbalance podía afectar negativamente tanto al entrenamiento como a la evaluación del modelo. Para cuantificar este desequilibrio se utilizó el *imbalance ratio* (IR), definido como la relación entre el número de imágenes de la clase mayoritaria y el número de imágenes de la clase minoritaria.

Como se observa en la Figura 7, en la distribución original del conjunto de datos el género mayoritario fue rosa, con aproximadamente 89 imágenes, mientras que numerosas clases contaban únicamente con una muestra. Esto da lugar a un *imbalance ratio* aproximado de 89:1, lo que evidencia un desbalance alto entre clases. Esta situación podía afectar negativamente tanto al entrenamiento como a la evaluación del modelo, ya que las clases con una o dos imágenes no proporcionan ejemplos suficientes para aprender rasgos visuales representativos ni permiten estimar de forma fiable el rendimiento del modelo.

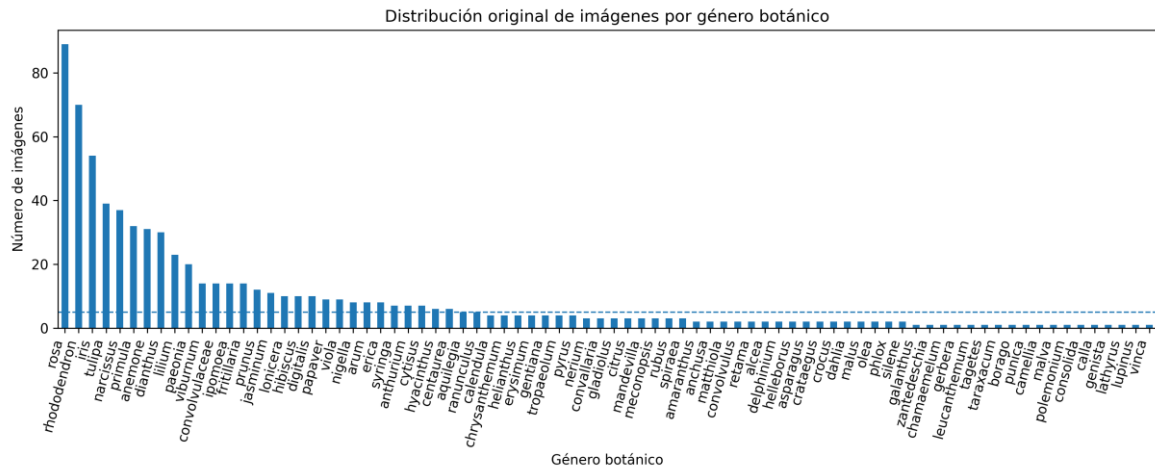


Figura 7. Distribución de imágenes por género botánico antes de aplicar el criterio mínimo de representación

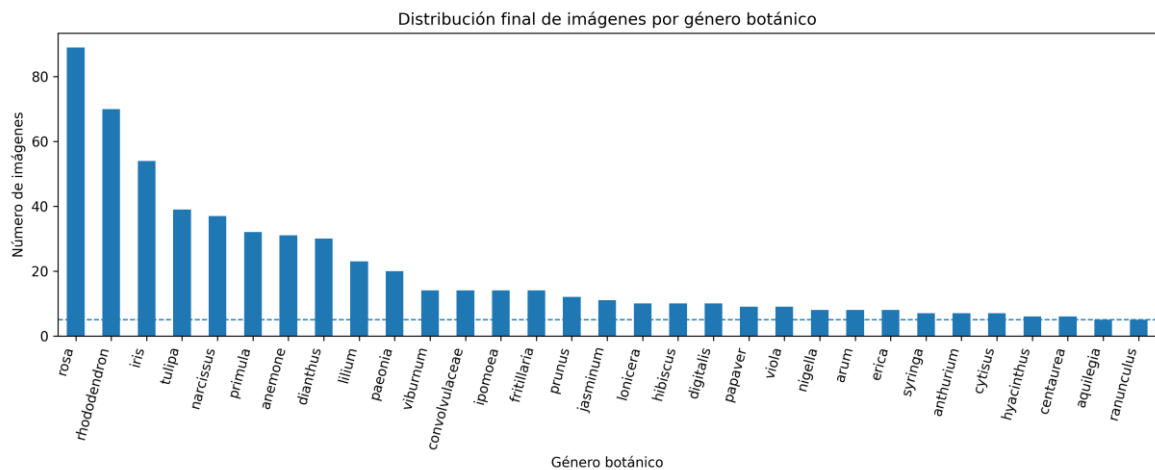


Figura 8. Distribución de imágenes por género botánico tras aplicar el criterio mínimo de representación

Para reducir el efecto del desbalance extremo, se estableció un criterio mínimo de representación por clase. En concreto, se mantuvieron en los subconjuntos principales aquellas clases que contaban con al menos cinco muestras, mientras que las clases con menor representación se excluyeron del entrenamiento y de la evaluación principales. Esta decisión supuso descartar parte de la información disponible, pero permitió trabajar con un conjunto de datos más consistente y con clases mínimamente representadas.

Tras aplicar este filtrado, el conjunto final quedó formado por géneros con al menos cinco imágenes, como se puede observar en la Figura 8. Aunque la distribución sigue siendo desigual, el *imbalance ratio* final se redujo aproximadamente a 17,8:1. Por tanto, el filtrado no elimina completamente el desbalance, pero sí evita los casos más extremos de clases con representación insuficiente.

Además, la evaluación se diseñó teniendo en cuenta este desbalance. Por ello, el mejor *checkpoint* se seleccionó mediante *balanced accuracy* en validación y los resultados finales se analizaron mediante métricas macro, como F1 macro, precisión macro y *recall* macro. El código también permite aplicar pesos de clase en la función de pérdida y utilizar un muestreador ponderado, aunque estos mecanismos no se activaron en la configuración final. Por tanto, el tratamiento principal del desbalance se basó en el filtrado de clases minoritarias, la selección mediante *balanced accuracy* y el uso de métricas equilibradas entre clases.

### 3.4 IMPLEMENTACIÓN

Tras definir el diseño de la solución, se procedió a implementar el sistema completo de construcción del *dataset*, entrenamiento y evaluación del modelo. El código se realizó en Python, siguiendo una estructura modular que permite separar claramente las fases de preparación de datos, definición de modelos, entrenamiento, evaluación y análisis de los resultados.

#### 3.4.1 TECNOLOGÍAS Y RECURSOS HW/SW EMPLEADOS

Como se ha mencionado en 3.4, el código del proyecto se escribió en Python, debido a su soporte de tareas de aprendizaje profundo, como la visión por ordenador.

Para el desarrollo de los modelos evaluados se utilizó la librería de `PyTorch` [16], que proporciona las herramientas necesarias para definir *datasets*, crear *DataLoaders*, ejecutar entrenamiento en GPU, calcular las pérdidas y optimizar redes neuronales.

También se usó `Torchvision` para realizar las transformaciones de imagen. Durante el entrenamiento se utilizaron técnicas de *data augmentation*, como redimensionado, recorte aleatorio, rotaciones y variaciones cromáticas controladas. Estas transformaciones permiten enriquecer el conjunto de entrenamiento y mejorar la capacidad de generalización del modelo, sin alterar de forma agresiva la morfología floral representada en las imágenes.

Para cargar los modelos preentrenados se utilizó la librería de `timm` [17], que permite utilizar modelos de visión por ordenador preentrenados, para adaptarlos a esta tarea de clasificación.

Para la evaluación y el análisis de resultados se utilizaron `Pandas`, `NumPy` y `scikit-learn`. `Pandas` y `NumPy` se emplearon para la gestión de metadatos, el tratamiento de resultados y la generación de tablas de análisis. Por su parte, `scikit-learn` [18] permitió calcular métricas de clasificación como *accuracy*, *balanced accuracy*, *precision*, *recall*, *F1-score*, *Cohen's kappa* y *Matthews correlation coefficient*.

Además se utilizó `TensorBoard` para registrar la evolución del entrenamiento.

### **3.4.2 ESTRUCTURA DEL CÓDIGO**

El código se organizó siguiendo una organización modular, en la que cada parte del código se emplea para una tarea concreta dentro del flujo completo del proyecto. Así, es más simple modificar cada componente y reproducir los resultados.

En la Figura 9, se puede ver la estructura funcional del código, y las distintas fases de la solución diseñada.

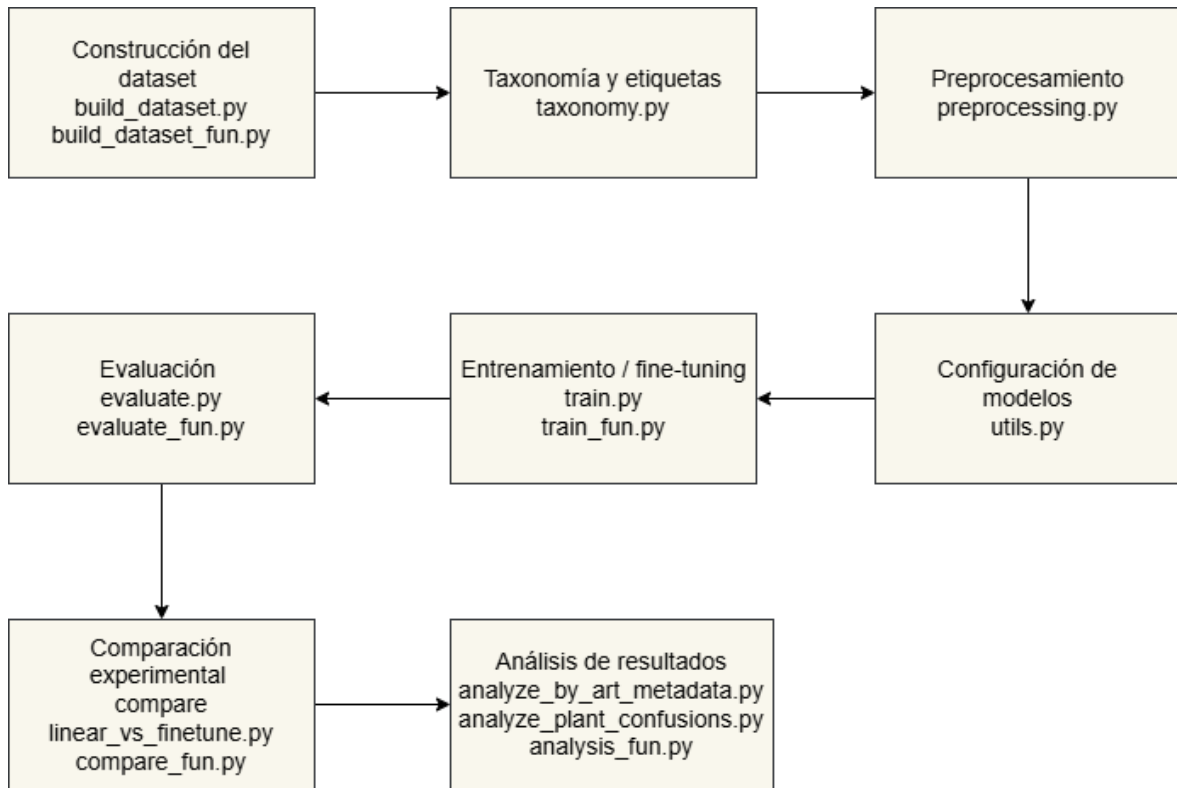


Figura 9. Esquema de la estructura del código desarrollado

Como se observa en la Figura 9, el flujo comienza con la construcción y normalización del *dataset*, continúa con el preprocesamiento y la configuración de los modelos, y finaliza con las fases de entrenamiento, evaluación, comparación experimental y análisis de resultados.

