

Konputazio Ingeniaritza eta
Sistema Adimentsuak Unibertsitate Masterra
Máster Universitario en Ingeniería Computacional
y Sistemas Inteligentes

Konputazio Zientziak eta Adimen Artifiziala Saila
Departamento de Ciencias de la Computación e Inteligencia Artificial

Master Tesia
Tesis de Máster

Preprocesado y clasificación de imágenes médicas
para el tratamiento SBRT

Unai Muñoz Aseguinolaza

Zuzendaritza
Dirección

Naiara Aginako Bengoa
Ciencia de la Computación e Inteligencia Artificial

Basilio Sierra Araujo
Ciencia de la Computación e Inteligencia Artificial

Trabajo Fin de Máster

Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes

Preprocesado y clasificación de imágenes médicas para el tratamiento SBRT

Unai Muñoz Aseguiolaza

Dirección

Naiara Aginako Bengoa

Basilio Sierra Araujo

2022ko uztailaren 6a/6 de julio de 2022

Esker onak / Agradecimientos

Master Amaierako Lana bukatu ostean pertsona batzuk bereziki eskerrak merezi dituzte.

Lehenik eta behin, Naiara eta Basiri. Lan hau posible egiteagatik, jarritako arretagatik eta aukera berriak eskeintzeagatik.

Etxekoei. Momentu latzenetan ere, entzuteagatik eta laguntzeko prest egoteagatik.

Eibarko lagunei. Nahiz eta aurten gutxi ikusi garen, gertu sentiarazteagatik.

Donostiko lagunei. Nerea, Gari, Miren, Jone, Álvaro eta Irati. Donostian egonaldia errazteagatik eta behar izan dudan edozein momentuan elkarregaz egoteagatik.

Azkenik, zuri. Nire alboan egonez jarraitzeko indarra emateagatik.

Resumen

En este proyecto de fin de máster se han desarrollado y estudiado herramientas para el preprocesado y clasificación de imágenes médicas para el tratamiento de Radiocirugía Estereotáctica (SBRT). La prevención y terapia contra el cáncer de pulmón sigue siendo un reto dentro del campo de la medicina y también de la visión por computador.

La idea de este proyecto de investigación ha supuesto un desafío en cuanto a la metodología de trabajo. A pesar de trabajar con imágenes tomográficas, las técnicas llevadas a cabo han sido muy diferentes entre sí. Por una parte, se ha propuesto un experimento relacionado con la predicción de frames con el fin de intentar anticiparse a la evolución de alguna anomalía dentro de la zona pulmonar. Por otra parte, se han desarrollado varios modelos capaces de identificar el contenido de las imágenes como la zona corporal o el tipo de plano tomográfico realizado. De esta manera se ha conseguido facilitar la selección de datos en futuros modelos a realizar.

En cuanto a los conjuntos de imágenes empleados podemos numerar diversas bases de datos de diferentes características y contenido. Por una parte, las fases previas de todos los experimentos han contado con conjuntos públicos más simples, tanto de carácter médico como de otros; Una vez realizadas las pruebas pertinentes, se ha trabajado con una base de datos (DB_SBRT) privada perteneciente a un proyecto real de tratamiento de cáncer. A lo largo de todo el proyecto, de forma paralela, se ha efectuado un trabajo de etiquetado manual en el que se ha cuestionado la posibilidad de que cada imagen pertenezca a un plano axial o no.

Índice de contenidos

Índice de contenidos	v
Índice de figuras	vii
Índice de tablas	ix
1 Introducción	1
2 Motivación y Objetivos	3
2.1. Anatomía del pulmón	3
2.2. Cáncer de pulmón	4
2.3. Radiocirugía estereotáctica (SRS/SBRT)	5
2.4. Radiómica y clasificación	7
2.5. Objetivos	8
3 Conceptos teóricos	9
3.1. Redes neuronales	9
3.1.1. Conceptos matemáticos	10
3.2. Redes neuronales convolucionales	11
4 Descripción de las bases de datos	15
4.1. MNIST	16
4.2. Lung-Pet-Ct-Dx	17
4.3. MosMedData	18
4.4. Base de datos del proyecto SBRT (DB_SBRT)	19
4.4.1. Aproximaciones en el etiquetado	20

5	Exploración de imágenes tumorales	25
5.1.	Exploración de imágenes cerebrales	25
5.2.	Exploración en 2D	27
5.3.	Exploración en 3D	30
6	Modelos de predicción de frames	33
6.1.	Construcción del modelo	34
6.2.	Comportamiento del modelo	36
6.2.1.	Predicción de frames	37
6.2.2.	Rendimiento del modelo	40
6.3.	Técnicas de mejora	43
7	Modelos de clasificación	45
7.1.	Clasificación de imágenes 3D	46
7.2.	Clasificación de imágenes 2D	48
7.3.	Resultados de los experimentos	50
8	Conclusiones y trabajo futuro	57
	Bibliografía	59

Índice de figuras

2.1. Lóbulos pulmonares y sus componentes	4
2.2. Lóbulos pulmonares afectados por cáncer de pulmón	5
2.3. Modalidades tomográficas	6
3.1. Deep Neural Network	10
3.2. Perceptron Model	10
3.3. Un ejemplo de operación de convolución con un kernel de tamaño 3×3	12
3.4. Un ejemplo de operación de max pooling con un filtro de tamaño 2×2	12
3.5. Un ejemplo de operación de Global Average Pooling con un filtro de tamaño 2×2	13
3.6. RGB matrix	13
4.1. Ejemplos de la base de datos MNIST	16
4.2. Cantidad de estudios extraídos por centro hospitalario	19
4.3. Proporción de imágenes por parte del cuerpo	21
4.4. Cantidad de imágenes según la parte del cuerpo y el etiquetado manual	22
4.5. Proporción de imágenes por tipo de imagen	23
4.6. Proporción de imágenes por tipo de imagen y el etiquetado manual	23
5.1. Imagen craneal original	26
5.2. Histograma de píxeles no-negros	26
5.3. Initial Binary Image, Final Binary Image and Skull Stripped Away	27
5.4. Resultado de la detección tumoral	28
5.5. Pulmón del paciente A0001	29
5.6. Pulmón del paciente A0002	29
5.7. Localización del tumor para el paciente NHC70D2E159 y el estudio ACCA2D0A707	30
5.8. Localización del tumor para el paciente NHC6731499A y el estudio ACCEB5D3D24	31
6.1. Arquitectura de la CNN para la predicción de vídeo de fotogramas siguientes	35

6.2. Predicción de frames para el caso 1.1	37
6.3. Predicción de frames para el caso 1.2	38
6.4. Predicción de frames para el caso 1.3	38
6.5. Predicción de frames para el caso 2.1	39
6.6. Predicción de frames para el caso 2.2	39
6.7. Rendimiento del caso 1.1	40
6.8. Rendimiento del caso 1.2	41
6.9. Rendimiento del caso 1.3	42
6.10. Rendimiento del caso 2.1	42
6.11. Rendimiento del caso 2.2	43
7.1. Arquitectura CNN para la clasificación de imágenes 3D	47
7.2. Montaje de los primeros 40 los cortes de un TAC	48
7.3. Arquitectura CNN para la clasificación de imágenes 2D	50
7.4. Rendimiento del caso 3.1	50
7.5. Primer bloque de ejemplos de clasificación (Experimento 4)	51
7.6. Segundo bloque de ejemplos de clasificación (Experimento 4)	51
7.7. Tercer bloque de ejemplos de clasificación (Experimento 4)	52
7.8. Cuarto bloque de ejemplos de clasificación (Experimento 4)	52
7.9. Quinto bloque de ejemplos de clasificación (Experimento 5)	53
7.10. Sexto bloque de ejemplos de clasificación (Experimento 5)	53
7.11. Séptimo bloque de ejemplos de clasificación (Experimento 5)	54
7.12. Octavo bloque de ejemplos de clasificación (Experimento 6)	54
7.13. Noveno bloque de ejemplos de clasificación (Experimento 5)	55

Índice de tablas

3.1. Tipos de datos y su rango de valores	14
4.1. Descripción detallada de la base de datos MNIST	16
4.2. Descripción detallada de la base de datos Lung-Pet-Ct-Dx	17
4.3. Descripción detallada de la base de datos DB_SBRT	20
6.1. Características de los modelos de predicción de frames	34
7.1. Características de los modelos de clasificación de imágenes 3D	46
7.2. Características de los modelos de clasificación	48
7.3. Características de los modelos de clasificación	49
7.4. Resultados de los modelos de clasificación	56

Introducción

Los recientes desarrollos en tecnología e investigación han ofrecido una gran variedad de nuevas técnicas para el análisis de imágenes y datos dentro del campo de la medicina. La investigación médica ayuda a que los médicos e investigadores conozcan más sobre la salud y las enfermedades en los seres humanos y también les permite conocer mejores maneras de prevenir y tratar enfermedades.

La cantidad de datos médicos recogidos hasta día de hoy es enorme, sin embargo, es crucial hacer un buen uso de éstos para contribuir al desarrollo de la medicina. Este tipo de datos es muy variado, y puede incluir mapas, textos, vídeos o imágenes debido a los diferentes equipos utilizados [1]. Además, la calidad de los datos varía mucho y pueden presentar características fluctuantes que pueden cambiar con el tiempo y eventos específicos. Todo avance metodológico en las ciencias médicas constituye una parte fundamental en la adquisición y generación de nuevos conocimientos. En este ámbito, a la hora de hacer diagnósticos o tomar decisiones referentes a la detección de cáncer, se requiere un conocimiento humano con suficiente experiencia. En consecuencia, en el caso de evaluar una gran cantidad de datos, el proceso puede ser muy lento y por lo tanto, ha aumentado la necesidad de contar con clasificadores y analizadores de imagen que automaticen el trabajo.

La investigación científica provee herramientas para el desarrollo de diversas disciplinas biomédicas y a través de ella se mejora la habilidad para entender y establecer de manera diferencial aspectos observacionales de la medicina basada en evidencia.

Para poder afrontar un reto tan ambicioso es necesario utilizar uno de los avances más notables de los últimos tiempos en el diagnóstico radiológico del cáncer, la radiómica [2]. Esta técnica tiene como objetivo extraer datos cuantitativos de las imágenes médicas (tomografía computarizada, resonancia magnética, etc.), con el fin de construir un modelo que sea capaz de mejorar el diagnóstico de enfermedades focales. Asimismo, con estos datos se pretende identificar y describir con precisión una entidad patológica.

Es una evidencia que en los últimos años la inteligencia artificial ha estado cada vez más presente en el campo de la radiología y debemos tener en cuenta que las ramas de Machine Learning y Deep Learning están muy presentes en este ámbito.

Por una parte, *Machine Learning* [3] es una rama de los algoritmos computacionales que están diseñados para simular la inteligencia humana aprendiendo del entorno. Esta técnica se ha aplicado en diversos campos como el reconocimiento de patrones, visión por computador, aplicaciones biomédicas y médicas... Especialmente en el campo de la medicina, el grado de complejidad de estos procesos puede ser elevado y puede implicar varias interacciones así como la toma de decisiones. Esto conlleva utilizar algoritmos de aprendizaje automático para optimizar y automatizar estos procesos. Los modelos generados tienen el potencial de mejorar a medida que pasa el tiempo mediante la incorporación de nuevos datos de entrenamiento y ajustando el algoritmo con el fin de obtener resultados más precisos.

