



Universidad de Jaén

Escuela Politécnica Superior
de Jaén

TRABAJO FIN DE MÁSTER

IDENTIFICACIÓN DE OBJETOS EN IMÁGENES POR SATÉLITE UTILIZANDO TÉCNICAS DEEP LEARNING

Alumno

Guillermo Martínez Lledó

Tutores

María Dolores Pérez Godoy

(Departamento de Informática)

Antonio Jesús Rivera Rivas

(Departamento de Informática)

Noviembre, 2020



Universidad de Jaén

Departamento de Informática

Doña María Dolores Pérez Godoy y Don Antonio Jesús Rivera Rivas, tutores del Trabajo Fin de Máster titulado: '**Identificación de objetos en imágenes por satélite utilizando técnicas Deep Learning**', que presenta Don Guillermo Martínez Lledó, otorgan el visto bueno para su entrega y defensa en la Escuela Politécnica Superior de Jaén.

Jaén, Noviembre de 2020

El alumno:

Los tutores:

Guillermo Martínez Lledó

María Dolores Pérez Godoy
Antonio Jesús Rivera Rivas

FICHA DEL TRABAJO FIN DE TÍTULO	
Titulación	Máster en Ingeniería Informática
Modalidad	Trabajo Teórico/Experimental
Idioma	Español
Tipo	General
TFT en equipo	No
Autor/a	Guillermo Martínez Lledó
Fecha de asignación	Curos 2019/20
Descripción corta	El objetivo de este trabajo fin de máster consiste en evaluar y comparar el comportamiento de diferentes métodos de Deep Learning para la identificación y reconocimiento de objetos en imágenes de satélite

NORMAS APLICADAS EN ESTE DOCUMENTO	
LOCALES	
TFT-UJA:2017	Normativa de Trabajos Fin de Grado, Fin de Máster y otros Trabajos Fin de Título de la Universidad de Jaén (Normativa marco UJA aprobada en Consejo de Gobierno)
TFT-EPSJ:2017	Normativa sobre Trabajos Fin de Grado y Fin de Máster en la Escuela Politécnica Superior de Jaén (Normativa EPSJ aprobada en Junta de Escuela)
TFT-EPSJ	Criterios de evaluación y normas de estilo para TFG y TFM de la Escuela Politécnica Superior de Jaén
NACIONALES E INTERNACIONALES	
ISO 2145:1978	Documentación - Numeración de divisiones y subdivisiones en documentos escritos
UNE 50132:1994	Traducción de la ISO 2145
APA 6ª edición	Estilo de referencias y citas de APA (American Psychological Association)

NORMAS UTILIZADAS COMO BASE O REFERENCIA	
NACIONALES	
UNE 157001:2014	Criterios generales para la elaboración formal de los documentos que constituyen un proyecto técnico
UNE 157801:2007	Criterios generales para la elaboración de proyectos de sistemas de información
<i>Estas normas se han utilizado como base o referencia para la inclusión de algunos contenidos y definiciones sobre elaboración de proyectos, entendiendo como proyecto la documentación consensuada entre una empresa y un cliente, que da lugar al perfeccionamiento de un contrato para la elaboración de una obra o la prestación de un servicio. Por consiguiente, no debe esperarse la aplicación de estas normas en cuanto a la completitud de los contenidos ni a la organización de los mismos.</i>	

Contenido

1	Introducción	11
1.1	Descripción del proyecto.....	11
1.2	Motivación	11
1.3	Objetivos del trabajo	13
1.4	Estructura del proyecto	13
1.5	Planificación temporal.....	14
2	Ciencia de datos.....	17
2.1	Proceso de extracción de conocimiento de bases de datos.....	17
2.1.1	Recopilación	18
2.1.2	Selección, limpieza y transformación	19
2.1.3	Minería de datos o “Data Mining”	20
2.1.3.1	Tareas en minería de datos.....	20
2.1.3.2	Técnicas de minería de datos	23
2.1.4	Interpretación y evaluación.....	25
3	Deep Learning	29
3.1	Redes neuronales artificiales	29
3.2	Redes neuronales convolucionales	33
3.2.1	Entrenamiento de una CNN	34
3.2.1.1	Propagación.....	36
3.2.2	Elementos de una Red Neuronal Convolucional.....	36
3.2.2.1	Capas convolucionales.....	37
3.2.2.2	Capas de reducción de muestreo o submuestreo (Pooling)	40
3.2.2.3	Funciones de activación	41
3.2.2.4	Configuración de hiperparámetros adicionales y consideraciones para evitar el sobreaprendizaje.....	43
3.2.2.5	Aumentación de datos o “Data Augmentation”	44
4	Arquitecturas y tecnologías usadas	46
4.1	Arquitecturas utilizadas	46
4.1.1	Resnet50	47
4.1.2	InceptionV3.....	49
4.1.3	Xception.....	50
4.1.4	VGG-16.....	51
4.2	Tecnologías utilizadas.....	52
4.2.1	Sistema de computación de alto rendimiento ADA.	53
4.2.2	Librerías utilizadas.....	54
5	Experimentación y resultados.....	56
5.1	Pruebas preliminares	60
5.2	Predicción de modelos entrenados.....	61
5.3	Ensemble de modelos.....	65
5.3.1	Ensemble de modelos con conjunto de datos de entrenamiento original.....	65
5.3.2	Ensembles de modelos por red y con muestreo del conjunto de datos de entrenamiento	

5.3.3	Ensemble general de modelos y datos muestreados	69
5.4	Comparativa final y análisis de resultados.....	71
6	Conclusiones y trabajos futuros.....	78
6.1	Conclusiones de la experimentación	78
6.2	Conclusiones personales	78
6.3	Propuestas de trabajo futuro.....	79
7	Definiciones y abreviaturas.....	81
8	Bibliografía	83

Índice de ilustraciones

Ilustración 1.1: Imagen por satélite [6]	12
Ilustración 1.2: Diagrama de Gantt (Fuente: Elaboración propia)	16
Ilustración 2.1: Esquema de las fases del KDD (Fuente: timeofsoftware.com)	18
Ilustración 2.2: Ejemplo de clustering (Fuente: Wikipedia).....	22
Ilustración 2.3: Ejemplo de árbol de decisión (Fuente: numerentur.org)	24
Ilustración 2.4: Representación gráfica de tipos de resultados, y sus medidas de precisión y exhaustividad (Fuente: Wikipedia).....	28
Ilustración 3.1: Estructura de una red neuronal [22]	30
Ilustración 3.2: Ejemplo de perceptrón [25].....	31
Ilustración 3.3: Red Hopfield de cuatro nodos (Fuente: Wikipedia).....	32
Ilustración 3.4: Ejemplo de red neuronal recurrente (Fuente: researchgate.net).....	33
Ilustración 3.5: Arquitectura de una CNN (Fuente: www.aprendemachinlearning.com)	34
Ilustración 3.6: Ejemplos de underfitting, entrenamiento correcto y overfitting [34]	35
Ilustración 3.7: Comparación entre capas completamente conectadas y capas convolucionales [35].....	37
Ilustración 3.8: Ejemplo de operación de convolución (Fuente anhreynolds.com)	38
Ilustración 3.9: Aplicación de padding o relleno (Fuente: datahackers.rs)	39
Ilustración 3.10: Ejemplos de Max y Avg Pooling (Fuente: www.quora.com)	40
Ilustración 3.11: Comparación entre función logística y de tangente hiperbólica (Fuente: www.oreilly.com)	42
Ilustración 3.12: Comparación entre función ReLU (izquierda) y leaky ReLU (derecha) [36]	42
Ilustración 3.13: Comparación entre modelos antes y después de aplicar dropout (Fuente: www.medium.com)	44
Ilustración 3.14: Ejemplo de Data Augmentation aplicando distintas técnicas a una imagen (Fuente: www.mygreatlearning.com)	45
Ilustración 4.1: Arquitectura ResNet [17]	48
Ilustración 4.2: Forma canónica de un módulo Inception[12]	49
Ilustración 4.3: Versión de módulo Inception para el modelo Xception [12]	50
Ilustración 4.4: Tabla de arquitecturas VGG [13]	52
Ilustración 5.1: Ejemplos de FMoW, de arriba abajo y de izquierda a derecha: Estación de bomberos, estadio, parque de atracciones, estación nuclear, aeropuerto, zona de arqueología, cementerio, concesionario de coches, represa, campo de golf, prisión, lugar de culto, asentamiento empobrecido y parque. (Fuente: Conjunto de imágenes FMoW).....	59
Ilustración 5.2: Esquema sobre el sistema de ensamblaje (Fuente: Elaboración propia).....	65
Ilustración 5.4: Esquema de ensamblaje de modelos por red (Fuente: Elaboración propia)	67

Índice de tablas

Tabla 1.1: Planificación temporal y descripción de tareas (Fuente: Elaboración propia).....	16
Tabla 4.1: Comparación de errores Top-1 y Top-5 de varios modelos y arquitecturas para el conjunto ImageNet [17]	48
Tabla 4.2: Comparación de errores Top-1 y Top-5 de varios modelos para el conjunto ILSVRC-2012 [10]	50
Tabla 4.3: Comparación de precisión Top-1 y Top-5 de varios modelos para el conjunto ImageNet [12].....	51
Tabla 4.4: ADA – Nodo principal [8].....	53
Tabla 4.5: ADA – Nodo de cómputo TESLA [8]	53
Tabla 4.6: ADA – Nodo de cómputo RTX [8]	54
Tabla 5.1: Categorías y número de imágenes que contienen en entrenamiento y validación (Fuente: Elaboración propia)	59
Tabla 5.2: Predicciones obtenidas durante el entrenamiento (Fuente: Elaboración propia) ..	61
Tabla 5.3: Precisión, exhaustividad y valor-F de los modelos entrenados (Fuente: Elaboración propia).....	63
Tabla 5.4: Precisión, exhaustividad y valor-F del ensemble de los modelos entrenados con el conjunto original (Fuente: Elaboración propia)	66
Tabla 5.5: Precisión, exhaustividad y valor-F del ensemble por red de los distintos modelos entrenados con un muestreo aleatorio del 70% del conjunto original (Fuente: Elaboración propia).....	68
Tabla 5.6: Precisión, exhaustividad y valor-F del ensamblaje de los modelos entrenados con un muestreo aleatorio del 70% del conjunto (Fuente: Elaboración propia)	70
Tabla 5.7: Mejores precisiones por categoría en tipo de imagen (Fuente: Elaboración propia)	71
Tabla 5.8: Mejores precisiones por categoría en terreno (Fuente: Elaboración propia)	71
Tabla 5.9: Mejores precisiones por categoría en tamaño (Fuente: Elaboración propia).....	71
Tabla 5.10: Mejor exhaustividad por categoría en tipo de imagen (Fuente: Elaboración propia)	72
Tabla 5.11: Mejor exhaustividad por categoría en terreno (Fuente: Elaboración propia)	72
Tabla 5.12: Mejor exhaustividad por categoría en tamaño (Fuente: Elaboración propia).....	72
Tabla 5.13: Mejor valor-F por categoría en tipo de imagen (Fuente: Elaboración propia)	73
Tabla 5.14: Mejor valor-F por categoría en terreno (Fuente: Elaboración propia)	73
Tabla 5.15: Mejor valor-F por categoría en terreno (Fuente: Elaboración propia)	73
Tabla 5.16: Recopilación de resultados de precisión (Fuente: Elaboración propia)	75
Tabla 5.17: Recopilación de resultados de exhaustividad (Fuente: Elaboración propia).....	76
Tabla 5.18: Recopilación de resultados de valor-F (Fuente: Elaboración propia)	77

1 INTRODUCCIÓN

En este capítulo se presenta la especificación del trabajo, con una estructura y contenidos inspirados en los criterios y recomendaciones que establece la norma UNE 157801:2007 - “*Criterios Generales para la elaboración de proyectos de Sistemas de Información*”.

A lo largo del documento se utilizarán términos y acrónimos cuya descripción aparecen en el apartado 7 (Definiciones y abreviaturas).

1.1 Descripción del proyecto

Hoy en día son muchas las fuentes abiertas que proporcionan imágenes por satélite. Esto ha provocado un aumento en el interés por el campo de identificación y clasificación de objetos sobre este tipo de imágenes. Los métodos basados en Deep Learning, en particular las redes neuronales convolucionales (CNN) están arrojando buenos resultados dentro del campo de la identificación y clasificación de objetos.

Se realizará un estudio de varias herramientas y modelos orientados a la identificación y clasificación de imágenes para su posterior análisis, evaluación y comparación de los diferentes métodos usados, valorando además qué conjunto de imágenes será más adecuado para realizar el estudio, centrándonos en conjuntos de datos de imagen por satélite.

Cada modelo que sea elegido para trabajar con el conjunto de imágenes seleccionado tendrá que ser adaptado a éste para poder obtener una o varias medidas con las que poder evaluar y comparar con otros modelos.

1.2 Motivación

Las CNN son una clase de redes neuronales de aprendizaje profundo (Deep Learning) que representan un gran avance en el reconocimiento y clasificación de

imágenes. Habitualmente se usan para analizar imágenes visuales, y juegan un papel primordial a la hora de clasificarlas. En la actualidad, podemos encontrar numerosos ejemplos de detección y clasificación de imágenes por una CNN en muchos lugares, desde el etiquetado automatizado de fotos en la red social Facebook, hasta en la conducción automatizada de algunos coches. La detección y clasificación de imágenes pueden representar un gran avance en diversos campos de investigación, como salud o seguridad. [2]



Ilustración 1.1: Imagen por satélite [6]

Alguna de las características que las hacen más apropiadas son su rapidez y eficacia, pudiendo analizar y clasificar imágenes con una precisión igual o superior a la de un humano. Por poner un ejemplo, si se nos presentase la imagen mostrada en la ilustración 1.1, y nos pidiesen identificar qué es la estructura del centro, posiblemente no acertemos con la solución, sin embargo, una red bien entrenada podría indicarnos, en menos de un segundo, que se trata de una gasolinera.

En este trabajo de fin de máster se pretende hacer un estudio de diferentes herramientas y redes neuronales a la hora de trabajar con imágenes a vista de satélite,

similares a las que se han mostrado anteriormente, valorando la capacidad y precisión a la hora de detectar edificios, accidentes geográficos, o puntos de interés en general, y de clasificarlos correctamente.

1.3 Objetivos del trabajo

En este proyecto se pretenden alcanzar los siguientes objetivos:

- Estudiar los conceptos de identificación, clasificación y reconocimiento de objetos en imágenes.
- Estudio de los métodos Deep Learning para el reconocimiento de objetos.
- Conocer las características de las bases de datos de imágenes por satélite, tales como: sellos de tiempo, georreferenciación, etc.
- Estudiar las herramientas y bibliotecas de Deep Learning existentes para el reconocimiento de objetos en imágenes.
- Evaluar, comparar y analizar los resultados obtenidos de la ejecución de diferentes métodos de Deep Learning aplicados en la identificación, clasificación y reconocimiento de imágenes.
- Redacción de la memoria correspondiente del TFM.

1.4 Estructura del proyecto

En este apartado se da una breve descripción de los capítulos recogidos en este documento y el contenido de los mismos.

En el primer y actual capítulo se detallará qué se pretende lograr con el desarrollo de este TFM y una introducción al mismo.

A continuación, en los dos siguientes capítulos se exponen las bases teóricas del proyecto. En el capítulo 2 se presenta el concepto de Ciencia de datos, como un

campo de investigación, obtención de información y conocimiento extraído de uno o varios conjuntos de datos. Posteriormente, en el capítulo 3 se explica en qué consiste el aprendizaje profundo (o Deep Learning) y las CNN, en las que se va a basar este proyecto para obtener la información de un conjunto de imágenes concretas.

En el capítulo 4 se da una explicación teórica de las diferentes arquitecturas de redes neuronales que serán utilizadas y una descripción de los equipos que van a usarse durante la realización del desarrollo de este TFM.

Seguidamente, en el capítulo 5 se detallan los experimentos realizados, así como su preparación, y la interpretación de los resultados obtenidos.

En el capítulo 6 se recopilan las conclusiones obtenidas tras la realización de este proyecto y se proponen algunas guías para un trabajo futuro.

En último lugar se incluye la bibliografía utilizada.

1.5 Planificación temporal

El TFM se define como una asignatura de 12 créditos, lo que supone que la duración total del proyecto será de 300 horas. Por consiguiente, la principal restricción aplicable es la limitación de la duración del trabajo.

Ya que el presente proyecto es un TFM, no existen restricciones de tipo económico, sino de tipo temporal (un número aproximado de horas). Por consiguiente, los cálculos de tamaño del proyecto están supeditados al tiempo disponible. En cuanto al esfuerzo, se dispone de tan solo un efectivo (la persona autora del trabajo).

Este proyecto pretende llevar a cabo un estudio sobre tecnologías existentes, por lo tanto, la mayor inversión temporal se centra en la investigación y experimentación. Para realizar una planificación temporal previa a la realización del proyecto, que mediremos mediante días, se tendrá en cuenta que cada día equivale

a una jornada diaria de 8 horas de trabajo. La tabla siguiente recoge las tareas que han sido planificadas para la realización de este estudio.

Tarea	Descripción	Estimación temporal (días)
Revisión teórica	Esta tarea busca recoger información en el campo tecnológico que fundamenta la detección y clasificación de imágenes, así como información de ésto aplicado a imágenes tomadas por satélite.	5
Búsqueda y descarga de conjuntos de imágenes	Se trata de encontrar y valorar conjuntos de datos de imágenes tomadas vía satélite para la realización del estudio, y, si es posible, que contengan metadatos para poder ser procesados y evaluados	5
Revisión y adaptación de modelos de redes neuronales	Para realizar el estudio necesitaremos valorar los diferentes modelos que puedan ser adaptados al trabajo con imágenes de nuestras características. En caso de encontrar un modelo que pueda interpretar estas imágenes, si es necesario, debemos adaptar su configuración al conjunto de imágenes con el que se haya decidido trabajar.	6
Entrenamiento de modelos y obtención de resultados	Una vez se haya conseguido uno o varios modelos capaces de trabajar con nuestras imágenes, se realizarán varias pruebas con éstos, que serán entrenados con distintas configuraciones y redes neuronales analizando y comparando los resultados.	19
Interpretación y ensamblaje de resultados	Se valorarán los resultados obtenidos en busca de datos de interés. Se procederá además al ensamblaje de los resultados obtenidos por varias redes, para comprobar si esto es capaz de mejorar los resultados obtenidos por cada red de forma individual.	3

Conclusiones	Finalmente, se hará una evaluación de todo el estudio, planteando posibles conclusiones que surjan de cara a los resultados obtenidos.	2
Redacción de memoria	Redacción del documento de texto que recopila todo lo relevante al estudio realizado y el desarrollo del mismo. Será realizada en el transcurso del resto de actividades.	Realizado paralelamente con el resto del proyecto
TOTAL: 37 días = 296 horas		

Tabla 1.1: Planificación temporal y descripción de tareas (Fuente: Elaboración propia)

La realización de estas tareas supone, salvo en el caso de la redacción de la memoria, una ejecución secuencial y ordenada, dependiendo la ejecución de cada tarea intermedia de la finalización de otra tarea anterior. El único paralelismo se podría dar, como se ha mencionado, en la redacción de este documento, siendo esto de gran importancia para poder plasmar todo el trabajo realizado en el estudio. Podemos ver un desglose de tareas, su orden y su coste temporal representados en el siguiente diagrama de Gantt (Ver ilustración 1.3)

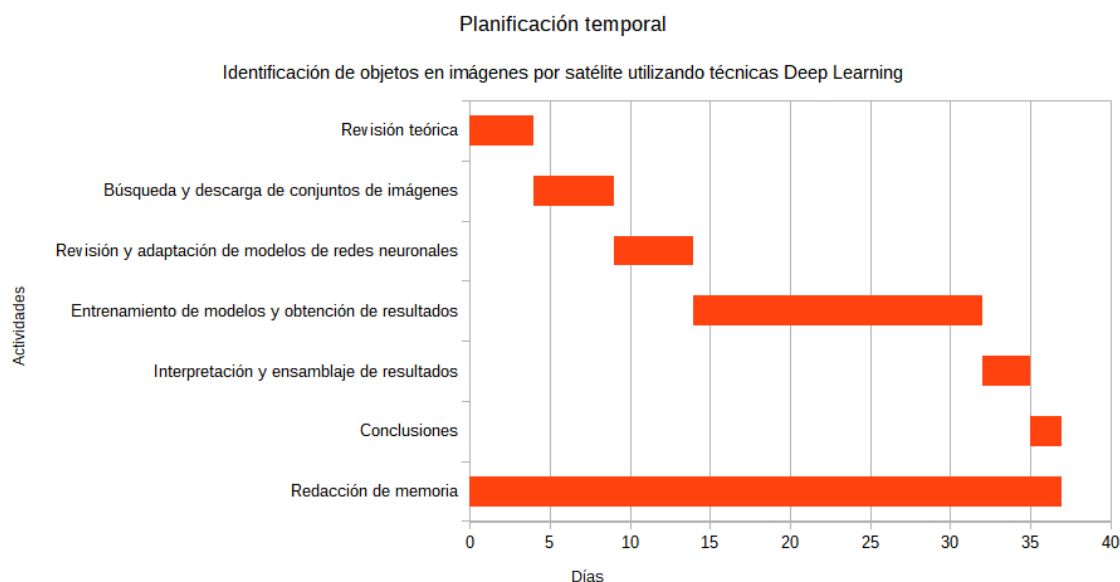


Ilustración 1.2: Diagrama de Gantt (Fuente: Elaboración propia)

2 CIENCIA DE DATOS

La ciencia de los datos es un campo interdisciplinario sobre procesos y sistemas para extraer conocimientos de grandes volúmenes de datos en diversas formas estructuradas o no estructuradas. Ésta es una continuación de algunos de los campos de análisis de datos como la minería de datos y el análisis predictivo, así como de la extracción de conocimientos de bases de datos (KDD). [7]

Aunque ambos términos suelen confundirse como si tratasen de lo mismo, hay que aclarar que la ciencia de datos es, de forma genérica, todo aquello que tiene que ver con los datos, por lo que la minería de datos, centrada en las técnicas usadas para la extracción de conocimiento a partir de datos o estadísticas, sería una parte importante, pero no la única.

El término “Ciencia de datos” ha surgido para designar a una nueva profesión que se espera que dé sentido a grandes cantidades de datos. Pero dar sentido a los datos tiene una larga historia y ha sido discutido por científicos, estadísticos, bibliotecarios, informáticos y otros durante años [19].

En los siguientes apartados se presentarán estos campos referentes a la ciencia de datos, en los que estará basado el desarrollo del TFM.

2.1 Proceso de extracción de conocimiento de bases de datos

Se define KDD [39] como el proceso de utilizar una base de datos junto con cualquier selección, preprocesamiento, submuestreo y transformación de la misma que se requiera; aplicar métodos de extracción de datos (algoritmos) para enumerar patrones a partir de ella, y evaluar los productos de extracción de datos para identificar el subconjunto de los patrones enumerados que se consideran “conocimiento”. Dichos patrones deben de cumplir cuatro propiedades: que sean válidos, no solamente para datos actuales, sino para otros datos futuros; que sea novedoso, es decir, que aporte

conocimiento nuevo, sin redundancias; que tenga potencial de aportar utilidad al usuario, y que sea comprensible y fácil de interpretar.

Este proceso está compuesto por cuatro fases principales, además de ser **iterativo**, es decir, la salida de cualquiera de éstas puede volver a una etapa anterior en caso de que tenga potencial de mejorar el posible conocimiento que se pueda extraer (ver ilustración 2.1).

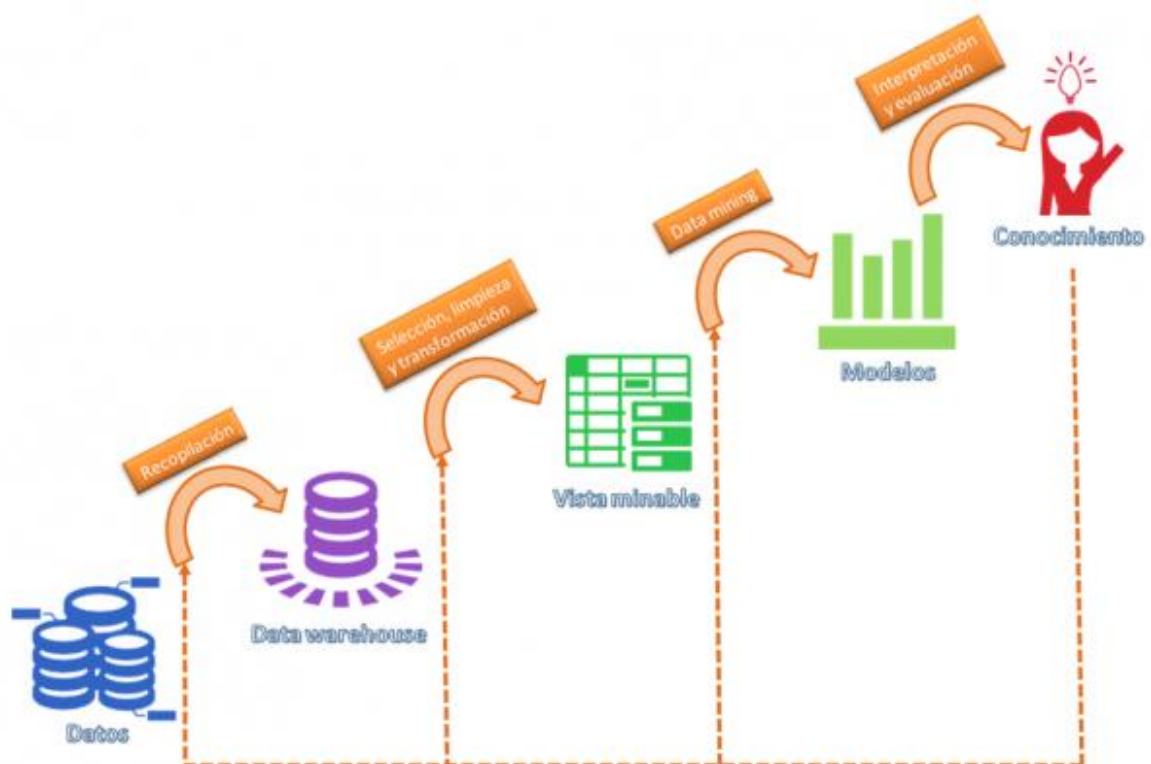


Ilustración 2.1: Esquema de las fases del KDD (Fuente: timeofsoftware.com)

2.1.1 Recopilación

Para realizar una extracción de conocimiento de calidad es habitual que tengamos que usar diferentes fuentes de datos, pues si nos limitamos a trabajar con una sola base de datos es posible que no podamos sacar ninguna información adicional a la que nos aporta la propia base de datos, por lo que incumpliría una de las propiedades que hemos nombrado al principio del capítulo: que sea novedoso.