Dentro de esta estructura se distinguen dos tipos de archivos. Por un lado, los scripts principales, como `train.py`, `evaluate.py` o `build_dataset.py`, actúan como puntos de entrada para ejecutar cada fase desde terminal. Por otro lado, los módulos auxiliares, como `train_fun.py`, `eval_fun.py`, `preprocessing.py` o `analysis_fun.py`, contienen la lógica reutilizable del proyecto. Esta separación evita duplicar código y permite reutilizar las mismas funciones en distintos experimentos.

### 3.4.3 PIPELINE DE PREPROCESADO

El pipeline de preprocesado se encarga de transformar las imágenes del *dataset* final en entradas compatibles con los modelos de visión por ordenador. Esta fase parte del archivo

metadata.csv, donde se almacena la información asociada a cada muestra, y permite construir los objetos *Dataset* y *DataLoader* utilizados durante el entrenamiento, la validación y la evaluación.

En primer lugar, cada imagen debe localizarse dentro de la estructura jerárquica del *dataset*. Para ello, el módulo de preprocesado reconstruye la ruta de cada muestra a partir de los campos *family*, *genus*, *species* y *filename*. De este modo, la organización física de las carpetas queda vinculada directamente con la información taxonómica almacenada en los metadatos.

```
def _resolve_image_path(self, row: pd.Series) -> Path:
    return (
        self.dataset_root
        / str(row["family"])
        / str(row["genus"])
        / str(row["species"])
        / str(row["filename"])
    )
```

Este procedimiento permite mantener la correspondencia entre cada imagen y su etiqueta botánica, evitando duplicar información sobre las rutas completas en el archivo de metadatos. Además, facilita que el mismo pipeline pueda trabajar con diferentes niveles taxonómicos de clasificación, ya que las etiquetas de familia, género y especie se encuentran disponibles para cada muestra.

Una vez localizada la imagen, se aplican las transformaciones necesarias para adaptarla al formato de entrada del modelo. En el conjunto de entrenamiento se utilizaron transformaciones de *data augmentation*, combinando operaciones geométricas y cromáticas. Estas transformaciones buscan introducir variabilidad visual durante el entrenamiento sin alterar de forma significativa la morfología de las flores representadas.

```
train_tfms = transforms.Compose([
    transforms.Resize((256, 256)),
    transforms.RandomResizedCrop(image_size, scale=(0.85, 1.0), ratio=(0.9,
1.1)),
    transforms.RandomHorizontalFlip(p=0.5),
    transforms.RandomRotation(degrees=10),
```

```
transforms.ColorJitter(  
    brightness=0.15,  
    contrast=0.15,  
    saturation=0.15,  
    hue=0.03  
),  
transforms.ToTensor(),  
transforms.Normalize(mean=IMAGENET_MEAN, std=IMAGENET_STD),  
])
```

Como se observa en el código superior, las imágenes de entrenamiento se redimensionan, se recortan aleatoriamente y se someten a pequeñas rotaciones, volteos horizontales y variaciones controladas de color. Por último, cada imagen se convierte a tensor y se normaliza utilizando los valores medios y desviaciones típicas de ImageNet.

Una vez hecho esto, el módulo genera los *DataLoader* correspondientes a entrenamiento, validación y test. El conjunto de entrenamiento se carga con mezcla aleatoria de muestras, pero validación y test tienen un orden fijo.

### 3.4.4 PIPELINE DE ENTRENAMIENTO

El pipeline de entrenamiento se encarga de adaptar los modelos preentrenados al conjunto de datos construido en este trabajo. Esta fase parte de los *Dataloader* generados durante el preprocesado y del modelo seleccionado, y ejecuta el proceso de optimización necesario para ajustar la red a la tarea de clasificación botánica.

El entrenamiento se implementó en los módulos `train.py` y `train_fun.py`, como se muestra en la Figura 9. En primer lugar, se cargan los *DataLoader* de entrenamiento y validación, junto con la información del *dataset*, como el número de clases, el nivel taxonómico utilizado y el diccionario que relaciona cada clase con su índice numérico. En la configuración principal del sistema, el nivel de clasificación utilizado es el género, aunque el código permite trabajar también a nivel de familia o especie.

Una vez preparados los datos, se construye el modelo seleccionado. Para ello, se carga una arquitectura preentrenada mediante la librería `timm` y se sustituye la cabeza de

clasificación original por una nueva cabeza adaptada al número de clases del *dataset*. Esta nueva cabeza está formada por una capa de *dropout* y una capa lineal final. En el caso de `plantclef_dinov2`, el sistema permite cargar opcionalmente un *checkpoint* externo especializado en botánica antes de iniciar el entrenamiento.

```
in_features = infer_in_features(model, cfg["classifier_attr"])
new_head = nn.Sequential(
    nn.Dropout(p=dropout),
    nn.Linear(in_features, num_classes),
)
set_nested_attr(model, cfg["classifier_attr"], new_head)
```

El proceso de optimización utiliza una función de pérdida de entropía cruzada, y como optimizador se utilizó `AdamW`, una variante de `Adam` que desacopla la regularización por decaimiento de pesos del cálculo adaptativo de los gradientes. Esta elección permite entrenar modelos profundos de forma estable y controlar el sobreajuste mediante el parámetro de *weight decay*. Además, se empleó un planificador de tasa de aprendizaje `CosineAnnealingLR`, que reduce progresivamente el *learning rate*. De esta forma, el entrenamiento comienza con actualizaciones más amplias y termina con ajustes más finos de los pesos del modelo

```
criterion = nn.CrossEntropyLoss(
    weight=class_weights,
    label_smoothing=args.label_smoothing,
)

optimizer = optim.AdamW(
    model.parameters(),
    lr=args.lr,
    weight_decay=args.weight_decay,
)

scheduler = CosineAnnealingLR(
    optimizer,
    T_max=args.epochs,
```

```
eta_min=args.lr / 100,  
)
```

Durante cada época, el modelo se entrena sobre el conjunto de entrenamiento y posteriormente se evalúa sobre el conjunto de validación. En ambas fases se calculan métricas que permiten controlar tanto el rendimiento global del modelo como su comportamiento en un dataset con clases representadas de forma desigual.

El sistema guarda automáticamente el mejor modelo obtenido durante el entrenamiento. Para ello, se utiliza una métrica de validación configurable, que en la configuración principal corresponde a `val_balanced_acc`. Esta elección permite priorizar modelos con un comportamiento más equilibrado entre clases, en lugar de seleccionar únicamente el modelo con mayor precisión global.

Además del *checkpoint*, el pipeline almacena el histórico de métricas de cada época y registra la evolución del entrenamiento en `TensorBoard`. Esto permite analizar posteriormente la convergencia del modelo, comparar ejecuciones y detectar posibles signos de sobreajuste.

### 3.4.5 PIPELINE DE EVALUACIÓN

El *pipeline* de evaluación permite medir el rendimiento de los modelos entrenados sobre un subconjunto concreto del *dataset*, normalmente validación o prueba. Esta fase parte del *checkpoint* guardado durante el entrenamiento y aplica un procedimiento común para todas las arquitecturas, de forma que los resultados obtenidos sean comparables.

La evaluación se implementó en los módulos `evaluate.py` y `eval_fun.py`. El primero actúa como punto de entrada desde terminal, mientras que `eval_fun.py` contiene la lógica principal del proceso: carga del *checkpoint*, reconstrucción del modelo, preparación del conjunto de evaluación, ejecución de inferencia y cálculo de métricas.

En primer lugar, se carga el archivo `best_model.pt` correspondiente al modelo seleccionado. Este archivo contiene los pesos entrenados, el número de clases, el diccionario `class_to_idx`, el nivel taxonómico utilizado y la época en la que se obtuvo el mejor resultado de validación. A partir de esta información, el sistema reconstruye la arquitectura

correspondiente, sustituye la cabeza de clasificación por la estructura utilizada durante el entrenamiento y carga los pesos almacenados.

El conjunto de evaluación se construye a partir del archivo *metadata.csv*, seleccionando las imágenes correspondientes al *split* indicado: entrenamiento, validación, prueba o conjunto completo. Las muestras marcadas como *unused* se excluyen de la evaluación principal. A diferencia del entrenamiento, en esta fase no se aplica *data augmentation*, sino transformaciones deterministas de redimensionamiento, conversión a tensor y normalización, garantizando así una evaluación estable y reproducible.

Una vez preparado el *DataLoader*, el modelo se ejecuta en modo evaluación. Para cada imagen se calculan los *logits*, se obtiene la distribución de probabilidades mediante *softmax* y se selecciona la clase con mayor probabilidad como predicción Top-1. Además, se conservan las probabilidades necesarias para calcular métricas Top-3 y Top-5.

El *pipeline* calcula métricas globales como *accuracy* Top-1, Top-3 y Top-5, *balanced accuracy*, F1 macro, F1 weighted, precisión macro, *recall* macro, Cohen's kappa y Matthews correlation coefficient. La inclusión de métricas macro y *balanced accuracy* resulta especialmente importante debido al desbalance entre clases, ya que permiten valorar el rendimiento del modelo sin que las clases mayoritarias dominen completamente la evaluación.

Además, el sistema genera un informe de clasificación por clase y, cuando se evalúan varios modelos, una tabla comparativa con las métricas principales. En conjunto, este *pipeline* permite evaluar todas las arquitecturas bajo las mismas condiciones: mismo conjunto de datos, mismas transformaciones de evaluación, mismo nivel taxonómico y mismas métricas.

### 3.4.6 REPRODUCIBILIDAD

La reproducibilidad del proyecto se abordó mediante una organización modular del código, el uso de archivos de metadatos estructurados y el almacenamiento de los resultados en formatos reutilizables. El objetivo es que el proceso completo, desde la construcción del

*dataset* hasta la evaluación de los modelos, pueda repetirse siguiendo una secuencia clara de ejecución.

Para reproducir el proceso desde la construcción del *dataset*, es necesario disponer de las imágenes recopiladas y etiquetadas, de los scripts del proyecto y de la configuración empleada en cada fase. Si se parte directamente del *dataset* ya construido, basta con disponer de su estructura final, del archivo `metadata.csv` y de los scripts correspondientes. El archivo de metadatos recoge para cada imagen el nombre del archivo, la especie, el género, la familia, la fuente de procedencia, el identificador original y la partición asignada, por lo que constituye el punto de conexión entre las imágenes almacenadas y los procesos de preprocesado, entrenamiento, evaluación y análisis.

Los *checkpoints* entrenados no son una entrada necesaria cuando el objetivo es reentrenar los modelos, ya que se generan durante la fase de entrenamiento. En cambio, sí son necesarios si se desea reproducir directamente la evaluación de un modelo previamente entrenado, dado que el *pipeline* de evaluación reconstruye la arquitectura a partir de los pesos almacenados y de la información contenida en el archivo `best_model.pt`.