Por otra parte, *Deep Learning* [4] es un subconjunto de modelos de Machine Learning basados en una estructura de una red neuronal (NN) cuya ventaja principal se basa en la posibilidad de integrar la extracción de características. Dentro del aprendizaje profundo, '*Convolutional Neural Networks*' (CNN) han supuesto un gran avance en el procesamiento de imágenes, vídeo, voz y audio, mientras que las redes neuronales tradicionales se han centrado particularmente en datos secuenciales como el texto y el habla.

Las CNN han alcanzado una precisión significativa en tareas de visión por computador como el reconocimiento de objetos, la clasificación y la segmentación. Esta última tarea se encarga de producir anotaciones que delimiten estructuras o regiones de interés, como por ejemplo, lesiones, tumores o deformaciones. La segmentación manual todavía está considerada como la técnica más fiable pero es muy lenta y presenta limitaciones a nivel del software y hardware utilizado. La Inteligencia Artificial muestra desarrollos más prometedores en la segmentación automatizada a base de reducir los errores, optimizar el tiempo y aumentar la eficacia de los métodos.

Motivación y Objetivos

El presente capítulo hace una breve descripción acerca de la motivación para el preproceso y clasificación de imágenes médicas a la vez que se presentan los objetivos que se han querido alcanzar a lo largo de todo el proyecto. A pesar de que se ha trabajado con numerosas bases de datos y herramientas, el proyecto tendrá como base algunos conceptos que serán indispensables entender para poder realizar el seguimiento de manera sencilla.

La detección de tumores y anomalías mediante el tratamiento de radiocirugía estereotáctica supone un gran reto y es imprescindible tener una buena base de información antes de empezar a trabajar en ella. Por lo tanto, debemos tener en cuenta algunos conceptos que nos ayudarán a entender de forma clara la composición y el significado de cada imagen.

2.1. Anatomía del pulmón

Los pulmones son estructuras anatómicas pertenecientes al sistema respiratorio que permiten a la sangre recibir oxígeno procedente del aire inspirado. De la misma manera, se consigue desprender del dióxido de carbono, el cual pasa al aire espirado. En la Figura 2.1 [5] se muestra el dibujo de un par de pulmones a la vez que se indica sus componentes más importantes.

Como norma general, el pulmón derecho es significativamente más grande que el izquierdo porque el corazón se desplaza más hacia la izquierda que hacia la derecha. Además, este par de organos está formado por lóbulos. El derecho tiene un lóbulo superior, un lóbulo medio y un lóbulo inferior; mientras que el izquierdo tiene un lóbulo superior y uno inferior.

Diversas son las enfermedades que afectan al pulmón; entre las más frecuentes se encuentran la neumonía, el cáncer de pulmón, el asma, o la bronquitis. El conocimiento de la posible aparición de los cánceres [5] ha conducido al desarrollo de nuevos enfoques de tratamientos. Sin embargo, la comprensión de la arquitectura básica de este órgano es fundamental para poder analizar su comportamiento.

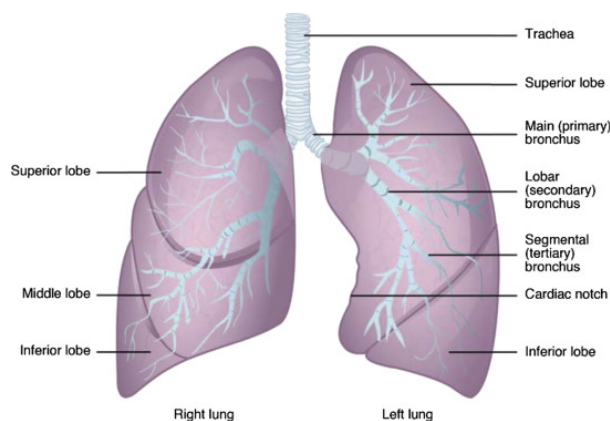


Figura 2.1: Lóbulos pulmonares y sus componentes

2.2. Cáncer de pulmón

El cáncer pulmonar es un tipo de cáncer que se inicia en los pulmones y se origina cuando las células del cuerpo comienzan a crecer sin control. El cáncer de pulmón es el segundo cáncer más común [6] que afecta tanto a hombres como a mujeres (Alrededor del 13%). Para el año 2022, los cálculos de la *Sociedad Americana Contra El Cáncer* en Estados Unidos son.

El cáncer de pulmón principalmente ocurre en las personas de edad avanzada. La mayoría de las personas diagnosticadas con cáncer de pulmón tiene 65 años o más y la edad media alcanza los 70 años. El cáncer de pulmón es por mucho la causa principal de muerte por cáncer entre hombres y mujeres, representando casi un 25 % de todas las muertes por cáncer.

El tabaco es la causa más importante de los cánceres de pulmón. Cada año se diagnostica la enfermedad a más de dos millones [7] de personas en todo el planeta y la mayoría tienen un historial de tabaquismo. Sin embargo, alrededor del 15 % nunca han fumado y ese cáncer de pulmón se da con más frecuencia en las mujeres y a una edad más temprana que el cáncer de pulmón en los fumadores. Existen importantes diferencias geográficas, raciales y de género en la incidencia y algunos informes sugieren que las mujeres pueden tener un mayor riesgo de padecer cáncer de pulmón por la exposición a los carcinógenos del humo del tabaco.

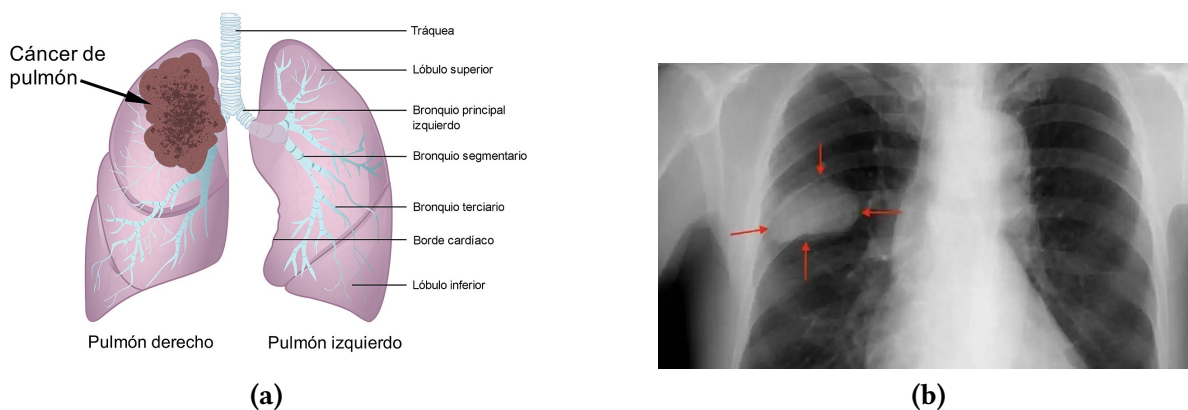


Figura 2.2: *Lóbulos pulmonares afectados por cáncer de pulmón*

2.3. Radiocirugía estereotáctica (SRS/SBRT)

La radiómica es una técnica derivada de la radiología y oncología, que consiste en obtener información relevante y cuantificable a partir de imágenes médicas. Es decir, el objetivo es extraer información en forma de valores numéricos, que corresponden a una cierta región de interés, por ejemplo, un tumor o una lesión. Los parámetros más utilizados para llevar a cabo la extracción son el tamaño, la forma o la textura de la región analizada. Las imágenes médicas, como todas las imágenes digitales, están compuestas por píxeles. Por lo tanto, podemos separar una región del resto de la imagen mediante técnicas de segmentación. De esta manera, en el caso de tumores, analizando los valores de los píxeles deseados, podríamos tener acceso de manera sencilla a los parámetros mencionados.

La gran mayoría de las imágenes utilizadas en este proyecto han sido adquiridas para formar series detalladas del interior del cuerpo de los pacientes. Estas imágenes son creadas por una máquina de rayos X mediante diferentes técnicas, y entre ellas podemos destacar la tomografía axial computarizada (TAC/TC) y la tomografía por emisión de positrones (PET).

Ambos dispositivos proporcionan una imagen tomográfica que representa un corte transversal del cuerpo, pero cada modalidad aporta diferente información al estudio. Por una parte, una TC crea una imagen detallada y sin movimiento de los órganos, huesos y tejidos; mientras que la PET muestra cómo funcionan los tejidos del cuerpo a nivel celular. Por consiguiente, la combinación de ambos sistemas ofrece tanto información metabólica, como la estructura anatómica [8], presentándose así como un sistema superior a cualquiera de los dos por separado. Además de poder combinar la información obtenida por los dos sistemas, se consigue un dispositivo que permite distinguir tejido maligno de benigno; y esto puede ayudar a la detección precoz de tumores que aún no son visibles en un escáner TC.

2. MOTIVACIÓN Y OBJETIVOS

Después de aplicar el tratamiento por primera vez se escanea el nódulo mediante una imagen TAC para ver si ha habido algún cambio en la estructura de los tejidos. En el caso de un resultado positivo, es decir, el nódulo sigue ahí, las siguientes imágenes de control serán de tipo PET. En el caso negativo, es decir, no hay tumor, los siguientes controles también serán hechos con imagen de tipo TAC. La Figura 2.3 muestra imágenes del estudio realizado en el artículo "Tomografía por emisión de positrones de cuerpo completo (PET/TC)" [9] mostrando un ejemplo de las tres modalidades tomográficas explicadas con anterioridad.

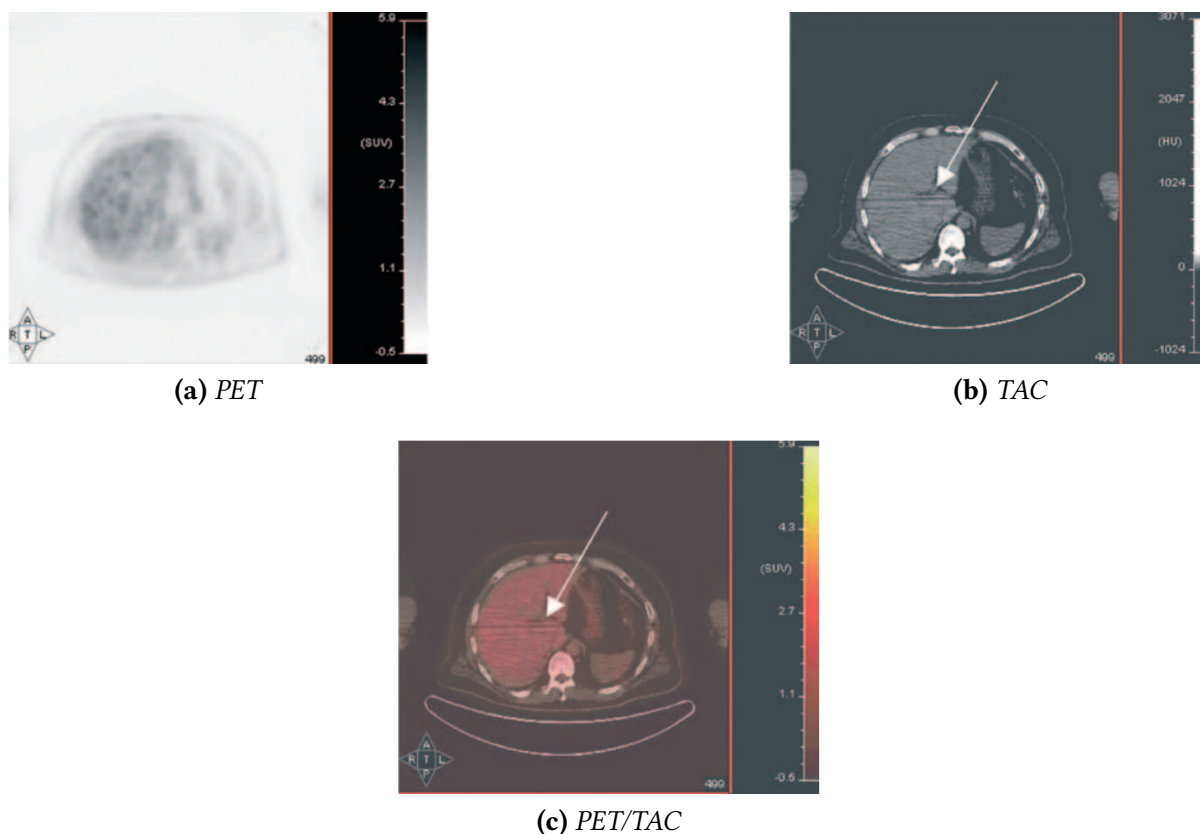


Figura 2.3: Modalidades tomográficas

La radiocirugía estereotáctica (SBRT) es una técnica de radioterapia empleada por médicos con el fin de tratar los cánceres de pulmón con mayor precisión. Además, se pretende reducir la exposición a la radiación de los tejidos sanos adyacentes al posible tumor. Este sistema se aplica para tratar los cánceres de pulmón en etapas tempranas cuando no se puede hacer la cirugía debido a la salud de la persona.

