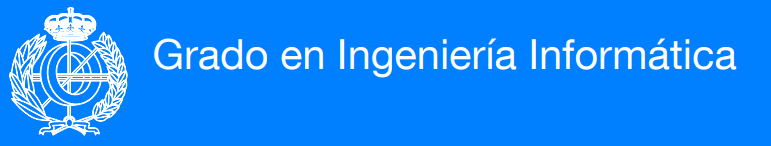
A blue sign with white text

Description automatically generated

Reconocimiento visual de aves con Deep Learning





Trabajo Fin de Grado

Autor: Juan García Martínez

Tutor/es:

José García Rodríguez

Esther Sebastián González

02/2024

2024

Contenido

[Palabras clave 3](#_Toc167614879)

[1. Introducción 3](#_Toc167614880)

[1.1 Justificación 3](#_Toc167614881)

[1.2 Estructura del TFG 3](#_Toc167614882)

[2. Objetivos 4](#_Toc167614883)

[3. Marco teórico 4](#_Toc167614884)

[3.1 Introducción a la inteligencia artificial 4](#_Toc167614885)

[3.2 Machine Learning 6](#_Toc167614886)

[3.3 Deep Learning 7](#_Toc167614887)

[3.4 Transfer Learning 8](#_Toc167614888)

[3.4.1 Transfer Learning en Deep Learning 8](#_Toc167614889)

[3.4.2 Transfer Learning en Machine Learning 8](#_Toc167614890)

[3.5 Redes Neuronales Artificiales 9](#_Toc167614891)

[3.5.1 Redes Neuronales Convolucionales 9](#_Toc167614892)

[3.5.2 Redes Neuronales Recurrentes (RNN) 11](#_Toc167614893)

[3.5.3 Redes Generativas Adversarias (GAN) 12](#_Toc167614894)

[4. Estado del arte 13](#_Toc167614895)

[5. Metodología 13](#_Toc167614896)

[6. Resultados 14](#_Toc167614897)

[7. Conclusión 14](#_Toc167614898)

[Bibliografía 14](#_Toc167614899)

[Referencias 16](#_Toc167614900)

[Lista de Acrónimos y Abreviaturas 16](#_Toc167614901)

# Palabras clave

Inteligencia artificial, Machine Learning, Deep Learning, Tranfer Learning, Redes neuronales Convolucionales, dataset, aves, arquitectura.

# 1. Introducción

En este proyecto vamos a estudiar el uso de la Inteligencia Artificial y las tecnologías para la detección de imágenes automáticamente, concretamente ente imágenes de aves.

En este apartado veremos la motivación que hizo interesarme por esta propuesta y los objetivos planteados para completarla.

## 1.1 Justificación

La elección de esta propuesta es mayoritariamente una motivación personal ya que me fascinan las aves, conocer su hábitat, su comportamiento y la conservación de las especies.

Considero que los parques naturales representan un escenario fundamental en la gestión para la sostenibilidad de la biodiversidad a largo plazo. Son espacios donde se concentra la protección de ecosistemas frágiles, la investigación científica y la educación ambiental, desempeñando un papel fundamental en la preservación de la diversidad biológica y en el desarrollo de prácticas sostenibles.

La decisión para contemplar este tema en el Trabajo Fin de Grado no se limita a satisfacer mis intereses personales, sino que también para ampliar mis conocimientos sobre un área que me interesa y en la que todavía tengo mucho que aprender.

## 1.2 Estructura del TFG

A continuación, se detallará la estructura del TFG y las tareas que se deben de realizar para su desarrollo:

* Objetivos: Se definirán los objetivos que se pretende alcanzar mediante el desarrollo del TFG
* Marco teórico: Se explicarán las bases y el contexto de la rama a la que pertenece el TFG, como la Inteligencia Artificial (IA), Machine Learning (ML), Deep Learning (DL), Redes Neuronales Artificiales (ANN) y las Arquitecturas.
* **Estado del arte:** Se mostrarán las tecnologías empleadas para el desarrollo de este TFG.
* **Metodología:** Se mostrarán las herramientas empleadas en este TFG.
* **Resultados:** Se analizarán los resultados con las diferentes pruebas sobre sus arquitecturas y parámetros.
* **Conclusión:** Se realizará un análisis de los resultados obtenidos para extraer conclusiones y determinar si se han alcanzado los objetivos propuestos.

# 2. Objetivos

El objetivo principal de este TFG es realizar un sistema capaz de reconocer y clasificar visualmente aves utilizando técnicas de Deep Learning. Este sistema podrá identificar automáticamente las aves mediante imágenes y para ello se empleará un conjunto de imágenes.