El flujo de reproducción puede resumirse en las siguientes fases:

#### ***3.4.6.1 Construcción del dataset***

En primer lugar, se genera la estructura jerárquica del conjunto de datos a partir de las imágenes recopiladas y etiquetadas. Para ello se ejecuta el script de construcción del *dataset*, que organiza las imágenes según la estructura `family/genus/species` y genera el archivo `metadata.csv`.

```
python build_dataset.py -apply
```

#### ***3.4.6.2 Creación de los conjuntos de datos y preprocesado.***

A partir del *dataset* final, el módulo de preprocesamiento crea los objetos *Dataset* y *DataLoader* correspondientes a entrenamiento, validación y prueba. En esta fase se define

también el nivel taxonómico utilizado como etiqueta de clasificación. En la configuración principal se utiliza el nivel de género.

```
python preprocessing.py --dataset-root dataset_final --label-level genus
```

### 3.4.6.3 Entrenamiento del modelo.

Una vez preparados los datos, se ejecuta el entrenamiento indicando el modelo que se desea ajustar. El entrenamiento genera un directorio de artefactos con el mejor *checkpoint*, el histórico de métricas y los registros de TensorBoard.

```
python train.py --model plantclef_dinov2 --dataset-root dataset_final --label-level genus
```

### 3.4.6.4 Evaluación del checkpoint.

Tras el entrenamiento, el modelo guardado como `best_model.pt` se evalúa sobre el subconjunto deseado, normalmente validación o test. Los resultados pueden almacenarse en formato JSON para su análisis posterior.

```
python evaluate.py --model plantclef_dinov2 --data-dir dataset_final --split test --save-json results/eval_test.json
```

### 3.4.6.5 Análisis posterior de resultados.

Finalmente, los resultados pueden analizarse por clase botánica, fuente, estilo artístico, técnica, autor u otras variables de metadatos.

```
python analyze_by_art_metadata.py --checkpoint artifacts/plantclef_dinov2/best_model.pt --data-dir dataset_final --art-metadata art_metadata.xlsx --split test --out-dir analysis_results
```

Además de esta secuencia de ejecución, el proyecto incorpora varios elementos que favorecen la reproducibilidad. La partición del *dataset* se genera mediante una semilla fija,

lo que permite mantener una división estable entre entrenamiento, validación y prueba. Los *checkpoints* almacenan información necesaria para reconstruir el modelo, como el número de clases, el diccionario `class_to_idx`, el nivel taxonómico y la métrica utilizada para seleccionar el mejor modelo.

## Capítulo 4. RESULTADOS

En este capítulo se presentan los resultados obtenidos en la evaluación de los modelos de clasificación de motivos florales en obras artísticas. El objetivo es validar la viabilidad del enfoque propuesto y analizar qué arquitecturas y estrategias de entrenamiento se adaptan mejor al reconocimiento de flores representadas en obras pictóricas.

El análisis se organiza en varios niveles. En primer lugar, se comparan las arquitecturas evaluadas mediante métricas globales de clasificación. A continuación, se estudia el impacto de la estrategia de entrenamiento, comparando *linear probing* y *full fine-tuning*. Posteriormente, se analiza el comportamiento del modelo final por clase botánica, fuente de procedencia y características artísticas de las obras. Finalmente, se realiza un análisis de los errores y confusiones más relevantes, con el fin de identificar las principales limitaciones del sistema.

### 4.1 CONFIGURACIÓN EXPERIMENTAL

La evaluación se realizó sobre el conjunto de datos construido en el proyecto, organizado mediante una estructura taxonómica jerárquica de familia, género y especie. Aunque el *dataset* conserva información en estos tres niveles, la configuración principal de clasificación se definió a nivel de género.

Se evaluaron los modelos recogidos en el apartado 3.3.3. La selección de estas arquitecturas permite comparar modelos convolucionales clásicos, modelos convolucionales modernos, arquitecturas basadas en *Vision Transformers* y modelos con preentrenamiento más próximo al dominio botánico.

Para garantizar una comparación lo más homogénea posible, todos los modelos se entrenaron y evaluaron siguiendo un mismo protocolo experimental: mismo conjunto de datos, mismo nivel taxonómico de clasificación, mismas particiones de entrenamiento, validación y

prueba, estrategia general de *full fine-tuning* y selección del mejor *checkpoint* a partir de la *balanced accuracy* en el conjunto de validación. No obstante, algunos parámetros variaron entre arquitecturas, especialmente el número de épocas, el *learning rate* inicial y el tamaño de entrada requerido por cada modelo.

La Tabla 5 recoge la configuración común utilizada en los experimentos principales.

Parámetro	Valor utilizado
Nivel taxonómico de clasificación	Género
Estrategia principal de entrenamiento	Full fine-tuning
Optimizador	AdamW
Función de pérdida	Entropía cruzada
Weight decay	1e-2
Dropout	0.4
Label smoothing	0.1
Planificador de tasa de aprendizaje	CosineAnnealingLR
Métrica de selección del mejor checkpoint	val_balanced_acc
Transformaciones	Data augmentation en entrenamiento y transformaciones deterministas en validación/test

Tabla 5. Configuración utilizada en los modelos evaluados

La Tabla 6 muestra la configuración específica de cada arquitectura en cuanto al número de épocas, el *learning rate* inicial y el tamaño de entrada empleado.

Modelo	Épocas	Learning rate inicial	Tamaño de entrada
ResNet-50	20	1e-4	224 × 224
EfficientNet-B3	35	1e-4	224 × 224
ConvNeXt-Small	35	1e-4	224 × 224
ViT-Base	35	1e-4	224 × 224

PlantCLEF-DINOv2	30	2e-5	518 × 518
------------------	----	------	-----------

Tabla 6. Configuración utilizada por modelo evaluado

La evaluación final se llevó a cabo sobre el subconjunto de prueba, utilizando transformaciones deterministas de imagen. A diferencia del entrenamiento, en esta fase no se aplicó *data augmentation*, con el fin de obtener resultados estables y comparables entre modelos. Las imágenes se redimensionaron, se convirtieron a tensor y se normalizaron con los valores estándar utilizados habitualmente en modelos preentrenados sobre ImageNet.

Dado que el conjunto de datos presenta una distribución desigual entre clases, se calcularon distintas métricas de clasificación: *accuracy* Top-1, *accuracy* Top-3, *accuracy* Top-5, *balanced accuracy*, F1 macro, F1 weighted, precisión macro, *recall* macro, Cohen's kappa y Matthews *correlation coefficient*. Estas métricas permiten analizar el rendimiento del modelo desde distintas perspectivas y valorar mejor su comportamiento en clases con diferente representación dentro del *dataset*.

En particular, la *balanced accuracy* y el F1 macro son importantes en este trabajo, ya que asignan un peso más equilibrado a las distintas clases y evitan que el rendimiento quede dominado únicamente por los géneros más frecuentes. Las métricas Top-3 y Top-5 también son útiles, ya que permiten comprobar si la clase correcta aparece entre las principales predicciones del modelo, incluso cuando no ocupa la primera posición.

## 4.2 COMPARACIÓN GLOBAL DE MODELOS

En este apartado se comparan los resultados obtenidos por las distintas arquitecturas evaluadas sobre el conjunto de *test*. El objetivo es analizar qué tipo de modelo se adapta mejor a la clasificación botánica de flores representadas en obras pictóricas y valorar si existen diferencias relevantes entre redes convolucionales clásicas, redes convolucionales modernas, *Vision Transformers* y modelos con preentrenamiento más próximo al dominio botánico.

Modelo	Acc. Top-1	Acc. Top-3	Acc. Top-5	Balanced Accuracy	F1 Macro	F1 Weighted	MCC
ResNet-50	0.281	0.447	0.579	0.209	0.141	0.241	0.258
EfficientNet-B3	0.596	0.754	0.807	0.468	0.451	0.586	0.573
ViT-Base	0.640	0.789	0.868	0.502	0.454	0.608	0.618
ConvNeXt-Small	0.719	0.842	0.877	0.598	0.541	0.690	0.701
PlantCLEF-DINOv2	0.754	0.807	0.816	0.704	0.644	0.767	0.744

Tabla 7. Resultados en el conjunto de test para los diferentes modelos entrenados

Como se observa en la Tabla 7, ResNet-50 obtiene el rendimiento más bajo, por lo que funciona únicamente como *baseline* convolucional clásico. EfficientNet-B3 mejora de forma clara este resultado, aunque sus métricas equilibradas siguen siendo moderadas, lo que indica un rendimiento desigual entre géneros.

Entre los modelos generalistas, ConvNeXt-Small obtiene el mejor comportamiento global, con una *accuracy* Top-1 de 0.719 y los valores más altos de Top-3 y Top-5. Esto sugiere que sus representaciones son útiles para esta tarea, aunque no siempre sitúe la clase correcta como primera predicción.

El mejor modelo global es PlantCLEF-DINOv2, que alcanza la mayor *accuracy* Top-1, *balanced accuracy*, F1 macro, F1 weighted y MCC. La mejora frente a ConvNeXt-Small es especialmente relevante en *balanced accuracy* y F1 macro, lo que indica un rendimiento más equilibrado entre géneros botánicos.

Además de las métricas recogidas en la Tabla 7, se calculó la curva ROC multiclase para comparar la capacidad de separación probabilística de los modelos. Como se observa en la Figura 10, ConvNeXt-Small obtiene el mayor valor de ROC-AUC, seguido de ViT-Base y EfficientNet-B3. Este resultado indica que ConvNeXt-Small presenta una buena capacidad para asignar puntuaciones elevadas a la clase correcta frente al resto de clases a lo largo de distintos umbrales de decisión. Esta observación es coherente con sus valores de *accuracy* Top-3 y Top-5, que son los más altos de la comparación.

Esta curva debe interpretarse como una métrica complementaria, ya que en este trabajo, el objetivo principal del sistema es asignar una etiqueta botánica concreta a cada imagen, por lo que las métricas más directamente relacionadas con la predicción final son la *accuracy* Top-1, la *balanced accuracy*, el F1 macro y el MCC. Desde esta perspectiva, PlantCLEF-DINOv2 ofrece el mejor rendimiento global como clasificador final, mientras que ConvNeXt-Small destaca por su capacidad de separación probabilística y por situar la clase correcta entre las primeras predicciones.

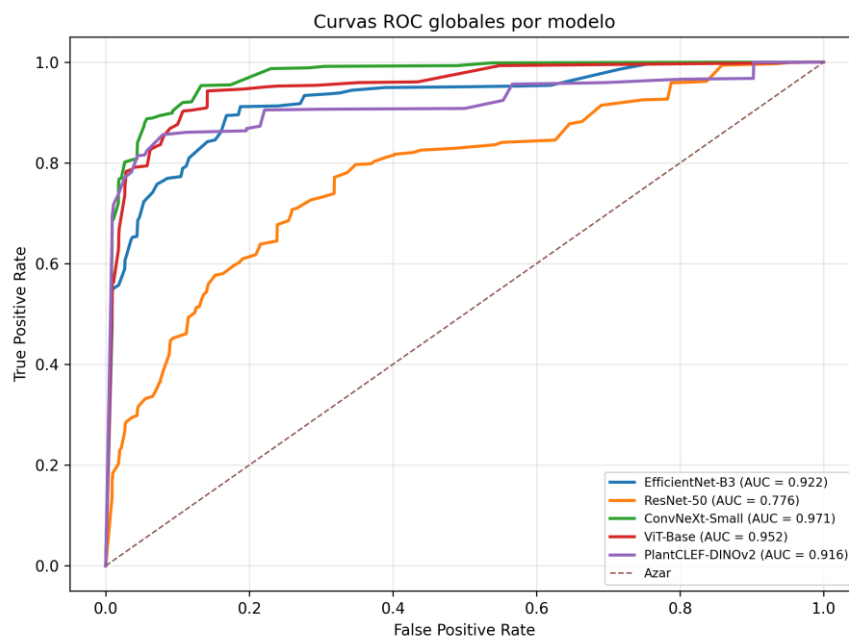


Figura 10. Curva ROC-AUC para los modelos evaluados

Los resultados sugieren que el preentrenamiento especializado en botánica aporta una ventaja clara cuando la predicción final se evalúa a nivel de género. Por este motivo, PlantCLEF-DINOv2 se selecciona como modelo principal para los análisis posteriores, especialmente por su mejor rendimiento en métricas equilibradas entre clases. ConvNeXt-Small, por su parte, se considera el *baseline* generalista más competitivo, ya que obtiene resultados elevados y destaca especialmente en Top-3, Top-5 y ROC-AUC.