2.4. Radiómica y clasificación

La aplicación de la radiómica en el campo de la medicina, concretamente en las enfermedades pulmonares, supone mejorar el flujo de trabajo en el pronóstico, el seguimiento y la monitorización de la respuesta al tratamiento.

La clasificación diagnóstica actual se basa en el tamaño y la tasa de crecimiento como principales diferenciadores entre los nódulos benignos y malignos. Sin embargo, este enfoque sigue siendo imperfecto y debe mejorarse, ya que la detección de nódulos pequeños es una tarea difícil que no se puede pasar por alto. Diferentes grupos de investigación están desarrollando modelos radiómicos para describir nódulos, basados en diferentes parámetros como el tamaño, la forma, los márgenes, la textura y la intensidad. Su objetivo es determinar si el tumor es maligno o benigno, y precisar su pronóstico con el fin de evaluar su respuesta al tratamiento.

Además de diferenciar el cáncer de pulmón de los nódulos pulmonares benignos, el análisis radiómico también amplía su perspectiva a diferentes evaluaciones del tratamiento. Por ejemplo, en el artículo [10] definieron una firma radiómica, mostrando una precisión del 84 % en la clasificación de nódulos pulmonares. El estudio del tiempo de duplicación del volumen (VTD) puede conducir a una mejor evaluación de la tasa de crecimiento del nódulo. Como norma general, un nódulo con un VTD superior a 400-600 días tiene una probabilidad menor de ser maligno respecto a un VTD inferior a 400 días.

Según el artículo [11], los cambios temporales en las características radiómicas (delta-radiomics) de los nódulos pulmonares indican una tendencia maligna frente a la benigna. Se ha demostrado que las características radiómicas de los nódulos benignos se mantienen estables a lo largo del tiempo, mientras que las características radiómicas de los nódulos malignos cambian significativamente. Además, las diferentes segmentaciones pueden afectar a la extracción de características radiómicas. Al realizar un estudio basado en un conjunto común de imágenes, los responsables del artículo [12] demostraron una alta correlación entre grupos de características, mientras que las características específicas no se correlacionaron.

El cáncer de pulmón sigue siendo la principal causa de muerte entre las enfermedades neoplásicas en hombres y mujeres de todo el mundo. A día de hoy, las biopsias siguen siendo necesarias para establecer el diagnóstico y el estado de los tumores. Gracias a los avances tecnológicos en IA, el análisis radiómico podría considerarse como una herramienta de biopsia virtual.

2.5. Objetivos

Después de haber estudiado y analizado lo que se ha desarrollado dentro del tema elegido, es hora de estructurar el proyecto que se va a llevar a cabo para el Trabajo de Fin de Máster. A pesar de ser un proyecto ambicioso debemos determinar de manera concreta los objetivos de este proyecto con el fin de concretar una meta. Es importante mencionar que nos encontramos ante un trabajo de investigación en el que se pretende estudiar diferentes comportamientos y herramientas. En consecuencia, a lo largo de esta memoria se han desarrollado varios experimentos para así poder analizar y comparar diversas cuestiones.

En cuanto al trabajo realizado podemos destacar varias metas que servirán para explorar el funcionamiento y rendimiento de los modelos creados. En base a las herramientas empleadas éstos son los objetivos propuestos en este trabajo:

- **Predicción.** Construir un modelo capaz de predecir la evolución de una posible anomalía es fundamental para la prevención y tratamiento de la misma. La idea de este objetivo se basa en coger como referencia una serie de imágenes que representan una línea temporal, y anticiparse a la evolución de ese comportamiento mediante la construcción de nuevos frames.
- **Clasificación.** También es importante identificar con qué imágenes estamos trabajando para ser eficientes y evitar trabajo innecesario. En consecuencia, se pretende crear diferentes modelos capaces de clasificar la parte del cuerpo analizada en cada frame. De la misma manera, es esencial detectar la composición e identificar el plano tomográfico de cada imagen para una aplicación efectiva.
- **Análisis.** Una vez construidos los modelos pertinentes se ha llevado a cabo la evaluación en cuanto al rendimiento. Valorar si un modelo ha sido entrenado de manera eficiente no siempre es tarea sencilla. Además, saber si el diseño de un experimento se ajusta a las características de una base de datos puede ser de gran ayuda para futuras aproximaciones y así no partir desde cero.

Conceptos teóricos

3.1. Redes neuronales

Las redes neuronales, también conocidas como redes neuronales artificiales, son un subconjunto del aprendizaje automático [13] y forman el núcleo de los algoritmos de *Deep Learning*. Se inspiran en el cerebro humano, imitando la forma en la que las neuronas se comunican entre sí con el fin de clasificar y agrupar datos a gran velocidad. Es decir, estos algoritmos nos ayudan a agrupar y clasificar datos no etiquetados mediante la búsqueda de similitudes de un conjunto de datos etiquetados de antemano.

Una de las características principales de las redes neuronales es su composición. Esta estructura se compone mediante tres tipos de capas [14]. La capa de entrada o *input layer* está formada por los datos iniciales mientras que la capa de salida o *ouput layer* produce el resultado para los datos de entrada. Las capas ocultas o *hidden layers* se encuentran en la mitad del proceso y se encargan de realizar los cálculos necesarios. Cada nodo está conectado con cada nodo de la capa siguiente y cada conexión tiene un peso específico que representa el impacto que tiene en el nodo de la siguiente capa.

En la Figura 3.1 encontramos tres círculos amarillos que representan la capa de entrada y normalmente se anotan como vector X . Los círculos azules y verdes forman las capas ocultas y representan los nodos de activación y normalmente se anotan como W o θ . La capa formada por círculos rojos es la capa de salida o el valor predicho.

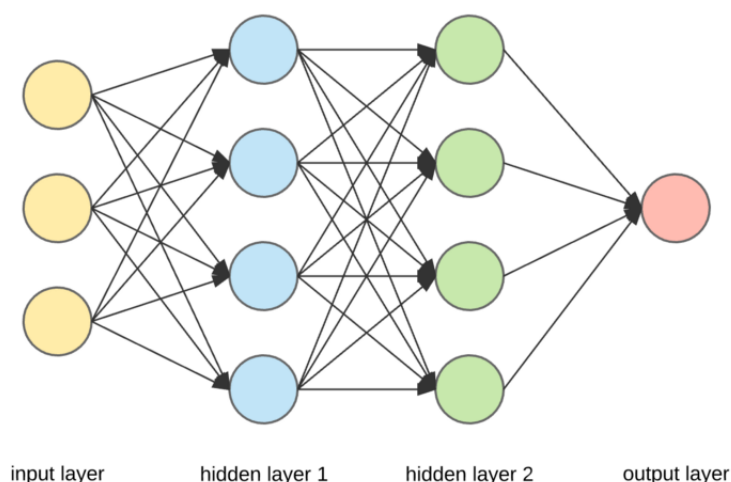


Figura 3.1: Deep Neural Network

3.1.1. Conceptos matemáticos

Con el objetivo de saber cómo funcionan las redes neuronales se han recogido las explicaciones de un artículo [15] para entender los conceptos más importantes de su metodología. Los *Perceptrones* son la red neuronal más sencilla que consta de un número M de entradas, una sola neurona y una salida, donde M es el número de características de nuestro conjunto de datos.

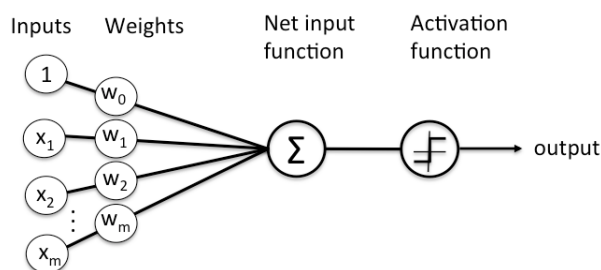


Figura 3.2: Perceptron Model

El primer paso consiste en multiplicar cada valor de entrada x_i con sus respectivo peso w_i y sumarlos con el fin de obtener un valor conjunto. Los pesos representan la fuerza de conexión entre las neuronas y decide cuánta influencia tiene cada valor de entrada en la salida de la neurona. Si el peso w_1 tiene un valor mayor que el peso w_2 , entonces la entrada w_1 tendrá una mayor influencia en la salida que w_2 .

$$\Sigma = x \cdot w = (x_1 \times w_1) + (x_2 \times w_2) + \dots + (x_m \times w_m)$$

El siguiente paso consiste en añadir un parámetro adicional, *bias*, a la red neuronal que se utiliza para ajustar los datos de salida junto con la suma ponderada de las entradas. Por lo tanto, es una constante que ayuda al modelo a ajustarse mejor a los datos dados.

$$z = x \cdot w + b$$

Por último, el proceso concluye pasando el valor de z a una función de activación no lineal. Estas funciones se utilizan para introducir la no linealidad en la salida de las neuronas, sin la cual la red neuronal será sólo una función lineal. Además, tienen un impacto significativo en la velocidad de aprendizaje de la red neuronal.

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

3.2. Redes neuronales convolucionales

La diferencia más notable entre las *Redes neuronales convolucionales* (CNN) y las tradicionales es que las primeras se utilizan principalmente dentro del campo del reconocimiento de patrones en las imágenes [16]. Esto nos permite codificar características más específicas, al mismo tiempo que se reducen los parámetros para configurar el modelo.

Las CNN se componen normalmente de tres tipos de capas: *convolution layer*, *pooling layer* y *fully connected layers*. Las dos primeras, realizan extracción de características, mientras que la tercera convierte las características extraídas en un resultado final, como la clasificación.