Además, se investigarán los procesos, arquitecturas, modelos y tecnologías más empleadas por la comunidad científica para el reconocimiento visual de aves. El objetivo final es desarrollar un modelo ligero y eficiente que puedan ser implementados en dispositivos como Raspberri Pi o Jetson Nano, permitiendo realizar un seguimiento continuo y automatizado de las aves y su comportamiento en los humedales de la provincia de Alicante.

# 3. Marco teórico

En este capítulo se introducirán las bases teóricas para el posterior entendimiento del proyecto, incluyendo las tecnologías, arquitecturas y herramientas de la Inteligencia artificial que se emplean hoy en día para el reconocimiento visual de imágenes, concretamente de aves.

## 3.1 Introducción a la inteligencia artificial

La **inteligencia artificial (IA)** es un campo de la informática donde su principal objetivo es la creación de sistemas que puedan realizar labores que normalmente requieren de la inteligencia humana, como el razonamiento, percepción y el aprendizaje.

Los orígenes de la inteligencia artificial se remontan a la década de 1950, cuando contempla la posibilidad de crear maquinas que pudiera simular la inteligencia humana.

En 1956 se realizó la Conferencia de Dartmouth College, organizada por Marvin Minsky, John McCarthy y Claude Shanon, en la cual se reunió a todos los que trabajaban en el campo de la inteligencia artificial.



Figura 2.1: Conferencia de Dartmouth College 1956. De derecha a izquierda: John McCarthy, Oliver Selfridge, Marvin Minsky, Trenchard More y Ray Solomonoff (<https://infinitomaizum.wordpress.com/2009/11/05/em-1956-nasce-a-inteligencia-artificial/>)

En esta conferencia apareció el termino de inteligencia artificial por primera vez y se prefijó que en 25 años los ordenadores serían capaces de hacer todo el trabajo de los humanos, algo que hoy en día podemos comprobar que fueron optimistas.

Dos años después, en 1958 John McCarthy inventa el lenguaje de programación LISP [1] que se convierte en uno de los principales lenguajes de programación para IA.



Figura 2.2: Libro LISP 1.5 Programmer’s Manual creado por John McCarthy (<https://onih.pastperfectonline.com/webobject/7D1DFC2C-E98B-4491-9EEA-891675764224>)

Tras las prometedoras declaraciones y avances iniciales en la inteligencia artificial, surgió “el primer invierno de la Inteligencia Artificial” que fue un periodo, generalmente ubicado en la década de 1970, en el cual hubo grandes fracasos por las promesas poco realistas por parte de los desarrolladores y las altas expectativas que tenían los usuarios finales.

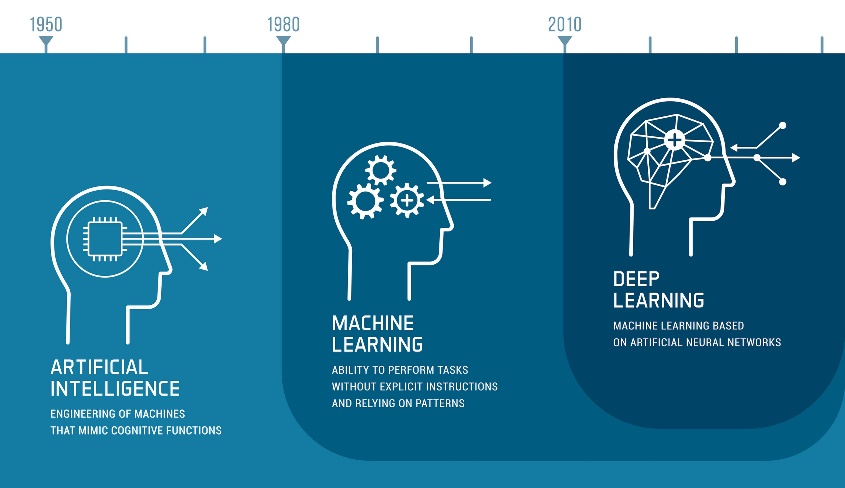
Principalmente, este estancamiento surgió por las limitaciones de esa época y los programadores e investigadores se encontraron con desafíos que no tenían una resolución ligera, como la lentitud de los procesos y la baja capacidad de memoria. Esto sentó precedente y empezó el declive del interés, inversión y la investigación de esta tecnología.

Sin embargo, el primer invierno de la IA llegó a su fin con la introducción de los sistemas expertos (SE) [2]. Estos sistemas, que comenzaron a ser adoptados por diversas organizaciones en todo el mundo, marcando un punto de inflexión que reavivó el interés y la inversión en la inteligencia artificial.

En las décadas posteriores, los investigadores tomaron nuevos enfoques respecto a la IA intentado abordar nuevos ámbitos como el comportamiento e inteligencia humana, la robótica, Procesamiento de Lenguaje Natural (PLN), Aprendizaje Automático Temprano y lógica y razonamiento.