### 4.3 COMPARACIÓN ENTRE LINEAR PROBING Y FULL FINE-TUNING

A continuación, se analizó el impacto de la estrategia de entrenamiento sobre el rendimiento del modelo PlantCLEF-DINOv2. Para ello, se compararon dos enfoques: *linear probing*, donde se mantiene congelado el *backbone* preentrenado y solo se entrena la cabeza de clasificación, y *full fine-tuning*, donde se actualizan tanto la cabeza final como los pesos del modelo completo.

Métrica	Linear probing	Full fine-tuning	Diferencia
Accuracy Top-1	0.6657	0.7544	+0.0887
Accuracy Top-3	0.8231	0.8070	-0.0161
Accuracy Top-5	0.8078	0.8158	+0.0080
Balanced accuracy	0.5742	0.7042	+0.1300
F1 macro	0.5335	0.6436	+0.1101
F1 weighted	0.6510	0.7673	+0.1163
Precision macro	0.5344	0.6295	+0.0951

Recall macro	0.5742	0.7042	+0.1300
Cohen Kappa	0.6400	0.7393	+0.0993
MCC	0.6456	0.7443	+0.0987

Tabla 8. Comparación de resultados entre *full fine-tuning* y *linear probing* para el modelo PlantCLEF-DINOv2

Los resultados de la Tabla 8 muestran que el *full fine-tuning* mejora de forma clara la mayoría de métricas. La mejora es especialmente relevante en *balanced accuracy* y F1 macro, con incrementos de 0.1300 y 0.1101 respectivamente. Esto indica que actualizar el modelo completo permite adaptar mejor las representaciones aprendidas al dominio artístico-botánico, especialmente en un conjunto de datos con clases desbalanceadas.

El *linear probing* obtiene resultados razonables, lo que sugiere que el modelo preentrenado ya contiene representaciones útiles para la identificación floral. Sin embargo, la adaptación completa mediante *full fine-tuning* permite mejorar la predicción Top-1, el rendimiento medio entre clases y la correlación global entre predicciones y etiquetas reales.

Dado que el *full fine-tuning* mejora las métricas más relevantes para este trabajo, especialmente *balanced accuracy*, F1 macro y MCC, esta estrategia se considera más adecuada para el modelo final.

#### 4.4 ANÁLISIS POR CLASE BOTÁNICA

Una vez seleccionado PlantCLEF-DINOv2 como modelo principal, se realizó un análisis detallado del rendimiento por clase botánica con el fin de identificar qué géneros vegetales resultaban más fáciles o más difíciles de reconocer. Este análisis permite ir más allá de las métricas globales y estudiar si el rendimiento del modelo se distribuye de forma homogénea entre los distintos géneros o si, por el contrario, existen clases especialmente problemáticas.

El conjunto de *test* utilizado en esta evaluación contiene 114 imágenes distribuidas en 31 géneros botánicos. Dado que el número de ejemplos por género no es uniforme, se presta especial atención al valor de *support*, es decir, al número de imágenes disponibles en test para cada clase. Las clases con muy pocos ejemplos deben interpretarse con cautela, ya que un único acierto o error puede modificar de forma notable sus métricas.

Género botánico	Support	Recall / Top-1	Top-3	Top-5	Interpretación
rhododendron	10	1.000	1.000	1.000	Reconocimiento excelente
narcissus	7	1.000	1.000	1.000	Reconocimiento excelente
tulipa	7	1.000	1.000	1.000	Reconocimiento excelente
iris	9	0.889	0.889	0.889	Rendimiento muy alto
primula	5	0.800	0.800	1.000	Buen rendimiento, con mejora en Top-5
dianthus	6	0.667	0.833	0.833	Rendimiento intermedio
rosa	22	0.545	0.591	0.591	Clase difícil y heterogénea

anemone	7	0.429	0.429	0.429	Clase especialmente problemática
---------	---	-------	-------	-------	----------------------------------

Tabla 9. Resultados por género botánico para las clases con mayor soporte

Como se observa en la Tabla 9, el rendimiento de PlantCLEF-DINOv2 varía de forma notable entre géneros botánicos. Los mejores resultados se obtienen en clases con rasgos morfológicos más distintivos y relativamente estables, como *rhododendron*, *narcissus*, *tulipa* e *iris*. En cambio, géneros como *rosa*, *anemone* y *dianthus* presentan un rendimiento inferior, a pesar de contar en algunos casos con un número apreciable de imágenes en el conjunto de prueba. Esto sugiere que la dificultad de la clasificación no depende únicamente del soporte disponible, sino también de factores visuales como la variabilidad pictórica, la ambigüedad morfológica y la similitud con otros géneros ornamentales. Por tanto, futuras ampliaciones del dataset deberían reforzar especialmente estas clases problemáticas, incorporando ejemplos más diversos y revisando posibles casos de ambigüedad en el etiquetado.

#### 4.5 ANÁLISIS POR FUENTE Y ESTILO ARTÍSTICO

Para complementar el análisis por clase botánica, se evaluó el rendimiento del modelo final PlantCLEF-DINOv2 en el conjunto de prueba en función de dos variables contextuales: la fuente de procedencia de la imagen y el estilo o movimiento artístico asociado a la obra. Este análisis permite estudiar si el rendimiento del modelo se mantiene estable entre distintos tipos de representación visual, especialmente entre ilustraciones botánicas históricas y obras pictóricas procedentes de colecciones museísticas.

El análisis se realizó sobre el conjunto de prueba, compuesto por 114 imágenes. Las imágenes procedentes de *Curtis's Botanical Magazine* se agruparon como ilustración botánica histórica, al tratarse de representaciones centradas en la descripción visual de la planta y no de obras adscritas a un movimiento pictórico convencional. Para las imágenes

procedentes de museos, se utilizaron los metadatos artísticos disponibles, en particular el movimiento artístico asociado a cada obra.

En primer lugar, se analizó el rendimiento por fuente original del *dataset*. La Tabla 10 muestra los resultados obtenidos por PlantCLEF-DINOv2 para cada fuente.

Fuente	Support	Recall / Top-1	Top-3	Top-5	Balanced accuracy	F1 macro
Curtis Botanical Magazine	56	0.768	0.768	0.786	0.729	0.659
Museo del Prado	53	0.755	0.849	0.849	0.698	0.730
Rijksmuseum	4	0.500	0.750	0.750	0.500	0.500
MET Museum	1	1.000	1.000	1.000	1.000	1.000

Tabla 10. Resultados del modelo por fuente

Como se puede observar en la Tabla 10, el modelo obtiene un rendimiento elevado tanto en las imágenes de Curtis Botanical Magazine como en las imágenes procedentes del Museo del Prado. En Curtis se alcanza una accuracy Top-1 de 0.768 y una balanced accuracy de 0.729, mientras que en el Prado se obtiene una accuracy Top-1 de 0.755 y una balanced accuracy de 0.698. La cercanía entre ambos resultados es relevante, ya que indica que el modelo no solo funciona sobre ilustraciones botánicas históricas, donde las plantas suelen aparecer de forma más aislada y morfológicamente clara, sino también sobre pintura artística, donde las flores pueden estar integradas en composiciones más complejas.

La comparación entre Curtis y Prado sugiere que el preentrenamiento botánico del modelo resulta útil en ambos contextos. En el caso de Curtis, las imágenes suelen presentar una representación más directa del motivo vegetal, lo que facilita la identificación de rasgos morfológicos. En el caso del Prado, el reto es mayor, ya que las flores aparecen en obras pictóricas con variaciones de estilo, iluminación, técnica y composición. Aun así, el modelo

mantiene un rendimiento sólido, lo que indica una capacidad razonable de generalización al dominio artístico.

Los resultados de Rijksmuseum y MET Museum deben interpretarse con cautela, ya que su soporte en test es muy reducido. Rijksmuseum cuenta únicamente con cuatro imágenes y MET Museum con una sola, por lo que sus métricas no permiten extraer conclusiones generales sobre el rendimiento del modelo en estas fuentes.

A continuación, se analizó el rendimiento en función del movimiento artístico. Las imágenes procedentes de la revista Curtis, al no tener un movimiento artístico asociado, se han dejado como “Ilustración botánica”.

Movimiento artístico	Support	Accuracy Top-1	Accuracy Top-5	Balanced Accuracy	F1 Macro	Error rate
Ilustración botánica	58	0.776	0.793	0.735	0.670	0.224
Baroque	44	0.773	0.864	0.702	0.739	0.227
Rococo	5	0.600	0.800	0.556	0.600	0.400
Rococo/Neoclassicism	2	1.000	1.000	1.000	1.000	0.000
Mannerism	1	1.000	1.000	1.000	1.000	0.000
Renaissance	1	1.000	1.000	1.000	1.000	0.000
Academic realism	1	0.000	0.000	0.000	0.000	1.000
Baroque/Neoclassicism	1	0.000	1.000	0.000	0.000	1.000
Impressionism	1	0.000	0.000	0.000	0.000	1.000

Tabla 11. Resultados del modelo por estilo artístico

Tal y como se puede observar en la Tabla 11 el modelo el modelo obtiene resultados altos tanto en ilustración botánica como en pintura barroca. En el grupo de ilustración botánica alcanza una *accuracy* Top-1 de 0.776 y una *balanced accuracy* de 0.735, lo que resulta coherente con la mayor claridad morfológica de este tipo de imágenes. Al tratarse de representaciones centradas en la planta y menor complejidad compositiva, el modelo dispone de rasgos visuales más directos para realizar la clasificación.

El Barroco presenta un rendimiento similar en *accuracy* Top-1, con 0.773, y alcanza un F1 macro de 0.739. Estos valores muestran un comportamiento consistente porque constituye el bloque pictórico con mayor soporte dentro del conjunto de test. La explicación más probable es que muchas representaciones florales barrocas aparecen en bodegones, guirnaldas o composiciones donde la flor tiene un papel visual destacado y se representa con suficiente detalle.

Por su parte, el Rococó obtiene una *accuracy* Top-1 de 0.600 y una Top-5 de 0.800. No obstante, este grupo cuenta únicamente con cinco imágenes en el conjunto de test, por lo que su impacto en el análisis global es limitado. El resto de movimientos artísticos no se analizan de forma individual debido a su escasa representación en test, con una o dos muestras por categoría, lo que impide realizar comparaciones significativas.

Estos resultados indican que el modelo no solo funciona sobre ilustraciones botánicas históricas, sino también sobre pintura artística, especialmente cuando los motivos florales aparecen con suficiente protagonismo y detalle visual. La comparación entre ilustración botánica y Barroco sugiere que el preentrenamiento botánico resulta útil en ambos contextos.