- **Convolution layer.** Una capa de convolución es un componente fundamental de las CNN que realiza la extracción de características [17] mediante la combinación de operaciones lineales y no lineales. Una pequeña matriz compuesta por números, llamada *kernel*, se aplica a los datos de entrada, que también es una matriz de números, llamada *tensor*. Se calcula el producto por elementos y se suma para obtener el valor de salida en la posición correspondiente del *tensor* de salida. Por ejemplo, si el dato de entrada es una imagen de tamaño $64 \times 64 \times 3$ y establecemos el *kernel* como una estructura de 6×6 , tendríamos un total de 108 pesos en cada neurona.

Para entender la importancia de esta técnica, se considera un gran avance comparándola con la cantidad de pesos que constaría una neurona estándar, en este caso serían 12288 pesos cada una.

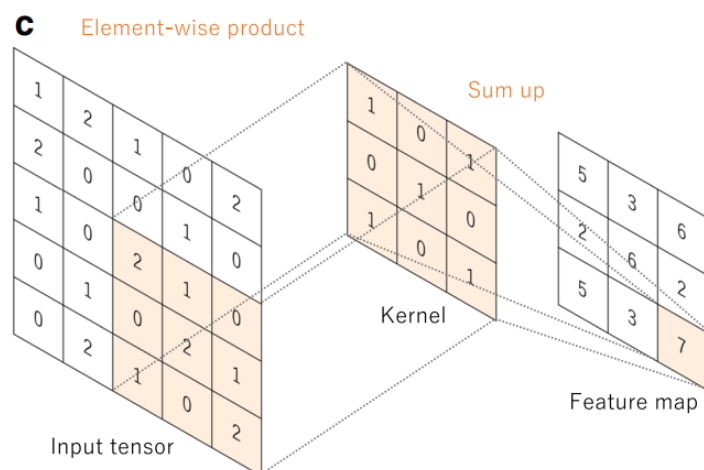


Figura 3.3: Un ejemplo de operación de convolución con un kernel de tamaño 3×3

- Pooling layer.** Las capas de agrupación se usan principalmente para disminuir las dimensiones del mapa de características con el fin de reducir la complejidad [16] de las capas posteriores. En el ámbito del procesamiento de imágenes, puede considerarse similar a la reducción de la resolución. Dentro de los métodos de *Pooling* se pueden destacar *Max-Pooling* y *Global Average Pooling*.

La primera de estas técnicas, *Max-Pooling*, consiste en dividir la imagen en subregiones mediante rectángulos como en la Figura 3.4, y devolver el valor máximo del interior de esa subregión. En este caso, al coger un filtro de dimensión 2×2 y aplicarlo a un tensor de 4×4 , la dimensión final se reduce a la mitad.

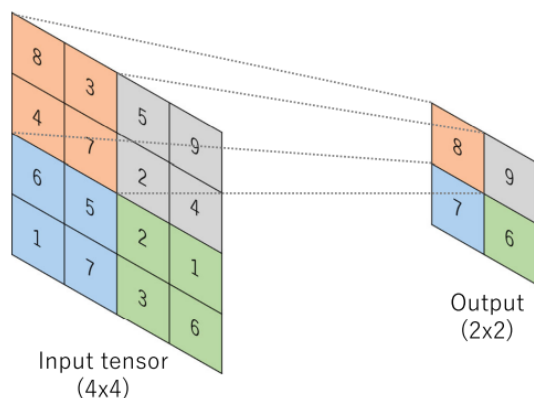


Figura 3.4: Un ejemplo de operación de max pooling con un filtro de tamaño 2×2

Global Average Pooling es una técnica empleada para vectorizar los mapas de características de la última capa convolucional. Las capas convolucionales también se pueden considerar extractores de características [18], y la resultante se clasifica de forma tradicional mediante capas totalmente conectadas, las cuales son propensas al sobreajuste.

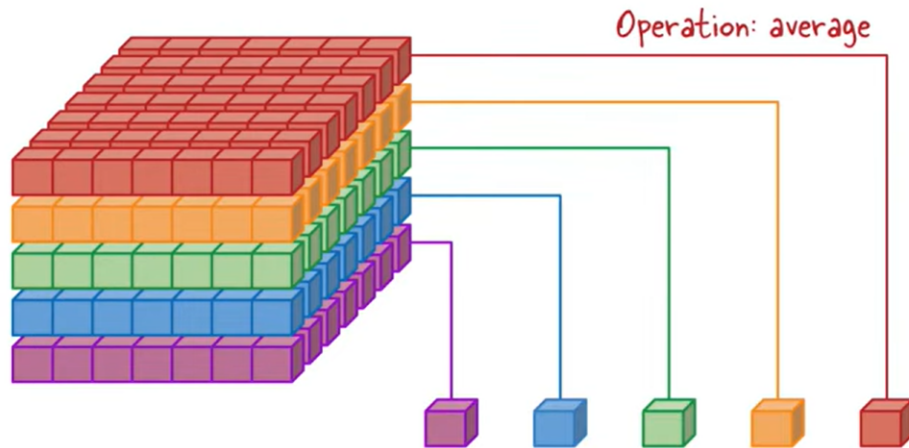


Figura 3.5: Un ejemplo de operación de Global Average Pooling con un filtro de tamaño 2×2

- Fully connected layer.** La capa totalmente conectada es similar a la forma en la que las neuronas están estructuradas [19] en una red neuronal tradicional. Por lo tanto, cada nodo está directamente conectado a todos los nodos de la capa anterior. La última capa suele tener el mismo número de nodos de salida que el número de clases. Por ejemplo, dada una imagen, la capa de convolución detecta características y las capas totalmente conectadas actúan como un clasificador sobre estas características y asignan una probabilidad de que la imagen de entrada sea de una determinada clase.

Debemos tener en cuenta que vamos a trabajar con diferentes imágenes y que por lo tanto, éstos serán nuestros datos de entrada. Como se ha explicado en el Subapartado 3.1.1 los datos que componen la primera capa de las redes neuronales deben ser valores numéricos. Todas las imágenes digitales también se pueden representar mediante tres matrices que guardan por cada canal RGB los valores de cada píxel. De esta manera, podemos crear por cada imagen un array colocando uno detrás de otro los valores que corresponden a cada fila y columna. Cabe mencionar que el orden en el que se coloca cada valor es muy importante y que siempre se debe seguir el mismo patrón.

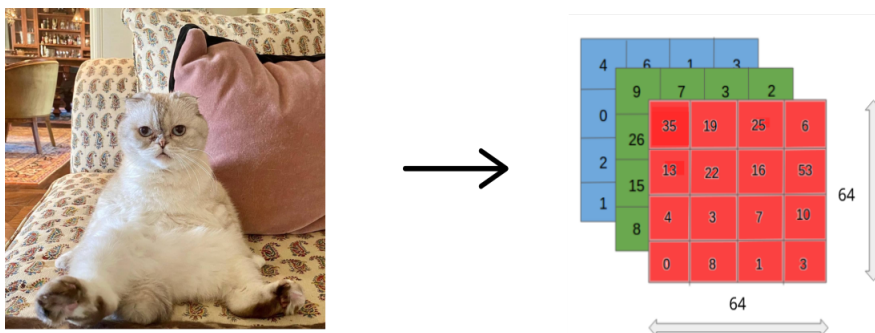


Figura 3.6: RGB matrix

En el caso de la Figura 3.6 se considera una imagen de 64 píxeles por cada columna y fila. Teniendo en cuenta que la imagen consta de tres capas, cada imagen de estas características contiene un total de 12288 píxeles. Los valores de cada píxel tienen un rango establecido el cual varía en base a la estructura por la que se compone. Como norma general, el valor mínimo siempre es el 0, pero el máximo depende de una pequeña operación. Es decir, el valor máximo siempre es $2^n - 1$ donde n corresponde a los bits de la estructura de datos por la que se compone.

$$X = \begin{bmatrix} 35 \\ 19 \\ \dots \\ 9 \\ 7 \\ \dots \\ 4 \\ 6 \\ \dots \end{bmatrix} \quad M = 64 \times 64 \times 3 = 12288$$

Tipo de datos	Min	Max
uint8	0	$2^8 - 1 = 255$
uint16	0	$2^{16} - 1 = 65,538$
uint32	0	$2^{32} - 1 = 4,294,967,296$

Tabla 3.1: Tipos de datos y su rango de valores

Descripción de las bases de datos

Una base de datos es un conjunto de datos almacenados que está organizado mediante una estructura de datos y cada una se diseña para satisfacer diferentes requisitos en base al contenido y objetivos que deben cumplirse [20]. A lo largo de todo el proyecto se han desarrollado diversas tareas, algunas de forma simultánea, con el fin de alcanzar los diferentes objetivos mencionados en la Sección 2.5. Cada uno de estos objetivos tiene un peso diferente dentro de este proyecto, a la vez que no todas las técnicas tienen el mismo efecto en los resultados globales. Por estas razones, se ha optado por trabajar con diferentes bases de datos compuestas por distintas características.

Se debe tener en cuenta que todas las bases de datos presentadas en este capítulo están formada por imágenes; sin embargo, no todas contienen tomografías ni material médico. Ésto se debe a que en algunas tareas a realizar con el contenido de las imágenes no siempre será tan interesante como el comportamiento que puede alcanzar.

A continuación encontramos una lista de todas las bases de datos con las que se ha trabajado a lo largo de este proyecto.

- MNIST
- Lung-Pet-Ct-Dx
- MosMedData
- DB_SBRT

4.1. MNIST

La base de datos MNIST (Modified National Institute of Standards and Technology) [23] se ha convertido en una referencia estándar para los sistemas de aprendizaje, clasificación y visión por computador. Este conjunto de datos está formado por dígitos escritos a mano y el reconocimiento de estos dígitos es un problema importante en el reconocimiento óptico de caracteres [21].

MNIST cuenta con 60,000 imágenes de entrenamiento y 10,000 imágenes de prueba, ambas extraídas de la misma distribución [22]. Todos estos dígitos negros y blancos están normalizados y centrados en una imagen de tamaño fijo contruida por 28×28 píxeles. Ésta base de datos es muy útil para las personas que quieren probar técnicas de aprendizaje y métodos de reconocimiento de patrones en datos del mundo real, dedicando un esfuerzo mínimo al preprocesamiento y al formateo. En la Figura 4.1 se muestra un ejemplo de esta base datos.

Imágenes	70,000
Tamaño (MB)	11,06

Tabla 4.1: Descripción detallada de la base de datos MNIST

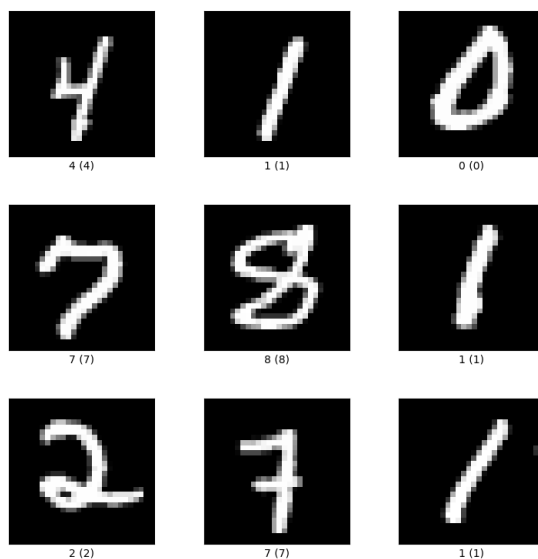


Figura 4.1: Ejemplos de la base de datos MNIST

4.2. Lung-Pet-Ct-Dx

Este conjunto de datos está formada por imágenes *DICOM* de pacientes con cáncer de pulmón y por archivos de anotación *XML* que indican la localización del tumor mediante cuadros de delimitación. Las imágenes fueron adquiridas de pacientes con sospecha de cáncer de pulmón, los cuales fueron sometidos a una biopsia pulmonar y a una PET/TC. Las localizaciones de los tumores en cada *DICOM* han sido anotadas por 5 radiólogos torácicos académicos para que este conjunto de datos sea útil para desarrollar algoritmos de diagnóstico médico.