Actualmente, gracias a los avances de las tecnológicos, como el aumento de potencia del cómputo y el almacenamiento, se han desarrollado nuevas técnicas y algoritmos que tienen la capacidad de procesar información derivada de datos y tomar decisiones para lograr un objetivo dado, dicho de otra manera, que puedan pensar y actuar como humanos e incluso superarlo en algunos ámbitos.

En este proyecto nos centraremos en las ramas de la Inteligencia Artificial, Machine Learning y Deep Learning.

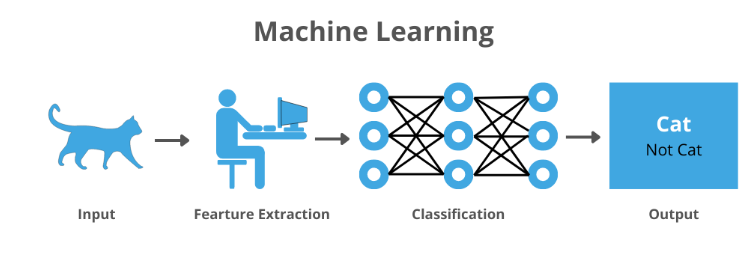


**Figura 3.1:** Inteligencia Artificial, Machine Learning y Deep Learning (<https://computerhoy.com/tecnologia/deep-learning-considera-revolucion-ia-1241180>)

3.2 Machine Learning

Machine Learning (ML), también conocido como aprendizaje automático, es una subdisciplina de la IA que permite aprender de los datos utilizando algoritmos para identificar patrones, realizar predicciones y mejorar el rendimiento sin estar programadas para ello.

El proceso de ML se constituye principalmente por 3 componentes: el modelo, la entrada y la salida de datos. El modelo es un algoritmo que aprende patrones a partir de los datos, la entrada de datos es el conjunto de datos que se emplea para entrenar el modelo y los datos de salida son los resultados generados por la predicción del modelo tras el procesamiento de los datos de entrada.



**Figura 3.2:** Esquema del proceso de Machine Learning (<https://intersog.co.il/blog/ais-powerful-duo-difference-between-machine-learning-and-deep-learning/>)

Existen diversos tipos de ML, dependiendo de la forma en la que aprenden: El aprendizaje supervisado, aprendizaje no supervisado y aprendizaje de refuerzo.

El aprendizaje supervisado entrena con una base de datos con información correctamente etiquetada que se utiliza posteriormente para predecir nuevos datos sin etiquetar. El aprendizaje no supervisado entrena con datos sin etiquetar detectando patrones o estructuras ocultas en los datos. Por último, el aprendizaje reforzado forma parte de lo que se conoce como Deep Learning y tiene la finalidad de construir modelos que aumenten el rendimiento y realicen predicciones tomando como base los resultados, ya sean recompensas o castigos, de cada iteración realizada en el modelo.

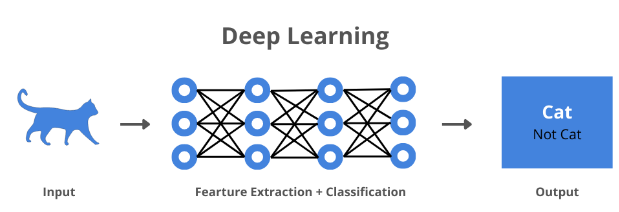
ML se emplea en diversos ámbitos, desde la detección de spam, completar texto y recomendación de contenido hasta servicios en línea, predicción del clima y diagnósticos médicos.

Existen gran cantidad y diversidad de algoritmos de ML, algunos de ellos incluyen Redes Neuronales, los Árboles de decisión (supervisado), K means (No supervisado) y Deep Q-Networks (Reforzado).

## 3.3 Deep Learning

Deep Learning (DP) es una subrama del ML que se enfoca en el uso de ANN para modelar y aprender parones complejos en gran cantidad de datos de entrada, que a diferencia de ML emplea capas de neuronas artificiales extrayendo automáticamente características de alto nivel directamente de los datos.

A las capas de entrada y salida se les denominan capas visibles, en la capa de entrada es ingiriere los datos para el procesado y la capa de salida es donde se realiza la predicción final. Entre estas capas se encuentran las neuronas que se denominan las capas ocultas, de ahí el termino “Deep” (profundo), donde capa tras capa se aprenden características jerárquicamente, donde las capas iniciales aprenden características simples y las más profundas características más abstractas y complejas dando lugar a dando lugar a representaciones complejas que mejoran la precisión de las predicciones en la capa de salida.