## 4.6 ANÁLISIS DE ERRORES

El análisis de errores confirma las tendencias observadas en los resultados del apartado 4.4. En el conjunto de test se registraron 28 errores Top-1, concentrados principalmente en géneros como *rosa*, *anemone* y *dianthus*. El caso de *rosa* resulta especialmente relevante, ya

que es una de las clases con mayor número de muestras, pero también una de las más variables desde el punto de vista pictórico, lo que dificulta su clasificación. Asimismo, algunas clases como *ranunculus* y *aquilegia* aparecen con frecuencia como predicciones incorrectas, lo que sugiere que pueden actuar como clases atractoras ante flores ornamentales, estilizadas o con morfología ambigua. Por tanto, estos errores son coherentes con la dificultad del dominio artístico, donde la simplificación visual, la pérdida de detalle morfológico y la similitud entre géneros florales pueden afectar a la predicción final.

La distribución completa de errores por clase real, clase predicha y pares de confusión se recoge en el Anexo III. Análisis de errores del modelo.

## Capítulo 5. CONCLUSIONES Y TRABAJOS FUTUROS

En este Trabajo Fin de Grado se ha estudiado la viabilidad de aplicar modelos de visión por ordenador preentrenados a la identificación automática de motivos florales representados en obras pictóricas e ilustraciones botánicas históricas. Para ello, se ha construido un conjunto de datos específico, se ha definido una estructura taxonómica coherente y se han evaluado distintas arquitecturas de clasificación de imágenes mediante técnicas de *transfer learning* y *fine-tuning*.

El primer objetivo del proyecto consistía en construir un *dataset* propio que combinase información botánica con imágenes artísticas. Este objetivo se ha cumplido mediante la recopilación, selección, recorte y etiquetado de imágenes procedentes del Museo del Prado, Rijksmuseum, The Metropolitan Museum of Art y *Curtis's Botanical Magazine*. El conjunto final empleado en los experimentos principales contiene 619 imágenes, organizadas en una estructura jerárquica de familia, género y especie. La tarea de clasificación se planteó a nivel de género botánico, con un total de 31 géneros considerados en los experimentos.

Una de las principales aportaciones del trabajo es precisamente la construcción de este conjunto de datos, ya que no existe una base de datos consolidada que relacione de forma directa representaciones florales en obras artísticas con etiquetas botánicas normalizadas. Además, el proceso desarrollado permite conservar tanto información taxonómica como metadatos artísticos, lo que facilita no solo el entrenamiento de modelos, sino también el análisis posterior de los resultados en función de la fuente, el estilo, el movimiento artístico o el tipo de representación.

El segundo objetivo consistía en evaluar diferentes modelos de visión por ordenador para analizar su capacidad de adaptación al dominio artístico. Para ello se compararon arquitecturas convolucionales clásicas y modernas, modelos basados en *Vision Transformers* y un modelo con preentrenamiento especializado en el dominio botánico. Los resultados muestran que PlantCLEF-DINOv2 obtiene el mejor rendimiento global, con una *accuracy*

Top-1 de 0.754, una *balanced accuracy* de 0.704 y un F1 macro de 0.644 en el conjunto de prueba. Estos resultados indican que el preentrenamiento botánico aporta una ventaja frente a modelos generalistas, especialmente en métricas equilibradas entre clases.

También se ha comprobado que la estrategia de entrenamiento influye de forma relevante en el rendimiento final. La comparación entre *linear probing* y *full fine-tuning* muestra que adaptar el modelo completo mejora la mayoría de métricas, especialmente la *balanced accuracy* y el F1 macro. Esto sugiere que, aunque el modelo preentrenado ya contiene representaciones útiles para la identificación floral, la actualización completa de sus pesos permite una mejor adaptación al dominio artístico-botánico.

El análisis por clase botánica ha permitido observar que el rendimiento del modelo no es homogéneo entre géneros. Algunas clases con rasgos visuales más distintivos, como rhododendron, narcissus, tulipa o iris, obtienen resultados muy altos. En cambio, géneros como rosa, anemone o dianthus concentran más errores, probablemente debido a su mayor variabilidad visual, a la similitud morfológica con otros géneros ornamentales y a la forma en que aparecen representados en pintura. Este resultado indica que la dificultad del problema no depende únicamente del número de imágenes por clase, sino también de la claridad morfológica y de la variabilidad pictórica de cada género.

El análisis por fuente y movimiento artístico muestra que el modelo obtiene buenos resultados tanto en ilustraciones botánicas históricas como en pintura barroca. Este resultado es relevante, ya que indica que el modelo no se limita a funcionar sobre imágenes botánicas más claras y aisladas, sino que también puede reconocer motivos florales en obras pictóricas de mayor complejidad compositiva. No obstante, las fuentes con menor representación, como Rijksmuseum y The Metropolitan Museum of Art, deben interpretarse con cautela debido al reducido número de muestras disponibles en el conjunto de prueba.

Los resultados obtenidos sugieren que la identificación automática de motivos florales en obras artísticas es viable mediante modelos preentrenados y técnicas de *fine-tuning*, al menos en el escenario experimental evaluado. El sistema desarrollado no pretende sustituir el

criterio experto en identificación botánica o análisis histórico-artístico, pero sí puede actuar como herramienta de apoyo para la exploración de colecciones digitales, la organización de imágenes y la generación de hipótesis preliminares sobre la presencia de determinados géneros vegetales en obras de arte.

A pesar de los resultados positivos, el proyecto presenta varias limitaciones. En primer lugar, el tamaño del conjunto de datos sigue siendo reducido en comparación con los conjuntos habitualmente utilizados en visión por ordenador. Además, la distribución entre clases es desigual, lo que afecta especialmente a los géneros menos representados. Aunque se aplicó un filtrado mínimo por número de imágenes y se utilizaron métricas adecuadas para datos desbalanceados, futuras ampliaciones del *dataset* permitirían mejorar la robustez del modelo.

En segundo lugar, la clasificación se ha realizado principalmente a nivel de género. Esta decisión resulta adecuada para mantener la fiabilidad del etiquetado, pero limita el grado de precisión taxonómica alcanzado. En algunas obras, distinguir especies próximas puede ser difícil incluso para un observador experto, debido a la estilización artística o a la ausencia de rasgos morfológicos suficientes. Por ello, la extensión del sistema a nivel de especie requeriría una revisión botánica más exhaustiva y un mayor número de ejemplos por clase.

Otra limitación importante es que el modelo trabaja sobre recortes previamente seleccionados, por lo que no resuelve todavía el problema completo de localizar automáticamente flores dentro de una obra. En una aplicación real sobre colecciones museísticas, sería necesario incorporar una etapa previa de detección o segmentación de motivos vegetales, capaz de identificar las regiones de interés antes de aplicar la clasificación taxonómica.

Como líneas de trabajo futuro, sería conveniente ampliar el conjunto de datos incorporando nuevas fuentes museísticas, más obras por género y una mayor variedad de periodos, técnicas y estilos artísticos. También sería útil reforzar específicamente las clases más problemáticas

detectadas en el análisis de errores, como *rosa*, *anemone* o *dianthus*, con el objetivo de mejorar la capacidad del modelo para distinguir géneros visualmente próximos.

Por último, el sistema podría evolucionar hacia una herramienta más completa para el análisis de patrimonio artístico, integrando detección automática de motivos florales, clasificación botánica y visualización de resultados por colección, periodo, autor o escuela. De este modo, el trabajo desarrollado constituye una primera aproximación exploratoria, pero abre la puerta a futuras investigaciones en la intersección entre visión por ordenador, botánica e historia del arte.

## Capítulo 6. BIBLIOGRAFÍA

- [1] E. Barba Gómez, El jardín del Prado: Un paseo botánico por las obras de los grandes maestros. Barcelona, España: Espasa, 2020.
- [2] E. Barba Gómez, El paraíso a pinceladas: Jardines en las obras de arte. Barcelona, España: Espasa, 2023.
- [3] Museo Nacional del Prado. “Un paseo botánico por el Prado”. Museo Nacional del Prado, 2024–2025. <https://www.museodelprado.es/actualidad/exposicion/un-paseo-botanico-por-el-prado/3f48df04-a1fb-d356-7ad5-56cbf5d3b2ce>.
- [4] Museo Nacional del Prado. “La botánica en el arte. Las plantas en las colecciones del Museo del Prado”. Museo Nacional del Prado, 2026. <https://www.museodelprado.es/actualidad/exposicion/la-botanica-en-el-arte/6ba69e8a-e79a-0d35-31ac-0fd7cb6415ed>.
- [5] He, K.; Zhang, X.; Ren, S.; Sun, J. “Deep residual learning for image recognition”. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [6] Tan, M.; Le, Q.V. “EfficientNet: Rethinking model scaling for convolutional neural networks”. Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 2019, pp. 6105–6114.
- [7] Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. “A ConvNet for the 2020s”. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 11966–11976.
- [8] Dosovitskiy, A. et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. International Conference on Learning Representations (ICLR), 2021.
- [9] Oquab, M. et al. “DINOv2: Learning robust visual features without supervision”. arXiv:2304.07193, 2023. <https://arxiv.org/abs/2304.07193>.

- [10] Goëau, H.; Lombardo, J.-C.; Affouard, A.; Espitalier, V.; Bonnet, P.; Joly, A. PlantCLEF 2024 pretrained models on the flora of the south western Europe based on a subset of Pl@ntNet collaborative images and a ViT base patch 14 DINOv2. Zenodo, 2024. <https://doi.org/10.5281/zenodo.10848263>.
- [11] Pan, S.J.; Yang, Q. “A survey on transfer learning”. IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, 2010, pp. 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>.
- [12] Crowley, E.J.; Zisserman, A. “In search of art”. Computer Vision – ECCV 2014 Workshops. Cham, Switzerland: Springer, 2015, pp. 54–70. [https://doi.org/10.1007/978-3-319-16178-5\\_4](https://doi.org/10.1007/978-3-319-16178-5_4).
- [13] Madhu, P.; Kosti, R.; Mührenberg, L.; Bell, P.; Maier, A.; Christlein, V. “Recognizing characters in art history using deep learning”. Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC ’19), Nice, France, 2019, pp. 15–22. <https://doi.org/10.1145/3347317.3357242>
- [14] Joly, A. et al. “A look inside the Pl@ntNet experience: The good, the bias and the hope”. Multimedia Systems, vol. 22, no. 6, 2016, pp. 751–766. <https://doi.org/10.1007/s00530-015-0462-9>
- [15] Müller, R.; Kornblith, S.; Hinton, G. “When does label smoothing help?”. Advances in Neural Information Processing Systems, vol. 32, 2019.
- [16] Paszke, A. et al. “PyTorch: An imperative style, high-performance deep learning library”. Advances in Neural Information Processing Systems, vol. 32, 2019, pp. 8024–8035.
- [17] Wightman, R. PyTorch Image Models. Zenodo, 2019. <https://doi.org/10.5281/zenodo.4414861>.
- [18] Pedregosa, F. et al. “Scikit-learn: Machine learning in Python”. Journal of Machine Learning Research, vol. 12, 2011, pp. 2825–2830.
- [19] The Metropolitan Museum of Art. “The Metropolitan Museum of Art Collection API”. <https://metmuseum.github.io/>.