La principal dificultad que puede presentarse en esta fase es la discrepancia que puede haber entre varias bases de datos, debido a que, frecuentemente, serán de distintas fuentes, y, por lo tanto, es posible que no utilicen los mismos formatos, unidades de medida, etc. Por lo cual debemos de homogeneizar todos los datos que utilicemos para que puedan estar relacionados entre sí, sin presentar incoherencias o información incorrecta.

2.1.2 Selección, limpieza y transformación

Una vez contamos con una fuente de datos bien integrada, el siguiente objetivo debe ser el tratamiento de los datos con la intención de poder aplicar algoritmos de minería de datos y obtener información útil, que pueda ayudarnos en nuestra tarea. Para ello, distinguimos entre tres procesos que pueden ser realizados para lograr este fin:

- **Selección** de los atributos que puedan resultar de interés para nuestro objetivo, es decir, elegir los datos con los que podamos crear un patrón de conocimiento, ya que muchas bases de datos contendrán algunos atributos, como por ejemplos, números de identificación, nombres propios, marcas... que no son de valor en nuestro propósito.
- Realizar una **limpieza** de aquellos valores que no cumplan un comportamiento genérico, comúnmente conocidos como datos anómalos o ruido. La inclusión de datos no genéricos puede llevar incluso al mal funcionamiento de los algoritmos de minería de datos, ya que se tratan de valores que han sido tomados erróneamente, o incluso donde algunos de sus atributos son desconocidos, como por ejemplo, los valores indeterminados (no sabe/no contesta) que pueden presentar algunas estadísticas. De todas formas, hay que realizar este proceso de forma cuidadosa, y saber distinguir cuando un dato anómalo se trata realmente de un caso excepcional o erróneo, o en cambio, puede ser un pico de interés y resultarnos de utilidad.

- Aplicar una **transformación** a aquellos datos que puedan presentar problemas de interpretación, centrándose especialmente, en la etapa de minería de datos. Esta necesidad puede darse de diferentes maneras, y debemos tener en cuenta que, por norma general, un valor numérico podría ser de mayor utilidad, por ejemplo, convirtiendo un formato de tiempo establecido en horas:minutos:segundos, por el total en segundos, para poder realizar operaciones matemáticas con facilidad, o asignando un valor numérico a atributos que no lo son. Sin embargo existen excepciones, con las que también debemos ser cuidadosos. Estas excepciones podría ser alguna de las siguientes: uso de atributos numéricos de forma errónea, por ejemplo, usar los meses de un año como valor numérico, cuando no tendría sentido realizar operaciones matemáticas con ellos; o la necesidad de discretizar ciertos valores, como por ejemplo, una longitud en diferentes tamaños según unos intervalos, intentando, como siempre, poder sacar la mayor utilidad de éstos.

2.1.3 Minería de datos o “Data Mining”

Esta fase está caracterizada por la utilización de procesos de exploración y análisis, mediante métodos automáticos o semiautomáticos, del conjunto de datos del que queremos extraer conocimiento. Nuestro objetivo principal será la extracción de patrones o modelos significativos. Para ello escogeremos qué tareas serán realizadas, qué modelo seguiremos, y qué algoritmos escogeremos para resolver la tarea usando el modelo elegido.

2.1.3.1 Tareas en minería de datos

Definimos como tareas de minería de datos [41] aquellos conjuntos de algoritmos, de mayor o menor complejidad, que son aplicados sobre un conjunto de datos para obtener unos resultados. Éstos pueden ser clasificados en predictivos, si tratan problemas donde hay que predecir valores para muestras desconocidas a priori, o descriptivos, que tratan de aportar información extra a datos ya existentes.

Y, en función del tipo de entrenamiento realizado en la obtención de información, podemos clasificarlos en dos grupos:

- Supervisados (o predictivos): Predicen valores desconocidos de un conjunto de datos a partir de otros conocidos.
- No supervisados (o del descubrimiento del conocimiento): Donde el algoritmo no recibe etiquetas de los datos, y debe aprender por sí mismo, cómo se estructuran los datos para poder clasificarlos.

En los siguientes apartados vamos a ver con más detalle en qué consiste cada uno de estos tipos de tareas, mostrando ejemplos de los más usados.

2.1.3.1.1 Tareas descriptivas

Las tareas descriptivas utilizan los datos históricos, tratando de identificar comportamientos y asignando características a partir de ellos. Su objetivo es presentar una clasificación de las distintas muestras, con la intención de poder tomar mejores decisiones. Los tipos de tareas descriptivas más comunes son las siguientes:

- Segmentación: Estas tareas se encargan de identificar grupos naturales dependiendo de un conjunto de atributos concretado en el conjunto de datos. Diferenciamos dos técnicas principales:
 - Clustering: Los segmentos se determinan durante la ejecución del algoritmo. Es capaz de procesar de manera eficaz tanto las variables cualitativas como cuantitativas. Como ejemplos de algoritmos de clustering nos encontramos K-means y K-medoids

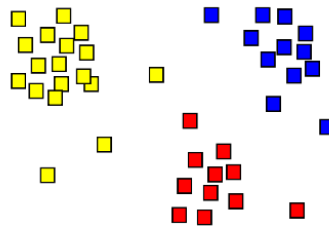


Ilustración 2.2: Ejemplo de clustering (Fuente: Wikipedia)

- Segmentación neuronal: en este caso, es necesario definir, antes de la ejecución del algoritmo, el número de segmentos y su distribución bidimensional. A diferencia del clustering, destaca más en el procesamiento de variables cualitativas que cuantitativas.
- Asociación: Se trata de organizar los datos según las relaciones que hay entre ellos a través de reglas lógicas de asociación. Podemos encontrar un patrón útil analizando, por ejemplo, los datos obtenidos en las compras de un supermercado, donde las compras de un producto pueden estar influenciadas por las compras de otros.
- Descubrimiento de subgrupos: Consiste describir problemas mediante la obtención de reglas que describan los datos para una variable objetiva.
- Correlaciones: Ayudan a determinar el grado de similitud de dos variables numéricas mediante un coeficiente de similitud, que toma valores entre -1 y 1, indicándonos si dos variables están más relacionadas, si toma valores cercanos a 1; si son opuestas, en el caso de tomar valores cercanos a -1, o si no tienen relación, si el valor es 0.

2.1.3.1.2 Tareas predictivas

Permiten la creación de modelos que puedan intentar obtener información desconocida a partir de muestras conocidas, las más utilizadas son:

- **Clasificación:** Son las más comunes en minería de datos. Su funcionamiento principal consiste en asignar una categoría a cada caso a partir de sus atributos, diferenciando a uno concreto como “atributo clase”. Tratan de buscar una función que pueda diferenciar, a partir del atributo clase designado, las distintas clases en las que se pretende dividir el conjunto de datos. Entre estos algoritmos diferenciamos los dos tipos más representativos: Los basados en redes neuronal, y los basados en árboles de decisión.
- **Regresión:** Tiene un planteamiento similar al de clasificación, con la diferencia de que se especializa en datos cuantitativos (valores continuos) en lugar de cualitativos (valores discretos). Trata de obtener una función, aplicable en el campo de los números reales, que asigne a cada muestra un resultado numérico.
- **Predicción:** Trata de pronosticar un valor de salida a partir de valores obtenidos con datos similares en el pasado.
- **Análisis de secuencia:** Intenta encontrar patrones en unos eventos concretos, denominados secuencias o transacciones.
- **Análisis de desviaciones:** Se especializa en la detección de valores anómalos. Como se ha explicado en la fase de limpieza, hay ciertos valores, que se salen de la norma respecto al resto, pero que pueden proporcionar información útil.
- **Análisis de similitud en series temporales:** Intenta detectar secuencias similares en colecciones de datos situados en distintos espacios temporales.

2.1.3.2 Técnicas de minería de datos

Existen varios métodos a la hora de llevar a cabo la minería de datos, y, por desgracia, no podemos señalar que una sea mejor que otra, pues ésto depende de muchos factores como: la cantidad de datos con los que queremos trabajar; el tipo de estos datos, o incluso el tipo de resultado que queremos obtener.

A continuación, vamos a ver algunas de las más comunes:

- Redes neuronales artificiales: Son un paradigma de aprendizaje y procesamiento automático basado en la forma en la que funciona el sistema nervioso de los animales. Es la técnica en la que se basa el desarrollo del TFM, por lo tanto, será vista en mayor profundidad a lo largo de este documento.
- Árboles de decisión: Se trata de un aprendizaje supervisado donde los datos se ordenan en una estructura de árbol, como la mostrada en la ilustración 2.3.

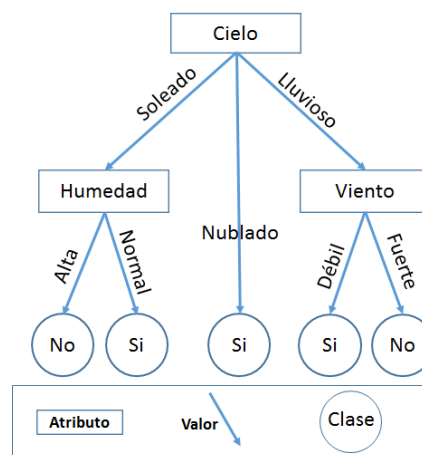


Ilustración 2.3: Ejemplo de árbol de decisión (Fuente: numerentur.org)

- Algoritmos genéticos: Éstos están inspirados en las leyes sobre la evolución de Darwin y los conceptos básicos de la biología genética. En este algoritmo, los datos pasan por procesos evolutivos, donde se da una selección, cruce, mutación aleatoria y reemplazo. Los valores de cada dato son equivalentes a sus genes, en los que, a través del proceso de evolución, se mantienen los más representativos y se eliminan los menos útiles.
- Modelos estadísticos: Se trata de expresiones lógicas que se emplean en modelos experimentales o en la regresión para identificar factores claves, es decir, los atributos que pueden modificar el resultado.
- Métodos bayesianos: Son aquellos que buscan, para cada muestra desconocida, la probabilidad de ser asignada a cada una de las posibles clases, con el fin de elegir la de mayor probabilidad como resultado de la

predicción. Uno de los métodos más conocidos es el de *Naïve Bayes*, caracterizado por realizar la predicción considerando que todos los atributos son independientes entre sí.

2.1.4 Interpretación y evaluación

Antes de usar los modelos obtenidos durante el KDD, éstos deben ser evaluados para medir su calidad. Se trata de un proceso no trivial, ya que algunos resultados pueden ser bastante subjetivos o estar enfrentados entre sí. Durante esta etapa debemos recordar que un buen modelo presenta patrones válidos, novedosos, potencialmente útiles y comprensibles.

Durante la etapa, usaremos conjuntos de datos que distinguiremos en dos grupos:

- Conjunto de entrenamiento: Se trata de todos aquellos datos que son utilizados en la construcción del modelo.
- Conjunto de validación: Son aquellas muestras que no han sido utilizadas durante la construcción del modelo y sirven para valorar la calidad de éste. Es común que dicho conjunto sea elaborado como una porción de la inicial, es decir, se divide éste en dos partes y se usa una como conjunto de entrenamiento y otra de validación. Este subconjunto es usado también durante el entrenamiento, donde después de haber completado una etapa con el conjunto de entrenamiento, evaluará este otro subconjunto, pudiendo rectificar pesos a partir de sus aciertos o fallos, además de poder informarnos sobre qué tasa de acierto aproximado tiene el modelo durante todo el entrenamiento. Normalmente, el conjunto de validación siempre será de menor tamaño que el de entrenamiento, pues interesa que nuestro modelo cuente con la mayor cantidad de información a la hora de elaborar patrones. Esta división puede realizarse de varias maneras:

- Validación simple: se divide aleatoriamente el conjunto de datos en dos subconjuntos. El de validación suele estar formado por un tamaño del 5 al 50% del total.
- Validación cruzada: también se realiza una división aleatoria, pero la característica de esta es se basa en entrenar el modelo, primero usando un subconjunto como entrenamiento y el otro como validación, a continuación se entrena al contrario, y por último, se construye el modelo y se valida con el conjunto completo. La precisión final se obtendrá como la media de las precisiones anteriores.
- Validación cruzada con k pliegues: Similar a la anterior, pero dividiendo el conjunto en k subconjuntos. El resto del proceso es similar, tomando en cada entrenamiento uno de los subconjuntos para validar y con el resto se construye el modelo. El proceso es repetido hasta que todos los subconjuntos hayan sido usados como validación. Finalmente, al igual que en la validación cruzada, se construye el modelo y se valida con todos los datos. La precisión final se obtiene de la misma manera que en la validación anterior.

Adicionalmente, podemos usar un conjunto de test, que se trata de imágenes que no han estado presentes en ninguno de los dos subconjuntos anteriores, y que usaremos para valorar un modelo con un entrenamiento ya finalizado.

A la hora de valorar la precisión de un modelo se tienen en cuenta los resultados obtenidos, clasificándolos, de forma binaria, de la siguiente manera:

- Verdaderos positivos (VP): Son aquellos elementos de la categoría que está siendo evaluada, a los que el modelo ha asignado correctamente su etiqueta.
- Falsos positivos (FP): Se trata de los elementos que no son de la categoría evaluada pero que el modelo si los ha identificado como tal.
- Falsos negativos (FN): Aquellos elementos pertenecientes a una categoría que no han sido acertados.

- Verdaderos negativos (VN): Son los elementos que, siendo de una categoría distinta, no han sido identificados con la categoría que está siendo evaluada.

A partir de estos datos podemos calcular tres medidas que informarán de la calidad de nuestro modelo. Como hemos mencionado antes, es un resultado por cada categoría, pero si queremos tener una visión general, podemos hacer una media entre todas las categorías. Las medidas son:

- Precisión: Es la división de los positivos verdaderos entre la suma de éstos con los falsos positivos. Con esto podemos medir la calidad del modelo en la clasificación de las muestras asignadas a cada categoría. Debemos tener en cuenta que es una medida que va a tener en cuenta los fallos que haya tenido en una categoría, pero no va a representar los elementos de una categoría que no han sido incluidos dentro de ésta. Su fórmula es la siguiente:

$$\text{Precisión} = \frac{VP}{VP + FP}$$

- Exhaustividad (Recall): Se calcula dividiendo los positivos verdaderos entre la suma de éstos y los falsos negativos. Nos informa sobre la cantidad de elementos de una categoría que nuestro modelo ha podido identificar. Al contrario que con la medida de precisión, ésta no penaliza los fallos al clasificar elementos incorrectos, sino cuando no incluye dentro de la categoría a elementos que sí lo son. Su fórmula es la siguiente:

$$\text{Exhaustividad} = \frac{VP}{VP + FN}$$

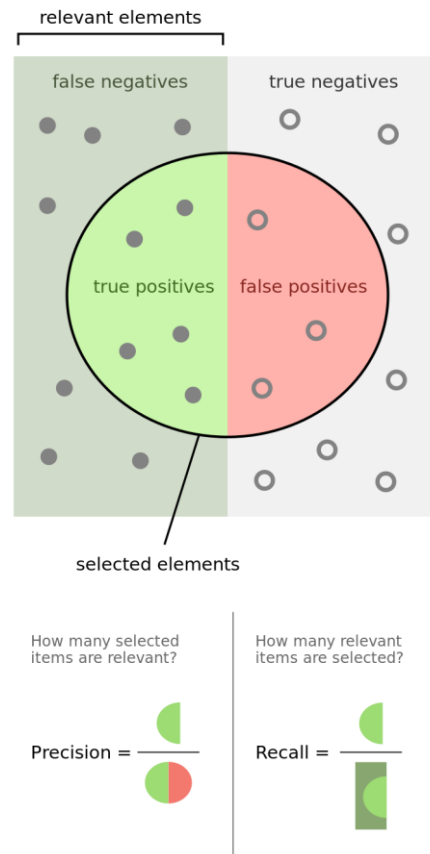


Ilustración 2.4: Representación gráfica de tipos de resultados, y sus medidas de precisión y exhaustividad (Fuente: Wikipedia)

- Valor-F (F-Score): Esta medida combina los valores de precisión y exhaustividad. Es, por tanto, la medida más representativa para medir la calidad de un modelo, ya que tendrá en cuenta tanto la información aportada por la precisión como por la exhaustividad. Su cálculo se realiza mediante la siguiente fórmula:

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

3 DEEP LEARNING

En este capítulo se tratan los aspectos fundamentales de las redes neuronales artificiales y arquitecturas utilizados en el campo del aprendizaje profundo.

Como introducción al capítulo, definimos el aprendizaje profundo (Deep Learning) como un conjunto de algoritmos de aprendizaje automático que intenta modelar abstracciones de alto nivel en datos usando arquitecturas computacionales [20].

Aunque muchas veces se confunde el término “Machine Learning” con el de “Deep Learning”, la diferencia entre ambos se establece en el tipo de algoritmos usados, más sencillos en Machine Learning y más abstractos en Deep Learning, y la cantidad de capas de abstracción, siendo más numerosa en Deep Learning, pero sin llegar a establecer un número mínimo de capas para separar ambos conceptos. Como similitud tienen que su fin es conseguir que una máquina pueda “aprender” sobre un campo concreto, pudiendo reconocer o interactuar dentro del mismo (Reconocimiento de imágenes, de voz, de texto...)

3.1 Redes neuronales artificiales

Las redes neuronales artificiales (ANN) han experimentado un renacimiento que no solo ha transformado el campo de la inteligencia artificial, sino que también proporcionan nuevos conocimientos sobre la computación neuronal en los sistemas biológicos [21].

El funcionamiento de estos sistemas, centrados en el aprendizaje, están inspirados en las propias redes neuronales biológicas formadas en el cerebro de los animales, donde las neuronas están interconectadas y donde la información que salga de una llegará a otras adyacentes.

De igual forma, y como vemos en la siguiente ilustración, cada nodo de la ANN, al que llamamos neurona artificial, está conectado con otros, formando un grafo dirigido.

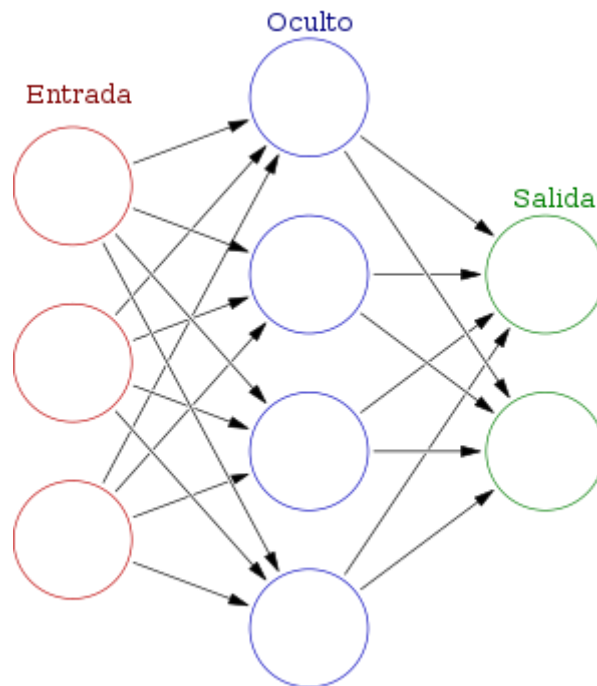


Ilustración 3.1: Estructura de una red neuronal [22]

En referencia a la anterior ilustración, podemos observar que las neuronas artificiales, o nodos, están divididos en tres grupos: Entrada, Oculto y Salida. Los tres grupos forman lo que se conoce como capas. La primera capa sería la de entrada, la última, de salida, y el resto de capas intermedias, oculto. A diferencia de los grupos de entrada y salida, formados por una capa cada uno, el grupo de capas ocultas puede presentar varias capas, con nodos del mismo grupo conectados entre sí.

Cada uno de estos nodos cuenta con lo que se conoce como función de activación. Ésta define la salida de un nodo dada una entrada o conjunto de entradas. Es común que esta función sea la misma para todos los nodos de una misma capa.

Según la configuración en la que se presenten las capas de una ANN nos encontramos con diferentes arquitecturas. Las más populares se presentan a continuación:

- Perceptrón simple: fue creado por F. Rosenblatt, siendo la red neuronal más antigua, y siendo utilizada en la actualidad como identificador de patrones. Este modelo es capaz de generalizar una serie de patrones, es decir, después de haberlos aprendido, poder reconocer otros similares que no hubiesen estado presentes en el entrenamiento. Aún así, contaba con ciertas limitaciones con algunas funciones lógicas (como la disyunción exclusiva o XOR) y, en general, era incapaz de clasificar clases no separables linealmente [24]. Puede verse a continuación un ejemplo de perceptrón, formados principalmente por dos o más nodos de entrada asociados con un peso, a la que se aplica una función de activación y devuelven una única salida (Ver ilustración 3.2),.

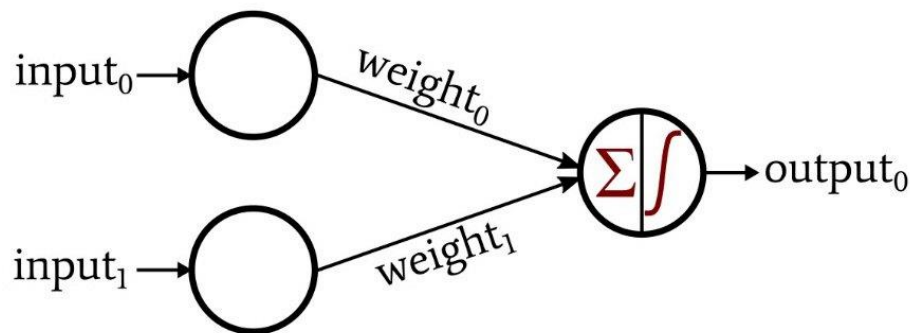


Ilustración 3.2: Ejemplo de perceptrón [25]

- Perceptrón multicapa: Esta arquitectura soluciona el principal inconveniente del perceptrón simple, resolver problemas que no son separables linealmente. Su configuración es una ampliación del perceptrón simple, añadiendo una serie de capas que realizan transformaciones sobre las variables de entrada, logrando resolver el problema mencionado. Como ejemplo de perceptrón multicapa, podemos fijarnos en el esquema que mostrábamos como ejemplo de red neuronal artificial (Ilustración 3.1), es decir, a diferencia del perceptrón simple, presenta una capa oculta.
- Red de Hopfield: Como indica su nombre, fue inventada por J. Hopfield. Se trata de una ANN que se presenta con una configuración monocapa, donde, al contrario que en el resto, los nodos de la misma capa están conectados entre sí. En algunas ocasiones, puede presentar una versión bicapa, donde

cada nodo de la primera capa está conectado con su nodo respectivo de la segunda capa, pero manteniendo la estructura de que todos los nodos de la misma capa estén conectados. El funcionamiento de esta red se basa en enviar un patrón en la red y actualizar las neuronas repetidamente hasta estabilizarlas.

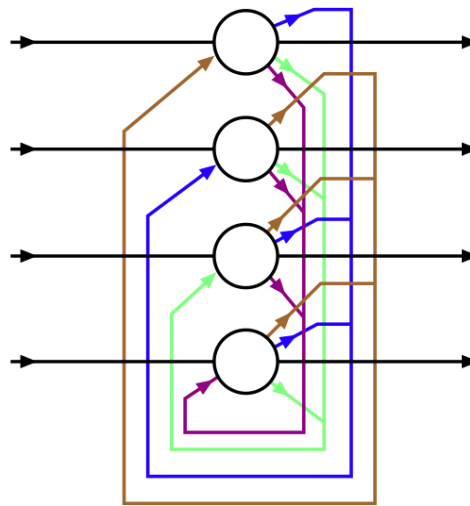


Ilustración 3.3: Red Hopfield de cuatro nodos (Fuente: Wikipedia)

- Redes neuronales profundas (DNN): Se denominan así a las ANN con una gran cantidad de capas ocultas y de neuronas por capa. Son capaces de resolver problemas muy complejos aunque también de mayor coste computacional. Sin embargo, puede presentar algunos problemas, como el sobreaprendizaje, haciendo que no sea capaz de trabajar con datos que no hayan sido presentados en el entrenamiento (o reduciendo su eficacia).
- Redes neuronales convolucionales (CNN): Estas redes, que son un tipo de DNN, están basadas en el concepto del Neocognitron, introducido por K. Fukushima en 1980 [26] y mejorado en 1998 por Y. LeCun et al. al introducir la propagación hacia atrás. Son las más utilizadas en la clasificación de imágenes, pues requieren de poco preprocesamiento en comparación con otros sistemas. Las capas ocultas que forman las CNN pueden ser: capas de convolución, capas ReLU, capa de Pool o Subsampling, y capa tradicional. Al ser estas redes uno de los objetos de estudio principales de este TFM, serán vistas con mayor detalle en el siguiente apartado.

- Redes neuronales recurrentes (RNR): son un grupo de redes neuronales especializadas en procesar datos secuenciales de longitud variable y con capacidad de incrementar el tamaño (longitud) de las secuencias [28]. A diferencia de las redes neuronales clásicas, tratan datos secuenciales de forma eficiente, recuerdan las salidas anteriores como entrada (es decir, sin mezclar información entre distintas ejecuciones), y son capaces de tratar secuencias muy largas, elemento a elemento, en vez de tratarlas al completo de una vez.