**Figura 3.2:** Esquema del proceso de Deep Learning (<https://intersog.co.il/blog/ais-powerful-duo-difference-between-machine-learning-and-deep-learning/>)

Respecto al uso del DL, es muy útil en áreas como VC, NLP, ciencias de la salud y reconocimiento de voz y audio, gracias a su capacidad para manejar y extraer información de grandes volúmenes de datos. Con el avance y mejoras sustanciales en los algoritmos, el DL se emplea en diversas áreas desde el marketing personalizado hasta la conducción autónoma y el diagnóstico médico.

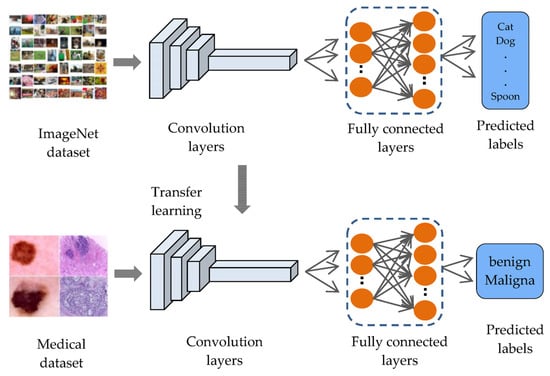
En resumen, actualmente el DP es una herramienta de vital importancia gracias a la capacidad de trabajar de manera autónoma. Sin embargo, presenta ciertas limitaciones como los altos requerimientos computacionales, necesidad de grandes conjuntos de datos y el riesgo de sobreajuste, más conocido como overfitting.

## 3.4 Transfer Learning

El Transfer Learning es un método de ML que aprovecha el entrenamiento de una red adquirido en una terea determinada para emplear como punto de partida en el desarrollo de otro modelo con otra terea diferente pero relacionada. En el contexto de DL. su uso es muy popular debido a su efectividad y capacidad de mejorar el rendimiento de los modelos y acelerar los procesos de entrenamiento.

### 3.4.1 Transfer Learning en Deep Learning

En DL, Transfer Learning se emplea principalmente en CNN preentrenadas. Este uso es especialmente útil cuando se parte de un conjunto de datos limitado para la nueva tarea, ya que el modelo empleado como punto de partida es entrenado con un conjunto de datos grandes y diversos.



**Figura 3.3:** Esquema Transfer Learning en DL (<https://www.mdpi.com/1424-8220/23/2/570>)

Algunas aplicaciones de Transfer Learning dentro de DL son las siguientes:

* **Modelo Preentrenado:** Emplear una CNN preentrenada con un conjunto de datos grande.
* **Fine-Tuning:** Reemplazar y entrenar las capas finales del modelo con el nuevo conjunto de datos específico para la nueva tarea.
* **Congelación de capas:** Se congela las primeras capas del modelo ya que aprenden características muy genéricas, mientas que las ultimas se ajustan a la tarea específica.

### 3.4.2 Transfer Learning en Machine Learning

Aunque Transfer Learning es más común emplearlo con DL gracias a la capacidad de aprendizaje en modelos complejos, también se emplea en ML mejorando modelos de NLP ajustándose a otras tareas de NLP.

Algunas aplicaciones de Transfer Learning en ML son las siguientes:

* NLP: En modelos preentrenados de grandes textos, pudiendo ser entrenados para tareas más específicas como la clasificación de texto y la traducción automática.
* Reconocimiento de Voz y Audio: Empleado en modelos con grandes datos de audios para ajustarse a tareas específicas como la detección de sonidos específicos o reconocimiento de voz en distintos idiomas.

## 3.5 Redes Neuronales Artificiales

Las redes neuronales artificiales (ANN) son modelos computacionales que permite dotar a los sistemas de la capacidad de aprender, interpretar y predecir de una manera muy similar a como lo hace el cerebro humano.

Las ANN están formadas por neuronas artificiales conectadas entre sí, las cuales reciben información de los datos de entrada o de otras neuronas, de manera muy similar a las neuronas del cerebro, que reciben impulsos nerviosos. Estas ANN generan un valor de salida que se transmite a otras neuronas o a la salida final de la red. En los siguientes apartados se introducirán los tipos más comunes de ANN, cada uno de ellos diseñados para resolver distintos tipos de desafíos.

### 3.5.1 Redes Neuronales Convolucionales

Las CNN son una subcategoría de las ANN implementadas para el procesamiento y clasificación de imágenes, es decir, para aplicaciones de VC. La estructura y procesamiento están inspiradas en el sistema visual de los animales.

La arquitectura de una CNN enfocadas a la clasificación de imágenes es prácticamente igual, aplicando filtrado de imágenes, extracciones de características de cada imagen y reducción de parámetros.