En la Tabla 4.2 se recoge la descripción detallada de esta base de datos, la cual está formada por un total de 355 pacientes. En la mayoría de los pulmones analizados se ha realizado un único estudio para descartar un posible cáncer. Sin embargo, en aquellos casos que se ha detectado alguna irregularidad, se ha requerido llevar a cabo dos o incluso tres estudios. De este modo, la suma total de estudios alcanza el número 436. Cada estudio se compone de un volumen de TC, un volumen de PET e imágenes fusionadas de PET y TC. La resolución de la TC es de 512×512 píxeles y la de PET de 200×200 píxeles.

Modalidades	PET, TC
Participantes	355
Estudios	436
Series	1,295
Imágenes	251,135
Tamaño (GB)	127,2

Tabla 4.2: Descripción detallada de la base de datos Lung-Pet-Ct-Dx

Lung-Pet-Ct-Dx está distribuida en dos bloques generales según su contenido. Por una parte, tenemos los datos de los pacientes, es decir, las imágenes pulmonares con las que vamos a trabajar y por otra, la anotaciones correspondientes a la ubicación del tumor detectado. El primer bloque está dividido por los pacientes que han sido sometidos a la recopilación de datos y por cada uno de ellos podemos encontrar los estudios que se le han realizado. En el interior de cada estudio, organizadas de manera cronológica, se encuentran las imágenes pulmonares con las que se va a trabajar.

Por lo general, cada estudio contiene como mínimo dos series diferentes. Cogiendo como ejemplo el caso del tercer paciente (*Lung_Dx-A0003*) encontramos tres estudios en su interior y cada uno de ellos está compuesto por dos series de imágenes:

- 07-07-2006-NA-ThoraxAThoraxRoutine Adult-24087
 - 2.000000-ThoraxRoutine 10.0 B70f-09713
 - 3.000000-ThoraxRoutine 10.0 B40f-30728
- 07-21-2006-NA-Chest-84219
- 07-25-2006-NA-Chest-37655

En el bloque correspondiente a las anotaciones, la base de datos también se divide por pacientes. Sin embargo, a diferencia del bloque anterior, los ficheros *XML* se encuentran en el mismo nivel de directorio en vez de organizarlos por estudios. Para terminar, *Lung-Pet-Ct-Dx* también contiene un fichero (*memtada.csv*) que contiene un resumen de los datos. En la Sección 5.2 se muestra una herramienta de visualización para esta base de datos.

4.3. MosMedData

Este conjunto de datos contiene exploraciones anónimas de tomografía computarizada (TC) de pulmón humano con hallazgos relacionados con COVID-19. Un pequeño subconjunto de estudios ha sido anotado con máscaras binarias de píxeles que representan regiones de interés. Las tomografías computarizadas se obtuvieron entre el 1 de marzo de 2020 y el 25 de abril de 2020, y fueron proporcionadas por hospitales municipales de Moscú (Rusia). MosMedData contiene 1110 estudios, los cuales el 42 % pertenecen a hombres, 56 % a mujeres y el resto se desconoce [24]. A continuación se presentan varias características de esta base de datos:

A. Criterios de inclusión [25]: 1. La edad del paciente es mayor de 55 años y menor de 75 años. 2. Experiencia de fumar más de 30 paquetes al año. 3. Fumar actualmente o haber dejado de fumar hace no más de 15 años.

B. Criterios de no inclusión: 1. Cáncer de pulmón detectado en los 2 años siguientes a la primera ronda de cribado. 2. Antecedentes de cáncer de pulmón y/o cirugía pulmonar. 3. Antecedentes de cáncer diagnosticado hace menos de 5 años, con la excepción del cáncer de piel y el cáncer de cuello de útero. 4. Presencia de patología pronunciada de los sistemas cardiovascular, inmunológico, espiratorio o endocrino, así como una esperanza de vida inferior a 5 años. 5. Enfermedad aguda del sistema respiratorio.

C. Criterios de exclusión: 1. Ausencia de focos pulmonares en la primera ronda de Moscú de cáncer de pulmón de Moscú.

4.4. Base de datos del proyecto SBRT (DB_SBRT)

Esta base de datos es el conjunto principal de este proyecto. Todos el proceso realiado en las bases de anteriores tiene como objetivo aplicarlo posteriorme a este conjunto. El contenido principal son imágenes *DICOM* y provienen de diferentes pacientes a los que se le ha realizado una terapia con el fin de tratar el cáncer de pulmón.

Para la recopilación de datos, se ha contado con la participación de 99 pacientes que han sido reclutados desde diferentes centros hospitalarios dentro del territorio del País Vasco. En total se extrajeron 943 estudios, de los cuales 475 fueron PET/TAC y 468 estudios TAC. Asimismo, en la Figura 4.2 se puede observar que los estudios fueron realizados en 5 centros diferentes: En total se obtuvieron 562 estudios del Hospital Universitario de Cruces, 279 del H.U. Araba (HUA), 59 del H. Basurto, 42 de H. Galdakao y 1 de H.U. Donostia. Lo que implica una diversidad en el proceso de preparación del paciente, adquisición y reconstrucción de imágenes, que van a ser tenidas en cuenta en el estudio.

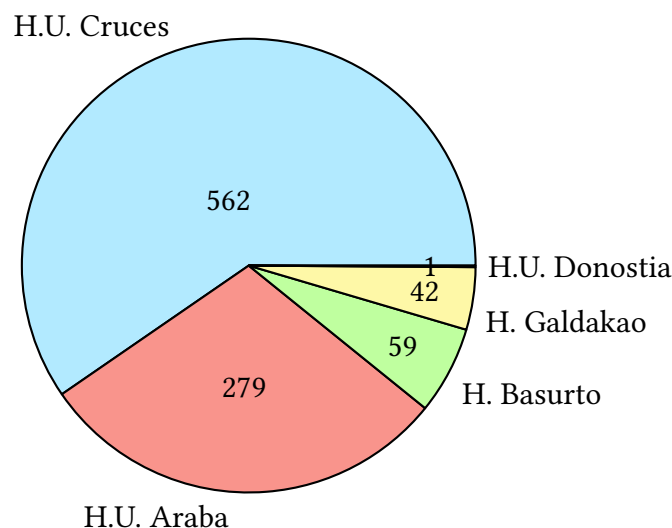


Figura 4.2: Cantidad de estudios extraídos por centro hospitalario

Con el fin de facilitar y organizar su extracción, las imágenes fueron agrupadas según el centro de realización y recogidas en diferentes fases para agilizar el proceso. En todos los estudios de este conjunto se ha realizado un proceso de eliminación de datos identificativos o personales del paciente y se procedió a su codificación y anonimización para poder ser entregado a los diferentes grupos de investigación. Sólo la investigadora principal mantiene la relación de la codificación para no perder la relación de la información con el paciente del estudio.

Tras la adquisición de las imágenes la siguiente tarea a realizar consiste en el proceso de reconstrucción y segmentación de las mismas. Por una parte, podemos determinar que cada estudio radiológico obtenido en la recopilación de datos presenta entre 5 y 10 series. En la mayoría de estos estudios se puede precisar que el número de imágenes en cada serie no supera la cantidad de 900. Por otra parte, es importante identificar la serie de mayor calidad para realizar en ella la reconstrucción y después la segmentación. Este proceso se ha llevado a cabo a través de la plataforma *3D Slicer* por una radióloga y una oncóloga radioterápica. Además, se ha combinando una segmentación manual (principalmente en la segmentación de nódulos) y una semi-manual según un rango de densidades y una posterior corrección manual.

Modalidades	PET/TAC, TC
Participantes	99
Estudios	943
Series	8640
Imágenes	1,668,747
Tamaño (GB)	794

Tabla 4.3: Descripción detallada de la base de datos DB_SBRT

4.4.1. Aproximaciones en el etiquetado

Un análisis preliminar ha permitido estudiar el contenido de esta base de datos en cuanto al etiquetado y así poder detectar posibles irregularidades. Para ello, antes de empezar a trabajar se han llevado a cabo varias aproximaciones en las que se ha querido analizar las proporciones en base a las etiquetas.

Debemos tener en cuenta que uno de nuestros objetivo se centra en analizar la zona pulmonar, por lo que no todas las imágenes nos serán de utilidad. Por lo tanto, se realizado una primera aproximación para tener una idea de a qué parte del cuerpo pertenece cada imagen y a raíz de eso se ha construido la Figura 4.3. Podemos encontrar siete partes del cuerpo diferentes en las que solo tres comprenden la zona del pulmón ('Abdomen', 'Chest' y 'Torax'). En total estas imágenes corresponden al 28,39 % del conjunto completo y se identifican mediante el termino 'ACT'. El resto de las partes ('Extremity', 'Head', 'Neck'y 'Shoulder'), clasificadas como 'OTHER', solamente forman el 0,75 % de esta base de datos. Para terminar, el resto de las imágenes, las que no están etiquetadas, componen el 70,86 % de la suma total.

Es evidente que ésto supone una desbalanceo de clases puesto que la gran parte de la base de datos está sin etiquetar. Uno de los principales retos que actualmente surge es la clasificación de bases de datos de alta dimensionalidad, con muy pocas muestras y alto desbalance de clases. El desbalanceo de clases es una característica que se genera cuando una o más clases se encuentra representada en una menor cantidad respecto a las otras [26]. Las clases minoritarias suelen representar el mayor interés del conjunto puesto que suelen ser valores atípicos asociados al problema. Los escenarios de desbalanceo generan grandes problemas para los algoritmos de clasificación, como la obtención del bajo rendimiento en la predicción de la clase minoritaria.

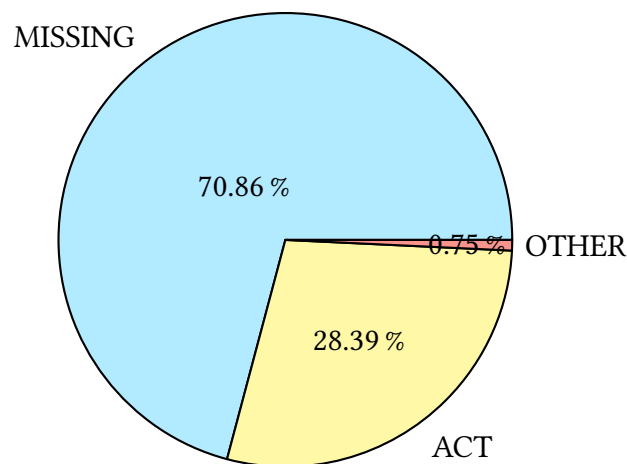


Figura 4.3: *Proporción de imágenes por parte del cuerpo*

Después de analizar la proporción de imágenes que corresponden a cada parte del cuerpo se ha hecho otra aproximación respecto al etiquetamiento manual que se ha llevado a cabo. En este proceso se ha estudiado cada imagen y considerado si realmente comprenden la zona del pulmón, es decir, si nos va a servir de utilidad o no. Las etiquetas 'Yes' corresponde a la valoración positiva y las 'No' a la negativa.