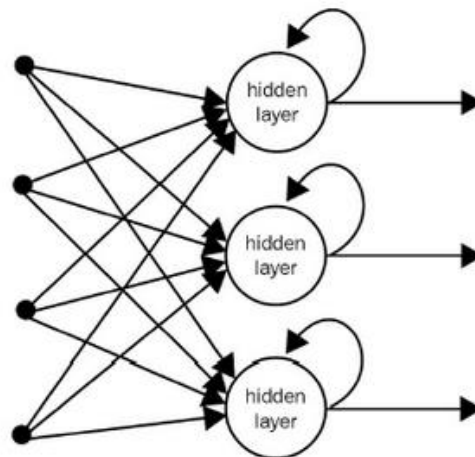


Ilustración 3.4: Ejemplo de red neuronal recurrente (Fuente: researchgate.net)

3.2 Redes neuronales convolucionales

Como hemos explicado anteriormente, son un tipo de ANN muy usadas en el procesamiento y clasificación de imágenes, con aprendizaje supervisado. Están formadas por varias capas ocultas especializadas y cuyas capas de convolución se agrupan formando una jerarquía, esto quiere decir que las primeras capas se encargarán de identificar aspectos más simples o genéricos, como líneas o curvas, y conforme avanzamos, las capas estarán especializadas en detalles o formas complejas, como rostros o siluetas. Estas capas, como indica su nombre, realizan convoluciones, que es la aplicación de un filtro mediante operaciones matemáticas realizada sobre la imagen. Para poder realizar estas operaciones, la imagen es interpretada por la CNN como una matriz formada por los píxeles de ésta.

A continuación, una vez se ha completado la fase de convolución, llegamos a una última capa oculta completamente conectada, donde se le aplica una función

encargada de clasificar cada imagen, por ejemplo, una de las más conocidas es *Softmax*, que da como salida un formato “one-hot-encoding” [29], según las clases con las que haya entrenado.

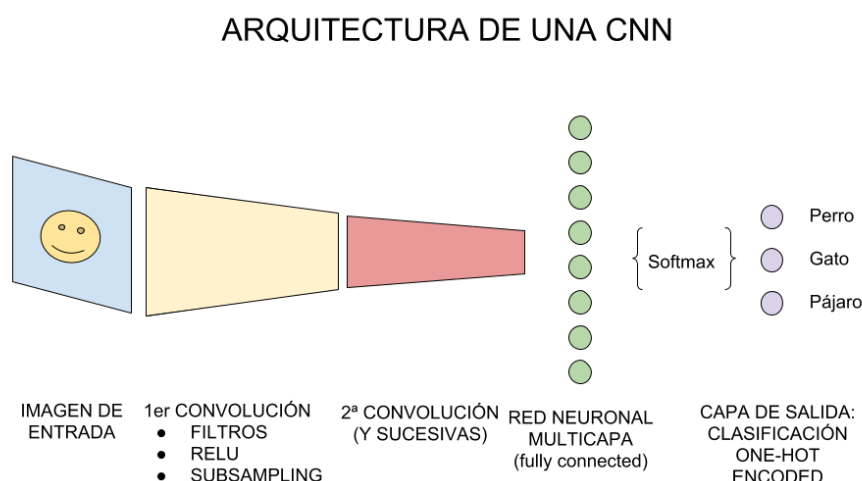


Ilustración 3.5: Arquitectura de una CNN (Fuente: www.aprendemachinellearning.com)

3.2.1 Entrenamiento de una CNN

Una vez se ha explicado cómo funciona una CNN a la hora de clasificar imágenes, vamos a ver en qué se basa el entrenamiento que hay previo y qué debemos tener en cuenta para entrenar una red de forma eficaz.

Una forma sencilla de explicar en qué consiste el entrenamiento de una red neuronal sería intentar que ésta pueda identificar algo (animal, edificio, rostro...), “enseñándole” muchas imágenes con ese concepto e “informándole” qué hay en cada imagen. Por ejemplo, si tomamos la red mostrada en la ilustración 3.5, en su entrenamiento se habrán usado imágenes de perros, gatos y pájaros. Normalmente, a mayor cantidad de imágenes mostradas mejor, pero hay que evitar caer en el **sobreaprendizaje**, como se ha mencionado anteriormente, en la que distinguimos entre sobreajuste u “overfitting”, y su antónimo, infrajuste o “underfitting”. Una recomendación para que esto no perjudique a una red, sería que las imágenes usadas en el entrenamiento sean variadas, es decir, imágenes tomadas con diferentes ángulos, y si el concepto puede presentar diversidad, intentar incluir la mayor variedad

posible. Por ejemplo, si a la red mostrada en la ilustración 3.5 solamente hubiera sido entrenada con imágenes de varios gatos, pero solo de color negro, se daría un caso de “underfitting”, donde posiblemente presente dificultades al identificar otras imágenes con gatos de un color distinto, ya que extraería, erróneamente, que el color es una cualidad de mucha importancia a la hora de identificar otro gato. Otro caso de sobreaprendizaje, en este caso “overfitting”, se daría cuando el entrenamiento ha sido entrenado con gran cantidad imágenes que pueden ser agrupadas entre sí, pero que no representan al conjunto completo que identifica el concepto en general, por poner un ejemplo similar al anterior, si la red a la que hemos hecho referencia, a la hora de entrenar perros, solo utilizamos imágenes de 5 razas de perros, podría fallar al intentar identificar una raza que no corresponda a ninguna del conjunto de entrenamiento. En la siguiente ilustración podemos ver, en forma de ecuación, que representaría el entrenamiento, un ejemplo de enternamiento correcto, de underfitting y de overfitting, donde los puntos marcados con un círculo representan una clase, y las equis, otra.

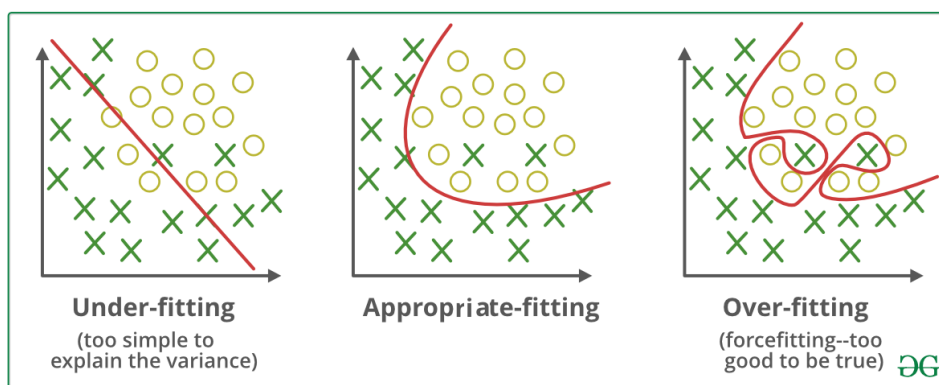


Ilustración 3.6: Ejemplos de underfitting, entrenamiento correcto y overfitting [34]

Como resultado de un entrenamiento, se obtiene lo que se conoce como un fichero de pesos, habitualmente con formato de datos jerárquico (HDF) [32], que representa la función que clasifica las distintas clases (Como por ejemplo las que se muestran en la ilustración 3.6). Estos ficheros son utilizados como parámetros de entrada cuando se desea clasificar un conjunto de imágenes de las mismas categorías que las del entrenamiento al que pertenece.

A continuación, se verá, de forma resumida, las dos fases principales de la que consta el entrenamiento de una red neuronal. En primer lugar, tenemos la fase de

propagación hacia adelante (forward propagation) y, en segundo lugar, propagación hacia atrás (backward propagation o backpropagation). Estos dos pasos son claves, pues son donde el algoritmo itera para “aprender”, ya que, al finalizar el segundo, se realiza la actualización de los pesos, que, como hemos hablado anteriormente, es lo que define el “aprendizaje” de una red.

3.2.1.1 Propagación

Como hemos indicado, consta de dos pasos:

1. Propagación hacia adelante: Como indica su nombre, los datos de entrada fluyen hacia adelante en la red (En dirección a la capa de salida). Cada capa oculta recibe los datos de entrada, los procesa según la función de activación y pasa a la capa sucesiva.
2. Propagación hacia atrás: En este paso es donde se reajustan los pesos de la red neuronal, basada en la tasa de error que se haya obtenido en la propagación hacia adelante. Dicho ajuste de pesos debe procurar que, al realizar otra iteración, se mejore los resultados.

3.2.2 Elementos de una Red Neuronal Convolutacional

Como hemos visto, una CNN es un tipo de DNN, por lo que mantiene sus tres grupos de capas característicos, donde lo característico de cada una se encuentra en las capas ocultas. Además de estos elementos principales, nos encontramos también con los parámetros e hiperparámetros, y dentro del primer grupo, las funciones de activación de las neuronas, que, como dijimos anteriormente, definen su salida. En las CNN, la función de activación que suele usarse es conocida como ReLU (Unidad lineal rectificada), que se tratará con más detalle en capítulos posteriores.

Hacemos distinción entre parámetros e hiperparámetros, considerando que dentro del primer grupo se encuentran todas las variables que son definidas durante

el entrenamiento de un modelo, por lo que son exclusivos para el propio modelo, ya que no serán modificados por ninguna persona. En cambio, entendemos hiperparámetros como la configuración a nivel usuario utilizada durante el proceso de entrenamiento. Es difícil conocer a priori cuales son los hiperparámetros óptimos, por lo que suelen establecerse valores genéricos, y optimizarlos a posteriori mediante ensayo y error.

3.2.2.1 Capas convolucionales

Como indica su nombre, son las que definen a las CNN. En ellas, se realizan convoluciones, que son operaciones matemáticas entre la capa de entrada y un filtro. La ventaja de esto, es que cada neurona, con su filtro correspondiente, puede extraer la misma característica en varias partes de la imagen de entrada, con lo que se reduce el coste computacional en comparación con una red profunda de otro tipo. A diferencia de otras capas, que están completamente conectadas, en las capas convolucionales cada unidad está conectada a un número constante de nodos de la capa anterior, compartiendo, algunas conexiones, los mismos pesos. Podemos hacernos una idea de esto con la siguiente ilustración, donde las líneas del mismo tipo comparten pesos.

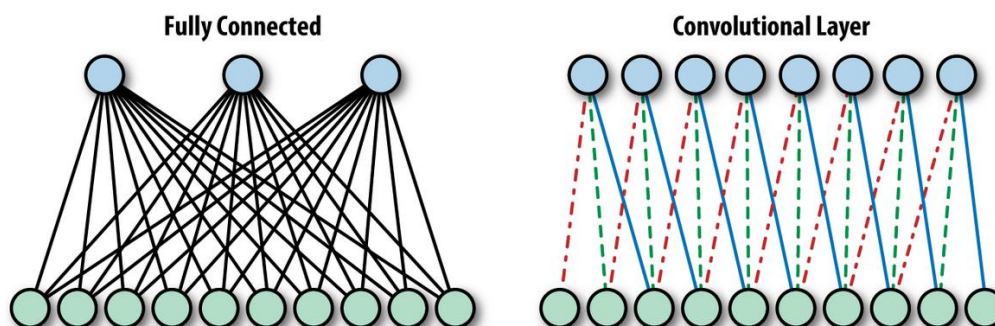


Ilustración 3.7: Comparación entre capas completamente conectadas y capas convolucionales [35]

Para este tipo de operaciones, y de las que veremos más adelante, hay que tener en cuenta que las imágenes pueden ser representadas en formato matricial, donde cada casilla representa a un pixel mediante un valor numérico entre el 0 y el 255, y teniendo en cuenta que, si la imagen es en blanco y negro, un valor inferior representaría un tono más cercano al blanco, y un valor más elevado, estaría más

cercano al negro. En cambio, en las imágenes a color, cada imagen suele dividirse en otras tres imágenes, representando, cada una, uno de los tres colores primarios en la luz (rojo, verde y azul), y comportándose de forma similar a las imágenes en blanco y negro (un número más elevado, simbolizaría una mayor saturación del color en cuestión). En lo referente al uso de imágenes en color para entrenamiento y clasificación, suele aplicarse, o bien un preprocesamiento, que divida y pueda trabajar en los tres canales de colores, o bien, realizando una media entre los tres canales de colores, para obtener una sola imagen en blanco y negro.

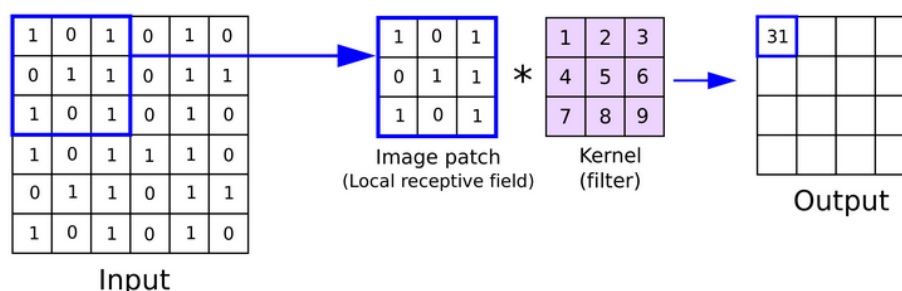


Ilustración 3.8: Ejemplo de operación de convolución (Fuente anhreynolds.com)

La operación de convolución recorre toda la imagen (en formato de matriz numérica), comenzando en una esquina de la imagen y trasladando el filtro hasta recorrerla al completo. Además de realizar la convolución, es posible configurar varios hiperparámetros que influirán, de forma positiva o negativa, sobre el resultado de la convolución, como las que analizaremos a continuación:

- **Tamaño y número de filtros:** El tamaño de un filtro indica cuántas casillas tiene la matriz del filtro que realiza la convolución, a mayor número de casillas, mayor número de píxeles se calcularán por convolución. El número de filas del filtro siempre tiene que ser el mismo que el de columnas, es decir, el filtro debe ser una matriz cuadrada. Por norma general, el filtro óptimo suele ser de 3x3, como el de la ilustración 3.8. Otros tamaños, como el 1x1, que podría ser utilizado para la reducción de dimensionalidad, es decir, para limitar el número de canales de una imagen, sin embargo, como no extrae características, suele descartarse. Otras opciones, como los filtros con tamaños pares (2x2, 4x4...) no suelen resultar de utilidad, porque, a

diferencia de los filtros de tamaño impar, no dividen simétricamente la imagen, por lo que puede crearse una distorsión en capas sucesivas. Debido a esto, en caso de querer un filtro distinto al 3x3, buscaremos unos distintos a los que hemos mencionado, como 5x5, 7x7... Del mismo modo, en algunos nodos puede aplicarse varios filtros en lugar de uno solo, que suelen dar mejor resultado que usar un filtro grande. Con esto conseguiríamos extraer un mayor número de características en un solo paso.

- Padding y Stride:** Uno de los mayores inconvenientes al recorrer la matriz de una imagen es que, en los límites de ésta, faltan datos, ya que, al contrario que en el resto de casillas, que disponen de 9 casillas vecinas (correlativas en horizontal, vertical y diagonal), solo disponen de 3 casillas vecinas en las esquinas, y 5 en los laterales. Para solucionarlo, tenemos los que se conoce como “Padding” (relleno), que consiste en expandir la matriz de la imagen con casillas que contengan un 0. Podemos ver como ejemplo la ilustración 3.9, donde una matriz de tamaño 6x6 se le aplica un *padding* de 1, transformándola en una matriz 8x8, y solucionando el cálculo de los píxeles en los extremos. Otra de las configuraciones posibles sería el “Stride” (paso), que simplemente define el número de casillas que se recorre en cada aplicación de filtro. Normalmente se usa un Stride de 1, es decir, recorriendo la matriz casilla a casilla, pero podemos cambiarlo a nuestra voluntad.

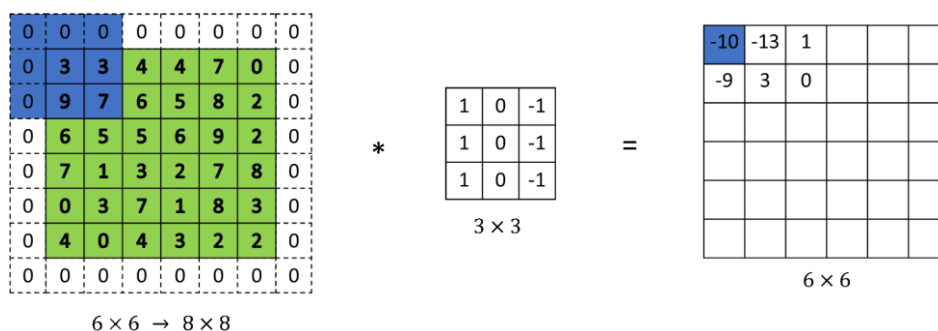


Ilustración 3.9: Aplicación de padding o relleno (Fuente: datahackers.rs)

3.2.2.2 Capas de reducción de muestreo o submuestreo (Pooling)

El objetivo de estas capas es, simplemente, reducir la resolución de una imagen, agrupando sus píxeles. El filtro más utilizado en esta operación es de tamaño 2x2, reduciendo, en cada agrupación, 3 píxeles, es decir, el resultado final será una imagen con una resolución equivalente a $\frac{1}{4}$ de la resolución original, pero seguirá manteniendo las características necesarias para su clasificación a través de variables. Esto lo convierte en una operación muy útil para reducir, en gran medida, el coste computacional. Las técnicas más comunes a la hora de elegir de qué manera se va a aplicar este filtro son: **max pooling**, donde el pixel resultante de cada grupo será el de mayor valor del grupo, y **avg pooling**, donde el resultado será la media de cada grupo. En la ilustración 3.10 podemos ver un ejemplo, tanto a nivel matricial como visual, el resultado de aplicar estas técnicas de *pooling* a la misma imagen.

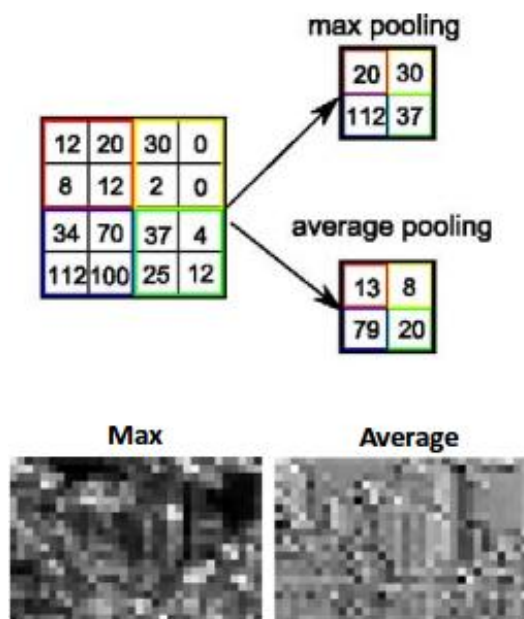


Ilustración 3.10: Ejemplos de Max y Avg Pooling (Fuente: www.quora.com)

Del mismo modo que hemos visto con los filtros de convolución, también es posible variar a través de hiperparámetros el tamaño del filtro, el relleno o *padding*, y los pasos o *stride*, y, al contrario que con los filtros de convolución, no disponen de parámetros que tengan que ser actualizados durante el entrenamiento.

3.2.2.3 Funciones de activación

Ya hemos visto en otros capítulos que son usadas para transformar la salida de un nodo, extrayendo características para poder entrenar o clasificar. A continuación, explicaremos en qué se basan las más frecuentes y su utilización:

- **Logística o Sigmoide:** Esta función, que presenta una curva en forma de letra “S”, transforma los valores de entrada en un número entre 0 y 1 (ambos inclusive). Suele ser utilizada en problemas de clasificación binarios, aunque su uso no es tan común como otras funciones, ya que puede causar que la red se atasque durante el proceso de entrenamiento al no contemplar casos de números negativos.

$$P(t) = \frac{1}{1 + e^{-t}}$$

- **Tanh:** Denomina a la función de **tangente hiperbólica**, definiendo su salida como el cociente entre seno hiperbólico y coseno hiperbólico del número de entrada. Como podemos ver en la ilustración 3.12, describe una curva muy similar a la función logística, siendo la principal diferencia entre éstas que la tangente hiperbólica da como salida un número entre -1 y 1, siendo más utilizada que la anterior en problemas de clasificación binarios, debido a que no presenta el inconveniente de atascar la red en el entrenamiento.

$$\tanh x = \frac{\sinh x}{\cosh x}$$

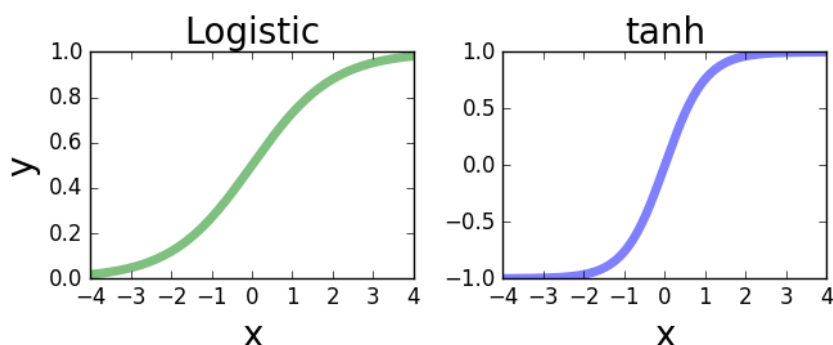


Ilustración 3.11: Comparación entre función logística y de tangente hiperbólica (Fuente: www.oreilly.com)

- **ReLU**: Actualmente es la función de activación más utilizada [36] en redes neuronales convolucionales. Su función da como salida el mismo número de entrada, si éste es mayor o igual a 0, y 0 en caso contrario, siendo éste también su punto negativo, pues no es capaz de mapear correctamente valores negativos. Existe una variante de esta función, llamada **Leaky ReLU**, que pretende corregir este problema. Las siguientes funciones hacen referencia a “ReLU” (izquierda) y “leaky ReLU” (derecha):

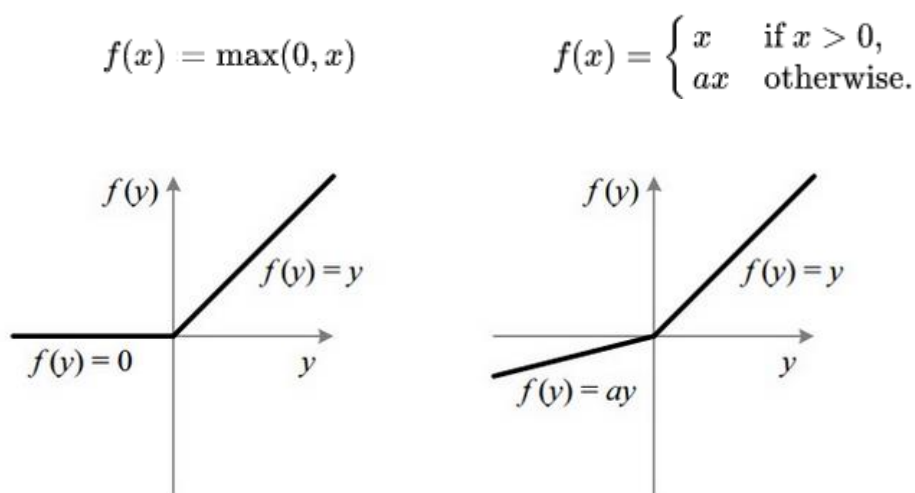


Ilustración 3.12: Comparación entre función ReLU (izquierda) y leaky ReLU (derecha) [36]

- **Softmax**: La regresión softmax (o regresión logística multinominal) es una generalización de la función logística, con la que se permite la clasificación de múltiples clases. Es por tanto más utilizada en la clasificación de redes neuronales que la función logística o la de tangente hiperbólica, pues éstas

solo permitían distinguir entre dos clases. Al igual que la función logística, el rango de salida de un número estará entre 0 y 1, sin embargo, se cumple que la suma de todas las salidas será igual a 1, por lo que podemos interpretarlas como **porcentajes**, es decir, la probabilidad que tiene una imagen de pertenecer a una clase u otra. Su función viene dada por:

$$\sigma : \mathbb{R}^K \rightarrow [0, 1]^K$$

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad \text{para } j = 1, \dots, K.$$

3.2.2.4 Configuración de hiperparámetros adicionales y consideraciones para evitar el sobreaprendizaje

Adicionalmente se pueden establecer otros parámetros que establecerá el funcionamiento de una red cada vez que la ejecutemos. Estos son:

- **Tamaño de lote (batch size):** hace referencia al número de imágenes que tomará el algoritmo en cada paso. Además de afectar al coste computacional, también puede alterar la precisión en el cálculo de los pesos, perjudicándola si los lotes son demasiado pequeños. Hay que tener también en cuenta que, si el tamaño de lote que establecemos no es divisor del número total de muestras, el último lote será de tamaño inferior al resto. Esto no significa necesariamente que pueda afectar demasiado al resultado final, pero podría hacerlo si la diferencia de tamaño entre el último lote y el resto es muy grande.
- **Etapas (epochs):** define el número de repeticiones que realizará el algoritmo de aprendizaje con el conjunto de aprendizaje completo. Cada vez que se complete una etapa los pesos serán actualizados, optimizándolo al conjunto con el que está entrenando. Sin embargo, no significa que a mayor número de etapas obtengamos un mejor resultado, pues podríamos caer en el sobreaprendizaje, donde un número muy bajo de etapas nos acercaría a un sobreajuste, mientras que un número muy alto nos dará un infrajuste.

Deberemos encontrar un número adecuado de etapas que pueda devolvernos un modelo cercano al óptimo (ver ilustración 3.6).

- **Dropout:** Es una técnica que consiste en que, de forma aleatoria, se desactivará un porcentaje (establecido por el usuario) de los nodos de la red. Los nodos desactivados variarán en cada etapa. A pesar de que puede parecer a priori una forma de reducir la calidad de la red, su uso ayuda a evitar el sobreajuste, además de reducir el coste computacional de ésta.