- [20] Rijksmuseum. “Rijksmuseum Data Services: Documentation”.  
<https://data.rijksmuseum.nl/docs/>
- [21] Biodiversity Heritage Library. “Curtis’s Botanical Magazine”. Bibliography no. 706.  
<https://www.biodiversitylibrary.org/bibliography/706>
- [22] Wikimedia Commons. “Commons:API/MediaWiki”.  
<https://commons.wikimedia.org/wiki/Commons:API/MediaWiki>
- [23] Mukherjee, S. “The Annotated ResNet-50”. Towards Data Science.  
<https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758/>.
- [24] Erdoğan, A. “ConvNeXt — Next Generation of Convolutional Networks”. Medium.  
<https://medium.com/@atakanerdogan305/convnext-next-generation-of-convolutional-networks-325607a08c46>.
- [25] GeeksforGeeks. “Vision Transformer (ViT) Architecture”. GeeksforGeeks.  
<https://www.geeksforgeeks.org/deep-learning/vision-transformer-vit-architecture/>

# ANEXO I. VOCABULARIO BOTÁNICO Y NORMALIZACIÓN TAXONÓMICA

Este anexo recoge los conceptos taxonómicos utilizados en la construcción del conjunto de datos, así como los criterios empleados para normalizar las etiquetas botánicas y organizar las imágenes en una estructura jerárquica de familia, género y especie. Su objetivo es complementar la explicación metodológica incluida en la memoria principal y facilitar la interpretación del vocabulario botánico utilizado en el proyecto.

## ***1.1 CONCEPTOS TAXONÓMICOS EMPLEADOS***

Para la organización del conjunto de datos se utilizaron varios conceptos básicos de taxonomía vegetal. Los más relevantes para este trabajo son los siguientes:

**Taxonomía:** ciencia encargada de clasificar y nombrar los organismos. En este proyecto se utiliza para organizar las imágenes florales en distintos niveles de clasificación.

**Familia:** categoría taxonómica que agrupa géneros con características comunes. Por ejemplo, la familia Rosaceae incluye géneros como *Rosa* o *Prunus*.

**Género:** conjunto de especies estrechamente emparentadas. En este trabajo, el género constituye el nivel principal utilizado para la clasificación experimental, ya que ofrece un equilibrio adecuado entre precisión taxonómica y fiabilidad visual.

**Especie:** unidad básica de clasificación biológica. En botánica, se expresa normalmente mediante nomenclatura binomial, formada por el género y el epíteto específico.

**Nomenclatura binomial:** sistema de denominación científica formado por dos términos: el nombre del género, escrito con mayúscula inicial, y el epíteto específico, escrito en

minúscula. Por ejemplo, en *Rosa canina*, *Rosa* corresponde al género y *canina* al epíteto específico.

**Epíteto específico:** segunda parte del nombre científico de una especie. Permite distinguir especies dentro de un mismo género.

**Sinonimia taxonómica:** existencia de diferentes nombres científicos que se han utilizado históricamente para referirse a un mismo taxón.

**Morfología floral:** estudio de la forma y estructura de las flores. En este proyecto resulta especialmente relevante porque la identificación visual de los motivos florales depende de rasgos como la forma de los pétalos, la simetría, la disposición floral, el color o la presencia de hojas asociadas.

## ***1.2 CONCEPTOS TAXONÓMICOS EMPLEADOS***

Durante la construcción del conjunto de datos fue necesario normalizar las etiquetas botánicas procedentes de distintas fuentes. Las denominaciones originales podían presentar diferencias de formato, variantes ortográficas, nombres comunes, errores de transcripción o distintos niveles de especificidad taxonómica.

Para mantener una estructura coherente, se definió un vocabulario controlado basado en tres niveles: familia, género y especie. Cada imagen del conjunto de datos conserva información asociada a estos tres niveles, aunque la clasificación experimental principal se realizó a nivel de género botánico.

Las etiquetas se normalizaron siguiendo los siguientes criterios:

- Se transformaron los nombres a minúsculas.
- Se sustituyeron espacios y guiones por guiones bajos.

- Se corrigieron errores ortográficos detectados en los nombres científicos.
- Se agruparon variantes equivalentes bajo una única etiqueta normalizada.
- Cuando la identificación a nivel de especie no era suficientemente fiable, la muestra se asignó al género correspondiente.
- Las imágenes sin rasgos morfológicos suficientes para una identificación fiable fueron descartadas del conjunto principal.

Por ejemplo, etiquetas como *Rosa centifolia*, *rosa-centifolia* o variantes con numeraciones añadidas al nombre del archivo se unificaron bajo la forma normalizada *rosa\_centifolia*. Del mismo modo, cuando una imagen permitía reconocer el género, pero no distinguir con seguridad la especie, se utilizó una etiqueta de género, como *rosa* o *tulipa*.

Este proceso permitió reducir la variabilidad artificial entre etiquetas y evitar que una misma clase apareciera representada con nombres distintos dentro del conjunto de datos.

### ***I.3. VOCABULARIO BOTÁNICO UTILIZADO***

El vocabulario botánico utilizado en el proyecto se organizó mediante una estructura jerárquica de familia, género y especie. Esta estructura permitió conservar información taxonómica detallada y, al mismo tiempo, formular la tarea principal como un problema de clasificación a nivel de género.

La Tabla 12 recoge el vocabulario taxonómico completo utilizado en el proyecto. Cada etiqueta normalizada se asocia a su género y familia correspondientes, de acuerdo con la estructura empleada para organizar el conjunto de datos.

<b>Familia</b>	<b>Género</b>	<b>Etiqueta normalizada</b>
Rosaceae	rosa	rosa

Rosaceae	rosa	rosa_alba_maxima
Rosaceae	rosa	rosa_albasemiplena
Rosaceae	rosa	rosa_centifolia
Rosaceae	rosa	rosa_eglanteria
Rosaceae	rosa	rosa_foetida
Rosaceae	rosa	rosa_gallica
Rosaceae	rosa	rosa_canina
Rosaceae	rosa	rosa_gigantea
Rosaceae	rosa	rosa_gros_choux_dhollande
Rosaceae	rosa	rosa_sertata
Rosaceae	rosa	rosa_setipoda
Rosaceae	rosa	rosa_semperflorens
Rosaceae	rosa	rosa_sp
Rosaceae	prunus	prunus
Rosaceae	prunus	prunus_dulcis
Rosaceae	prunus	prunus_mume
Liliaceae	lilium	lilium
Liliaceae	lilium	lilium_candidum

Liliaceae	lilium	lilium_bulbiferum
Liliaceae	lilium	lilium_chalcedonicum
Liliaceae	lilium	lilium_henryi
Liliaceae	lilium	lilium_pumilum
Liliaceae	lilium	lilium_sp
Liliaceae	tulipa	tulipa
Liliaceae	tulipa	tulipa_sp
Liliaceae	fritillaria	fritillaria
Liliaceae	fritillaria	fritillaria_imperialis
Liliaceae	fritillaria	fritillaria_sp
Iridaceae	iris	iris
Iridaceae	iris	iris_germanica
Iridaceae	iris	iris_pseudacorus
Iridaceae	iris	iris_xiphium
Iridaceae	iris	iris_barnumiae
Iridaceae	iris	iris_bracteata
Iridaceae	iris	iris_caroliniana
Iridaceae	iris	iris_chrysographes

Iridaceae	iris	iris_clarkei
Iridaceae	iris	iris_cretensis
Iridaceae	iris	iris_florentina
Iridaceae	iris	iris_hoogiana
Iridaceae	iris	iris_iberica
Iridaceae	iris	iris_kashmiriana
Iridaceae	iris	iris_meda
Iridaceae	iris	iris_mellita
Iridaceae	iris	iris_minuta
Iridaceae	iris	iris_pallida
Iridaceae	iris	iris_persica
Iridaceae	iris	iris_pumila
Iridaceae	iris	iris_reichenbachii
Iridaceae	iris	iris_scariosa
Iridaceae	iris	iris_spuria
Iridaceae	iris	iris_urumovii
Iridaceae	iris	iris_variegata
Iridaceae	iris	iris_versicolor

Iridaceae	iris	iris_wilsonii
Iridaceae	iris	iris_sp
Iridaceae	gladiolus	gladiolus
Iridaceae	gladiolus	gladiolus_sp
Amaryllidaceae	narcissus	narcissus
Amaryllidaceae	narcissus	narcissus_jonquilla
Amaryllidaceae	narcissus	narcissus_poeticus
Amaryllidaceae	narcissus	narcissus_tazetta
Amaryllidaceae	narcissus	narcissus_cyclamineus
Amaryllidaceae	narcissus	narcissus_italicus
Amaryllidaceae	narcissus	narcissus_minor
Amaryllidaceae	narcissus	narcissus_triandrus
Amaryllidaceae	narcissus	narcissus_pseudonarcissus
Amaryllidaceae	narcissus	narcissus_sp
Amaryllidaceae	galanthus	galanthus
Amaryllidaceae	galanthus	galanthus_nivalis
Asparagaceae	hyacinthus	hyacinthus
Asparagaceae	hyacinthus	hyacinthus_orientalis

Asparagaceae	hyacinthus	hyacinthus_sp
Asparagaceae	convallaria	convallaria
Asparagaceae	convallaria	convallaria_majalis
Oleaceae	syringa	syringa
Oleaceae	syringa	syringa_vulgaris
Asteraceae	chrysanthemum	chrysanthemum
Asteraceae	chrysanthemum	chrysanthemum_leucanthemum
Asteraceae	chrysanthemum	chrysanthemum_sp
Asteraceae	leucanthemum	leucanthemum
Asteraceae	leucanthemum	leucanthemum_maximum
Asteraceae	calendula	calendula
Asteraceae	calendula	calendula_officinalis
Asteraceae	centaurea	centaurea
Asteraceae	centaurea	centaurea_cyanus
Asteraceae	helianthus	helianthus
Asteraceae	helianthus	helianthus_annuus
Asteraceae	tagetes	tagetes
Asteraceae	tagetes	tagetes_erecta

Asteraceae	chamaemelum	chamaemelum
Asteraceae	chamaemelum	chamaemelum_nobile
Asteraceae	gerbera	gerbera
Asteraceae	gerbera	gerbera_jamesonii
Asteraceae	taraxacum	taraxacum
Asteraceae	taraxacum	taraxacum_officinale
Boraginaceae	anchusa	anchusa
Boraginaceae	anchusa	anchusa_azurea
Boraginaceae	borago	borago
Boraginaceae	borago	borago_officinalis
Brassicaceae	matthiola	matthiola
Brassicaceae	matthiola	matthiola_incana
Brassicaceae	erysimum	erysimum
Brassicaceae	erysimum	erysimum_cheiri
Caprifoliaceae	lonicera	lonicera
Caprifoliaceae	lonicera	lonicera_sp
Caprifoliaceae	viburnum	viburnum
Caprifoliaceae	viburnum	viburnum_opulus

Caprifoliaceae	viburnum	viburnum_tinus
Caprifoliaceae	viburnum	viburnum_betulifolium
Caprifoliaceae	viburnum	viburnum_henryi
Caprifoliaceae	viburnum	viburnum_rhytidophyllum
Caryophyllaceae	dianthus	dianthus
Caryophyllaceae	dianthus	dianthus_caryophyllus
Caryophyllaceae	dianthus	dianthus_chinensis
Caryophyllaceae	dianthus	dianthus_deltoides
Caryophyllaceae	dianthus	dianthus_sp
Convolvulaceae	convolvulaceae	convolvulaceae
Convolvulaceae	convolvulaceae	convolvulus
Convolvulaceae	ipomoea	ipomoea
Convolvulaceae	ipomoea	ipomoea_sp
Ericaceae	rhododendron	rhododendron
Ericaceae	rhododendron	rhododendron_periclymenoides
Ericaceae	rhododendron	rhododendron_sp
Fabaceae	retama	retama
Fabaceae	retama	retama_sphaerocarpa