En la Figura 4.4 se muestra la proporción que guarda cada uno de los tres casos posibles. Por una parte, la porción dibujada de color azul muestra la cantidad de imágenes que fueron etiquetadas como 'ACT' y realmente muestran la zona pulmonar: 253183. Por otra parte, el color amarillo muestra el caso opuesto, es decir, desde un principio 109101 imágenes fueron clasificadas como 'ACT' pero realmente esas imágenes no nos muestran la parte del cuerpo que deseamos. Para terminar, el último caso, dibujado de color rojo, lo componen las imágenes que de antemano no pertenecen a la clase 'ACT' y efectivamente no lo son.

Cabe destacar que los dos últimos casos también se contempla la posibilidad de que alguna imagen esté dañada y que por lo tanto no nos va a ser útil en nuestro proyecto. Además, se debe tener en cuenta que la suma total de las imágenes repartidas en la Figura 4.4 corresponden a las clasificadas como 'ACT' y 'OTHER' de la figura Figura 4.3; por lo que las imágenes no etiquetadas no forman parte de esta nueva aproximación. Para terminar, con este análisis también existía la posibilidad de que alguna imagen clasificada como externa a la parte pulmonar se haya etiquetado como tal. Sin embargo, no se ha encontrado ningún caso con esta condición.

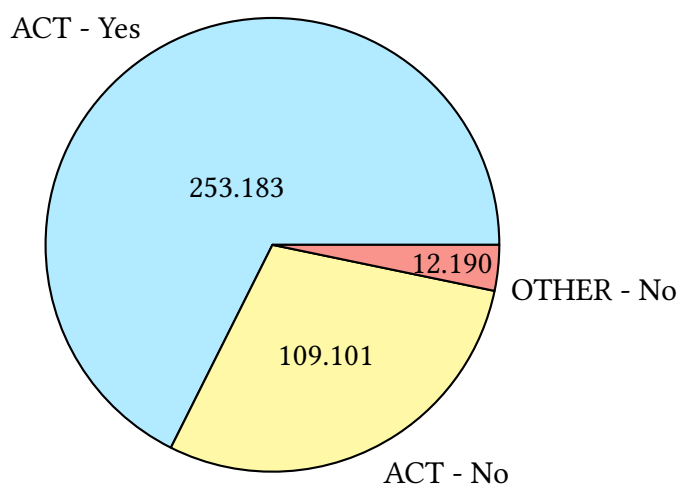


Figura 4.4: Cantidad de imágenes según la parte del cuerpo y el etiquetado manual

Durante este proyecto se han llevado a cabo algunos procesos de clasificación en los que también se ha intentado clasificar las imágenes por su tipo en cuanto a los planos tomográficos. Por consiguiente, también se ha hecho el cálculo de la cantidad de imágenes relacionadas con esta fase. Por una parte, la Figura 4.5 muestra la proporción de imágenes que componen las tomografías axiales, es decir, las tomadas perpendiculares al eje longitudinal del cuerpo [27].

En este caso se ha contruido la Figura 4.5 en la que se puede observar la proporción que corresponde a cada tipo de imagen. Concretamente, el 57,76 %, dibujado de color azul, pertenece a la tomografías axiales y el 42,2 %, dibujado de color amarillo al resto. Comparando este gráfico con la aproximación anterior, podemos asegurar que no existe un desbalanceo de clases tan notable y que por lo tanto partimos de una base más estable a la hora de trabajar con estos datos.

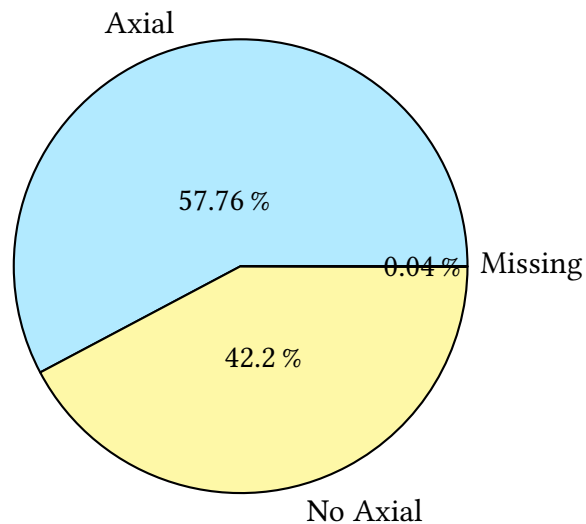


Figura 4.5: Proporción de imágenes por tipo de imagen

En la siguiente aproximación se ha querido estudiar si realmente las imágenes del caso anterior corresponden a los planos axiales en base al etiquetado manual. Por una parte podemos destacar que el 54,85 % de las imágenes fueron clasificadas como 'Axial' y realmente así se ha etiquetado. Al igual que el 33,66 %, dibujado de color amarillo, corresponden al caso en el que las imágenes no axiales han sido etiquetadas como tal. Es decir, casi el 90 % han sido etiquetadas de manera correcta. Por otra parte, los casos restantes, dibujados de color rojo y verde, son imágenes que el etiquetado no corresponde a la clase real.

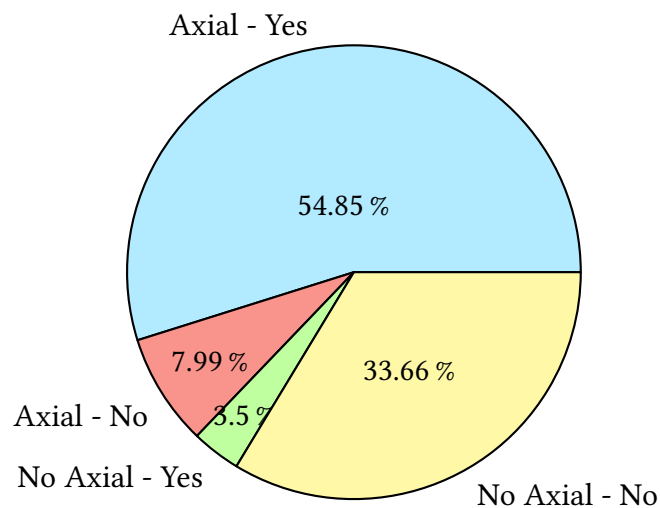


Figura 4.6: Proporción de imágenes por tipo de imagen y el etiquetado manual

Exploración de imágenes tumorales

En algunas ocasiones, varios métodos de preprocesamiento son necesarios antes de poder emplear las imágenes y técnicas definitivas. La base de datos que se pretende analizar con mayor detenimiento es excesivamente grande como para usarla en su totalidad a la hora de hacer pruebas. De la misma manera, al ser tan grande también nos vamos a encontrar con una variedad de casos más amplia. Estos casos serán analizados con detenimiento con técnicas más avanzadas.

Antes de empezar a trabajar con imágenes pulmonares, se ha hecho una primera exploración con imágenes cerebrales para poder entender uno de los grandes objetivos propuestos en este proyecto. Se trata de la detección de anomalías.

5.1. Exploración de imágenes cerebrales

Las imágenes tridimensionales del cerebro son cada vez más populares en las aplicaciones médicas [28] y se utilizan para la investigación y así poder elaborar su diagnóstico y tratamiento. De la misma manera, este tipo de imágenes también son empleadas como entrenamiento en otras aplicaciones médicas. Debido a la sencillez que el cerebro proporciona, han sido utilizadas como previa a nuestros objetivos.

Con el fin de facilitar el proceso, se va a aplicar la técnica propuesta por *Image Analyst* [29] en la Figura 5.1. Esta imagen en escala de grises muestra el escáner de un cráneo realizado a un paciente con el fin de diagnosticar un posible tumor. A diferencia las demás técnicas desarrolladas en este proyecto, se ha trabajado solamente con una única imagen para poder ver de manera clara el objetivo que se quiere alcanzar.

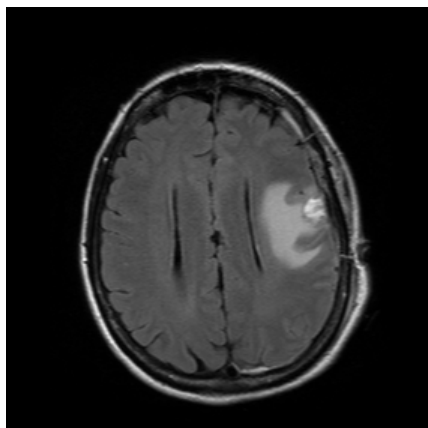


Figura 5.1: *Imagen craneal original*

Por una parte, esta imagen tiene unas dimensiones de 288×288 de ancho y largo, por lo que esta compuesta por un total de 82,944 píxeles. Por otra parte, su composición se basa en una estructura *uint16*, por lo que los valores de sus píxeles se va a mantener dentro del rango $[0, 65,535]$, dónde el color negro corresponde al valor 0 y el color blanco al valor 65,535. La Figura 5.2 es un histograma en el que se muestra la cantidad de píxeles que determina cada uno de los valores en la escala de grises. En esta imagen, debido al contorno de color negro, hay un gran número de píxeles con un nivel de gris inferior a 11, lo que crea un gran pico en la primera casilla. Se ha decidido ignorar esos píxeles para poder obtener un histograma sólo con píxeles distintos a cero. Por una parte, se puede observar que la mayoría de los valores se agrupan en el rango $[200, 300]$, lo que equivale al color gris que compone la gran parte del cerebro. Por otra parte, debido a la escasez del color claro, es decir, los mas parecidos al color blanco, apenas hay valores superiores a 700.

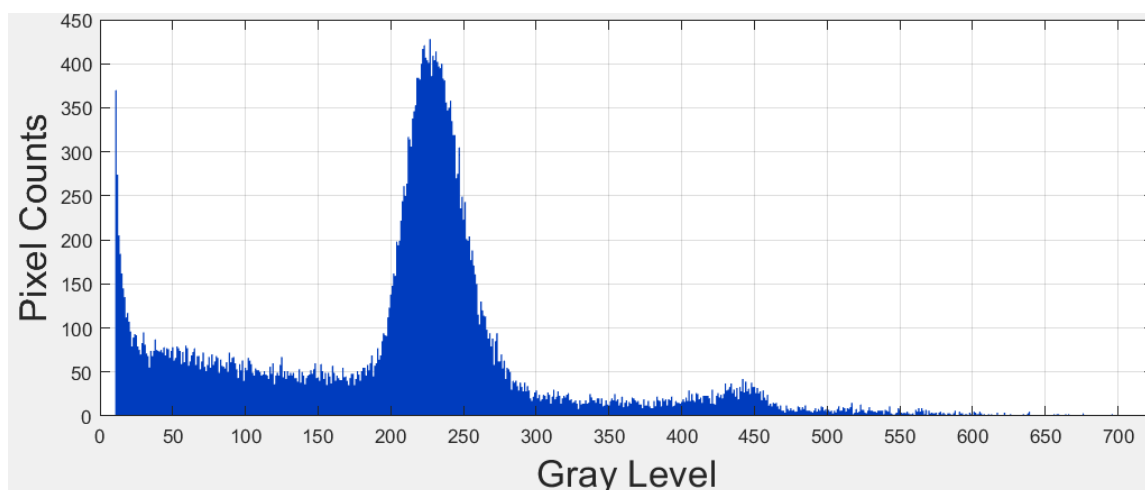


Figura 5.2: *Histograma de píxeles no-negros*

A partir de este gráfico, se va a trabajar con la figura original (Figura 5.1) con el fin de eliminar el cráneo de la imagen y así poder conseguir un resultado más preciso. Podemos observar que los valores más claros, dicho de otra manera, los de mayor valor en la escala de grises corresponden al cráneo. De esta manera, se va a establecer un umbral en 260 para prescindir de los píxeles inferiores a este valor (Figura 5.3a). A partir de esta nueva imagen, se va a etiquetar los componentes conectados entre sí, mediante identificadores y se va a detectar el bloque más grande (Figura 5.3b), para así poder utilizarlo como máscara y extraerla de la imagen original (Figura 5.3c).