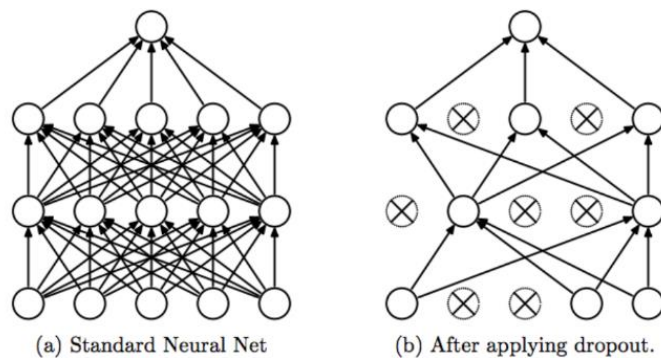


Ilustración 3.13: Comparación entre modelos antes y después de aplicar dropout (Fuente: www.medium.com)

3.2.2.5 Aumentación de datos o “Data Augmentation”

Por último, vamos a ver una técnica bastante utilizada para mejorar el entrenamiento de una red. Se trata de “crear” nuevas muestras a partir del propio conjunto de imágenes con el que estemos trabajando. A pesar de que las imágenes creadas son modificaciones de las originales, suele conseguirse una mejora de la precisión del modelo a la hora de clasificar nuevas imágenes, además, también puede ayudar a evitar el sobreaprendizaje.



Ilustración 3.14: Ejemplo de Data Augmentation aplicando distintas técnicas a una imagen
(Fuente: www.mygreatlearning.com)

Para realizar este aumento de datos, a las imágenes se le aplican pequeños cambios, como rotaciones, translaciones, cambios en el color, texturas o contrastes... Estos cambios son más eficaces si son aplicados en cada entrenamiento o en cada etapa, evitando que las nuevas imágenes puedan perjudicar el resultado final.

4 ARQUITECTURAS Y TECNOLOGÍAS USADAS

4.1 Arquitecturas utilizadas

A continuación, se da un repaso teórico de las cuatro CNN que han sido elegidas para formar parte de esta investigación. Si bien es cierto que no se puede considerar que una red sea mejor que otra, pues entran en juego muchísimos factores, como, por ejemplo, el propio conjunto de imágenes con el que trabajan, se intenta buscar que sus resultados sean competentes en conjuntos de imágenes muy conocidos en lo referente a Deep Learning, como ImageNet.

Se destaca como curiosidad la comparación que se muestra en el artículo “Xception: Deep Learning With Depthwise Separable Convolutions” por F. Chollet, 2017, que involucra las cuatro redes seleccionadas en este TFM, y donde el modelo Xception da el mejor resultado (Ver tabla 4.3). Adicionalmente, puede verse una comparación de precisión Top-1 y Top-5 de todos los modelos disponibles en la librería Keras, para el conjunto ImageNet, en la siguiente página: <https://keras.io/api/applications/>.

Para entender a qué se refiere los valores Top-1 y Top-5 como referencia para medir cómo de bueno es un modelo, vamos a tomar como ejemplo una red que categoriza imágenes de animales. Imaginemos que, valorando un conjunto de imágenes, al predecir una imagen de un perro, obtenemos la siguiente salida:

- Gato: 0,5
- Ratón: 0,25
- Perro: 0,2
- Oso: 0,03
- León: 0,01
- Caballo: 0,006
- Jirafa: 0,004

Este resultado muestra cuánto se aproxima la imagen a cada categoría con la que ha entrenado la red, dicho de otra forma, la medida en que la red “piensa” que la imagen se corresponde con cada categoría.

En este caso, usando la medida Top-1, esta predicción sería errónea, pues la de mayor porcentaje no se corresponde con la categoría real de la muestra, sin embargo, si usamos la medida Top-5, la predicción sería correcta, puesto que dentro de las cinco categorías con mayor porcentaje se encuentra la correcta.

Como razón adicional de seleccionar varias CNN para la investigación proponemos realizar un ensamblaje de resultados de los distintos modelos, es decir, sacar resultados conjuntos seleccionando las predicciones en las que coincidan el mayor número de redes.

4.1.1 Resnet50

Las redes ResNet (Acrónimo de “Residual Net”, o Red Residual) se caracterizan por ser de las que más capas usan, llegando a encontrar entre sus distintos modelos el ResNet152, que usa 152 capas, distanciándose de otras redes conocidas, como la GoogLeNet, que usa 22 [9].

Para los diferentes modelos que usan esta arquitectura, se parte de una imagen 112x112, que pasa por 5 etapas de extracción de características hasta reducirse a una imagen 7x7, y, por último, por una capa *softmax*. Podemos observar esta arquitectura en la siguiente ilustración.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Ilustración 4.1: Arquitectura ResNet [17]

En este estudio usaremos el modelo ResNet50 (De 50 capas), que es una versión reducida del ResNet152, pues las versiones de 101 y 152 requieren de una capacidad de cómputo muy superior. Aún sin ser el que mejor resultados da, sigue dando valores muy aceptables. En la siguiente tabla podemos ver una comparación de varios modelos y arquitecturas con las medidas de error Top-1 y error Top-5 con el conjunto de datos ImageNet.

Red	Error Top-1	Error Top-5
VGG	24,4%	7,1%
PReLU-net	21,59%	5,71%
BN-Inception	21,99%	5,81%
ResNet-34 B	21,84%	5,71%
ResNet-34 C	21,53%	5,60%
ResNet-50	20,74%	5,25%
ResNet-101	19,87%	4,60%
ResNet-152	19,38%	4,49%

Tabla 4.1: Comparación de errores Top-1 y Top-5 de varios modelos y arquitecturas para el conjunto ImageNet [17]

4.1.2 InceptionV3

Esta CNN surge como propuesta de investigadores de Google a partir de la arquitectura VGG (Se profundizará sobre esta arquitectura en el siguiente punto) como un nuevo estilo de red, a la que apodan GoogLeNet (También conocida como InceptionV1), posteriormente es perfeccionada y reconocida como la siguiente versión, InceptionV2[16], y más adelante, como InceptionV3[10]. La CNN Inception es conocida por pertenecer a una de las familias de modelos con mejor rendimiento en el conjunto de imágenes “ImageNet” [14], así como en los conjuntos internos en uso de Google [15].

La característica principal de los modelos que se basan en esta arquitectura se encuentra en su módulo “Inception”, del cual existen diferentes versiones. En la siguiente ilustración se muestra la forma canónica de un módulo “Inception” [12]

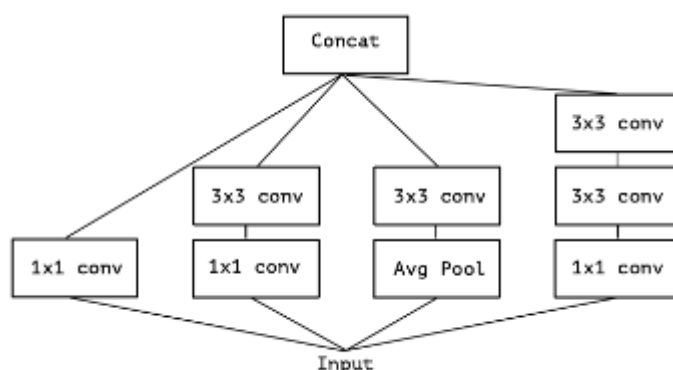


Ilustración 4.2: Forma canónica de un módulo Inception[12]

Para este estudio, hemos elegido la versión InceptionV3, ya que de las conocidas dentro de la familia Inception, suele ser la que mejores resultados da, tal y como demuestra, por ejemplo, la experimentación realizada en el artículo “*Rethinking the Inception Architecture for Computer Vision*” por C. Szegedy et al., 2015, donde, para valorar varios modelos (VGGNet, GoogLeNet, PReLU, BN-Inception, e InceptionV3), se utilizaron 50000 imágenes del conjunto ILSVRC-2012, y valoraron el porcentaje de error de Top-1 y Top-5 de los resultados obtenidos.

Red	Modelos Evaluados	Error Top-1	Error Top-5
VGGNet	2	23,7%	6,8%
GoogLeNet	7	-	6,67%
PReLU	-	-	4,94%
BN-Inception	6	20,1%	4,9%
Inception-v3	4	17,2%	3,46%

Tabla 4.2: Comparación de errores Top-1 y Top-5 de varios modelos para el conjunto ILSVRC-2012 [10]

4.1.3 Xception

La CNN Xception es propuesta por investigadores de Google, su nombre significa “Extreme Inception” [12], en referencia a que usa una versión “extrema” del módulo Inception, usado en la arquitectura que hemos explicado anteriormente.

Podemos apreciar en la ilustración 4.3 que, en comparación con el módulo Inception presentado en el apartado correspondiente al modelo Inception (Ilustración 4.2), se aumenta el número de módulos convolucionales, razón por la que recibe el apodo de “Inception extremo”.

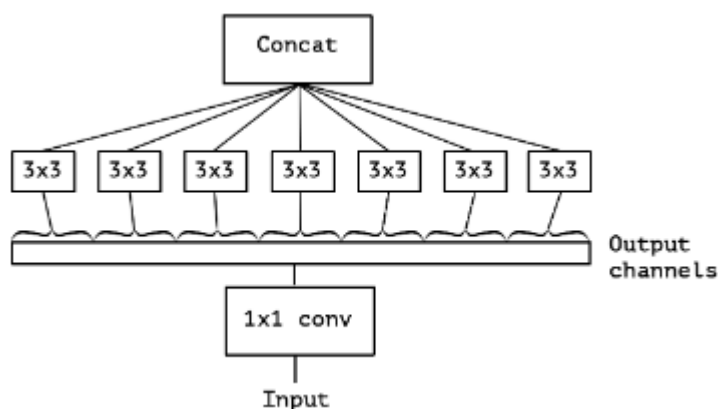


Ilustración 4.3: Versión de módulo Inception para el modelo Xception [12]

Como dato adicional, se incluye una comparación realizada en el mismo artículo de F. Chollet, donde comparan las cuatro CNN que se presentan en este capítulo: ResNet (En este caso la versión de 152 capas), InceptionV3, Xception y

VGG-16, usando las medidas de precisión Top-1 y Top-5, dando como ganador a la red Xception.

Red	Precisión Top-1	Precisión Top-5
VGG-16	71,5%	90,1%
ResNet-152	77,0%	93,3%
Inception V3	78,2%	94,1%
Xception	79,0%	94,5%

Tabla 4.3: Comparación de precisión Top-1 y Top-5 de varios modelos para el conjunto ImageNet [12]

4.1.4 VGG-16

Este modelo de CNN fue propuesto por K. Simonyan y A. Zisserman de la Unidad de Oxford, con el que alcanzaron una precisión de 92,7% en la competición *ImageNet* en el año 2014. Supuso una mejora respecto a la red AlexNet al sustituir algunos filtros. VGG-16 está dentro de un conjunto de redes similares donde varían el número de capas convolucionales, desde la VGG-11 hasta la VGG-19 [13].

Esta arquitectura se asemeja a la que vimos anteriormente con la arquitectura ResNet, como veremos a continuación.

En todas las versiones VGG, se parte de una imagen de tamaño 224x224, que pasa por 5 etapas de extracción de características hasta reducirse a un tamaño de 7x7. Una vez la imagen se ha reducido, pasa por otras tres capas interconectadas, y, por último, por la capa de salida *softmax*. Todo este proceso se puede observar en la siguiente ilustración.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Ilustración 4.4: Tabla de arquitecturas VGG [13]

En este proyecto se trabajará con la arquitectura mostrada en la columna D, de 16 capas, que es la más común dentro de las arquitecturas VGG y la que mejores resultados suele obtener.

4.2 Tecnologías utilizadas

Como recurso principal para la realización de este estudio se ha utilizado, para la realización tanto del entrenamiento del conjunto de imágenes como de la prueba “test”, el sistema de computación de altas prestaciones ADA, pues debido al tamaño del conjunto, y al de las propias imágenes, dificulta en gran medida la realización de diferentes pruebas en un equipo doméstico común, donde el tiempo tomado para la realización de pruebas llegaba a superar la semana. En cambio, el equipo doméstico

ha sido utilizado para la descarga y valoración de conjuntos de imágenes, estudio de las redes a utilizar, conexión remota segura (SSH) al sistema de computación ADA, el ensamblaje de resultados, elaboración de estadísticas y redacción de la memoria.

4.2.1 Sistema de computación de alto rendimiento ADA.

Ada es un sistema de computación de altas prestaciones basado en procesadores Intel Xeon Silver 4208 junto con tarjetas gráficas NVIDIA Tesla V100 y NVIDIA GeForce RTX 2080Ti [8].

Está dividido en 7 nodos, un nodo principal, donde se sitúan las carpetas personales de los usuarios, y que ejecutará los comandos lanzados directamente por éstos; 4 nodos de cómputo RTX, y 2 nodos de cómputo TESLA. Éstos últimos organizan las tareas mandadas por cada usuario mediante colas de trabajo con el sistema de gestión de tareas y clústeres “Slurm”. Las características de cada nodo se muestran en las siguientes tablas:

Nodo principal (1 nodo)	
Plataforma	R281-3CO
Procesador	2x Intel(R) Xeon(R) Silver 4210 CPU @ 2.20GHz
Memoria RAM	96 GB DDR4 @ 2933 Mhz
Almacenamiento	60 TB (almacenamiento compartido)

Tabla 4.4: ADA – Nodo principal [8]

Nodos de cómputo TESLA (2 nodos)	
Plataforma	G291-281
Procesador	2x Intel(R) Xeon(R) Silver 4208 CPU @ 2.10GHz
GPU	2x NVIDIA Tesla V100 (32GB, Volta)
Memoria RAM	192 GB DDR4 @ 2933 Mhz
Almacenamiento	1.8 TB

Tabla 4.5: ADA – Nodo de cómputo TESLA [8]

Nodos de cómputo RTX (4 nodos)	
Plataforma	G291-281
Procesador	2x Intel(R) Xeon(R) Silver 4208 CPU @ 2.10GHz
GPU	6x* NVIDIA GeForce RTX 2080Ti (11GB, Turing) *De los cuatro nodos, dos disponen de 6 GPUs, y los otros dos de 7 GPUs. Sumando un total de 26 GPUs RTX.
Memoria RAM	192 GB DDR4 @ 2933 Mhz
Almacenamiento	1.8 TB

Tabla 4.6: ADA – Nodo de cómputo RTX [8]

Debido al conjunto de imágenes utilizado en el estudio, se opta por el uso de los nodos TESLA, ya que la potencia ofrecida por los nodos RTX era insuficiente para el procesamiento del conjunto al completo.

4.2.2 Librerías utilizadas

Para el desarrollo de esta investigación se ha utilizado principalmente el lenguaje de programación Python (A excepción del modelo YOLO, programado en C, que fue descartado, como se explica en el capítulo 5), que ofrece gran cantidad de recursos a la hora de trabajar en el campo del Deep Learning. A continuación, se muestra una lista de las librerías más importantes con las que ha sido posible las diferentes operaciones de entrenamiento y prueba de modelos:

- Tensorflow: Se ha usado la versión que trabaja con GPUs. Junto con la siguiente, esta librerías de bajo nivel, desarrollada por Google, es una principales a la hora de trabajar en deep learning. Ofrece un rango de tareas para satisfacer las necesidades de sistemas orientados a la construcción y entrenamiento de redes neuronales artificiales.
- Keras: Como hemos mencionado, esta librería de alto nivel, desarrollada por el ingeniero de Google F. Chollet, es otra de las principales en el campo

del aprendizaje profundo. Representa además un gran apoyo para las CNN y RNR.

- Scikit-learn: También se trata de una biblioteca centrada en el aprendizaje automático, que incluye algoritmos de clasificación, regresión y análisis de grupos. Está diseñada para interoperar con las bibliotecas numéricas y científicas Numpy y Scipy, que nombraremos a continuación.
- Numpy: Esta librería es una de las más conocidas en el lenguaje Python, pues permite crear vectores y matrices multidimensionales, además de incluir numerosas funciones matemáticas.
- Scipy: Se compone de más herramientas y algoritmos matemáticos, en concreto, enfocados a la optimización, álgebra lineal, integración, funciones especiales, y procesamiento de imágenes.
- H5PY: Gracias a esta librería, podemos guardar los pesos generados al realizar el entrenamiento de un modelo en formato HDF, pudiendo realizar predicciones con conjuntos que no pertenezcan al entrenamiento.
- Pillow: Nos facilita la apertura y manipulación de imágenes en Python, necesario para el procesamiento que sea requerido al usar modelos de CNN.

5 EXPERIMENTACIÓN Y RESULTADOS

A continuación, se explicará: qué conjunto de imágenes se ha trabajado y cómo; la experimentación llevada a cabo; qué resultados hemos obtenido, y qué podemos interpretar de éstos.

Durante la revisión teórica encontramos dos conjuntos de imágenes tomadas vía satélite, con una gran cantidad de imágenes que creemos de interés para la realización de las pruebas. La primera es FMoW (Functional Map of the World) [6], creada por la agencia norteamericana de Actividad de Proyectos de Investigación Avanzados de Inteligencia (IARPA), y la segunda, xView [42]. Ambas nos ofrecen un modelo base, para poder empezar a utilizar sus imágenes, lo que es un aspecto de gran interés, pues tendríamos un modelo capacitado para empezar a trabajar con cada conjunto de imágenes sin tener que empezar de cero. Sin embargo, finalmente nos decantamos por llevar a cabo la experimentación con el primer conjunto, FMoW. La razón por la que elegimos éste y no el otro es que, a diferencia de xView, cada imagen viene acompañada de un fichero JSON, que nos proporciona información relativa a la imagen y podremos usarla, en caso de que fuese relevante, para el entrenamiento y clasificación, como, por ejemplo:

- Sus coordenadas en el mundo (latitud y longitud).
- La región de interés, es decir, dónde se sitúa el elemento identificado, mediante su posición dentro de la imagen.
- La menor o mayor cantidad de nubes de forma numérica, donde 0 sería despejado, y un número mayor implicaría más concentración de nubes.
- La fecha y hora, de cuando se tomó la fotografía.
- País donde se sitúa.
- El ángulo azimut (0-360°) respecto al sol, es decir, su ángulo de orientación en la superficie de la tierra hacia éste.
- El ángulo de elevación del sol (0-90°), es decir, la distancia angular vertical entre el horizonte y el sol desde el centro de la imagen.

Este conjunto dispone de 1.037.238 imágenes y, al ser un conjunto utilizado para competiciones de Deep Learning, ya se encuentra dividido en tres subconjuntos, uno de entrenamiento, otro de validación, y otro de prueba o test. Los dos primeros subconjuntos están divididos por categorías, mientras que el último no, sin embargo, con la finalización del concurso, podemos obtener un fichero con las categorías de todas las imágenes de este subconjunto, que nos servirá para poder valorar los resultados.

Las 62 categorías que forman este conjunto de imágenes, junto con el número de imágenes que contiene cada una en el conjunto de entrenamiento y validación se muestran en la siguiente tabla:

Categoría	Número de imágenes
Aeropuerto	3.532
Hangar de aeropuerto	21.008
Terminal de aeropuerto	19.272
Parque de atracciones	13.430
Acuicultura	3.774
Zona de arqueología	6.088
Granero	23.794
Puesto fronterizo	3.506
Cementerio	9.694
Concesionario de coches	11.998
Zona de construcción	13.892
Campo de cultivo	56.890
Represa	9.892
Escombros	4.866
Institución educativa	27.750
Subestación eléctrica	12.972
Fábrica o planta eléctrica	8.420
Estación de bomberos	12.134
Camino inundado	8.788
Fuente	13.232

Gasolinera	12.834
Campo de golf	24.200
Estación de transporte terrestre	46.132
Helipuerto	10.180
Hospital	12.288
Cruce	16.324
Lago o estanque	1.996
Faro	8.534
Instalación militar	29.458
Residencial colectivo	17.870
Estación nuclear	510
Edificio de oficinas	13.352
Instalación petrolera	20.660
Parque	11.454
Aparcamiento o garaje	30.808
Lugar de culto	44.396
Estación de policía	12.126
Puerto	4.685
Prisión	10.128
Pista de carreras	12.804
Puente ferroviario	19.912
Instalaciones recreativas	66.228
Asentamiento empobrecido	3.118
Puente de carretera	22.436
Pista de aterrizaje	10.560
Astillero	2.096
Centro comercial	13.588
Vivienda unifamiliar	24.772
Chimenea	11.476
Granja solar	11.582
Instalación espacial	516
Estadio	13.556
Tanque de almacenamiento	12.676
Mina de superficie	11.754

Piscina	19.406
Cabina de peaje	22.468
Torre	16.974
Apertura de túnel	18.851
Vertedero	9.834
Instalación de tratamiento de aguas	12.526
Granja Eólica	11.842
Zoológico	3.724

Tabla 5.1: Categorías y número de imágenes que contienen en entrenamiento y validación
(Fuente: Elaboración propia)



Ilustración 5.1: Ejemplos de FMoW, de arriba abajo y de izquierda a derecha: Estación de bomberos, estadio, parque de atracciones, estación nuclear, aeropuerto, zona de arqueología, cementerio, concesionario de coches, represa, campo de golf, prisión, lugar de culto, asentamiento empobrecido y parque. (Fuente: Conjunto de imágenes FMoW)

Además de usar el modelo propuesto por FMoW, se intenta experimentar con un modelo YOLO en versión 3, cuya principal característica es el etiquetado y señalización de posibles objetos de interés en la imagen. Sin embargo, tuvimos que descartar esta experimentación, pues, bien por el tipo de imágenes o bien por el tipo de arquitectura, nos vemos obligados a usar un entrenamiento algo simplificado debido a que el coste computacional de este modelo, respecto al equipo disponible,

era incompatible, y, por ello, no era posible realizar un entrenamiento estándar. Como resultado, no era capaz de identificar nada en la mayoría de imágenes, o marcaba muchos falsos positivos en otras.

5.1 Pruebas preliminares

En primer lugar, nuestra intención es identificar qué configuraciones y arquitecturas pueden resultar más interesantes para realizar diferentes experimentaciones.

Para empezar, decidimos que en estas pruebas y en las siguientes, el conjunto usado para construir el modelo estará formado tanto por el conjunto de entrenamiento como por el de validación, y que sea durante el propio entrenamiento donde se realice una división para formar ambos subconjuntos de forma aleatoria en cada etapa, consiguiendo una validación cruzada con k pliegues, donde k será el número de etapas. Definimos, por tanto, que, en cada etapa, el conjunto de validación estará formado un número de imágenes que será igual a la división del número de imágenes entre el tamaño del lote que sea establecido.

En cuanto a las redes que planteamos para la experimentación, elegimos las siguientes: Densenet, ResNet50, InceptionV3, Xception y Vgg-16 (Ver capítulo 4.1), sin embargo, nos vemos en la obligación de descartar la red Densenet, debido a que nos pide un coste computacional que no nos es posible alcanzar con los recursos de los que disponemos, impidiéndonos finalizar sus entrenamientos.

En cuanto al resto de modelos, para decidir un número apropiado de etapas, realizamos diferentes entrenamientos entre 8 y 16 etapas (epochs), aunque solamente con etapas pares para poder acotar un número sin tener que consumir demasiado tiempo. Decidimos establecer 12 etapas para todos los modelos, siendo, por tanto, también el número de pliegues en la validación cruzada, pues según los resultados, se comportan de manera similar, y, a partir de 12, la diferencia en la precisión en las siguientes etapas durante el entrenamiento es mínima, pudiendo ser señal de que el modelo pueda sufrir un sobreajuste. Obtenemos entonces, que cada subconjunto de

validación de cada etapa contendrá unas 86.376 imágenes. Para decidir el tamaño del lote, nos guiamos por encontrar un equilibrio entre poder realizar un buen entrenamiento sin alcanzar un coste computacional difícil de conseguir. Por tanto, se decide que el tamaño del lote sea 80. En cuanto a otros parámetros, utilizaremos, como optimizador, Adam [43], y, como métrica para los resultados, la precisión predictiva. Los resultados que obtenemos en el entrenamiento de cada red son los siguientes:

Red	Precisión (Accuracy)
Resnet50	0,9727
Xception	0,9935
InceptionV3	0,9558
Vgg-16	0,8614

Tabla 5.2: Predicciones obtenidas durante el entrenamiento (Fuente: Elaboración propia)

Debemos interpretar estos resultados teniendo en cuenta que son obtenidos con las imágenes usadas durante el entrenamiento, por lo tanto, podrían diferir al predecir otros conjuntos distintos (como el de prueba, que usaremos más adelante), incluso es muy probable que varíen si realizamos varios entrenamientos, pues los conjuntos de validación se forman, como ya dijimos anteriormente, de forma aleatoria.

5.2 Predicción de modelos entrenados

Una vez hemos entrenado los 4 modelos distintos, procedemos a la obtención de resultados con el conjunto de prueba, usando como medidas la precisión, la exhaustividad, y el valor-F. Uno de los principales inconvenientes que tenemos en este aspecto es que el conjunto de test dispone de una categoría adicional con la que no hemos podido entrenar, se trata de la categoría de “falsa detección”, cuyo conjunto de imágenes no hemos podido obtener, asignada a imágenes borrosas o erróneas, por lo que esto afectará negativamente la medida de precisión final, que es perjudicada cuando asignamos una categoría de forma errónea a elementos que no pertenecen a ella (Ver capítulo 2.1.4). Intentaremos minimizar este impacto, en la

medida de lo posible, en el ensamblaje de resultados, que explicaremos en el siguiente apartado.

Además, a la hora de presentar estos resultados, hemos asignado, a cada categoría, tres características distintas, según el tipo de imagen, la superficie en la que se encuentra la región de interés, y el tamaño que suelen tener estas regiones respecto al resto de la imagen. Éstas son las siguientes:

- Tipo de imagen: Edificio, edificio con carretera, carretera u otros.
- Terreno: Tierra, agua, tierra y agua o variable.
- Tamaño: Pequeño, mediano, grande, tamaños intermedios entre éstos o variable.