Diagrama, Dibujo de ingeniería

Descripción generada automáticamente

**Figura 3.4:** Esquema de una arquitectura CNN (<https://apiumhub.com/es/caso-de-uso-deep-learning/>)

En la figura 3.4 se observa las diferentes capas fundamentales. Cada capa realiza un propósito concreto:

* La capa convolucional es la capa encargada de analizar los datos de entrada detectando patrones, texturas, bordes y otras características mediante un filtro *Kernel* elegido y que recorrerá todos los píxeles de la imagen y se obtendrá una nueva matriz de salida.

Imagen que contiene teclado

Descripción generada automáticamente

**Figura 3.5:** Funcionamiento capa convolucional (<https://www.diegocalvo.es/red-neuronal-convolucional/>)

* La capa ReLU implementa una función de activación transformando cada valor en la salida de la convolución. Si el valor de salida es negativo, lo cambia a cero, si es positivo lo deja igual.
* La capa pooling reduce la resolución de las características detectadas en las capas convolucionales, perseverando así las características más esenciales. Esto genera una red más robusta y disminuye la cantidad de parámetros y el coste computacional. Existen varios tipos de pooling:
  + **Avg-Pooling:** Calcula la media de los valores en cada región del mapa de características.
  + **Min-Pooling:** Calcula el valor mínimo en cada región del mapa de características.
  + **Max-pooling:** Calcula el valor máximo en cada región del mapa de características, siendo una de las más empleadas debido a su efectividad para extraer las características más destacadas.



**Figura 3.6:** Funcionamiento Max-Pooling (<https://paperswithcode.com/method/max-pooling>)

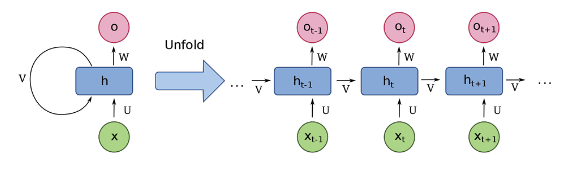
* La capa Fully Conected o Dense están ubicadas al final de la red, donde cada neurona en una capa dense, está conectada a todas las neuronas de la capa anterior, obteniendo como parámetro de entrada las características extraídas, realizando la clasificación y determinando la categoría o clase mediante una función de activación convirtiendo las salidas en probabilidades o a la clase que pertenece la entrada original.

Actualmente, este tipo de redes se emplean en aplicaciones cotidianas como Google Photos, Face ID, Instagram y Google Search.

### 3.5.2 Redes Neuronales Recurrentes

Las RNN está diseñadas principalmente para analizar secuencias de datos, donde la ordenación y dependencia tanto espacial como temporal son importantes. La información en las RNN fluye en bucle, es decir, la salida en un momento dado se procesa como entrada en el siguiente momento.

Una característica clave en las RNN es su capacidad de almacenar la información de los pasos anteriores mediante al denominado, estado oculto. Existen dos Redes Neuronales, de Corto y Largo plazo.



**Figura 3.7:** Arquitectura RNN (<https://kwfoundation.org/blog/2021/07/13/redes-neuronales-recurrentes/>)

Las Redes Neuronales de Memoria a Corto Plazo permite capturar dependencias en los datos, donde la influencia del estado oculto decae a medida que aumenta el tiempo, por lo que los pasos más próximos generarán un mayor impacto en la salida.

Las Redes Neuronales de Memoria a Largo Plazo aparecen para eliminar la limitación de capturar dependencias en los datos en las RNN estándar debido al desvanecimiento del gradiente, que se vuelven pequeños y dificulta el aprendizaje a medida que transcurre el tiempo. Aparecen dos soluciones: LSTM y GRU.

Para entender la Figuras 3.9 y la Figura 3.10, debemos de tener en cuenta la siguiente figura (Figura 3.7):

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**Figura 3.8:** Leyenda para arquitecturas de la Figura 3.9 y Figura 3.10 (<https://www.themachinelearners.com/modelos-secuencia/>)

* **LSTM:** Diseñadas para manejar dependencias a largo plazo mediante celdas de memoria, que almacenan información durante un periodo prolongado. Esta celda está controlada por 3 puertas: Puerta de entrada, puerta de olvido y puerta de salida, que información agregan, eliminan o generan de la celda de memoria.

Diagrama, Esquemático

Descripción generada automáticamente

**Figura 3.9:** Arquitectura LSTM (<https://www.themachinelearners.com/modelos-secuencia/>)

* **GRU:** Diseñada para resolver el problema del desvanecimiento de gradiente en las RNN, considerándose una simplificación de LSTM. Emplean dos puertas, la puerta de actualización que decide cuanta información pasada debe de transmitir en el siguiente paso y la puerta de reinicio que se emplea para decidir cuanta información pasada se debe olvidar.