Gentianaceae	gentiana	gentiana
Gentianaceae	gentiana	gentiana_sp
Lythraceae	punica	punica
Lythraceae	punica	punica_granatum
Malvaceae	alcea	alcea
Malvaceae	alcea	alcea_rosea
Theaceae	camellia	camellia
Theaceae	camellia	camellia_sp
Malvaceae	hibiscus	hibiscus
Malvaceae	hibiscus	hibiscus_sp
Malvaceae	hibiscus	hibiscus_syriacus
Malvaceae	malva	malva
Malvaceae	malva	malva_sylvestris
Oleaceae	jasminum	jasminum
Oleaceae	jasminum	jasminum_grandiflorum
Apocynaceae	nerium	nerium
Apocynaceae	nerium	nerium_oleander
Plantaginaceae	digitalis	digitalis

Plantaginaceae	digitalis	digitalis_purpurea
Plantaginaceae	digitalis	digitalis_sp
Paeoniaceae	paeonia	paeonia_suffruticosa
Paeoniaceae	paeonia	paeonia_sp
Paeoniaceae	paeonia	paeonia
Papaveraceae	papaver	papaver_somniferum
Papaveraceae	papaver	papaver
Papaveraceae	papaver	papaver_rhoeas
Papaveraceae	papaver	papaver_sp
Polemoniaceae	polemonium	polemonium
Polemoniaceae	polemonium	polemonium_caeruleum
Primulaceae	primula	primula
Primulaceae	primula	primula_vulgaris
Primulaceae	primula	primula_auricula
Ranunculaceae	anemone	anemone_coronaria
Ranunculaceae	anemone	anemone_hortensis
Ranunculaceae	anemone	anemone_sp
Ranunculaceae	anemone	anemone

Ranunculaceae	ranunculus	ranunculus
Ranunculaceae	ranunculus	ranunculus_asiaticus
Ranunculaceae	ranunculus	ranunculus_acris
Ranunculaceae	ranunculus	ranunculus_pedatus
Ranunculaceae	aquilegia	aquilegia
Ranunculaceae	aquilegia	aquilegia_vulgaris
Ranunculaceae	nigella	nigella
Ranunculaceae	nigella	nigella_damascena
Ranunculaceae	consolida	consolida
Ranunculaceae	consolida	consolida_ajacis
Ranunculaceae	delphinium	delphinium
Ranunculaceae	delphinium	delphinium_sp
Ranunculaceae	helleborus	helleborus
Ranunculaceae	helleborus	helleborus_sp
Rutaceae	citrus	citrus
Rutaceae	citrus	citrus_aurantium
Tropaeolaceae	tropaeolum	tropaeolum
Tropaeolaceae	tropaeolum	tropaeolum_majus

Amaranthaceae	amaranthus	amaranthus
Amaranthaceae	amaranthus	amaranthus_caudatus
Araceae	zantedeschia	zantedeschia
Araceae	zantedeschia	zantedeschia_aethiopica
Violaceae	viola	viola
Violaceae	viola	viola_tricolor
Magnoliaceae	magnolia	magnolia
Magnoliaceae	magnolia	magnolia_sp

*Tabla 12. Vocabulario botánico utilizado para la organización taxonómica del conjunto de datos*

En los experimentos principales se utilizó el género como etiqueta de clasificación, mientras que la familia y la especie se conservaron como información complementaria para la organización del dataset y el análisis posterior.

## ANEXO II. ARQUITECTURA DE LOS MODELOS

### EVALUADOS

Este anexo recoge una descripción ampliada de las arquitecturas de visión por ordenador evaluadas en el proyecto. En concreto, se presentan las características principales de ResNet-50, EfficientNet-B3, ConvNeXt-Small, ViT-Base y PlantCLEF-DINOv2, así como el motivo de su inclusión en la comparación experimental. Su objetivo es complementar la explicación metodológica incluida en la memoria principal y facilitar la interpretación de las diferencias entre modelos convolucionales, modelos basados en transformadores visuales y modelos con preentrenamiento específico en el dominio botánico.

#### II.1. RESNET-50

Esta arquitectura se basa en el uso de bloques residuales, que introducen conexiones de salto entre capas para facilitar el entrenamiento de las redes neuronales profundas (*Deep neural networks*). Estas conexiones permiten que la información fluya de manera más estable a lo largo de la red y soluciona problemas relacionados con el desvanecimiento del gradiente.

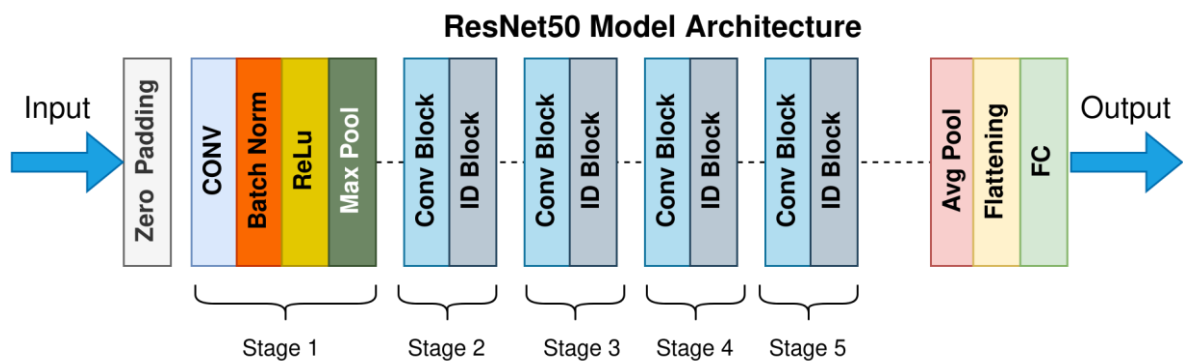


Figura 11 Arquitectura general de ResNet-50. Fuente: adaptado de Towards Data Science [23]

La Figura 11 muestra la arquitectura general de ResNet-50. Como se observa, el modelo está organizado en diferentes etapas que alternan bloques convolucionales y bloques residuales, lo que permite aumentar la profundidad de la red manteniendo un entrenamiento estable [5].

En este proyecto, la arquitectura ResNet-50 se evalúa porque las flores representadas en obras pictóricas contienen rasgos visuales locales que pueden ser capturados eficazmente mediante capas convolucionales. Por ejemplo, los bordes de los pétalos, la disposición de las hojas, los cambios de color o las texturas de la pincelada pueden actuar como patrones discriminativos para distinguir entre diferentes géneros botánicos.

Las primeras capas de la red convolucional tienden a aprender características básicas, como bordes, orientaciones, contrastes y pequeñas texturas. Las capas intermedias combinan estos patrones para representar formas más complejas, mientras que las capas profundas integran la información en descriptores visuales de mayor nivel. Esta jerarquía de características resulta adecuada para analizar imágenes artísticas, donde la identificación de la flor puede depender tanto de detalles morfológicos como de patrones cromáticos o compositivos.

Por este motivo, ResNet-50 se utiliza como modelo convolucional de referencia dentro de la comparación experimental. Aunque se trata de una arquitectura generalista, su uso permite establecer una base sólida frente a modelos más recientes o especializados, como ConvNeXt, ViT o PlantCLEF-DINOv2.

## ***II.2. EFFICIENTNET-B3***

EfficientNet-B3 se incorporó por su diseño orientado a optimizar la relación entre rendimiento y eficiencia computacional. La familia EfficientNet propone un escalado compuesto de la arquitectura, ajustando de forma equilibrada la profundidad de la red, la anchura de los canales y la resolución de entrada, en lugar de mejorar el modelo aumentando únicamente una de estas dimensiones [6].

Desde el punto de vista arquitectónico, este modelo está formado por una sucesión de bloques convolucionales optimizados, conocidos como MBConv (*Mobile Inverted Bottleneck Convolution*). Estos bloques utilizan convoluciones separables en profundidad (*depthwise separable convolutions*), que reducen el número de operaciones necesarias al separar la convolución espacial de la combinación entre canales. Además, incorporan mecanismos de atención sobre los canales, como los bloques *squeeze-and-excitation*, que permiten ponderar la importancia relativa de distintas características visuales. Esta combinación permite construir una red con buena capacidad de representación manteniendo un coste computacional controlado.

Estas características hacen que EfficientNet-B3 sea un modelo interesante para este proyecto, especialmente al trabajar con un conjunto de datos relativamente limitado. En tareas con pocas imágenes por clase, resulta importante utilizar modelos con buena capacidad de generalización y un número de parámetros razonable, reduciendo el riesgo de sobreajuste frente a arquitecturas excesivamente grandes.

En el caso de la identificación de flores en pintura, EfficientNet-B3 puede beneficiarse de su capacidad para capturar detalles visuales a diferentes escalas. Las flores aparecen en recortes de tamaño variable y presentan diferencias importantes de color, forma, textura y composición. Por ello, este modelo constituye una alternativa adecuada para analizar la clasificación de motivos florales artísticos, combinando capacidad de representación y eficiencia computacional.

### ***II.3. CONVNEXT-SMALL***

ConvNeXt-Small se incluyó en la comparación como arquitectura convolucional moderna. Aunque mantiene la base de las redes convolucionales, su diseño incorpora modificaciones inspiradas en la evolución reciente de los modelos de visión por ordenador, especialmente en los *Vision Transformers* [7]. El objetivo de ConvNeXt es actualizar el diseño tradicional de las CNN para aumentar su capacidad representacional sin abandonar el procesamiento

convolucional.

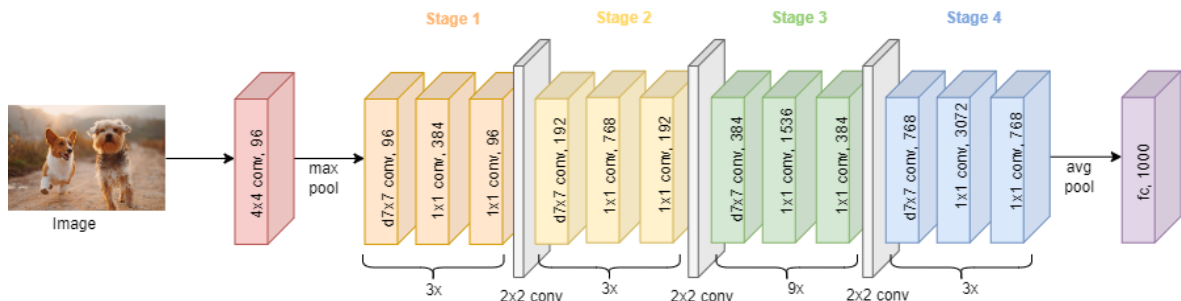


Figura 12. Esquema general de la arquitectura ConvNeXt, organizada en etapas jerárquicas que reducen progresivamente la resolución espacial y aumentan el número de canales. Fuente: adaptado de Medium [24]

La Figura 12 muestra una representación esquemática de su arquitectura general. El modelo se organiza en varias etapas sucesivas: entre una etapa y la siguiente se reduce la resolución espacial de la imagen y aumenta el número de canales. De este modo, la red pasa progresivamente de representaciones de bajo nivel, como bordes, texturas y patrones locales, a descriptores visuales más abstractos. Esta estructura jerárquica permite procesar la imagen a diferentes escalas, una característica especialmente útil en tareas de clasificación visual.

A nivel arquitectónico, este modelo introduce varios cambios respecto a redes convolucionales clásicas como ResNet-50. Entre ellos destacan el uso de bloques convolucionales modernizados, convoluciones de mayor tamaño para ampliar el campo receptivo y una organización interna más próxima a la de los modelos basados en transformadores. Además, incorpora mecanismos de normalización y capas lineales que favorecen la estabilidad del entrenamiento y mejoran la capacidad de aprendizaje.