Debido a la naturaleza de las diferentes categorías que contiene el conjunto, usaremos el término “variable” en los grupos de terreno y tamaño, pues algunas muestras pueden encontrarse en distintos tipos de terreno, como por ejemplo, los puentes de carretera, o distintos tipos de tamaño, como en el caso de los campos de cultivo.

A continuación, se mostrarán los resultados obtenidos para cada uno de los cuatro modelos entrenados con el conjunto completo con las medidas tanto de precisión, de exhaustividad, como valor-F, para cada categoría, así como su media total.

Identificación de objetos en imágenes por satélite utilizando técnicas Deep Learning

Guillermo Martínez Lledó

CATEGORIA	PRECISION				RECALL				F-SCORE				TIPO DE IMAGEN	TERRENO	TAMANO
	Resnet50	InceptionV3	Xception	Vgg16	Resnet50	InceptionV3	Xception	Vgg16	Resnet50	InceptionV3	Xception	Vgg16			
Aeropuerto	52,00 %	34,00 %	39,00 %	34,00 %	97,00 %	93,00 %	93,00 %	90,00 %	67,00 %	50,00 %	55,00 %	50,00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	72,00 %	74,00 %	75,00 %	73,00 %	69,00 %	68,00 %	66,00 %	62,00 %	71,00 %	71,00 %	70,00 %	67,00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	75,00 %	80,00 %	76,00 %	77,00 %	71,00 %	66,00 %	67,00 %	65,00 %	73,00 %	72,00 %	71,00 %	70,00 %	Edificio	Tierra	Variable
Parque de atracciones	57,00 %	73,00 %	78,00 %	81,00 %	83,00 %	88,00 %	90,00 %	89,00 %	67,00 %	80,00 %	84,00 %	85,00 %	Edificio	Variable	Mediano-Grande
Acuicultura	56,00 %	64,00 %	68,00 %	65,00 %	88,00 %	86,00 %	81,00 %	84,00 %	68,00 %	73,00 %	74,00 %	74,00 %	Otros	Agua	Mediano
Zona de Arqueología	40,00 %	54,00 %	54,00 %	58,00 %	59,00 %	59,00 %	60,00 %	57,00 %	47,00 %	56,00 %	57,00 %	58,00 %	Otros	Tierra	Variable
Granero	64,00 %	65,00 %	65,00 %	65,00 %	76,00 %	77,00 %	77,00 %	71,00 %	69,00 %	70,00 %	70,00 %	68,00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	42,00 %	28,00 %	42,00 %	50,00 %	31,00 %	22,00 %	33,00 %	25,00 %	36,00 %	24,00 %	37,00 %	34,00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	73,00 %	78,00 %	77,00 %	72,00 %	86,00 %	85,00 %	90,00 %	88,00 %	79,00 %	81,00 %	83,00 %	80,00 %	Edificio	Tierra	Variable
Concesionario de coches	70,00 %	74,00 %	70,00 %	65,00 %	84,00 %	81,00 %	81,00 %	82,00 %	76,00 %	77,00 %	75,00 %	73,00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	33,00 %	39,00 %	37,00 %	35,00 %	29,00 %	27,00 %	39,00 %	20,00 %	31,00 %	32,00 %	39,00 %	26,00 %	Otros	Tierra	Variable
Campo de cultivo	73,00 %	78,00 %	82,00 %	81,00 %	95,00 %	94,00 %	94,00 %	93,00 %	83,00 %	86,00 %	88,00 %	87,00 %	Otros	Tierra	Variable
Represa	63,00 %	67,00 %	69,00 %	58,00 %	90,00 %	87,00 %	88,00 %	86,00 %	74,00 %	76,00 %	77,00 %	69,00 %	Otros	Tierra y Agua	Grande
Escombros	34,00 %	22,00 %	18,00 %	26,00 %	60,00 %	60,00 %	58,00 %	46,00 %	43,00 %	32,00 %	28,00 %	33,00 %	Otros	Tierra	Variable
Institución educativa	41,00 %	44,00 %	46,00 %	47,00 %	59,00 %	63,00 %	53,00 %	60,00 %	48,00 %	51,00 %	49,00 %	53,00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	80,00 %	82,00 %	85,00 %	82,00 %	86,00 %	86,00 %	90,00 %	86,00 %	83,00 %	84,00 %	87,00 %	84,00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	63,00 %	64,00 %	74,00 %	55,00 %	42,00 %	46,00 %	43,00 %	38,00 %	50,00 %	53,00 %	55,00 %	45,00 %	Edificio	Variable	Variable
Estación de bomberos	47,00 %	41,00 %	52,00 %	39,00 %	52,00 %	43,00 %	50,00 %	35,00 %	49,00 %	42,00 %	51,00 %	37,00 %	Edificio	Tierra	Pequeño
Camino inundado	12,00 %	19,00 %	15,00 %	29,00 %	60,00 %	71,00 %	59,00 %	68,00 %	20,00 %	29,00 %	24,00 %	41,00 %	Carretera	Tierra	Grande
Fuente	49,00 %	31,00 %	31,00 %	19,00 %	86,00 %	87,00 %	86,00 %	85,00 %	62,00 %	45,00 %	46,00 %	30,00 %	Otros	Tierra	Pequeño
Gasolinera	70,00 %	72,00 %	72,00 %	60,00 %	88,00 %	85,00 %	86,00 %	79,00 %	78,00 %	78,00 %	78,00 %	69,00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	89,00 %	90,00 %	81,00 %	89,00 %	90,00 %	90,00 %	87,00 %	89,00 %	90,00 %	90,00 %	83,00 %	89,00 %	Otros	Tierra	Grande
Estación de transporte terrestre	68,00 %	64,00 %	67,00 %	59,00 %	71,00 %	73,00 %	76,00 %	67,00 %	70,00 %	68,00 %	71,00 %	63,00 %	Edificio con carretera	Tierra	Variable
Helipuerto	71,00 %	67,00 %	65,00 %	51,00 %	82,00 %	84,00 %	82,00 %	79,00 %	76,00 %	74,00 %	73,00 %	62,00 %	Carretera	Tierra	Mediano-Grande
Hospital	45,00 %	47,00 %	50,00 %	47,00 %	36,00 %	36,00 %	47,00 %	35,00 %	40,00 %	41,00 %	49,00 %	40,00 %	Edificio	Tierra	Mediano-Grande
Cruce	66,00 %	79,00 %	77,00 %	74,00 %	94,00 %	92,00 %	91,00 %	89,00 %	78,00 %	85,00 %	83,00 %	81,00 %	Carretera	Tierra	Grande
Lago o estanque	53,00 %	72,00 %	62,00 %	49,00 %	64,00 %	54,00 %	64,00 %	39,00 %	58,00 %	61,00 %	63,00 %	44,00 %	Otros	Agua	Pequeño-Mediano
Faro	51,00 %	44,00 %	31,00 %	35,00 %	83,00 %	77,00 %	77,00 %	68,00 %	63,00 %	56,00 %	45,00 %	46,00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	53,00 %	56,00 %	47,00 %	61,00 %	59,00 %	64,00 %	66,00 %	63,00 %	55,00 %	60,00 %	55,00 %	62,00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	48,00 %	48,00 %	48,00 %	52,00 %	41,00 %	43,00 %	44,00 %	39,00 %	44,00 %	45,00 %	46,00 %	44,00 %	Edificio	Tierra	Variable
Estación nuclear	42,00 %	43,00 %	24,00 %	61,00 %	100,00 %	91,00 %	100,00 %	100,00 %	59,00 %	59,00 %	39,00 %	76,00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	21,00 %	25,00 %	27,00 %	25,00 %	12,00 %	19,00 %	18,00 %	10,00 %	15,00 %	22,00 %	21,00 %	15,00 %	Edificio	Tierra	Variable
Instalación petrolera	73,00 %	73,00 %	72,00 %	68,00 %	86,00 %	85,00 %	81,00 %	78,00 %	79,00 %	78,00 %	76,00 %	73,00 %	Otros	Tierra	Pequeño-Mediano
Parque	49,00 %	53,00 %	51,00 %	44,00 %	68,00 %	72,00 %	72,00 %	66,00 %	57,00 %	61,00 %	59,00 %	53,00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaie	73,00 %	79,00 %	79,00 %	78,00 %	81,00 %	80,00 %	78,00 %	78,00 %	77,00 %	80,00 %	78,00 %	78,00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	62,00 %	69,00 %	63,00 %	68,00 %	72,00 %	72,00 %	72,00 %	68,00 %	67,00 %	71,00 %	67,00 %	68,00 %	Edificio	Tierra	Variable
Estación de policía	31,00 %	32,00 %	37,00 %	34,00 %	26,00 %	26,00 %	34,00 %	23,00 %	28,00 %	29,00 %	36,00 %	28,00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	33,00 %	25,00 %	22,00 %	22,00 %	84,00 %	81,00 %	73,00 %	83,00 %	48,00 %	38,00 %	34,00 %	35,00 %	Otros	Agua	Variable
Prisión	66,00 %	71,00 %	60,00 %	67,00 %	65,00 %	66,00 %	71,00 %	68,00 %	66,00 %	69,00 %	65,00 %	68,00 %	Edificio	Tierra	Grande
Pista de carreras	90,00 %	89,00 %	90,00 %	85,00 %	91,00 %	91,00 %	90,00 %	91,00 %	90,00 %	90,00 %	90,00 %	88,00 %	Carretera	Tierra	Grande
Puente ferroviario	49,00 %	53,00 %	63,00 %	41,00 %	83,00 %	82,00 %	79,00 %	79,00 %	62,00 %	64,00 %	70,00 %	54,00 %	Otros	Variable	Variable
Instalaciones recreativas	85,00 %	84,00 %	86,00 %	86,00 %	92,00 %	93,00 %	91,00 %	91,00 %	89,00 %	88,00 %	89,00 %	88,00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	26,00 %	36,00 %	37,00 %	48,00 %	85,00 %	85,00 %	91,00 %	88,00 %	40,00 %	51,00 %	52,00 %	62,00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	60,00 %	64,00 %	73,00 %	69,00 %	75,00 %	72,00 %	72,00 %	70,00 %	66,00 %	68,00 %	73,00 %	70,00 %	Carretera	Variable	Variable
Pista de aterrizaje	76,00 %	66,00 %	65,00 %	56,00 %	85,00 %	88,00 %	86,00 %	88,00 %	80,00 %	76,00 %	74,00 %	68,00 %	Carretera	Variable	Grande
Astillero	38,00 %	24,00 %	28,00 %	22,00 %	44,00 %	59,00 %	63,00 %	41,00 %	41,00 %	34,00 %	39,00 %	28,00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	66,00 %	66,00 %	64,00 %	58,00 %	65,00 %	67,00 %	63,00 %	66,00 %	65,00 %	66,00 %	64,00 %	61,00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	69,00 %	68,00 %	66,00 %	68,00 %	74,00 %	80,00 %	79,00 %	62,00 %	72,00 %	74,00 %	72,00 %	65,00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	62,00 %	50,00 %	58,00 %	35,00 %	78,00 %	73,00 %	75,00 %	68,00 %	69,00 %	59,00 %	66,00 %	46,00 %	Edificio	Tierra	Pequeño
Granja solar	89,00 %	90,00 %	90,00 %	92,00 %	94,00 %	92,00 %	89,00 %	92,00 %	91,00 %	91,00 %	90,00 %	92,00 %	Otros	Tierra	Variable
Instalación espacial	78,00 %	44,00 %	29,00 %	47,00 %	78,00 %	89,00 %	100,00 %	89,00 %	78,00 %	59,00 %	45,00 %	62,00 %	Edificio con carretera	Tierra	Grande
Estadio	81,00 %	86,00 %	83,00 %	85,00 %	83,00 %	83,00 %	85,00 %	77,00 %	82,00 %	85,00 %	84,00 %	81,00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	86,00 %	82,00 %	83,00 %	89,00 %	97,00 %	93,00 %	87,00 %	91,00 %	91,00 %	87,00 %	85,00 %	90,00 %	Otros	Tierra	Grande
Mina de superficie	84,00 %	83,00 %	88,00 %	82,00 %	88,00 %	86,00 %	86,00 %	86,00 %	86,00 %	84,00 %	87,00 %	84,00 %	Otros	Tierra	Grande
Piscina	82,00 %	76,00 %	76,00 %	53,00 %	94,00 %	94,00 %	94,00 %	93,00 %	87,00 %	84,00 %	84,00 %	68,00 %	Otros	Agua	Pequeño
Cabina de peaje	64,00 %	66,00 %	62,00 %	62,00 %	98,00 %	97,00 %	96,00 %	97,00 %	77,00 %	79,00 %	76,00 %	75,00 %	Edificio con carretera	Tierra	Pequeño
Torre	58,00 %	47,00 %	51,00 %	42,00 %	77,00 %	75,00 %	78,00 %	67,00 %	66,00 %	58,00 %	61,00 %	52,00 %	Edificio	Tierra	Pequeño
Apertura de túnel	61,00 %	55,00 %	53,00 %	52,00 %	98,00 %	97,00 %	94,00 %	97,00 %	75,00 %	71,00 %	67,00 %	67,00 %	Otros	Tierra	Pequeño
Vertedero	63,00 %	64,00 %	62,00 %	52,00 %	58,00 %	62,00 %	63,00 %	57,00 %	60,00 %	63,00 %	63,00 %	54,00 %	Otros	Tierra	Mediano
Instalación de tratamiento de agua	87,00 %	88,00 %	90,00 %	83,00 %	85,00 %	86,00 %	85,00 %	85,00 %	86,00 %	87,00 %	88,00 %	84,00 %	Otros	Tierra y Agua	Mediano-Grande
Granja ecológica	85,00 %	89,00 %	90,00 %	90,00 %	97,00 %	96,00 %	93,00 %	94,00 %	91,00 %	92,00 %	91,00 %	92,00 %	Otros	Tierra	Variable
Zoológico	49,00 %	40,00 %	54,00 %	33,00 %	74,00 %	61,00 %	61,00 %	48,00 %	59,00 %	48,00 %	57,00 %	39,00 %	Edificio	Variable	Variable
Promedio	60,45 %	60,16 %	57,82 %	57,82 %	74,16 %	73,87 %	70,45 %	70,45 %	65,27 %	64,63 %	61,74 %	61,74 %			

Tabla 5.3: Precisión, exhaustividad y valor-F de los modelos entrenados (Fuente: Elaboración propia)

Analizando estos resultados podemos sacar las siguientes conclusiones:

- En primer lugar, vemos que a veces se produce una diferencia muy grande entre la precisión y la exhaustividad. Esto se debe al problema que hemos comentado antes, la falta de la categoría “falsa detección”. El no haber podido entregar una categoría que luego se presenta en el conjunto de validación afecta a estos resultados perjudicando los valores de precisión, pues esta medida solo castiga cuando se clasifican erróneamente los elementos (Falsos positivos), mientras que la exhaustividad sale beneficiada, pues solo se centra en la inclusión de los valores relevantes (Sin castigar la clasificación de falsos positivos).
- En segundo lugar, vemos que, entre las distintas categorías, la mayoría obtienen resultados muy buenos en prácticamente todas las redes, sin embargo, hay algunas concretas, con un porcentaje muy bajo. En este caso, las razones pueden ser diversas, por ejemplo, que las características varíen mucho entre las muestras y/o que el número de imágenes de esta categoría usadas en el entrenamiento no hayan sido suficientes. Analizando las categorías más perjudicadas, que en este caso son: puesto fronterizo, zona de construcción, edificios de oficinas y estación de policía, comprobamos que respectivamente existen en el conjunto usado para la construcción del modelo: 3.506, 13.892, 13.352 y 12.126. Comparado con una categoría que tenga una puntuación aceptable, como la de tanque de almacenamiento, que tiene 12.676 imágenes, podemos deducir que, en el caso del puesto fronterizo si ha podido influir la falta de imágenes, mientras que en el resto, puede haber sido la propia naturaleza de las imágenes o los elementos en sí, por lo menos, respecto a las redes que hemos usado.
- En general, el modelo que mejores resultados ha dado ha sido ResNet50, aunque, entre las categorías, no siempre es el que mejor resultado da, en algunas, como, por ejemplo, la de parque de atracciones, es de hecho la que nos da un porcentaje de acierto más bajo. Esta razón nos indica que, si queremos obtener mejores resultados, debemos contar con varias redes.

5.3 Ensemble de modelos

Uno de los experimentos que queremos realizar, como hemos hablado anteriormente, consiste en “unificar” varios modelos mediante un sistema de votación, donde cada modelo aportará un voto, representado por la categoría seleccionada por el modelo para cada imagen según sus predicciones, usando como método de agregación de resultados la clase más votada, con el fin de comprobar si un resultado conjunto es mejor que cada uno de los resultados dados por los modelos de forma individual. En el caso de producirse un empate, la categoría asignada sería “falsa detección”, con lo que intentamos reducir el problema del que hablamos en el apartado 5.2.

5.3.1 Ensemble de modelos con conjunto de datos de entrenamiento original

Para este primer ensamblaje vamos a usar los modelos que han realizado un entrenamiento con el conjunto de imágenes original, tal como hemos mostrado en el capítulo 5.2

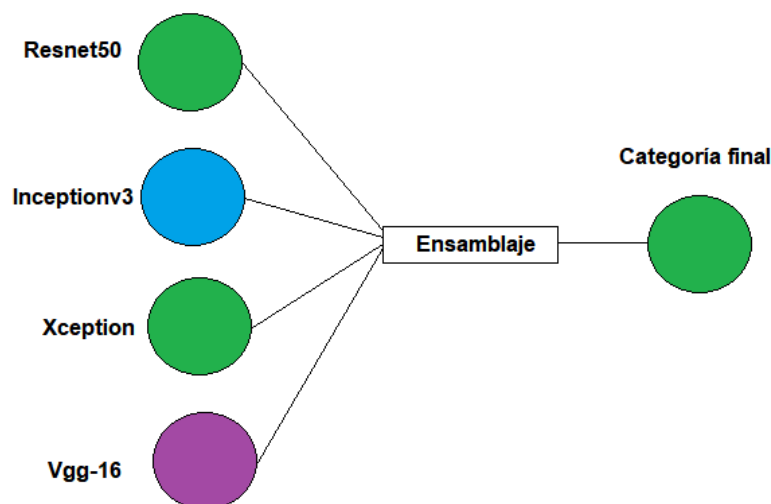


Ilustración 5.2: Esquema sobre el sistema de ensamblaje (Fuente: Elaboración propia)

Como resultado, obtenemos la siguiente tabla, en la que podemos apreciar que los resultados obtenidos en las tres medidas distintas son, en general, mejores que los resultados individuales de cualquiera de las redes anteriores.

CATEGORIA	PRECISION	RECALL	F-SCORE	TIPO DE IMAGEN	TERRENO	TAMANO
	CONJUNTO COMPLETO	CONJUNTO COMPLETO	CONJUNTO COMPLETO			
Aeropuerto	38,00 %	94,00 %	54,00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	82,00 %	65,00 %	72,00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	84,00 %	69,00 %	76,00 %	Edificio	Tierra	Variable
Parque de atracciones	79,00 %	89,00 %	83,00 %	Edificio	Variable	Mediano-Grande
Acuicultura	68,00 %	88,00 %	76,00 %	Otros	Agua	Mediano
Zona de Arqueología	70,00 %	62,00 %	66,00 %	Otros	Tierra	Variable
Granero	73,00 %	78,00 %	75,00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	63,00 %	33,00 %	44,00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	85,00 %	89,00 %	87,00 %	Edificio	Tierra	Variable
Concesionario de coches	77,00 %	84,00 %	80,00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	52,00 %	29,00 %	37,00 %	Otros	Tierra	Variable
Campo de cultivo	82,00 %	95,00 %	88,00 %	Otros	Tierra	Variable
Represa	71,00 %	89,00 %	79,00 %	Otros	Tierra y Agua	Grande
Escombros	30,00 %	62,00 %	40,00 %	Otros	Tierra	Variable
Institución educativa	50,00 %	61,00 %	55,00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	89,00 %	88,00 %	89,00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	79,00 %	44,00 %	57,00 %	Edificio	Variable	Variable
Estación de bomberos	54,00 %	46,00 %	50,00 %	Edificio	Tierra	Pequeño
Camino inundado	21,00 %	67,00 %	32,00 %	Carretera	Tierra	Grande
Fuente	32,00 %	87,00 %	47,00 %	Otros	Tierra	Pequeño
Gasolinera	79,00 %	88,00 %	83,00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	91,00 %	91,00 %	91,00 %	Otros	Tierra	Grande
Estación de transporte terrestre	75,00 %	72,00 %	74,00 %	Edificio con carretera	Tierra	Variable
Helipuerto	73,00 %	82,00 %	77,00 %	Carretera	Tierra	Mediano-Grande
Hospital	60,00 %	37,00 %	45,00 %	Edificio	Tierra	Mediano-Grande
Cruce	79,00 %	92,00 %	85,00 %	Carretera	Tierra	Grande
Lago o estanque	73,00 %	58,00 %	65,00 %	Otros	Agua	Pequeño-Mediano
Faro	46,00 %	82,00 %	59,00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	66,00 %	66,00 %	66,00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	58,00 %	39,00 %	47,00 %	Edificio	Tierra	Variable
Estación nuclear	48,00 %	100,00 %	65,00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	29,00 %	13,00 %	18,00 %	Edificio	Tierra	Variable
Instalación petrolera	81,00 %	87,00 %	84,00 %	Otros	Tierra	Pequeño-Mediano
Parque	61,00 %	70,00 %	65,00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaie	82,00 %	80,00 %	81,00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	75,00 %	73,00 %	74,00 %	Edificio	Tierra	Variable
Estación de policía	42,00 %	27,00 %	33,00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	24,00 %	78,00 %	37,00 %	Otros	Agua	Variable
Prisión	74,00 %	68,00 %	71,00 %	Edificio	Tierra	Grande
Pista de carreras	93,00 %	91,00 %	91,00 %	Carretera	Tierra	Grande
Puente ferroviario	61,00 %	82,00 %	70,00 %	Otros	Variable	Variable
Instalaciones recreativas	89,00 %	92,00 %	90,00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	42,00 %	85,00 %	56,00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	72,00 %	72,00 %	72,00 %	Carretera	Variable	Variable
Pista de aterrizaje	67,00 %	86,00 %	75,00 %	Carretera	Variable	Grande
Astillero	29,00 %	52,00 %	37,00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	73,00 %	67,00 %	70,00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	74,00 %	79,00 %	76,00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	60,00 %	74,00 %	66,00 %	Edificio	Tierra	Pequeño
Granja solar	94,00 %	94,00 %	94,00 %	Otros	Tierra	Variable
Instalación espacial	53,00 %	89,00 %	67,00 %	Edificio con carretera	Tierra	Grande
Estadio	87,00 %	85,00 %	86,00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	91,00 %	95,00 %	93,00 %	Otros	Tierra	Grande
Mina de superficie	88,00 %	87,00 %	88,00 %	Otros	Tierra	Grande
Piscina	80,00 %	95,00 %	87,00 %	Otros	Agua	Pequeño
Cabina de peaje	66,00 %	97,00 %	78,00 %	Edificio con carretera	Tierra	Pequeño
Torre	58,00 %	77,00 %	66,00 %	Edificio	Tierra	Pequeño
Apertura de túnel	57,00 %	98,00 %	72,00 %	Otros	Tierra	Pequeño
Vertedero	72,00 %	62,00 %	67,00 %	Otros	Tierra	Mediano
Instalación de tratamiento de aguas	91,00 %	86,00 %	89,00 %	Otros	Tierra y Agua	Mediano-Grande
Granja eólica	92,00 %	96,00 %	94,00 %	Otros	Tierra	Variable
Zoológico	55,00 %	63,00 %	59,00 %	Edificio	Variable	Variable
Promedio	66,74 %	74,61 %	68,55 %			

Tabla 5.4: Precisión, exhaustividad y valor-F del ensemble de los modelos entrenados con el conjunto original (Fuente: Elaboración propia)

Echando un vistazo a la tabla, podemos determinar dos cosas: en primer lugar, que la mejora más significativa se da en la precisión, lo que nos demuestra que el método de agregación que hemos usado, donde, en caso de empate en la votación, asignamos la categoría “detección falsa” ha dado buenos resultados. En segundo lugar, aunque la mejora en exhaustividad haya sido menor al 1% con respecto a la mejor red (ResNet50), sigue siendo un aumento, lo que nos demuestra no solo que las redes con peores resultados no son suficientemente influyentes para perjudicar la calidad general, sino que se consigue cierta mejora en algunas categorías en las que la mejor red tenía peor puntuación que el resto.

5.3.2 Ensembles de modelos por red y con muestreo del conjunto de datos de entrenamiento

Adicionalmente a los entrenamientos que ya hemos realizado, realizamos, de nuevo, otros 5 con cada una de estas redes, obteniendo en total 20 nuevos resultados, con el añadido de que ahora, del conjunto de imágenes inicial, cada uno solo usará un muestreo aleatorio del 70, que será distinto para cada modelo, y descartando el 30% restante.