Diagrama, Esquemático

Descripción generada automáticamente

**Figura 3.10:** Arquitectura GRU (<https://www.themachinelearners.com/modelos-secuencia/>)

Esta capacidad de procesamiento de datos secuenciales hace que este tipo de red sea útil para aplicaciones como reconocimiento de voz, traducción de texto y generación de texto. Algunos ejemplos de aplicaciones son Google Translate, Asistentes virtuales y OpenAI GPT-2.

### 3.5.3 Redes Generativas Adversarias

Las GAN son un tipo de redes introducidas en 2014 y se emplean para que un cierto dato se adopte para otro conjunto de datos, es decir, busca que dada una entrada se asemeje demasiado a la distribución deseada y que el dato generado sea indistinguible comparado con un dato real de dicha distribución.

Las GAN están compuestas por dos ANN que compiten entre sí:

* Red generativa: Emplea el ruido aleatorio como entrada y genera datos falsos muy semejantes a los reales, es la que se encarga del trabajo creativo y se ve obligada a mejorar para conseguir engañar a la segunda ANN (Red discriminadora) o que le apruebe la tarea. Esta red es capaz de realizar millones de pruebas hasta que la Red discriminadora acepta el resultado.
* Red discriminadora: Tom datos reales como dato de entrada y trata de distinguir entre ellos, revisando las creaciones de la primera red y moderándola. Esta es mucho más precisa ya que a la IA actualmente se le da mejor retocar y reconocer que crear.

En la figura 3.10 se muestra la arquitectura de una GAN:

Diagrama

Descripción generada automáticamente

**Figura 3.10:** Arquitectura de una GAN (<https://www.xataka.com/inteligencia-artificial/conceptos-inteligencia-artificial-que-gans-redes-generativas-antagonicas>)

Al inicio del entrenamiento los datos generados por la Red generativa serán significantemente distintos a los datos reales, entonces no será complicado para la Red discriminadora, distinguir los datos reales de los generados y el error en ambos casos será diminuto.

Sin embargo, en la etapa final del entrenamiento, la Red generativa puede ser capaz de generar datos muy realistas y la Red discriminadoras se vuelve extremadamente eficiente, aunque ya no puede distinguir claramente entre datos reales y generados, debido a la alta capacidad creativa de la Red generativa. Este equilibrio entre la generación y discriminación hace que este tipo de redes sean muy útiles en aplicaciones para generar datos realistas, como imágenes, videos, texto, sonido y otros varios.

# 4. Estado del arte

En la actualidad el reconocimiento visual de aves utilizando DL ha avanzado notablemente gracias a los desarrollos de las CNN, aplicaciones y conjuntos de imágenes etiquetados. En este apartado se abordarán los métodos, tecnologías, arquitecturas y aplicaciones que se emplean en la actualidad para el reconocimiento visual de aves mediante imágenes.

## 4.1 Cornell Lab of Ornithology

El laboratorio de ornitología de Cornell se fundó en 1910 en Ithaca, Nueva York, Estados Unidos por unos miembros de la universidad que se centra en el estudio de aves y otros animales salvajes, asi como la naturaleza.

El principal objetivo de Birds Cornell es apoyar iniciativas que ayuden a la conservación de las aves y la educación en las ciencias y acciones por el medio ambiente para que impacten positivamente en la biodiversidad e inspirar a las personas, sin importar las edades y el sector que desempeñen, a cuidar la naturaleza y su entorno. Para ello, ofrece una gran cantidad variada de recursos públicos con los que se podrán obtener gran cantidad de datos, desde imágenes, audios, videos y seminarios, hasta aplicaciones para identificar aves mediante imágenes y sonido.

Unos ejemplos de aplicaciones sobre el reconocimiento de aves mediante imágenes son los siguientes:

* **eBird:** base de datos que contiene millones de datos sobre aves.
* **Merlin Bird ID:** herramienta diseñada para identificar aves.
* **Bird Academy:** Imparten seminarios creados para poder entender y conectar con las aves.
* **Living Bird:** es una revista trimestral que publica artículos sobre investigaciones y conservación de las aves.

En los siguientes apartados se profundizará en eBird y Merlin Bird ID.

### 4.1.1 eBird

ebird empezó como una idea simple en el laboratorio de Ornitología de Cornell hasta convertirse en uno de los proyectos científicos de biodiversidad más grande del mundo, gracias a su expansión en 2010.

Actualmente con el Gobierno de España, específicamente con el Centro Nacional de Educación Ambiental (CENEAM) donde el objetivo principal es la conservación y educación de las especies.