En el contexto de este trabajo, ConvNeXt-Small permite evaluar si una CNN modernizada se adapta mejor al dominio pictórico que una arquitectura convolucional clásica. Esta comparación resulta relevante porque las flores representadas en pintura pueden presentar una gran variabilidad de escala, textura, color y grado de detalle. Por ello, una arquitectura capaz de combinar patrones locales con representaciones más abstractas puede resultar adecuada para reconocer motivos florales en imágenes artísticas.

Además, ConvNeXt-Small ofrece un equilibrio razonable entre capacidad del modelo y coste computacional. Esto es importante en un problema con un conjunto de datos limitado y riesgo de sobreajuste. Su evaluación permite comparar el comportamiento de una arquitectura convolucional actual frente a modelos más clásicos, como ResNet-50, y frente a enfoques basados en transformadores, como ViT o DINOv2.

#### ***II.4. ViT-BASE***

ViT-Base se incorporó al estudio para evaluar el comportamiento de una arquitectura basada en *Vision Transformers* frente a los modelos convolucionales. A diferencia de las CNN, que procesan la imagen mediante filtros locales aplicados sobre regiones pequeñas, los *Vision Transformers* dividen la imagen en parches de tamaño fijo. Cada parche se transforma en un vector mediante una proyección lineal y posteriormente se procesa mediante bloques de atención [8].

La Figura 13 muestra el funcionamiento general de un *Vision Transformer*. La imagen de entrada se divide en una secuencia de parches, a los que se añade información posicional para conservar la estructura espacial de la imagen. Esta secuencia se introduce en un codificador *transformer*, formado por capas de normalización, mecanismos de atención multi-cabeza y redes MLP. Finalmente, la representación obtenida se utiliza para realizar la clasificación mediante una cabeza final.

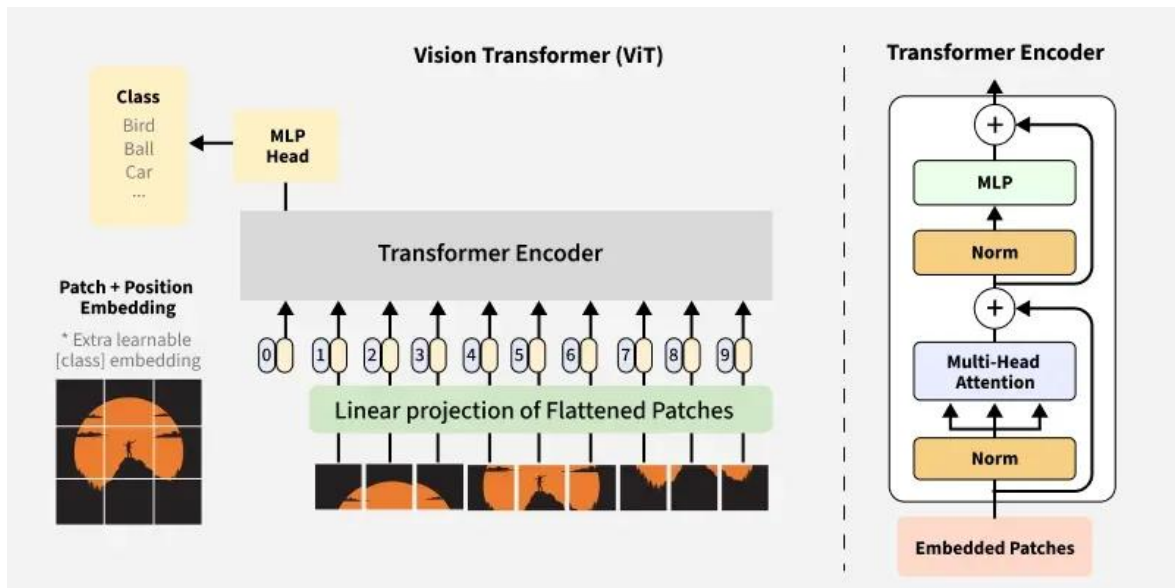


Figura 13 Esquema general de la arquitectura Vision Transformer, basada en la división de la imagen en parches y el uso de mecanismos de atención para modelar relaciones globales entre regiones de la imagen.

Fuente: adaptado de GeeksforGeeks [25]

La principal diferencia respecto a una CNN es la forma en que se modela la información visual. En una red convolucional, el aprendizaje se basa en filtros locales que capturan patrones próximos en el espacio, como bordes, texturas o formas parciales. En cambio, un *Vision Transformer* puede establecer relaciones entre parches alejados de la imagen mediante el mecanismo de atención. Esto permite capturar dependencias globales, como la forma completa del motivo vegetal o la relación entre distintas partes de la flor.

En el contexto de este trabajo, esta característica resulta relevante porque las flores representadas en pintura pueden mantener una estructura global reconocible incluso cuando algunos detalles locales aparecen simplificados, estilizados o alterados por la técnica artística. La identificación puede depender no solo de rasgos concretos, como el contorno de los pétalos, sino también de la disposición general de la flor, su simetría, su color dominante o la relación entre distintas partes del motivo floral.

## II.5. DINOv2 Y PLANTCLEF-DINOv2

DINOv2 es un modelo fundacional de visión por ordenador desarrollado por Meta AI y basado en aprendizaje auto supervisado. A diferencia de los modelos entrenados directamente para una tarea cerrada de clasificación, DINOv2 aprende representaciones visuales generales a partir de grandes colecciones de imágenes sin depender de etiquetas manuales [9]. Estas representaciones pueden reutilizarse posteriormente en distintas tareas, como clasificación, segmentación, recuperación de imágenes o estimación de profundidad.

Desde el punto de vista arquitectónico, DINOv2 se basa en *Vision Transformers*. La imagen de entrada se divide en parches de tamaño fijo, que se transforman en vectores y se procesan mediante bloques de atención. Además, el modelo utiliza un *token* global de clasificación, habitualmente denominado [CLS], que resume la información de la imagen y puede emplearse como representación visual para tareas posteriores. Por tanto, su diferencia principal respecto a un *ViT* convencional no reside tanto en la estructura de inferencia, sino en el procedimiento de preentrenamiento auto supervisado a gran escala.

En el ámbito botánico, DINOv2 ha sido utilizado como base para modelos de identificación vegetal, entre ellos enfoques desarrollados en el contexto de PlantCLEF [10]. Esta línea resulta especialmente relevante para el presente proyecto, ya que permite combinar una arquitectura moderna de visión con un preentrenamiento más próximo a la morfología vegetal. Frente a un modelo generalista, un modelo con pesos ajustados al dominio botánico puede haber aprendido representaciones más útiles para reconocer rasgos como la forma de las flores, la disposición de los pétalos, o la estructura general de la planta.

En este proyecto se utilizó PlantCLEF-DINOv2 como modelo con preentrenamiento especializado en el dominio botánico. Este modelo parte de una arquitectura basada en DINOv2, pero incorpora pesos orientados a la identificación de especies vegetales, por lo que resulta especialmente adecuado para una tarea centrada en motivos florales. La diferencia principal respecto a un modelo visual generalista no reside tanto en la estructura

arquitectónica, sino en el conocimiento previo incorporado durante el preentrenamiento, más próximo a la morfología de plantas reales.

En la implementación, se empleó su variante con tamaño de entrada de  $518 \times 518$  píxeles. Al igual que en el resto de arquitecturas evaluadas, la cabeza de clasificación final se sustituyó por una nueva capa adaptada al número de clases botánicas del *dataset*. Esta configuración permitió aprovechar las representaciones aprendidas previamente por el modelo y ajustarlas mediante *fine-tuning* a la identificación de flores representadas en obras pictóricas.

La elección de PlantCLEF-DINOv2 resulta útil para evaluar si las representaciones aprendidas por un modelo con preentrenamiento botánico pueden adaptarse a la identificación de motivos florales en un contexto visual alejado de la fotografía naturalista convencional.

## ANEXO III. ANÁLISIS DE ERRORES DEL MODELO

En este anexo se recoge el análisis detallado de los errores cometidos por el modelo final PlantCLEF-DINOv2 sobre el conjunto de prueba. Mientras que en la memoria principal se presentan las conclusiones más relevantes del análisis de errores, este anexo incluye las tablas completas utilizadas para identificar qué géneros botánicos concentran más fallos, qué clases aparecen con mayor frecuencia como predicciones incorrectas y qué pares de confusión se repiten con mayor frecuencia.

El análisis se realizó a partir de las predicciones Top-1 del modelo sobre el conjunto de test. Un error se contabiliza cuando la clase predicha por el modelo no coincide con la etiqueta real asociada a la imagen. Estos resultados permiten interpretar las limitaciones del sistema desde el punto de vista visual y botánico, especialmente en géneros con alta variabilidad pictórica o morfologías similares.

### III.1. ERRORES POR CLASE

En primer lugar, se agruparon los errores según la clase real. Como se observa en la Tabla 13, *rosa* es el género que concentra más fallos, con 10 errores, seguido de *anemone*, con 4. Esta distribución indica que las clases más problemáticas no son necesariamente las menos representadas, sino aquellas que presentan mayor variabilidad visual o una representación pictórica más ambigua.

Clase real	Número de errores
rosa	10
anemone	4
dianthus	2
viburnum	2
centaurea	1

syringa	1
iris	1
primula	1
hibiscus	1
aquilegia	1
nigella	1
ranunculus	1
viola	1

Tabla 13. Número de errores Top-1 por clase real en el conjunto de test

### III.2. CLASES PREDICHAS INCORRECTAMENTE

También se analizó qué clases aparecían con mayor frecuencia como predicciones incorrectas, como se puede observar en la Tabla 14. En este caso, *ranunculus* destaca como la etiqueta errónea más frecuente, con 7 casos, seguida de *aquilegia*, con 4. Esto sugiere que estas clases actúan como focos de atracción para determinadas flores ornamentales o de morfología compleja, especialmente cuando la representación artística simplifica o estiliza los rasgos florales.

Clase predicha incorrectamente	Número de veces
ranunculus	7
aquilegia	4
prunus	2
hyacinthus	2
viburnum	1
rosa	1
narcissus	1
anthurium	1
jasminum	1
ipomoea	1

papaver	1
paeonia	1
hibiscus	1
arum	1
anemone	1
fritillaria	1

Tabla 14. Clases predichas incorrectamente con mayor frecuencia

### **III.3. PARES DE CONFUSIÓN REPETIDOS**

Por último, se revisaron los pares de confusión repetidos. La mayoría de los errores aparecen una sola vez, por lo que no todos los fallos forman patrones recurrentes claros. Los únicos pares repetidos fueron *dianthus* clasificado como *ranunculus* y *rosa* clasificada como *ranunculus*, ambos con dos casos. Estos errores pueden relacionarse con la similitud visual entre flores ornamentales con pétalos numerosos, formas redondeadas o contornos estilizados.

### **III.4. INTERPRETACIÓN GENERAL**

El análisis detallado de errores complementa el análisis por clase botánica presentado en la memoria principal. En conjunto, los resultados muestran que los fallos del modelo se concentran especialmente en *rosa*, *anemone* y *dianthus*, géneros que presentan mayor variabilidad pictórica o similitud morfológica con otras flores ornamentales.

Asimismo, la aparición frecuente de *ranunculus* y *aquilegia* como predicciones incorrectas sugiere que el modelo tiende a asociar ciertas flores estilizadas, radiales o con pétalos numerosos con estas clases. Estos resultados refuerzan la necesidad de ampliar el conjunto de datos en las clases más problemáticas y de revisar con especial atención los casos de mayor ambigüedad visual o botánica.