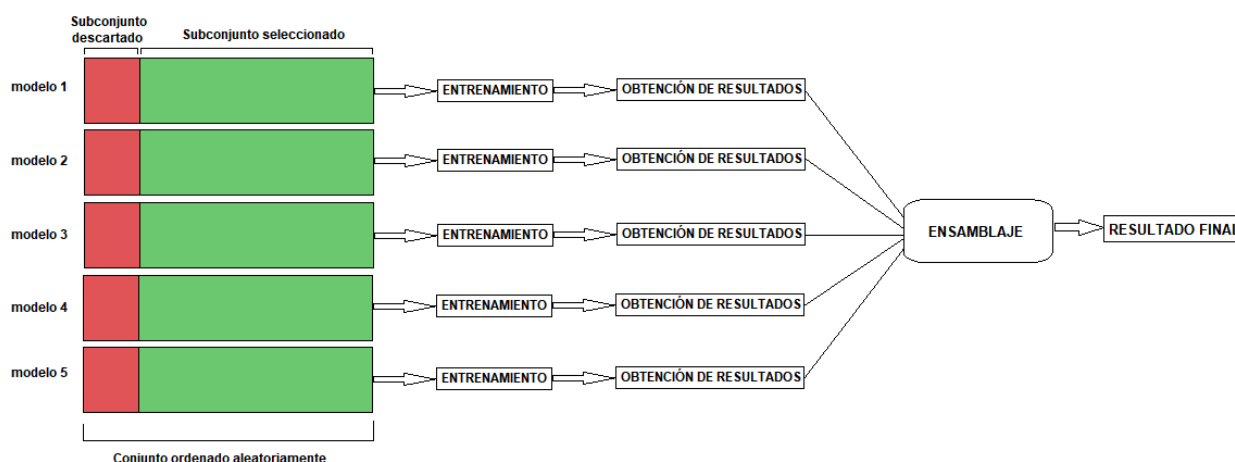


Ilustración 5.3: Esquema de ensamblaje de modelos por red (Fuente: Elaboración propia)

El funcionamiento del resto de los modelos será el mismo que cuando usaron el conjunto al completo. La precisión obtenida sigue siendo similar a los resultados obtenidos con el conjunto de imágenes completo, sin embargo, lo que nos interesa de esta experimentación, una vez obtenida la predicción de cada modelo con el conjunto de prueba, es ensamblar los resultados obtenidos para cada tipo de red (Ver ilustración 5.3). La realización de este ensamblaje es similar a la explicada en el punto 5.3.1, con la diferencia de que, en lugar de ensamblar cuatro modelos, cada uno con una arquitectura distinta, se haría con cinco modelos con la misma arquitectura, pero que han sido entrenados con conjuntos de imágenes distintas.

En la siguiente tabla se muestran los resultados obtenidos tras el ensamblaje de los 5 modelos, correspondientes a cada una de las cuatro arquitecturas entrenadas de esta forma:

Identificación de objetos en imágenes por satélite utilizando técnicas Deep Learning

Guillermo Martínez Lledó

CATEGORIA	PRECISION				RECALL				F-SCORE				TIPO DE IMAGEN	TERRENO	TAMANO
	Ensemble de Resnet50	Ensemble de InceptionV3	Ensemble de Xception	Ensemble de Vgg16	Ensemble de Resnet50	Ensemble de InceptionV3	Ensemble de Xception	Ensemble de Vgg16	Ensemble de Resnet50	Ensemble de InceptionV3	Ensemble de Xception	Ensemble de Vgg16			
Aeropuerto	59.00 %	34.00 %	38.00 %	32.00 %	96.00 %	94.00 %	96.00 %	96.00 %	73.00 %	50.00 %	54.00 %	48.00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	77.00 %	74.00 %	74.00 %	75.00 %	71.00 %	71.00 %	69.00 %	68.00 %	74.00 %	72.00 %	71.00 %	72.00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	80.00 %	86.00 %	81.00 %	82.00 %	72.00 %	68.00 %	72.00 %	67.00 %	76.00 %	76.00 %	76.00 %	74.00 %	Edificio	Tierra	Variable
Parque de atracciones	60.00 %	75.00 %	75.00 %	80.00 %	85.00 %	92.00 %	91.00 %	90.00 %	70.00 %	83.00 %	83.00 %	85.00 %	Edificio	Variable	Mediano-Grande
Acuicultura	56.00 %	60.00 %	63.00 %	64.00 %	90.00 %	89.00 %	86.00 %	85.00 %	69.00 %	72.00 %	73.00 %	73.00 %	Otros	Agua	Mediano
Zona de Arqueología	44.00 %	45.00 %	44.00 %	38.00 %	57.00 %	68.00 %	66.00 %	65.00 %	50.00 %	54.00 %	53.00 %	48.00 %	Otros	Tierra	Variable
Granero	67.00 %	69.00 %	72.00 %	64.00 %	71.00 %	74.00 %	77.00 %	76.00 %	69.00 %	71.00 %	75.00 %	70.00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	55.00 %	50.00 %	66.00 %	36.00 %	35.00 %	45.00 %	41.00 %	35.00 %	43.00 %	47.00 %	51.00 %	36.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	79.00 %	82.00 %	83.00 %	78.00 %	87.00 %	93.00 %	92.00 %	92.00 %	83.00 %	87.00 %	88.00 %	84.00 %	Edificio	Tierra	Variable
Concesionario de coches	75.00 %	80.00 %	79.00 %	77.00 %	86.00 %	80.00 %	82.00 %	81.00 %	80.00 %	80.00 %	81.00 %	79.00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	49.00 %	40.00 %	46.00 %	39.00 %	33.00 %	38.00 %	44.00 %	33.00 %	39.00 %	39.00 %	45.00 %	36.00 %	Otros	Tierra	Variable
Campo de cultivo	74.00 %	78.00 %	77.00 %	76.00 %	96.00 %	96.00 %	96.00 %	96.00 %	84.00 %	86.00 %	85.00 %	85.00 %	Otros	Tierra	Variable
Represa	66.00 %	73.00 %	74.00 %	62.00 %	90.00 %	89.00 %	90.00 %	89.00 %	76.00 %	80.00 %	81.00 %	73.00 %	Otros	Tierra y Agua	Grande
Escombros	38.00 %	32.00 %	37.00 %	22.00 %	58.00 %	63.00 %	58.00 %	58.00 %	45.00 %	42.00 %	45.00 %	32.00 %	Otros	Variable	Variable
Institución educativa	44.00 %	43.00 %	46.00 %	44.00 %	60.00 %	62.00 %	64.00 %	67.00 %	51.00 %	51.00 %	53.00 %	53.00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	86.00 %	87.00 %	87.00 %	89.00 %	89.00 %	87.00 %	87.00 %	88.00 %	87.00 %	87.00 %	87.00 %	89.00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	73.00 %	69.00 %	75.00 %	66.00 %	47.00 %	49.00 %	54.00 %	47.00 %	57.00 %	57.00 %	63.00 %	55.00 %	Edificio	Variable	Variable
Estación de bomberos	55.00 %	58.00 %	62.00 %	49.00 %	55.00 %	52.00 %	49.00 %	52.00 %	55.00 %	54.00 %	55.00 %	50.00 %	Edificio	Tierra	Pequeño
Camino inundado	13.00 %	13.00 %	14.00 %	15.00 %	62.00 %	71.00 %	63.00 %	69.00 %	21.00 %	22.00 %	23.00 %	25.00 %	Carretera	Tierra	Grande
Fuente	55.00 %	56.00 %	57.00 %	55.00 %	90.00 %	90.00 %	87.00 %	89.00 %	68.00 %	69.00 %	69.00 %	68.00 %	Otros	Tierra	Pequeño
Gasolinera	76.00 %	84.00 %	82.00 %	80.00 %	87.00 %	88.00 %	84.00 %	84.00 %	81.00 %	86.00 %	83.00 %	82.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	87.00 %	92.00 %	88.00 %	92.00 %	93.00 %	92.00 %	90.00 %	92.00 %	90.00 %	92.00 %	89.00 %	92.00 %	Otros	Tierra	Grande
Estación de transporte terrestre	71.00 %	70.00 %	70.00 %	68.00 %	73.00 %	75.00 %	73.00 %	74.00 %	72.00 %	73.00 %	72.00 %	71.00 %	Edificio con carretera	Tierra	Variable
Helipuerto	77.00 %	76.00 %	79.00 %	74.00 %	84.00 %	85.00 %	85.00 %	79.00 %	80.00 %	80.00 %	82.00 %	77.00 %	Carretera	Tierra	Mediano-Grande
Hospital	56.00 %	55.00 %	50.00 %	49.00 %	35.00 %	40.00 %	39.00 %	38.00 %	43.00 %	46.00 %	44.00 %	43.00 %	Edificio	Tierra	Mediano-Grande
Cruce	70.00 %	76.00 %	75.00 %	72.00 %	86.00 %	93.00 %	94.00 %	94.00 %	81.00 %	83.00 %	83.00 %	82.00 %	Carretera	Tierra	Grande
Lago o estanque	61.00 %	58.00 %	62.00 %	66.00 %	58.00 %	62.00 %	61.00 %	59.00 %	59.00 %	60.00 %	62.00 %	62.00 %	Otros	Agua	Pequeño-Mediano
Faro	50.00 %	51.00 %	48.00 %	48.00 %	84.00 %	79.00 %	79.00 %	81.00 %	63.00 %	62.00 %	59.00 %	60.00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	57.00 %	59.00 %	59.00 %	54.00 %	61.00 %	64.00 %	66.00 %	67.00 %	59.00 %	61.00 %	62.00 %	60.00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	51.00 %	48.00 %	53.00 %	57.00 %	41.00 %	44.00 %	45.00 %	43.00 %	45.00 %	46.00 %	49.00 %	49.00 %	Edificio	Tierra	Variable
Estación nuclear	53.00 %	32.00 %	37.00 %	43.00 %	91.00 %	100.00 %	91.00 %	91.00 %	67.00 %	49.00 %	53.00 %	59.00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	28.00 %	26.00 %	24.00 %	27.00 %	12.00 %	20.00 %	15.00 %	13.00 %	16.00 %	23.00 %	19.00 %	18.00 %	Edificio	Tierra	Variable
Instalación petrolera	79.00 %	77.00 %	73.00 %	77.00 %	85.00 %	88.00 %	88.00 %	80.00 %	81.00 %	82.00 %	79.00 %	80.00 %	Otros	Tierra	Pequeño-Mediano
Parque	52.00 %	55.00 %	53.00 %	55.00 %	69.00 %	72.00 %	70.00 %	71.00 %	59.00 %	63.00 %	61.00 %	62.00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaje	77.00 %	79.00 %	78.00 %	78.00 %	84.00 %	82.00 %	84.00 %	81.00 %	80.00 %	80.00 %	81.00 %	80.00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	64.00 %	69.00 %	68.00 %	68.00 %	73.00 %	73.00 %	77.00 %	72.00 %	68.00 %	71.00 %	72.00 %	70.00 %	Edificio	Tierra	Variable
Estación de policía	41.00 %	40.00 %	42.00 %	36.00 %	23.00 %	30.00 %	27.00 %	25.00 %	30.00 %	34.00 %	33.00 %	30.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	34.00 %	29.00 %	30.00 %	25.00 %	86.00 %	83.00 %	82.00 %	81.00 %	49.00 %	43.00 %	44.00 %	38.00 %	Otros	Agua	Variable
Prisión	71.00 %	76.00 %	76.00 %	79.00 %	66.00 %	69.00 %	67.00 %	67.00 %	69.00 %	72.00 %	71.00 %	73.00 %	Edificio	Tierra	Grande
Pista de carreras	93.00 %	91.00 %	93.00 %	93.00 %	91.00 %	91.00 %	95.00 %	93.00 %	92.00 %	91.00 %	94.00 %	93.00 %	Carretera	Tierra	Grande
Puente ferroviario	50.00 %	52.00 %	56.00 %	52.00 %	84.00 %	81.00 %	83.00 %	84.00 %	63.00 %	63.00 %	67.00 %	64.00 %	Otros	Variable	Variable
Instalaciones recreativas	87.00 %	90.00 %	87.00 %	88.00 %	93.00 %	93.00 %	93.00 %	93.00 %	90.00 %	91.00 %	90.00 %	90.00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	29.00 %	33.00 %	31.00 %	33.00 %	82.00 %	91.00 %	88.00 %	91.00 %	43.00 %	48.00 %	45.00 %	49.00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	61.00 %	64.00 %	60.00 %	61.00 %	75.00 %	71.00 %	72.00 %	74.00 %	67.00 %	67.00 %	66.00 %	67.00 %	Carretera	Variable	Variable
Pista de aterrizaje	82.00 %	72.00 %	73.00 %	61.00 %	88.00 %	88.00 %	88.00 %	87.00 %	85.00 %	79.00 %	80.00 %	72.00 %	Carretera	Variable	Grande
Astillero	46.00 %	33.00 %	40.00 %	28.00 %	44.00 %	52.00 %	59.00 %	41.00 %	45.00 %	40.00 %	48.00 %	33.00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	67.00 %	66.00 %	68.00 %	69.00 %	69.00 %	66.00 %	65.00 %	71.00 %	68.00 %	66.00 %	66.00 %	70.00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	73.00 %	75.00 %	77.00 %	77.00 %	75.00 %	72.00 %	73.00 %	71.00 %	74.00 %	73.00 %	75.00 %	74.00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	67.00 %	73.00 %	72.00 %	66.00 %	78.00 %	80.00 %	75.00 %	76.00 %	72.00 %	76.00 %	74.00 %	71.00 %	Edificio	Tierra	Pequeño
Granja solar	93.00 %	92.00 %	92.00 %	93.00 %	94.00 %	93.00 %	94.00 %	93.00 %	93.00 %	92.00 %	93.00 %	93.00 %	Otros	Tierra	Variable
Instalación espacial	86.00 %	62.00 %	64.00 %	60.00 %	67.00 %	89.00 %	100.00 %	100.00 %	75.00 %	73.00 %	78.00 %	75.00 %	Edificio con carretera	Tierra	Grande
Estadio	84.00 %	83.00 %	85.00 %	88.00 %	82.00 %	83.00 %	82.00 %	85.00 %	83.00 %	83.00 %	84.00 %	87.00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	91.00 %	86.00 %	89.00 %	90.00 %	97.00 %	94.00 %	95.00 %	93.00 %	94.00 %	90.00 %	92.00 %	91.00 %	Otros	Tierra	Grande
Mina de superficie	83.00 %	85.00 %	86.00 %	83.00 %	87.00 %	87.00 %	90.00 %	88.00 %	85.00 %	86.00 %	88.00 %	86.00 %	Otros	Tierra	Grande
Piscina	84.00 %	84.00 %	84.00 %	83.00 %	96.00 %	94.00 %	95.00 %	94.00 %	90.00 %	89.00 %	89.00 %	88.00 %	Otros	Agua	Pequeño
Cabina de peaje	74.00 %	75.00 %	75.00 %	72.00 %	98.00 %	98.00 %	98.00 %	98.00 %	84.00 %	85.00 %	85.00 %	83.00 %	Edificio con carretera	Tierra	Pequeño
Torre	68.00 %	64.00 %	65.00 %	59.00 %	76.00 %	78.00 %	81.00 %	76.00 %	72.00 %	70.00 %	72.00 %	66.00 %	Edificio	Tierra	Pequeño
Apertura de túnel	59.00 %	63.00 %	64.00 %	63.00 %	98.00 %	98.00 %	97.00 %	97.00 %	74.00 %	77.00 %	77.00 %	76.00 %	Otros	Tierra	Pequeño
Vertedero	68.00 %	66.00 %	71.00 %	59.00 %	60.00 %	65.00 %	64.00 %	62.00 %	64.00 %	66.00 %	67.00 %	61.00 %	Otros	Tierra	Mediano
Instalación de tratamiento de agua	91.00 %	93.00 %	92.00 %	91.00 %	84.00 %	86.00 %	87.00 %	88.00 %	88.00 %	89.00 %	90.00 %	89.00 %	Otros	Tierra y Agua	Mediano-Grande
Granja edícola	89.00 %	91.00 %	91.00 %	92.00 %	97.00 %	97.00 %	97.00 %	96.00 %	93.00 %	94.00 %	94.00 %	94.00 %	Otros	Tierra	Variable
Zoológico	57.00 %	59.00 %	51.00 %	49.00 %	65.00 %	59.00 %	72.00 %	72.00 %	61.00 %	59.00 %	59.00 %	58.00 %	Edificio	Variable	Variable
Promedio	65.19 %	64.65 %	65.53 %	62.92 %	74.37 %	76.06 %	76.19 %	75.18 %	68.02 %	68.26 %	69.06 %	66.98 %			

Tabla 5.5: Precisión, exhaustividad y valor-F del ensemble por red de los distintos modelos entrenados con un muestreo aleatorio del 70% del conjunto original (Fuente: Elaboración propia)

Comparando esta tabla con la mostrada en el capítulo 5.2 (Tabla 5.3), podemos apreciar una mejora del promedio en las tres métricas para todas las redes, aun existiendo algunos casos en determinadas categorías y para redes concretas donde sí se produce un descenso en el porcentaje de acierto, éstas son situaciones poco comunes, además de que dicho descenso de calidad no es muy pronunciado.

Podemos entonces indicar que, en nuestro caso, el entrenamiento y posterior ensamblamiento de varias redes con arquitectura común pero distinto conjunto de entrenamiento ha significado una mejora de calidad global para los resultados.

Como curiosidad a destacar en comparación a lo que hemos analizado en el apartado anterior, podemos apreciar que, en relación a las peores categorías que indicamos anteriormente, las de puesto fronterizo y zona de construcción, aún sin presentar todavía resultados que puedan ser considerados buenos, han mejorado enormemente, mientras que, en cambio, edificio de oficinas y estación de policía siguen presentando resultados muy similares.

5.3.3 Ensemble general de modelos y datos muestreados

Como última experimentación, procederemos a ensamblar, de nuevo, mediante el sistema de agregación de la clase más votada, todos los modelos anteriores entrenados con un muestreo aleatorio del 70%, lo que equivaldría a un nuevo ensemble formado por 20 modelos.

De nuevo, se vuelve a producir una mejora respecto a los resultados individuales anteriores, aunque esta vez no es tan pronunciada como en el primer ensamblaje que realizamos (Capítulo 5.3.1). De hecho, la medida de exhaustividad esta vez es mejorada por las redes Inception y Xception ensambladas de forma individual. Sin embargo, la precisión vuelve a ser bastante mejor en este caso, al igual que también mejora el valor-F, que nos indica que, a pesar del descenso en la exhaustividad, los resultados, en general, son mejores.

CATEGORIA	PRECISION	RECALL	F-SCORE	TIPO DE IMAGEN	TERRENO	TAMANO
	MODELOS ENSAMBLADOS	MODELOS ENSAMBLADOS	MODELOS ENSAMBLADOS			
Aeropuerto	36.00 %	96.00 %	52.00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	77.00 %	69.00 %	72.00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	87.00 %	69.00 %	77.00 %	Edificio	Tierra	Variable
Parque de atracciones	77.00 %	92.00 %	84.00 %	Edificio	Variable	Mediano-Grande
Acuicultura	64.00 %	88.00 %	74.00 %	Otros	Agua	Mediano
Zona de Arqueología	48.00 %	65.00 %	55.00 %	Otros	Tierra	Variable
Granero	72.00 %	75.00 %	73.00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	61.00 %	39.00 %	48.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	87.00 %	92.00 %	90.00 %	Edificio	Tierra	Variable
Concesionario de coches	84.00 %	84.00 %	84.00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	51.00 %	40.00 %	45.00 %	Otros	Tierra	Variable
Campo de cultivo	78.00 %	96.00 %	86.00 %	Otros	Tierra	Variable
Represa	72.00 %	90.00 %	80.00 %	Otros	Tierra y Agua	Grande
Escombros	35.00 %	60.00 %	44.00 %	Otros	Tierra	Variable
Institución educativa	47.00 %	63.00 %	54.00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	90.00 %	88.00 %	89.00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	81.00 %	49.00 %	61.00 %	Edificio	Variable	Variable
Estación de bomberos	64.00 %	52.00 %	57.00 %	Edificio	Tierra	Pequeño
Camino inundado	15.00 %	68.00 %	24.00 %	Carretera	Tierra	Grande
Fuente	59.00 %	90.00 %	71.00 %	Otros	Tierra	Pequeño
Gasolinera	87.00 %	88.00 %	87.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	95.00 %	93.00 %	94.00 %	Otros	Tierra	Grande
Estación de transporte terrestre	72.00 %	73.00 %	73.00 %	Edificio con carretera	Tierra	Variable
Helipuerto	81.00 %	84.00 %	83.00 %	Carretera	Tierra	Mediano-Grande
Hospital	59.00 %	37.00 %	45.00 %	Edificio	Tierra	Mediano-Grande
Cruce	78.00 %	94.00 %	85.00 %	Carretera	Tierra	Grande
Lago o estanque	70.00 %	61.00 %	65.00 %	Otros	Agua	Pequeño-Mediano
Faro	51.00 %	81.00 %	63.00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	62.00 %	63.00 %	62.00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	58.00 %	41.00 %	48.00 %	Edificio	Tierra	Variable
Estación nuclear	38.00 %	91.00 %	54.00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	30.00 %	14.00 %	19.00 %	Edificio	Tierra	Variable
Instalación petrolera	81.00 %	88.00 %	84.00 %	Otros	Tierra	Pequeño-Mediano
Parque	58.00 %	72.00 %	64.00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaje	82.00 %	83.00 %	82.00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	71.00 %	74.00 %	72.00 %	Edificio	Tierra	Variable
Estación de policía	46.00 %	26.00 %	34.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	29.00 %	81.00 %	43.00 %	Otros	Agua	Variable
Prisión	78.00 %	67.00 %	72.00 %	Edificio	Tierra	Grande
Pista de carreras	94.00 %	93.00 %	94.00 %	Carretera	Tierra	Grande
Puente ferroviario	59.00 %	84.00 %	69.00 %	Otros	Variable	Variable
Instalaciones recreativas	90.00 %	93.00 %	92.00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	31.00 %	85.00 %	46.00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	65.00 %	72.00 %	69.00 %	Carretera	Variable	Variable
Pista de aterrizaje	74.00 %	87.00 %	80.00 %	Carretera	Variable	Grande
Astillero	37.00 %	41.00 %	39.00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	72.00 %	67.00 %	69.00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	78.00 %	73.00 %	75.00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	74.00 %	79.00 %	77.00 %	Edificio	Tierra	Pequeño
Granja solar	94.00 %	93.00 %	93.00 %	Otros	Tierra	Variable
Instalación espacial	75.00 %	100.00 %	86.00 %	Edificio con carretera	Tierra	Grande
Estadio	88.00 %	84.00 %	86.00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	91.00 %	95.00 %	93.00 %	Otros	Tierra	Grande
Mina de superficie	88.00 %	88.00 %	88.00 %	Otros	Tierra	Grande
Piscina	86.00 %	95.00 %	90.00 %	Otros	Agua	Pequeño
Cabina de peaje	76.00 %	98.00 %	86.00 %	Edificio con carretera	Tierra	Pequeño
Torre	69.00 %	78.00 %	73.00 %	Edificio	Tierra	Pequeño
Apertura de túnel	64.00 %	97.00 %	77.00 %	Otros	Tierra	Pequeño
Vertedero	72.00 %	62.00 %	67.00 %	Otros	Tierra	Mediano
Instalación de tratamiento de aguas	94.00 %	87.00 %	90.00 %	Otros	Tierra y Agua	Mediano-Grande
Granja eólica	93.00 %	97.00 %	95.00 %	Otros	Tierra	Variable
Zoológico	63.00 %	72.00 %	67.00 %	Edificio	Variable	Variable
Promedio	68.35 %	75.74 %	70.16 %			

Tabla 5.6: Precisión, exhaustividad y valor-F del ensamblaje de los modelos entrenados con un muestreo aleatorio del 70% del conjunto (Fuente: Elaboración propia)

5.4 Comparativa final y análisis de resultados

A partir de los datos obtenidos en apartados anteriores, elaboramos las siguientes tablas, donde destacamos el mejor resultado según las características que hemos asignado, indicando entre paréntesis el número de categorías que pertenecen a cada asignación (Nota: Diferenciamos entre los modelos que usan el conjunto completo de los que no, añadiendo la palabra “Parcial” entre paréntesis):

- **Precisión**

Tipo de imagen	Mejor resultado
Edificio (20)	Ensamblaje (Parcial) (11)
Carretera (7)	Ensamblaje (Parcial) (3)
Edificio con carretera (10)	Ensamblaje (Parcial) (5)
Otros (25)	Ensamblaje (Parcial) (13)

Tabla 5.7: Mejores precisiones por categoría en tipo de imagen (Fuente: Elaboración propia)

Terreno	Mejor resultado
Tierra (47)	Ensamblaje (Parcial) (27)
Agua (4)	Ensamblaje (2)
Tierra y Agua (5)	Ensamblaje (Parcial) (3)
Variable (6)	Xception y Ensamblaje (Parcial) (2)

Tabla 5.8: Mejores precisiones por categoría en terreno (Fuente: Elaboración propia)

Tamaño	Mejor resultado
Pequeño (9)	Ensamblaje (Parcial) (9)
Pequeño-mediano (11)	Ensamblaje (6)
Mediano (3)	Ensamblaje (3)
Mediano-grande (10)	Ensamblaje (Parcial) (5)
Grande (12)	Resnet50 (Parcial) (5)
Variable (16)	Ensamblaje (Parcial) (8)

Tabla 5.9: Mejores precisiones por categoría en tamaño (Fuente: Elaboración propia)

- **Exhaustividad**

Tipo de imagen	Mejor resultado
Edificio (20)	InceptionV3 (Parcial) (6)
Carretera (7)	Resnet50 (Parcial), Xception (Parcial) (4)
Edificio con carretera (10)	Xception (4)
Otros (25)	Resnet50 (Parcial) (13)

Tabla 5.10: Mejor exhaustividad por categoría en tipo de imagen (Fuente: Elaboración propia)

Terreno	Mejor resultado
Tierra (47)	InceptionV3 (Parcial) (20)
Agua (4)	Resnet50 (Parcial) (3)
Tierra y Agua (5)	Resnet50, Resnet50 (Parcial) y Xception (2)
Variable (6)	Resnet50 (Parcial) (3)

Tabla 5.11: Mejor exhaustividad por categoría en terreno (Fuente: Elaboración propia)

Tamaño	Mejor resultado
Pequeño (9)	Resnet50 (Parcial) (6)
Pequeño-mediano (11)	InceptionV3 (Parcial) (7)
Mediano (3)	Resnet50 (Parcial), InceptionV3 (Parcial) (2)
Mediano-grande (10)	Vgg16 (Parcial) (5)
Grande (12)	Resnet50 (Parcial), Xception (Parcial) (5)
Variable (16)	Xception (Parcial) (9)

Tabla 5.12: Mejor exhaustividad por categoría en tamaño (Fuente: Elaboración propia)

- **Valor-F**

Tipo de imagen	Mejor resultado
Edificio (20)	Ensamblaje (Parcial) (8)
Carretera (7)	Ensamblaje (Parcial) (4)
Edificio con carretera(10)	Ensamblaje (Parcial) (4)
Otros (25)	Ensamblaje (Parcial) (12)

Tabla 5.13: Mejor valor-F por categoría en tipo de imagen (Fuente: Elaboración propia)

Terreno	Mejor resultado
Tierra (47)	Ensamblaje (Parcial) (22)
Agua (4)	Resnet50 (Parcial), Ensamblaje y Ensamblaje (Parcial) (2)
Tierra y Agua (5)	Xception (Parcial) (3)
Variable (6)	Resnet50 (Parcial), Xception, Xception (Parcial), Vgg16, Vgg16 (Parcial) y Ensamblaje (Parcial) (1)

Tabla 5.14: Mejor valor-F por categoría en terreno (Fuente: Elaboración propia)

Tamaño	Mejor resultado
Pequeño (9)	Ensamblaje (Parcial) (9)
Pequeño-mediano (11)	Ensamblaje (Parcial) (5)
Mediano (3)	Ensamblaje (2)
Mediano-grande (10)	Vgg16 (Parcial), Ensamblaje y Ensamblaje (Parcial) (3)
Grande (12)	Ensamblaje (Parcial) (5)
Variable (16)	Ensamblaje (6)

Tabla 5.15: Mejor valor-F por categoría en terreno (Fuente: Elaboración propia)

Para poder interpretar todos los resultados anteriores vamos a centrarnos en la medida de valor-F, pues, como hemos explicado anteriormente, el problema de la “falsa detección” va a perjudicar los valores de precisión y a beneficiar los de exhaustividad. Como primera interpretación, podemos afirmar que el mejor resultado, no solo a nivel general, sino en la mayoría de características, viene dado por el ensamblaje de los distintos ensembles de varios modelos entrenados con un conjunto fragmentado aleatoriamente. Si bien es cierto que en los últimos ensembles la mejora ha sido menor, sí que podemos afirmar que, a nivel global, siempre hemos obtenido mejores resultados a mayor número de ensamblajes, siendo los resultados de los modelos individuales los peores.