Imagen de la pantalla de un celular con la imagen de un ave

Descripción generada automáticamente con confianza baja

Figura 4.1: Inicio de la página web de eBird (<https://ebird.org/explore>)

eBird tiene una base de datos de observaciones sobre las aves, con un conteo de más de 100 millones de imágenes aportadas por científicos, ornitólogos y aficionados de todo el mundo. Estos datos no son solo imágenes, audios y videos, sino también la distribución, abundancia, habitat, comportamientos, localización y muchos más.

Esta base de datos es abierta y gratuita a aficionados, ebird cuenta con expertos en aves que realizan una rigurosa tarea de chequeo para garantizar la precisión, relevancia y la veracidad de los datos proporcionada por observadores.

En resumen, gracias a este proyecto se ha revolucionado la forma en la que se obtienen y emplean los datos sobre las aves y ha generado una comunidad global de observadores y científicos que contribuyen a la conservación de las especies y la biodiversidad.

### 4.1.2 Merlin Bird ID

Merlin Bird ID es una aplicación móvil gratuita implementada por el Laboratorio de Ornitología de Cornell que permite reconocer aves mediante imágenes, audios y descripciones de su aspecto, tamaño y comportamiento.



**Figura 4.2:** Logo de Merlin Bird ID (<https://earthsky.org/earth/media-we-love-merlin-bird-id-app-cornell-lab-review/>)

En las siguientes Figura 4.3, Figura 4.4, Figura 4.5 y Figura 4.6 mostramos el funcionamiento de la aplicación y los pasos que sigue:

* **Paso a paso:** reconoce aves mediante tamaño, color y comportamiento y genera un listado de las posibles aves que tienen esas características.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Figura 4.2: Proceso para identificar un ave mediante la funcionalidad Paso a paso.

Un pájaro con la boca abierta

Descripción generada automáticamente con confianza media

Figura 4.3: Resultado de Paso a paso.

* **Audio:** reconoce aves mediante un audio en tiempo real y aparece una alerta cuando lo ha detectado.

Interfaz de usuario gráfica, Sitio web

Descripción generada automáticamente

Figura 4.4: Proceso para identificar un ave mediante audio.

* **Foto:** reconoce aves mediante imágenes en tu dispositivo o en tiempo real, puedes indicar la ubicación y la fecha o que no la sabes.

Imagen de la pantalla de un ave

Descripción generada automáticamente con confianza media

Figura 4.5: Proceso de reconocimiento de aves mediante fotos.

Actualmente, Merlin Bird ID es una de las mejores aplicaciones para el reconocimiento de aves, tanto visualmente como por audios. Es muy completa y precisa en sus análisis, gracias a su extensa base de datos y tecnología avanzada de IA.

## 4.2 iNaturalist

iNaturalist es una plataforma tanto web como móvil que fue desarrollada en el año 2008 como proyecto final de postgrado de Nate Agrin, Jessica Kline y Ken-ichi para la Academia de Ciencias de California en Berkeley. Permite a los usuarios llevar un registro de la biodiversidad y compartirlas a nivel mundial.

Los objetivos de esta plataforma es promover la ciencia ciudadana, recopilar datos para crear una base de datos global para proteger la biodiversidad y a su vez emplearse para conservacionistas, educadores e investigadores, aumentando el conocimiento público sobre la concienciación ambiental.

El reconocimiento automático de aves no es su única funcionalidad, ofrece una amplia cantidad de funcionalidades que no están limitadas a aves, pero es la que interesa en este TFG. En el Apartado (por definir) se profundizará en las funcionalidades específicas de reconocimiento visual de aves y a continuación, mencionaré algunas de las funcionalidades que tiene esta plataforma:

* **Registrar observaciones:** permite que los usuarios suban imágenes, audios y videos desde plantas y hongos hasta insectos o animales, con una serie de características como podría ser la ubicación precisa.
* **Comentarios:** los usuarios tienen la posibilidad de comentar y discutir sobre las observaciones registradas, proporcionando correcciones o ampliando la información sobre la especie observada.
* **Proyectos:** los usuarios tienen laposibilidad de crear y unirse a proyectos sobre cualquier especie en específico.

**Interfaz de usuario gráfica

Descripción generada automáticamente**

Figura 4.6: Funcionalidad de creación de un proyecto en la web de iNaturalist.

* **Herramienta de exportación y análisis:** los usuarios pueden exportar datos en formato CSV con gran cantidad de información pudiendo filtrar para limitar los datos extraídos y se puede visualizar un mapa de distribución geográfica de las observaciones registradas por los usuarios.

En resumen, iNaturalist es una plataforma gratuita muy poderosa para cualquier persona interesada en la conservación y el estudio de la biodiversidad, abarcando gran variedad de especies, no solo de aves.