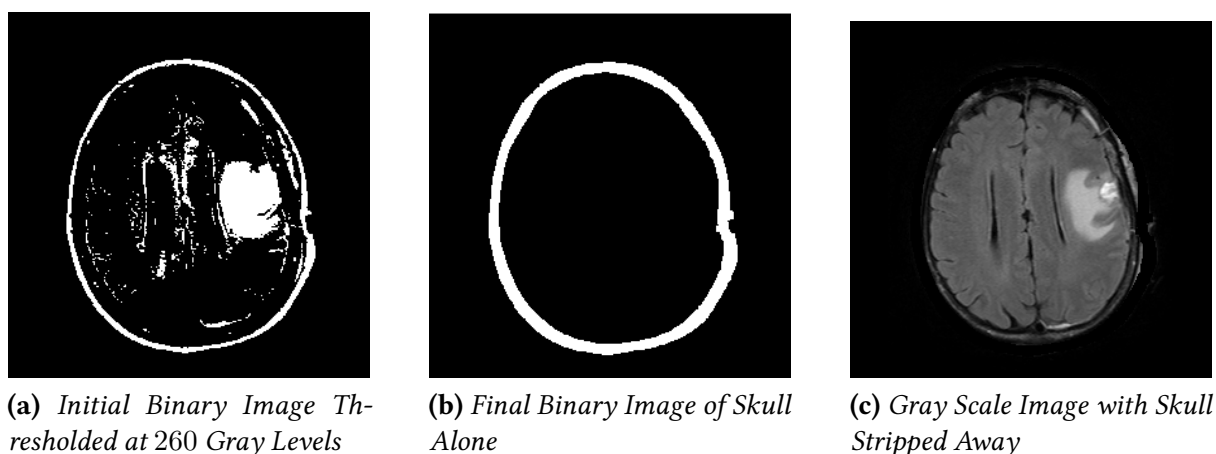


Figura 5.3: *Initial Binary Image, Final Binary Image and Skull Stripped Away*

Esta técnica concluye aplicando de nuevo el mismo desarrollo con algunas diferencias. En este caso, el proceso de binarización se va a emplear a la imagen obtenida (Figura 5.3c) tras la extracción del cráneo, pero con un umbral de 270 (Figura 5.4a). De esta manera, conseguimos filtrar más los píxeles que no corresponden al cerebro. Después de etiquetar los componentes conectados entre sí, damos por hecho que el bloque más grande corresponde al tumor (Figura 5.4b) y utilizando esta imagen como máscara, podemos detectar y dibujar el tumor en la imagen original (Figura 5.4c y Figura 5.4d) mediante el color rojo.

5.2. Exploración en 2D

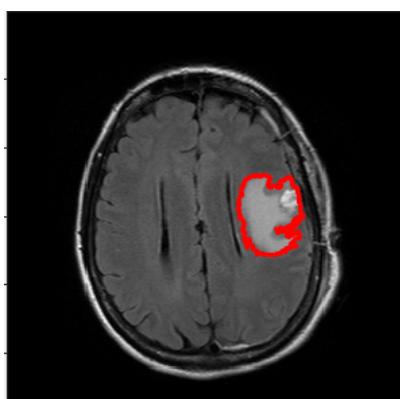
Una vez explicado y analizado visualmente lo que se quiere conseguir para un caso concreto como es un cerebro, vamos a hacer lo propio con una base de datos completa con el fin de cumplir nuestros objetivos. Se trata de la base de datos *Lung-Pet-Ct-Dx* [30] y su contenido está resumido en la Sección 4.2. En este apartado se pretende mostrar algunos ejemplos de los pulmones para saber con qué tipo de imágenes vamos a trabajar. Además, las imágenes pulmonares van acompañadas de un recuadro que gracias a su posición nos permite ubicar el nódulo.



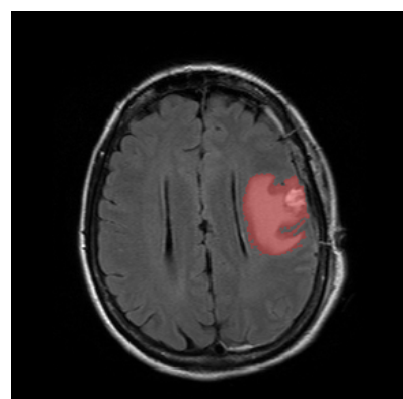
(a) *Initial Binary Image Thresholded at 270 Gray Levels*



(b) *Tumor Alone*



(c) *Tumor Outlined in red in the overlay*



(d) *Tumor Solid & tinted red in overlay*

Figura 5.4: *Resultado de la detección tumoral*

Cabe mencionar que cada fichero *XML* contiene en su interior las coordenadas del rectángulo que nos permite señalar el nódulo dentro del pulmón correspondiente. De la misma manera, cada uno de estos ficheros se relaciona con una sola imagen de la base de datos mediante un identificador. Debemos tener en cuenta que no todas las imágenes tienen un fichero con esta información. Mediante un programa se ha conseguido obtener cada una de las imágenes *DICOM* al que corresponde cada fichero para poder dibujar el rectángulo en su posición correcta.

En la Figura 5.5 se muestra el ejemplo de los pulmones del paciente *A0001* donde el tumor se posiciona en el lado derecho. Si comparamos su tamaño con el del pulmón completo, podemos determinar que nos encontramos ante un nódulo bastante voluminoso. En el caso de la Figura 5.6 se presentan los pulmones del paciente *A0002*. Nos encontramos ante un caso interesante debido a que se muestra el mismo nódulo con distinta forma. Debemos tener en cuenta que en cada estudio se toman una gran serie de tomografías con el escáner moviéndose a lo largo de la sesión. Por esta razón, dependiendo el ángulo en el que se toman las imágenes, el posible nódulo y el resto de órganos se visualizan de manera diferente.

Esta herramienta se ha aplicado de manera iterativa a todos los ficheros *XML* con el fin de visualizar todos los casos en los que un nódulo ha sido detectado. A partir de ahora se pretende trabajar con los casos en los que la posición exacta del tumor es conocida para poder entrenar un modelo capaz de detectarlo de manera automática.

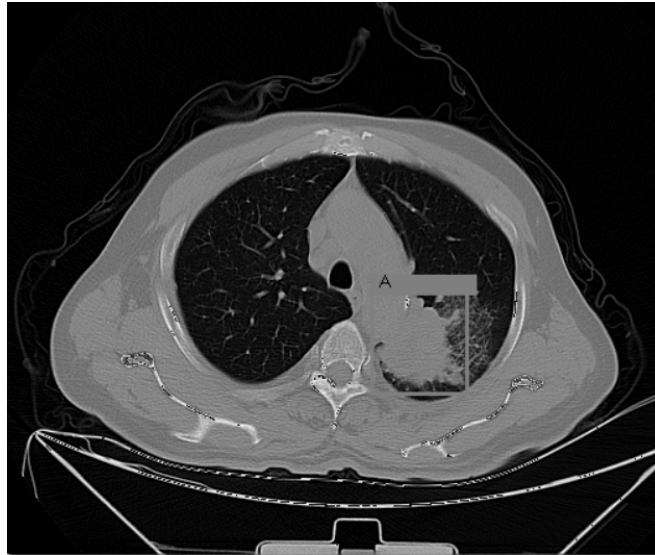
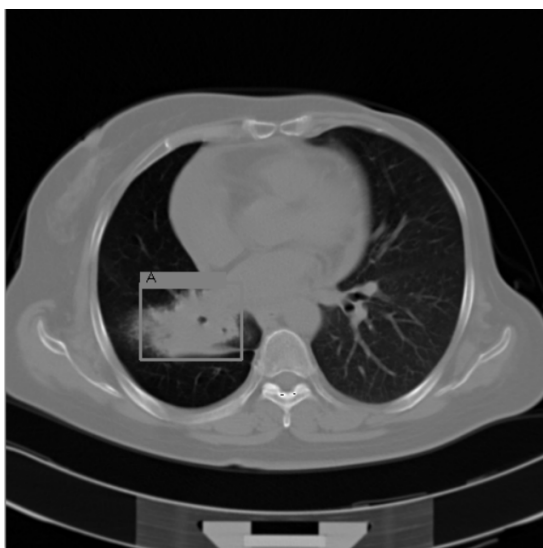
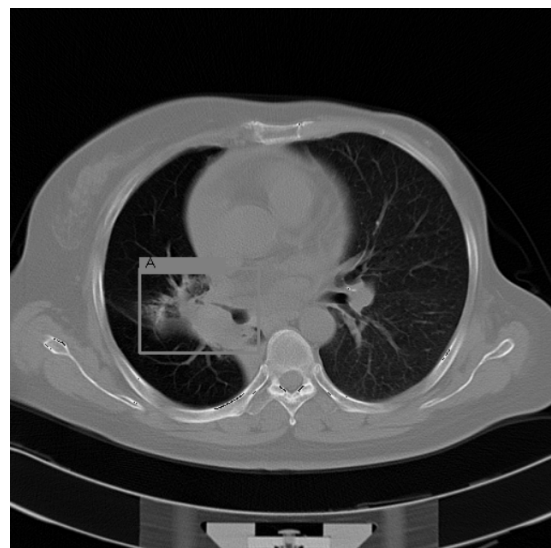


Figura 5.5: *Pulmón del paciente A0001*



(a) *Case 1*



(b) *Case 2*

Figura 5.6: *Pulmón del paciente A0002*

5.3. Exploración en 3D

En este apartado se presenta una herramienta de visualización de pulmones en la que también se representa un tumor existente. Dentro de la base de datos *DB_SBRT* existen 21 estudios con anotaciones sobre el tumor. Ambos casos expuestos a continuación pertenecen a este subconjunto y a base de unir las máscaras correspondientes al nódulo se ha conseguido representar de manera visual su forma y posición.

Por una parte, la Figura 5.7 muestra el primero de los dos casos. Se puede observar claramente la posición del tumor, dibujado de color rojo, en los tres diferentes ejes. El tumor se posiciona en la parte central del pulmón izquierdo. En cuanto a su tamaño, podemos determinar que nos encontramos ante un tumor de grandes dimensiones respectado a todo el pulmón.