Como apunte adicional en referencia a las características que hemos definido, podemos afirmar que ni el tipo de imagen, ni el terreno han sido claves a la hora de hacer destacar un modelo sobre otro, sin embargo, en lo referente al tamaño, uno menor suele obtener mejores resultados para el ensemble de redes.

Finalmente, para facilitar la comparación de todos los resultados obtenidos en este proyecto, se muestran tres tablas donde recopilamos todos los de las tablas anteriores, representando, cada una, una medida distinta:

Identificación de objetos en imágenes por satélite utilizando técnicas Deep Learning

Guillermo Martínez Lledó

PRECISION	CONJUNTO COMPLETO					ENTRENAMIENTOS PARCIALES ENSAMBLADOS					TIPO DE IMAGEN	TERRENO	TAMANO
CATEGORIA	Redes					Redes							
	Resnet50	InceptionV3	Xception	Vgg16	Ensamblaje	Resnet50	InceptionV3	Xception	Vgg16	Ensamblaje			
Aeropuerto	52.00 %	34.00 %	39.00 %	34.00 %	38.00 %	59.00 %	34.00 %	38.00 %	32.00 %	36.00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	72.00 %	74.00 %	75.00 %	73.00 %	82.00 %	77.00 %	74.00 %	74.00 %	75.00 %	77.00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	75.00 %	80.00 %	76.00 %	77.00 %	84.00 %	80.00 %	86.00 %	81.00 %	82.00 %	87.00 %	Edificio	Tierra	Variable
Parque de atracciones	57.00 %	73.00 %	78.00 %	81.00 %	79.00 %	60.00 %	75.00 %	75.00 %	80.00 %	77.00 %	Edificio	Variable	Mediano-Grande
Acuicultura	56.00 %	64.00 %	68.00 %	65.00 %	68.00 %	56.00 %	60.00 %	63.00 %	64.00 %	64.00 %	Otros	Agua	Mediano
Zona de Arqueologia	40.00 %	54.00 %	54.00 %	58.00 %	70.00 %	44.00 %	45.00 %	44.00 %	38.00 %	48.00 %	Otros	Tierra	Variable
Granero	64.00 %	65.00 %	65.00 %	65.00 %	73.00 %	67.00 %	69.00 %	72.00 %	64.00 %	72.00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	42.00 %	28.00 %	42.00 %	50.00 %	63.00 %	55.00 %	50.00 %	66.00 %	36.00 %	61.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	73.00 %	78.00 %	77.00 %	72.00 %	85.00 %	79.00 %	82.00 %	83.00 %	78.00 %	87.00 %	Edificio	Tierra	Variable
Concesionario de coches	70.00 %	74.00 %	70.00 %	65.00 %	77.00 %	75.00 %	80.00 %	79.00 %	77.00 %	84.00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	33.00 %	39.00 %	37.00 %	35.00 %	52.00 %	49.00 %	40.00 %	46.00 %	39.00 %	51.00 %	Otros	Tierra	Variable
Campo de cultivo	73.00 %	78.00 %	82.00 %	81.00 %	82.00 %	74.00 %	78.00 %	77.00 %	76.00 %	78.00 %	Otros	Tierra	Variable
Represa	63.00 %	67.00 %	69.00 %	58.00 %	71.00 %	66.00 %	73.00 %	74.00 %	62.00 %	72.00 %	Otros	Tierra y Agua	Grande
Escombros	34.00 %	22.00 %	18.00 %	26.00 %	30.00 %	38.00 %	32.00 %	37.00 %	22.00 %	35.00 %	Otros	Tierra	Variable
Institución educativa	41.00 %	44.00 %	46.00 %	47.00 %	50.00 %	44.00 %	43.00 %	46.00 %	44.00 %	47.00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	80.00 %	82.00 %	85.00 %	82.00 %	89.00 %	86.00 %	87.00 %	87.00 %	89.00 %	90.00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	63.00 %	64.00 %	74.00 %	55.00 %	79.00 %	73.00 %	69.00 %	75.00 %	66.00 %	81.00 %	Edificio	Variable	Variable
Estación de bomberos	47.00 %	41.00 %	52.00 %	39.00 %	54.00 %	55.00 %	58.00 %	62.00 %	49.00 %	64.00 %	Edificio	Tierra	Pequeño
Camino inundado	12.00 %	19.00 %	15.00 %	29.00 %	21.00 %	13.00 %	13.00 %	14.00 %	15.00 %	15.00 %	Carretera	Tierra	Grande
Fuente	49.00 %	31.00 %	31.00 %	19.00 %	32.00 %	55.00 %	56.00 %	57.00 %	55.00 %	59.00 %	Otros	Tierra	Pequeño
Gasolinera	70.00 %	72.00 %	72.00 %	60.00 %	79.00 %	76.00 %	84.00 %	82.00 %	80.00 %	87.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	89.00 %	90.00 %	81.00 %	89.00 %	91.00 %	87.00 %	92.00 %	88.00 %	92.00 %	95.00 %	Otros	Tierra	Grande
Estación de transporte terrestre	68.00 %	64.00 %	67.00 %	59.00 %	75.00 %	71.00 %	70.00 %	70.00 %	68.00 %	72.00 %	Edificio con carretera	Tierra	Variable
Helipuerto	71.00 %	67.00 %	65.00 %	51.00 %	73.00 %	77.00 %	76.00 %	79.00 %	74.00 %	81.00 %	Carretera	Tierra	Mediano-Grande
Hospital	45.00 %	47.00 %	50.00 %	47.00 %	60.00 %	56.00 %	55.00 %	50.00 %	49.00 %	59.00 %	Edificio	Tierra	Mediano-Grande
Cruce	66.00 %	79.00 %	77.00 %	74.00 %	79.00 %	70.00 %	76.00 %	75.00 %	72.00 %	78.00 %	Carretera	Tierra	Grande
Lago o estanque	53.00 %	72.00 %	62.00 %	49.00 %	73.00 %	61.00 %	58.00 %	62.00 %	66.00 %	70.00 %	Otros	Agua	Pequeño-Mediano
Faro	51.00 %	44.00 %	31.00 %	35.00 %	46.00 %	50.00 %	51.00 %	48.00 %	48.00 %	51.00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	53.00 %	56.00 %	47.00 %	61.00 %	66.00 %	57.00 %	59.00 %	59.00 %	54.00 %	62.00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	48.00 %	48.00 %	48.00 %	52.00 %	58.00 %	51.00 %	48.00 %	53.00 %	57.00 %	58.00 %	Edificio	Tierra	Variable
Estación nuclear	42.00 %	43.00 %	24.00 %	61.00 %	48.00 %	53.00 %	32.00 %	37.00 %	43.00 %	38.00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	21.00 %	25.00 %	27.00 %	25.00 %	29.00 %	28.00 %	26.00 %	24.00 %	27.00 %	30.00 %	Edificio	Tierra	Variable
Instalación petrolera	73.00 %	73.00 %	72.00 %	68.00 %	81.00 %	79.00 %	77.00 %	73.00 %	77.00 %	81.00 %	Otros	Tierra	Pequeño-Mediano
Parque	49.00 %	53.00 %	51.00 %	44.00 %	61.00 %	52.00 %	55.00 %	53.00 %	55.00 %	58.00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaie	73.00 %	79.00 %	79.00 %	78.00 %	82.00 %	77.00 %	79.00 %	78.00 %	78.00 %	82.00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	62.00 %	69.00 %	63.00 %	68.00 %	75.00 %	64.00 %	69.00 %	68.00 %	68.00 %	71.00 %	Edificio	Tierra	Variable
Estación de policía	31.00 %	32.00 %	37.00 %	34.00 %	42.00 %	41.00 %	40.00 %	42.00 %	36.00 %	46.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	33.00 %	25.00 %	22.00 %	22.00 %	24.00 %	34.00 %	29.00 %	30.00 %	25.00 %	29.00 %	Otros	Agua	Variable
Prisión	66.00 %	71.00 %	60.00 %	67.00 %	74.00 %	71.00 %	76.00 %	76.00 %	79.00 %	78.00 %	Edificio	Tierra	Grande
Pista de carreras	90.00 %	89.00 %	90.00 %	85.00 %	92.00 %	93.00 %	91.00 %	93.00 %	93.00 %	94.00 %	Carretera	Tierra	Grande
Puente ferroviario	49.00 %	53.00 %	63.00 %	41.00 %	61.00 %	50.00 %	52.00 %	56.00 %	52.00 %	59.00 %	Otros	Variable	Variable
Instalaciones recreativas	85.00 %	84.00 %	86.00 %	86.00 %	89.00 %	87.00 %	90.00 %	87.00 %	88.00 %	90.00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	26.00 %	36.00 %	37.00 %	48.00 %	42.00 %	29.00 %	33.00 %	31.00 %	33.00 %	31.00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	60.00 %	64.00 %	73.00 %	69.00 %	72.00 %	61.00 %	64.00 %	60.00 %	61.00 %	65.00 %	Carretera	Variable	Variable
Pista de aterrizaje	76.00 %	66.00 %	65.00 %	56.00 %	67.00 %	82.00 %	72.00 %	73.00 %	61.00 %	74.00 %	Carretera	Variable	Grande
Astillero	38.00 %	24.00 %	28.00 %	22.00 %	29.00 %	46.00 %	33.00 %	40.00 %	28.00 %	37.00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	66.00 %	66.00 %	64.00 %	58.00 %	73.00 %	67.00 %	66.00 %	68.00 %	69.00 %	72.00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	69.00 %	68.00 %	66.00 %	68.00 %	74.00 %	73.00 %	75.00 %	77.00 %	77.00 %	78.00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	62.00 %	50.00 %	58.00 %	35.00 %	60.00 %	67.00 %	73.00 %	72.00 %	66.00 %	74.00 %	Edificio	Tierra	Pequeño
Granja solar	89.00 %	90.00 %	90.00 %	92.00 %	94.00 %	93.00 %	92.00 %	92.00 %	93.00 %	94.00 %	Otros	Tierra	Variable
Instalación espacial	78.00 %	44.00 %	29.00 %	47.00 %	53.00 %	86.00 %	62.00 %	64.00 %	60.00 %	75.00 %	Edificio con carretera	Tierra	Grande
Estadio	81.00 %	86.00 %	83.00 %	85.00 %	87.00 %	84.00 %	83.00 %	85.00 %	88.00 %	88.00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	86.00 %	82.00 %	83.00 %	89.00 %	91.00 %	91.00 %	86.00 %	89.00 %	90.00 %	91.00 %	Otros	Tierra	Grande
Mina de superficie	84.00 %	83.00 %	88.00 %	82.00 %	88.00 %	83.00 %	85.00 %	86.00 %	83.00 %	88.00 %	Otros	Tierra	Grande
Piscina	82.00 %	76.00 %	76.00 %	53.00 %	80.00 %	84.00 %	84.00 %	84.00 %	83.00 %	86.00 %	Otros	Agua	Pequeño
Cabina de peaje	64.00 %	66.00 %	62.00 %	62.00 %	66.00 %	74.00 %	75.00 %	75.00 %	72.00 %	76.00 %	Edificio con carretera	Tierra	Pequeño
Torre	58.00 %	47.00 %	51.00 %	42.00 %	58.00 %	68.00 %	64.00 %	65.00 %	59.00 %	69.00 %	Edificio	Tierra	Pequeño
Apertura de túnel	61.00 %	55.00 %	53.00 %	52.00 %	57.00 %	59.00 %	63.00 %	64.00 %	63.00 %	64.00 %	Otros	Tierra	Pequeño
Vertedero	63.00 %	64.00 %	62.00 %	52.00 %	72.00 %	68.00 %	66.00 %	71.00 %	59.00 %	72.00 %	Otros	Tierra y Agua	Mediano
Instalación de tratamiento de agua	87.00 %	88.00 %	90.00 %	83.00 %	91.00 %	91.00 %	93.00 %	92.00 %	91.00 %	94.00 %	Otros	Tierra y Agua	Mediano-Grande
Granja eólica	85.00 %	89.00 %	90.00 %	90.00 %	92.00 %	89.00 %	91.00 %	91.00 %	92.00 %	93.00 %	Otros	Tierra	Variable
Zoológico	49.00 %	40.00 %	54.00 %	33.00 %	55.00 %	57.00 %	59.00 %	51.00 %	49.00 %	63.00 %	Edificio	Variable	Variable
Promedio	60.45 %	60.16 %	57.82 %	57.82 %	66.74 %	65.19 %	64.65 %	65.53 %	62.92 %	68.35 %			

Tabla 5.16: Recopilación de resultados de precisión (Fuente: Elaboración propia)

Identificación de objetos en imágenes por satélite utilizando técnicas Deep Learning

Guillermo Martínez Lledó

RECALL	CONJUNTO COMPLETO					ENSEMBLE POR RED					TIPO DE IMAGEN	TERRENO	TAMANO
CATEGORIA	Resnet50	InceptionV3	Xception	Vgg16	Ensamble	Resnet50	InceptionV3	Xception	Vgg16	Ensamble			
Aeropuerto	97.00 %	93.00 %	93.00 %	90.00 %	94.00 %	96.00 %	94.00 %	96.00 %	96.00 %	96.00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	69.00 %	68.00 %	66.00 %	62.00 %	65.00 %	71.00 %	71.00 %	69.00 %	68.00 %	69.00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	71.00 %	66.00 %	67.00 %	65.00 %	69.00 %	72.00 %	68.00 %	72.00 %	67.00 %	69.00 %	Edificio	Tierra	Variable
Parque de atracciones	83.00 %	88.00 %	90.00 %	89.00 %	89.00 %	85.00 %	92.00 %	91.00 %	90.00 %	92.00 %	Edificio	Variable	Mediano-Grande
Acuicultura	88.00 %	86.00 %	81.00 %	84.00 %	88.00 %	90.00 %	89.00 %	86.00 %	85.00 %	88.00 %	Otros	Agua	Mediano
Zona de Arqueologia	59.00 %	59.00 %	60.00 %	57.00 %	62.00 %	57.00 %	68.00 %	66.00 %	65.00 %	65.00 %	Otros	Tierra	Variable
Granero	76.00 %	77.00 %	77.00 %	71.00 %	78.00 %	71.00 %	74.00 %	77.00 %	76.00 %	75.00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	31.00 %	22.00 %	33.00 %	25.00 %	33.00 %	35.00 %	45.00 %	41.00 %	35.00 %	39.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	86.00 %	85.00 %	90.00 %	88.00 %	89.00 %	87.00 %	93.00 %	92.00 %	92.00 %	92.00 %	Edificio	Tierra	Variable
Concesionario de coches	84.00 %	81.00 %	81.00 %	82.00 %	84.00 %	86.00 %	80.00 %	82.00 %	81.00 %	84.00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	29.00 %	27.00 %	39.00 %	20.00 %	29.00 %	33.00 %	38.00 %	44.00 %	33.00 %	40.00 %	Otros	Tierra	Variable
Campo de cultivo	95.00 %	94.00 %	94.00 %	93.00 %	95.00 %	96.00 %	96.00 %	96.00 %	96.00 %	96.00 %	Otros	Tierra	Variable
Represa	90.00 %	87.00 %	88.00 %	86.00 %	89.00 %	90.00 %	89.00 %	90.00 %	89.00 %	90.00 %	Otros	Tierra y Agua	Grande
Escombros	60.00 %	60.00 %	58.00 %	46.00 %	62.00 %	58.00 %	63.00 %	58.00 %	58.00 %	60.00 %	Otros	Tierra	Variable
Institución educativa	59.00 %	63.00 %	53.00 %	60.00 %	61.00 %	60.00 %	62.00 %	64.00 %	67.00 %	63.00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	86.00 %	86.00 %	90.00 %	86.00 %	88.00 %	89.00 %	87.00 %	87.00 %	88.00 %	88.00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	42.00 %	46.00 %	43.00 %	38.00 %	44.00 %	47.00 %	49.00 %	54.00 %	47.00 %	49.00 %	Edificio	Variable	Variable
Estación de bomberos	52.00 %	43.00 %	50.00 %	35.00 %	46.00 %	55.00 %	52.00 %	49.00 %	52.00 %	52.00 %	Edificio	Tierra	Pequeño
Camino inundado	60.00 %	71.00 %	59.00 %	68.00 %	67.00 %	62.00 %	71.00 %	63.00 %	69.00 %	68.00 %	Carretera	Tierra	Grande
Fuente	86.00 %	87.00 %	86.00 %	85.00 %	87.00 %	90.00 %	90.00 %	87.00 %	89.00 %	90.00 %	Otros	Tierra	Pequeño
Gasolinera	88.00 %	85.00 %	86.00 %	79.00 %	88.00 %	87.00 %	88.00 %	84.00 %	84.00 %	88.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	90.00 %	90.00 %	87.00 %	89.00 %	91.00 %	93.00 %	92.00 %	90.00 %	92.00 %	93.00 %	Otros	Tierra	Grande
Estación de transporte terrestre	71.00 %	73.00 %	76.00 %	67.00 %	72.00 %	73.00 %	75.00 %	73.00 %	74.00 %	73.00 %	Edificio con carretera	Tierra	Variable
Helipuerto	82.00 %	84.00 %	82.00 %	79.00 %	82.00 %	84.00 %	85.00 %	85.00 %	79.00 %	84.00 %	Carretera	Tierra	Mediano-Grande
Hospital	36.00 %	36.00 %	47.00 %	35.00 %	37.00 %	35.00 %	40.00 %	39.00 %	38.00 %	37.00 %	Edificio	Tierra	Mediano-Grande
Cruce	94.00 %	92.00 %	91.00 %	89.00 %	92.00 %	96.00 %	93.00 %	94.00 %	94.00 %	94.00 %	Carretera	Tierra	Grande
Lago o estanque	64.00 %	54.00 %	64.00 %	39.00 %	58.00 %	58.00 %	62.00 %	61.00 %	59.00 %	61.00 %	Otros	Agua	Pequeño-Mediano
Faro	83.00 %	77.00 %	77.00 %	68.00 %	82.00 %	84.00 %	79.00 %	79.00 %	81.00 %	81.00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	59.00 %	64.00 %	66.00 %	63.00 %	66.00 %	61.00 %	64.00 %	66.00 %	67.00 %	63.00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	41.00 %	43.00 %	44.00 %	39.00 %	39.00 %	41.00 %	44.00 %	45.00 %	43.00 %	41.00 %	Edificio	Tierra	Variable
Estación nuclear	100.00 %	91.00 %	100.00 %	100.00 %	100.00 %	91.00 %	100.00 %	91.00 %	91.00 %	91.00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	12.00 %	19.00 %	18.00 %	10.00 %	13.00 %	12.00 %	20.00 %	15.00 %	13.00 %	14.00 %	Edificio	Tierra	Variable
Instalación petrolera	86.00 %	85.00 %	81.00 %	78.00 %	87.00 %	85.00 %	88.00 %	88.00 %	83.00 %	88.00 %	Otros	Tierra	Pequeño-Mediano
Parque	68.00 %	72.00 %	72.00 %	66.00 %	70.00 %	69.00 %	72.00 %	70.00 %	71.00 %	72.00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaie	81.00 %	80.00 %	78.00 %	78.00 %	80.00 %	84.00 %	82.00 %	84.00 %	81.00 %	83.00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	72.00 %	72.00 %	72.00 %	68.00 %	73.00 %	73.00 %	73.00 %	77.00 %	72.00 %	74.00 %	Edificio	Tierra	Variable
Estación de policía	26.00 %	26.00 %	34.00 %	23.00 %	27.00 %	23.00 %	30.00 %	27.00 %	25.00 %	26.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	84.00 %	81.00 %	73.00 %	83.00 %	78.00 %	86.00 %	83.00 %	82.00 %	81.00 %	81.00 %	Otros	Agua	Variable
Prisión	65.00 %	66.00 %	71.00 %	68.00 %	68.00 %	66.00 %	69.00 %	67.00 %	67.00 %	67.00 %	Edificio	Tierra	Grande
Pista de carreras	91.00 %	91.00 %	90.00 %	91.00 %	91.00 %	91.00 %	91.00 %	95.00 %	93.00 %	93.00 %	Carretera	Tierra	Grande
Puente ferroviario	83.00 %	82.00 %	79.00 %	79.00 %	82.00 %	84.00 %	81.00 %	83.00 %	84.00 %	84.00 %	Otros	Variable	Variable
Instalaciones recreativas	92.00 %	93.00 %	91.00 %	91.00 %	92.00 %	93.00 %	93.00 %	93.00 %	93.00 %	93.00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	85.00 %	85.00 %	91.00 %	88.00 %	85.00 %	82.00 %	91.00 %	88.00 %	91.00 %	85.00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	75.00 %	72.00 %	72.00 %	70.00 %	72.00 %	75.00 %	71.00 %	72.00 %	74.00 %	72.00 %	Carretera	Variable	Variable
Pista de aterrizaje	85.00 %	88.00 %	86.00 %	88.00 %	86.00 %	88.00 %	88.00 %	88.00 %	87.00 %	87.00 %	Carretera	Variable	Grande
Astillero	44.00 %	59.00 %	63.00 %	41.00 %	52.00 %	44.00 %	52.00 %	59.00 %	41.00 %	41.00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	65.00 %	67.00 %	63.00 %	66.00 %	67.00 %	69.00 %	66.00 %	65.00 %	71.00 %	67.00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	74.00 %	80.00 %	79.00 %	62.00 %	79.00 %	75.00 %	72.00 %	73.00 %	71.00 %	73.00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	78.00 %	73.00 %	75.00 %	68.00 %	74.00 %	78.00 %	80.00 %	75.00 %	76.00 %	79.00 %	Edificio	Tierra	Pequeño
Granja solar	94.00 %	92.00 %	89.00 %	92.00 %	94.00 %	94.00 %	93.00 %	94.00 %	93.00 %	93.00 %	Otros	Tierra	Variable
Instalación espacial	78.00 %	89.00 %	100.00 %	89.00 %	89.00 %	67.00 %	89.00 %	100.00 %	100.00 %	100.00 %	Edificio con carretera	Tierra	Grande
Estadio	83.00 %	83.00 %	85.00 %	77.00 %	85.00 %	82.00 %	83.00 %	82.00 %	85.00 %	84.00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	97.00 %	93.00 %	87.00 %	91.00 %	95.00 %	97.00 %	94.00 %	95.00 %	93.00 %	95.00 %	Otros	Tierra	Grande
Mina de superficie	88.00 %	86.00 %	86.00 %	86.00 %	87.00 %	87.00 %	87.00 %	90.00 %	88.00 %	88.00 %	Otros	Tierra	Grande
Piscina	94.00 %	94.00 %	94.00 %	93.00 %	95.00 %	96.00 %	94.00 %	95.00 %	94.00 %	95.00 %	Otros	Agua	Pequeño
Cabina de peaje	98.00 %	97.00 %	96.00 %	97.00 %	97.00 %	98.00 %	98.00 %	98.00 %	98.00 %	98.00 %	Edificio con carretera	Tierra	Pequeño
Torre	77.00 %	75.00 %	78.00 %	67.00 %	77.00 %	76.00 %	78.00 %	81.00 %	76.00 %	78.00 %	Edificio	Tierra	Pequeño
Apertura de túnel	98.00 %	97.00 %	94.00 %	97.00 %	98.00 %	98.00 %	97.00 %	97.00 %	97.00 %	97.00 %	Otros	Tierra	Pequeño
Vertedero	58.00 %	62.00 %	63.00 %	57.00 %	62.00 %	60.00 %	65.00 %	64.00 %	62.00 %	62.00 %	Otros	Tierra	Mediano
Instalación de tratamiento de agua	85.00 %	86.00 %	85.00 %	85.00 %	86.00 %	84.00 %	86.00 %	87.00 %	88.00 %	87.00 %	Otros	Tierra y Agua	Mediano-Grande
Granja edica	97.00 %	96.00 %	93.00 %	94.00 %	96.00 %	97.00 %	97.00 %	97.00 %	96.00 %	97.00 %	Otros	Tierra	Variable
Zoológico	74.00 %	61.00 %	61.00 %	48.00 %	63.00 %	65.00 %	59.00 %	72.00 %	72.00 %	72.00 %	Edificio	Variable	Variable
Promedio	74.16 %	73.87 %	70.45 %	70.45 %	74.61 %	74.37 %	76.06 %	76.19 %	75.18 %	75.74 %			