# 5. Metodología

# 6. Resultados

# 7. Conclusión

# Bibliografía

Inteligencia artificial:

<https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr>

<https://datascientest.com/es/inteligencia-artificial-definicion#:~:text=La%20historia%20de%20la%20inteligencia%20artificial%20comenz%C3%B3%20en%201943%20con,creaci%C3%B3n%20de%20una%20red%20neuronal>

<https://www.youtube.com/watch?v=_pcQnFz7wAI&ab_channel=RaicesdeEuropa>

<https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial_14419>

<https://medium.com/@natisr/historia-de-la-inteligencia-artificial-63277f78fe2c>

<https://cse.unl.edu/~choueiry/S09-476-876/Documents/whatisai.pdf>

<https://www.algotive.ai/es-mx/blog/historia-de-la-inteligencia-artificial-el-machine-learning-y-el-deep-learning#ancla_3>

<https://www.bbva.com/es/innovacion/machine-learning-que-es-y-como-funciona/>

IA, DEEP LEARNING Y MACHINE LEARNIG:

<https://intersog.co.il/blog/ais-powerful-duo-difference-between-machine-learning-and-deep-learning/>

Tipos algoritmos DEEP learning:

<https://www.tokioschool.com/noticias/tipos-deep-learning/>

Desventajas del Deep learning:

<https://prompt.uno/redes-neuronales-profundas/desventajas-de-las-redes-neuronales-profundas/>

Redes neuronales artificiales:

<https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/>

Transfer learning:

<https://www.tokioschool.com/noticias/transfer-learning/>

<https://www.mdpi.com/1424-8220/23/2/570>

<https://www.geeksforgeeks.org/transfer-learning-with-fine-tuning-in-nlp/>

CNN: <https://datascientest.com/es/convolutional-neural-network-es>

<https://medium.com/latinxinai/redes-neuronales-convolucionales-cnn-dd4b50ceb24>

<https://paperswithcode.com/method/max-pooling>

RNN: <https://torres.ai/redes-neuronales-recurrentes/>

<https://datascientest.com/es/recurrent-neural-network-rnn-de-que-se-trata>

<https://keepcoding.io/blog/que-son-las-redes-neuronales-recurrentes/>

<https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>

https://kwfoundation.org/blog/2021/07/13/redes-neuronales-recurrentes/

<https://gamco.es/glosario/red-neuronal-recurrente/>

GAN:

<https://arxiv.org/pdf/2302.09346>

<https://blog.damavis.com/redes-neuronales-generativas-adversarias-gans/>

<https://arxiv.org/abs/1406.2661>

<https://www.deeplearningbook.org/>

[https://www.allaboutbirds.org/news/living-bird-latest-issue/#](https://www.allaboutbirds.org/news/living-bird-latest-issue/)

Estado del arte:

<https://keras.io/api/applications/efficientnet/>

<https://merlin.allaboutbirds.org/es/pagina-de-inicio/>

Cornell:

<https://www.birds.cornell.edu/latam/programas-y-proyectos/>

<https://www.solucionescosteras.org/recursos-del-cornell-lab-of-ornithology/>

<https://www.miteco.gob.es/es/ceneam/quienes-somos.html>

<https://www.miteco.gob.es/es/ceneam/recursos/pag-web/ebird.html>

<https://www.miteco.gob.es/es/ceneam/recursos.html>

Dataset Birds:

<https://www.inaturalist.org/>

<https://ebird.org/spain/home>

<https://www.vision.caltech.edu/datasets/cub_200_2011/>

# Referencias

[1] John McCarthy 1962. *LISP 1.5 programmer's manual.*

[2] [Online] INER (29/03/2022). Procedente de <https://www.unir.net/ingenieria/revista/sistema-experto/#:~:text=Los%20sistemas%20expertos%20(SE)%20son,un%20profesional%20en%20la%20materia>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Codificando Bits. (2023, 13 de febrero). Autoencoders: explicación y tutorial en Python. Codificando Bits. Recuperado de <https://www.codificandobits.com/blog/autoencoders-explicacion-y-tutorial-python/>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[Online] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946v5. Recuperado de https://arxiv.org/abs/1905.11946v5

# Lista de Acrónimos y Abreviaturas

ANN Redes Neuronales Artificiales

CNN Redes Neuronales Convolucionales

CSV Valores Separados Por Comas

DL Deep Learning

DNN Redes Neuronales Profundas

GAN Redes Generativas Adversarias

GRU Unidades Recurrentes Cerradas

IA Inteligencia Artificial

LSTM Memoria a largo plazo

ML Machine Learning

NLP Procesamiento de Lenguaje Natural

ReLU Unidad Lineal Rectificada

RNN Redes Neuronales Recurrentes

TFG Trabajo de Fin de Grado

VC Visión por computadora