Tabla 5.17: Recopilación de resultados de exhaustividad (Fuente: Elaboración propia)

Identificación de objetos en imágenes por satélite utilizando técnicas Deep Learning

Guillermo Martínez Lledó

F1-Score	CONJUNTO COMPLETO					ENSEMBLE POR RED					TIPO DE IMAGEN	TERRENO	TAMANO
CATEGORIA	Redes												
	Resnet50	InceptionV3	Xception	Vgg16	Ensamblaje	Resnet50	InceptionV3	Xception	Vgg16	Ensamblaje			
Aeropuerto	67.00 %	50.00 %	55.00 %	50.00 %	54.00 %	73.00 %	50.00 %	54.00 %	48.00 %	52.00 %	Edificio con carretera	Tierra	Grande
Aeropuerto hangar	71.00 %	71.00 %	70.00 %	67.00 %	72.00 %	74.00 %	72.00 %	71.00 %	72.00 %	72.00 %	Edificio	Tierra	Mediano
Aeropuerto terminal	73.00 %	72.00 %	71.00 %	70.00 %	76.00 %	76.00 %	76.00 %	76.00 %	74.00 %	77.00 %	Edificio	Tierra	Variable
Parque de atracciones	67.00 %	80.00 %	84.00 %	85.00 %	83.00 %	70.00 %	83.00 %	83.00 %	85.00 %	84.00 %	Edificio	Variable	Mediano-Grande
Acuicultura	68.00 %	73.00 %	74.00 %	74.00 %	76.00 %	69.00 %	72.00 %	73.00 %	73.00 %	74.00 %	Otros	Agua	Mediano
Zona de Arqueología	47.00 %	56.00 %	57.00 %	58.00 %	66.00 %	50.00 %	54.00 %	53.00 %	48.00 %	55.00 %	Otros	Tierra	Variable
Granero	69.00 %	70.00 %	70.00 %	68.00 %	75.00 %	69.00 %	71.00 %	75.00 %	70.00 %	73.00 %	Edificio	Tierra	Pequeño-Mediano
Puesto fronterizo	36.00 %	24.00 %	37.00 %	34.00 %	44.00 %	43.00 %	47.00 %	51.00 %	36.00 %	48.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Cementerio	79.00 %	81.00 %	83.00 %	80.00 %	87.00 %	83.00 %	87.00 %	88.00 %	84.00 %	90.00 %	Edificio	Tierra	Variable
Concesionario de coches	76.00 %	77.00 %	75.00 %	73.00 %	80.00 %	80.00 %	80.00 %	81.00 %	79.00 %	84.00 %	Edificio con carretera	Tierra	Mediano-Grande
Zona de construcción	31.00 %	32.00 %	38.00 %	26.00 %	37.00 %	39.00 %	39.00 %	45.00 %	36.00 %	45.00 %	Otros	Tierra	Variable
Campo de cultivo	83.00 %	86.00 %	88.00 %	87.00 %	88.00 %	84.00 %	86.00 %	85.00 %	85.00 %	86.00 %	Otros	Tierra	Variable
Represa	74.00 %	76.00 %	77.00 %	69.00 %	79.00 %	76.00 %	80.00 %	81.00 %	73.00 %	80.00 %	Otros	Tierra y Agua	Grande
Escombros	43.00 %	32.00 %	28.00 %	33.00 %	40.00 %	45.00 %	42.00 %	45.00 %	32.00 %	44.00 %	Otros	Tierra	Variable
Institución educativa	48.00 %	51.00 %	49.00 %	53.00 %	55.00 %	51.00 %	51.00 %	53.00 %	53.00 %	54.00 %	Edificio	Tierra	Pequeño-Mediano
Subestación eléctrica	83.00 %	84.00 %	87.00 %	84.00 %	89.00 %	87.00 %	87.00 %	87.00 %	89.00 %	89.00 %	Edificio	Tierra	Pequeño
Fábrica o planta eléctrica	50.00 %	53.00 %	55.00 %	45.00 %	57.00 %	57.00 %	57.00 %	63.00 %	55.00 %	61.00 %	Edificio	Variable	Variable
Estación de bomberos	49.00 %	42.00 %	51.00 %	37.00 %	50.00 %	55.00 %	54.00 %	55.00 %	50.00 %	57.00 %	Edificio	Tierra	Pequeño
Camino inundado	20.00 %	29.00 %	24.00 %	41.00 %	32.00 %	21.00 %	22.00 %	23.00 %	25.00 %	24.00 %	Carretera	Tierra	Grande
Fuente	62.00 %	45.00 %	46.00 %	30.00 %	47.00 %	68.00 %	69.00 %	69.00 %	68.00 %	71.00 %	Otros	Tierra	Pequeño
Gasolinera	78.00 %	78.00 %	78.00 %	69.00 %	83.00 %	81.00 %	86.00 %	83.00 %	82.00 %	87.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Campo de golf	90.00 %	90.00 %	83.00 %	89.00 %	91.00 %	90.00 %	92.00 %	89.00 %	92.00 %	94.00 %	Otros	Tierra	Grande
Estación de transporte terrestre	70.00 %	68.00 %	71.00 %	63.00 %	74.00 %	72.00 %	73.00 %	72.00 %	71.00 %	73.00 %	Edificio con carretera	Tierra	Variable
Helipuerto	76.00 %	74.00 %	73.00 %	62.00 %	77.00 %	80.00 %	80.00 %	82.00 %	77.00 %	83.00 %	Carretera	Tierra	Mediano-Grande
Hospital	40.00 %	41.00 %	49.00 %	40.00 %	45.00 %	43.00 %	46.00 %	44.00 %	43.00 %	45.00 %	Edificio	Tierra	Mediano-Grande
Cruce	78.00 %	85.00 %	83.00 %	81.00 %	85.00 %	81.00 %	83.00 %	83.00 %	82.00 %	85.00 %	Carretera	Tierra	Grande
Lago o estanque	58.00 %	61.00 %	63.00 %	44.00 %	65.00 %	59.00 %	60.00 %	62.00 %	62.00 %	65.00 %	Otros	Agua	Pequeño-Mediano
Faro	63.00 %	56.00 %	45.00 %	46.00 %	59.00 %	63.00 %	62.00 %	59.00 %	60.00 %	63.00 %	Edificio	Tierra y Agua	Pequeño
Instalación militar	55.00 %	60.00 %	55.00 %	62.00 %	66.00 %	59.00 %	61.00 %	62.00 %	60.00 %	62.00 %	Otros	Tierra	Mediano-Grande
Residencial colectivo	44.00 %	45.00 %	46.00 %	44.00 %	47.00 %	45.00 %	46.00 %	49.00 %	49.00 %	48.00 %	Edificio	Tierra	Variable
Estación nuclear	59.00 %	59.00 %	39.00 %	76.00 %	65.00 %	67.00 %	49.00 %	53.00 %	59.00 %	54.00 %	Edificio	Tierra y Agua	Pequeño-Mediano
Edificio de oficinas	15.00 %	22.00 %	21.00 %	15.00 %	18.00 %	16.00 %	23.00 %	19.00 %	18.00 %	19.00 %	Edificio	Tierra	Variable
Instalación petrolera	79.00 %	78.00 %	76.00 %	73.00 %	84.00 %	81.00 %	82.00 %	79.00 %	80.00 %	84.00 %	Otros	Tierra	Pequeño-Mediano
Parque	57.00 %	61.00 %	59.00 %	53.00 %	65.00 %	59.00 %	63.00 %	61.00 %	62.00 %	64.00 %	Otros	Tierra	Pequeño-Mediano
Aparcamiento o garaje	77.00 %	80.00 %	78.00 %	78.00 %	81.00 %	80.00 %	80.00 %	81.00 %	80.00 %	82.00 %	Carretera	Tierra	Pequeño-Mediano
Lugar de culto	67.00 %	71.00 %	67.00 %	68.00 %	74.00 %	68.00 %	71.00 %	72.00 %	70.00 %	72.00 %	Edificio	Tierra	Variable
Estación de policía	28.00 %	29.00 %	36.00 %	28.00 %	33.00 %	30.00 %	34.00 %	33.00 %	30.00 %	34.00 %	Edificio con carretera	Tierra	Pequeño-Mediano
Puerto	48.00 %	38.00 %	34.00 %	35.00 %	37.00 %	49.00 %	43.00 %	44.00 %	38.00 %	43.00 %	Otros	Agua	Variable
Prisión	66.00 %	69.00 %	65.00 %	68.00 %	71.00 %	69.00 %	72.00 %	71.00 %	73.00 %	72.00 %	Edificio	Tierra	Grande
Pista de carreras	90.00 %	90.00 %	90.00 %	88.00 %	91.00 %	92.00 %	91.00 %	94.00 %	93.00 %	94.00 %	Carretera	Tierra	Grande
Puente ferroviario	62.00 %	64.00 %	70.00 %	54.00 %	70.00 %	63.00 %	63.00 %	67.00 %	64.00 %	69.00 %	Otros	Variable	Variable
Instalaciones recreativas	89.00 %	88.00 %	89.00 %	88.00 %	90.00 %	90.00 %	91.00 %	90.00 %	90.00 %	92.00 %	Otros	Tierra	Pequeño-Mediano
Asentamiento empobrecido	40.00 %	51.00 %	52.00 %	62.00 %	56.00 %	43.00 %	48.00 %	45.00 %	49.00 %	46.00 %	Otros	Tierra	Mediano-Grande
Puente de carretera	66.00 %	68.00 %	73.00 %	70.00 %	72.00 %	67.00 %	67.00 %	66.00 %	67.00 %	69.00 %	Carretera	Variable	Variable
Pista de aterrizaje	80.00 %	76.00 %	74.00 %	68.00 %	75.00 %	85.00 %	79.00 %	80.00 %	72.00 %	80.00 %	Carretera	Variable	Grande
Astillero	41.00 %	34.00 %	39.00 %	28.00 %	37.00 %	45.00 %	40.00 %	48.00 %	33.00 %	39.00 %	Edificio con carretera	Tierra y Agua	Grande
Centro comercial	65.00 %	66.00 %	64.00 %	61.00 %	70.00 %	68.00 %	66.00 %	66.00 %	70.00 %	69.00 %	Edificio	Tierra	Mediano-Grande
Vivienda unifamiliar	72.00 %	74.00 %	72.00 %	65.00 %	76.00 %	74.00 %	73.00 %	75.00 %	74.00 %	75.00 %	Edificio con carretera	Tierra	Mediano-Grande
Chimenea	69.00 %	59.00 %	66.00 %	46.00 %	66.00 %	72.00 %	76.00 %	74.00 %	71.00 %	77.00 %	Edificio	Tierra	Pequeño
Granja solar	91.00 %	91.00 %	90.00 %	92.00 %	94.00 %	93.00 %	92.00 %	93.00 %	93.00 %	93.00 %	Otros	Tierra	Variable
Instalación espacial	78.00 %	59.00 %	45.00 %	62.00 %	67.00 %	75.00 %	73.00 %	78.00 %	75.00 %	86.00 %	Edificio con carretera	Tierra	Grande
Estadio	82.00 %	85.00 %	84.00 %	81.00 %	86.00 %	83.00 %	83.00 %	84.00 %	87.00 %	86.00 %	Otros	Tierra	Mediano-Grande
Tanque de almacenamiento	91.00 %	87.00 %	85.00 %	90.00 %	93.00 %	94.00 %	90.00 %	92.00 %	91.00 %	93.00 %	Otros	Tierra	Grande
Mina de superficie	86.00 %	84.00 %	87.00 %	84.00 %	88.00 %	85.00 %	86.00 %	88.00 %	86.00 %	88.00 %	Otros	Tierra	Grande
Piscina	87.00 %	84.00 %	84.00 %	68.00 %	87.00 %	90.00 %	89.00 %	89.00 %	88.00 %	90.00 %	Otros	Agua	Pequeño
Cabina de peaje	77.00 %	79.00 %	76.00 %	75.00 %	78.00 %	84.00 %	85.00 %	85.00 %	83.00 %	86.00 %	Edificio con carretera	Tierra	Pequeño
Torre	66.00 %	58.00 %	61.00 %	52.00 %	66.00 %	72.00 %	70.00 %	72.00 %	66.00 %	73.00 %	Edificio	Tierra	Pequeño
Apertura de túnel	75.00 %	71.00 %	67.00 %	67.00 %	72.00 %	74.00 %	77.00 %	77.00 %	76.00 %	77.00 %	Otros	Tierra	Pequeño
Vertedero	60.00 %	63.00 %	63.00 %	54.00 %	67.00 %	64.00 %	66.00 %	67.00 %	61.00 %	67.00 %	Otros	Tierra	Mediano
Instalación de tratamiento de agua	86.00 %	87.00 %	88.00 %	84.00 %	89.00 %	88.00 %	89.00 %	90.00 %	89.00 %	90.00 %	Otros	Tierra y Agua	Mediano-Grande
Granja edícola	91.00 %	92.00 %	91.00 %	92.00 %	94.00 %	93.00 %	94.00 %	94.00 %	94.00 %	95.00 %	Otros	Tierra	Variable
Zoológico	59.00 %	48.00 %	57.00 %	39.00 %	59.00 %	61.00 %	59.00 %	59.00 %	58.00 %	67.00 %	Edificio	Variable	Variable
Promedio	65.27 %	64.63 %	61.74 %	61.74 %	68.55 %	68.02 %	68.26 %	69.06 %	66.98 %	70.16 %			

Tabla 5.18: Recopilación de resultados de valor-F (Fuente: Elaboración propia)

6 CONCLUSIONES Y TRABAJOS FUTUROS

Finalmente, en este capítulo se presentan las conclusiones que han sido extraídas a lo largo de este trabajo teórico, además de proponer algunas ideas, dentro de la misma línea de investigación, para trabajos futuros.

6.1 Conclusiones de la experimentación

Haciendo un repaso a todo lo explicado en el presente documento, y recuperando los objetivos principales que tenía este proyecto, hemos analizado el campo de la clasificación de imágenes mediante Deep Learning y redes neuronales convolucionales. Usando algunas de las redes más conocidas en este campo, se ha realizado una valoración de la capacidad que tienen las redes seleccionadas a la hora de clasificar imágenes tomadas por satélite. Finalmente, una vez obtenido estos resultados, se ha conseguido una forma de mejorarlos, sin elevar el coste computacional, mediante un sistema de ensamblaje basado en la agregación de resultados mediante votación.

En mi opinión, esto último ha significado la mayor mejora en los resultados, en especial el ensamblaje del conjunto completo y el ensemble general de los modelos con datos con muestreo aleatorio. Aunque cualquier ensemble haya conseguido reducir la problemática de las falsas detecciones, creo que cuando hacemos un ensemble entre varias arquitecturas, conseguimos reducir el bajo porcentaje de acierto que tienen algunas redes concretas sobre ciertas categorías particulares.

6.2 Conclusiones personales

Desde prácticamente el comienzo del desarrollo de este trabajo de fin de máster propuesto hemos tenido que afrontar un inconveniente principal, el coste computacional. Éste se debe principalmente a que los conjuntos de imágenes tomadas por satélite que hemos encontrado disponían de un gran número de imágenes, y éstas solían tener una resolución alta, en resumen, estábamos intentando

trabajar con unos 400 Gigabytes de imágenes. A pesar de tener otras opciones, como reducir este conjunto, o la calidad de las imágenes, éstas limitarían el posible potencial del estudio, además de que el coste seguía siendo demasiado elevado para un equipo personal, por lo que hemos optado por usar el clúster de alta computación de la Universidad de Jaén, ADA, que, a pesar de haber carecido de algunos recursos a la hora de realizar ciertos experimentos, nos daba la oportunidad de trabajar con el conjunto FMoW sin ser alterado.

Uno de los puntos que considero más importantes en esta investigación sería todo lo referente al ensamblaje de resultados. Durante la revisión teórica, no he encontrado demasiada información en lo referente a esto, la mayoría de artículos enfocados a la clasificación de imágenes por Deep Learning solo presenta los resultados de las arquitecturas de forma individual. Sin embargo, en este trabajo hemos podido demostrar que unos resultados conjuntos entre distintas redes son capaces de proporcionar mejores resultados que los de una red particular.

6.3 Propuestas de trabajo futuro

Muchas de las mejoras que se podrían aportar a esta investigación pasarían por el impedimento de tener recursos suficientes para poder llevarlas a cabo. Como se ha mencionado en otros apartados, este estudio se centró especialmente en la clasificación de imágenes, debido a que, al intentar usar modelos enfocados a la identificación, como YOLO, nos encontrábamos que los recursos con los que contábamos eran insuficientes para poder realizar un entrenamiento apropiado.

Una de mis propuestas iría en esa línea, intentar seguir en la línea de la identificación en imágenes de características similares a éstas, pues éstos modelos suelen ser bastante eficientes y podrían llegar a mejorar los resultados obtenidos en éste estudio, además de permitirnos una identificación en tiempo real, pues son capaces de identificar objetos en video además de imagen. Además, es un campo que avanza con rapidez, y es posible que otros modelos más novedosos de detección no requieran de un coste computacional tan elevado. Como curiosidad en lo referente a esta última frase, cuando se empezó este trabajo, la versión 3 de YOLO era la

novedad en el campo de la detección en imágenes, y a día de hoy, ya podemos encontrarnos modelos que usan la versión 5.

Una última propuesta que me gustaría aportar, que además no requeriría a priori elevar el coste computacional, sería probar a modificar el sistema de ensamblaje, intentando mejorar estos resultados conjuntos. Una idea que se me ocurre respecto a esto, sería utilizar un sistema de votos ponderado, en lugar de uno sencillo como el que se ha utilizado aquí. Dicho sistema se podría realizar de varias formas, por ejemplo, que el voto de cada modelo se pondere en relación al porcentaje de acierto que haya obtenido cada modelo durante el entrenamiento, sin embargo, habría que tener cuidado con esto, pues si un modelo sufre sobreaprendizaje, el resultado final sería peor. Otros posibles métodos sería incluir los resultados de la capa *Softmax*, usando el porcentaje de “confianza” que se emite en cada imagen para determinar la clasificación como valor de ponderación, o incluso, que cada modelo emita n votos ponderados para cada imagen con los n primeros resultados obtenidos por la capa *Softmax*.

7 DEFINICIONES Y ABREVIATURAS

- CNN: Redes neuronales convolucionales (Convolutional neural network).
- SSH: Acceso remoto seguro mediante intérprete de comandos (Secure Shell).
- KDD: Descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases).
- ANN: Red neuronal artificial (Artificial Neural Network).
- DNN: Red neuronal profunda (Deep Neural Network).
- GPU: Unidad de procesamiento gráfico (Graphics Processing Unit).
- RAM: Memoria de acceso aleatorio (Random-Access Memory).
- HDF: Formato de datos jerárquico (Hierarchical Data Format).
- ReLU: Unidad lineal rectificadora (Rectified Linear Unit).
- VP: Verdaderos positivos.
- FP: Falsos positivos.
- FN: Falsos negativos.
- VN: Verdaderos negativos.

8 BIBLIOGRAFÍA

- [1] R. Pressman (2010): *Ingeniería del Software*. McGraw-Hill.
- [2] A. Bonner (2019): *The Complete Beginner's Guide to Deep Learning: Convolutional Neural Networks and Image Classification*: <https://towardsdatascience.com/wtf-is-image-classification-8e78a8235acb>
- [3] R. Gandhi. (2018): *R-CNN, Fast R-CNN, Faster R-CNN, YOLO – Object Detection Algorithms*: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>
- [4] G. Garza. (2019): *Mask R-CNN for Ship Detection & Segmentation*: <https://towardsdatascience.com/mask-r-cnn-for-ship-detection-segmentation-a1108b5a083>
- [5] S. Rawat. (2019): *Airplanes Detection for Satellite using Faster RCNN*: <https://towardsdatascience.com/airplanes-detection-for-satellite-using-faster-rcnn-d307d58353f1>
- [6] G. Christie, N. Fendley, J. Wilson, R. Mukherjee. (2018). *Conjunto de imágenes FMoW (Functional Map of the World)*: <https://arxiv.org/pdf/1711.07846.pdf>
<https://www.iarpa.gov/challenges/fmow.html>
- [7] Dr. Alex Liu (2015). *Data Science and Data Scientist*: <http://www.researchmethods.org/DataScienceDataScientists.pdf>
- [8] Universidad de Jaén. *Sistema de supercomputación ADA*: <https://www.ujaen.es/centros/ceatic/servicios/supercomputacion/cluster-ada>

- [9] K. He, X. Zhang, S. Ren, J. Sun (2015). *Deep Residual Learning for Image Recognition*: <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>
- [10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens (2015). *Rethinking the Inception Architecture for Computer Vision*: <https://arxiv.org/pdf/1512.00567.pdf>
- [11] C. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu (2014): *Deeply-Supervised Nets*: <https://arxiv.org/pdf/1409.5185.pdf>.
- [12] F. Chollet (2017). *Xception: Deep Learning With Depthwise Separable Convolutions*
https://openaccess.thecvf.com/content_cvpr_2017/papers/Chollet_Xception_Deep_Learning_CVPR_2017_paper.pdf
- [13] K. Simonyan, A. Zisserman (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*: <https://arxiv.org/pdf/1409.1556.pdf>
- [14] T. Tieleman and G. Hinton (2012). *Divide the gradient by a run-ning average of its recent magnitude. COURSERA: NeuralNetworks for Machine Learning*,
- [14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al (2014). *Imagenet large scale visual recognition challenge*.
- [15] G. Hinton, O. Vinyals, and J. Dean (2015). *Distilling the knowledgein a neural network*.
- [16] S. Ioffe and C. Szegedy. Batch normalization (2015): *Accelerating deep network training by reducing internal covariate shift*. En las actas de la 32ª Conferencia Internacional sobre Machine Learning, páginas 448–456.
- [17] A. Kaushik: *Understanding ResNet50 architecture*: <https://iq.opengenus.org/resnet50-architecture/>

- [18] Conjunto de imágenes ImageNet: <http://www.image-net.org/>
- [19] G. Press, (2013): *A Very Short History of Data Science*: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=ae08c3b55cfc>
- [20] Y. Bengio, A. Courville, and P. Vincent (2013): *Representation Learning: A Review and New Perspectives*, *IEEE Trans. PAMI*, special issue Learning Deep Architectures
- [21] M. van Gerven, S. Bohte (2018): *Artificial Neural Networks as models of Neural Information Processing*: <https://www.frontiersin.org/research-topics/4817/artificial-neural-networks-as-models-of-neural-information-processing>
- [22] Wikipedia: *Artificial Neural Network*: https://en.wikipedia.org/wiki/Artificial_neural_network
- [23] F. Rosenblatt (1958): *The Perceptron: A Probabilistic Model for Information Storage And Organization In The Brain*.
- [24] D. J. Matich (2001): *Redes Neuronales: Conceptos Básicos y Aplicaciones*: https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monografias/matich-redesneuronales.pdf
- [25] R. Keim (2019): *How to Use a Simple Perceptron Neural Network Example to Classify Data*: <https://www.allaboutcircuits.com/technical-articles/how-to-perform-classification-using-a-neural-network-a-simple-perceptron-example/>
- [26] K. Fukushima (1980): *Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*: <https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf>

- [27] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner (1998): *Gradient-Based Learning Applied to Document Recognition*: <http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>
- [28] *Redes neuronales recurrentes*: <http://numerentur.org/rnr/>
- [29] Wikipedia: *One-hot*: <https://en.wikipedia.org/wiki/One-hot>
- [30] *¿Cómo funcionan las Convolutional Neural Networks? Visión por Ordenador*: <https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/>
- [31] S. Saha (2018): *A Comprehensive Guide to Convolutional Neural Networks*: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [32] *Hierarchical Data Formats – What is HDF5?*: <https://www.neonscience.org/about-hdf5>
- [33] *Convolutional Neural Networks (CNNs)*: <https://anhreynolds.com/blogs/cnn.html>
- [34] A. Rubiales (2020): *¿Qué es underfitting y overfitting?*: <https://rubialesalberto.medium.com/qu%C3%A9-es-underfitting-y-overfitting-c73d51ffd3f9>
- [35] *Convolutional Neural Networks*: <https://www.oreilly.com/library/view/learning-tensorflow/9781491978504/ch04.html>
- [36] S. Sharma (2017): *Activation functions in neural networks*: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>
- [37] Y. LeCun, L. Bottou, G. Orr y K. Müller (1998): *Efficient BackProp*: <http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf>

- [38] *Understanding Data Augmentation | What is Data Augmentation & how it works?:*
<https://www.mygreatlearning.com/blog/understanding-data-augmentation/>
- [39] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth (1996): *Knowledge Discovery and Data Mining: Towards a Unifying Framework:*
<https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf>
- [40] *Descubriendo información en bases de datos:*
<http://timeofsoftware.com/descubriendo-informacion/>
- [41] Y. Belinchón (2010): *Minería de datos:*
<http://www.it.uc3m.es/jvillena/irc/practicas/10-11/15mem.pdf>
- [42] D. Lam, R. Kuzma, K. McGee, S. Dooley, M. Laielli (2018): *xView: Objects in Context in Overhead Imagery:* <http://xviewdataset.org/>
<https://arxiv.org/pdf/1802.07856.pdf>
- [43] D. Kingma, J. Lei Ba (2017): *Adam: A Method for Stochastic Optimization:*
<https://arxiv.org/pdf/1412.6980.pdf>