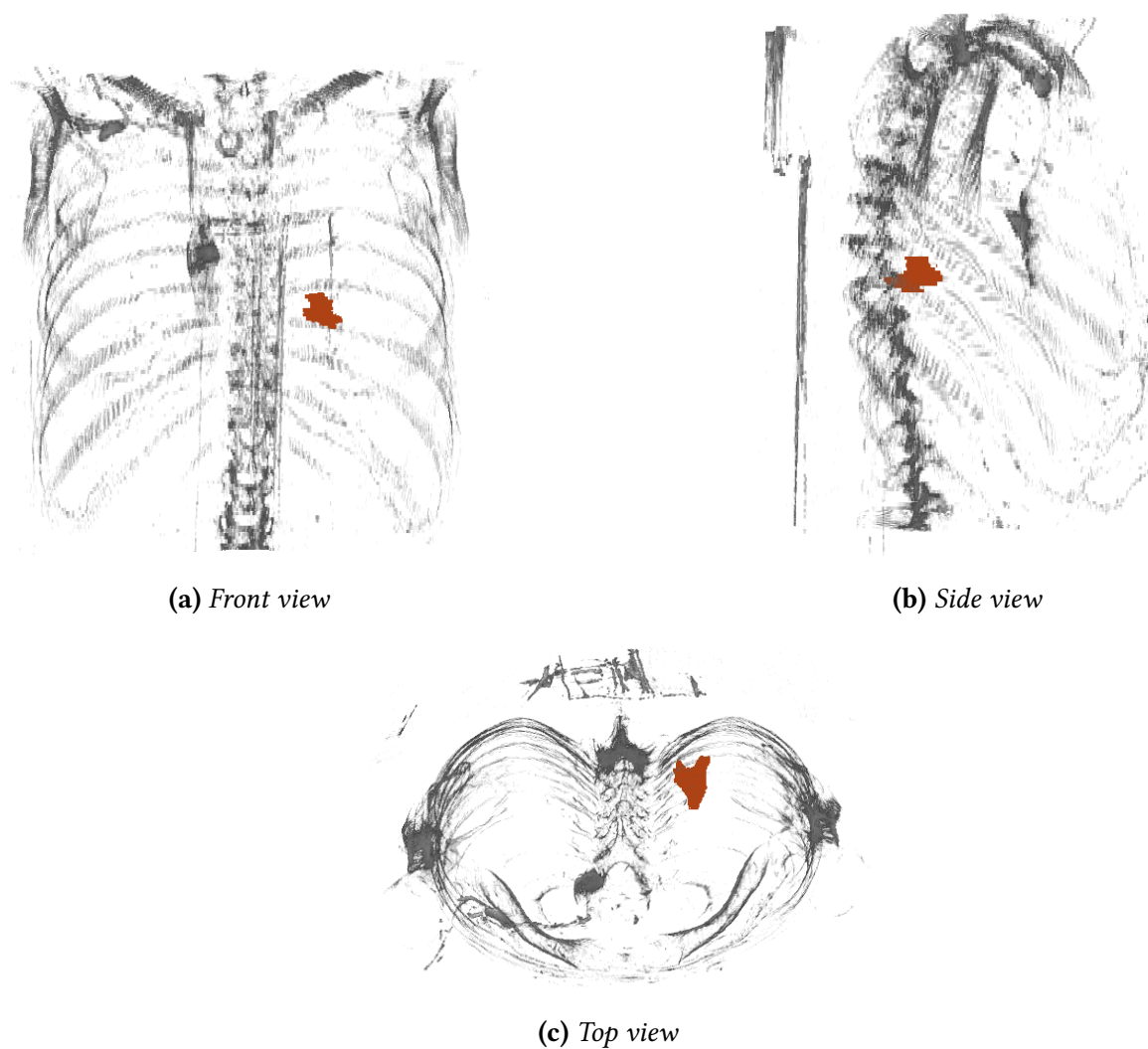


Figura 5.7: Localización del tumor para el paciente NHC70D2E159 y el estudio ACCA2D0A707

Por otra parte, la Figura 5.8 corresponde al segundo caso escogido para esta visualización. En esta ocasión podemos apreciar como las dimensiones del tumor son notablemente más pequeñas y de distinta forma. Por lo tanto, queda demostrado que los tumores no siempre tienen la misma forma ni tamaño y que dependiendo de su localización su detección puede suponer un gran reto.

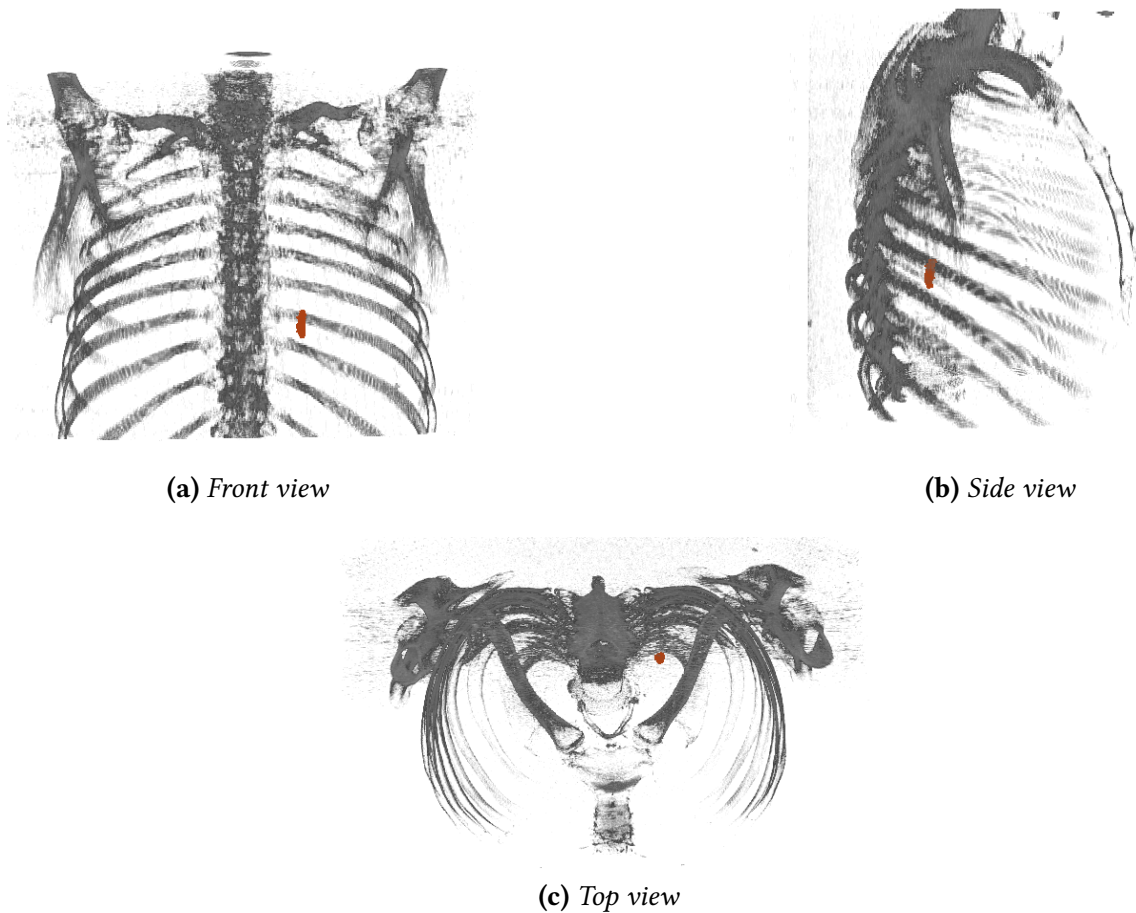


Figura 5.8: Localización del tumor para el paciente NHC6731499A y el estudio ACCEB5D3D24

Modelos de predicción de frames

La exploración de imágenes tumorales ha proporcionado una idea de lo que se quiere conseguir en este proyecto. No obstante, los casos estudiados en el apartado anterior se han desempeñado para ejemplos muy concretos que solo son útiles para entrar en contexto. Además, las herramientas empleadas no siguen un patrón de generalización y todas se han realizado manualmente para esas muestras en particular. Se han obtenido buenos resultados de visualización para esos casos en concreto, en los que la posición del posible tumor se resalta mediante diferentes técnicas. La clave de este proyecto se centra en hacer lo propio para cualquier caso proporcionado. En consecuencia, se ha querido construir diferentes modelos en los que cada caso particular ayuda a obtener de manera general un clasificador para la detección de estas anomalías.

No es tarea sencilla construir un modelo en el que sus características se adaptan a cualquier base de datos independientemente de su contenido y composición. Cada base de datos está estructurada de una manera distinta, a la vez que la información recogida varía en relación al uso que se le da. Por lo tanto, un mismo proceso tiende a obtener resultados muy diferentes para cada conjunto. De la misma manera, si a una base de datos se le aplican diversos procesos con diferentes componentes, también se deben esperar resultados variados.

En este apartado se han implementado diferentes experimentos para poder analizar y estudiar múltiples comportamientos dependiendo del contenido y sus características. Se ha tenido en cuenta la posibilidad de establecer algunas variaciones dentro de cada uno de ellos. La Tabla 6.1 recoge los diferentes casos en los que cada pequeña variación ha supuesto un factor clave a la hora de obtener los resultados.

Experimento	Caso	DB	Method	Examples	Ciclos
1	1.1	MNIST	Next-Frame	100	5
	1.2	MNIST	Next-Frame	1000	5
	1.3	MNIST	Next-Frame	1000	20
2	2.1	Lung-Pet-Ct-Dx	Next-Frame	383	5
	2.2	Lung-Pet-Ct-Dx	Next-Frame	383	20

Tabla 6.1: Características de los modelos de predicción de frames

Los cinco casos trabajados en esta parte del proyecto son muy similares entre sí pero un pequeño cambio en su composición condiciona todo el proceso y por lo tanto el resultado obtenido. En cada experimento se trabaja con una base de datos diferente para poder estudiar el comportamiento del mismo método en distintos contextos. En el primer experimento se emplea la base de datos *MNIST* [23] como precedente al segundo experimento donde se pretende construir un modelo capaz de predecir la evolución de una posible anomalía en imágenes pulmonares.

Dentro de cada experimento se han establecido diversas variaciones como el número de ejemplos empleados en el entrenamiento, y el número de datos que tiene cada iteración de un ciclo (epoch).

6.1. Construcción del modelo

Cada uno de los casos propuestos en estos dos experimentos se han creado en base al mismo método, *Next-Frame Video Prediction* [31], el cual hace uso de redes neuronales convolucionales. La predicción de fotogramas ha conseguido mucho interés en la visión por computador y el aprendizaje profundo [32] debido a su aplicación en varios campos como la detección de anomalías.

Estas aplicaciones utilizan la predicción de fotogramas de vídeo o secuencias de imágenes para comprender la evolución de un suceso. El cerebro humano tiene la capacidad de predecir los cambios previstos mediante la percepción sensorial y la intuición [33] pero la detección manual de anomalías generalmente supone una tarea más complicada. Esto ha provocado la necesidad de automatizar la detección y la segmentación de secuencias de interés.

Este apartado hace una contribución en el desarrollo de un modelo generativo, basado en una arquitectura de red neuronal convolucional capaz de codificar una secuencia de imágenes de entrada, reconstruirla y predecir su futuro próximo [33]. Nuestro modelo utiliza unidades *ConvLSTM* que permiten aprender mejor las representaciones espacio-temporales.

El modelo propuesto se basa en una arquitectura LSTM convolucional [34] que parte de un conjunto de datos compuesto por 20 fotogramas (10 de entrada y 10 de predicción) dentro de un marco de tamaño 64×64 píxeles. Con el propósito de realizar la predicción del siguiente fotograma, nuestro modelo utiliza un frame f_n , para predecir el siguiente, f_{n+1} . Se pretende que nuestro modelo sea capaz de crear las predicciones mediante un proceso específico: Se necesita procesar los datos de tal manera que hayamos desplazado las entradas y salidas, donde cada dato de entrada x_n se utiliza para predecir un dato de salida y_{n+1} .

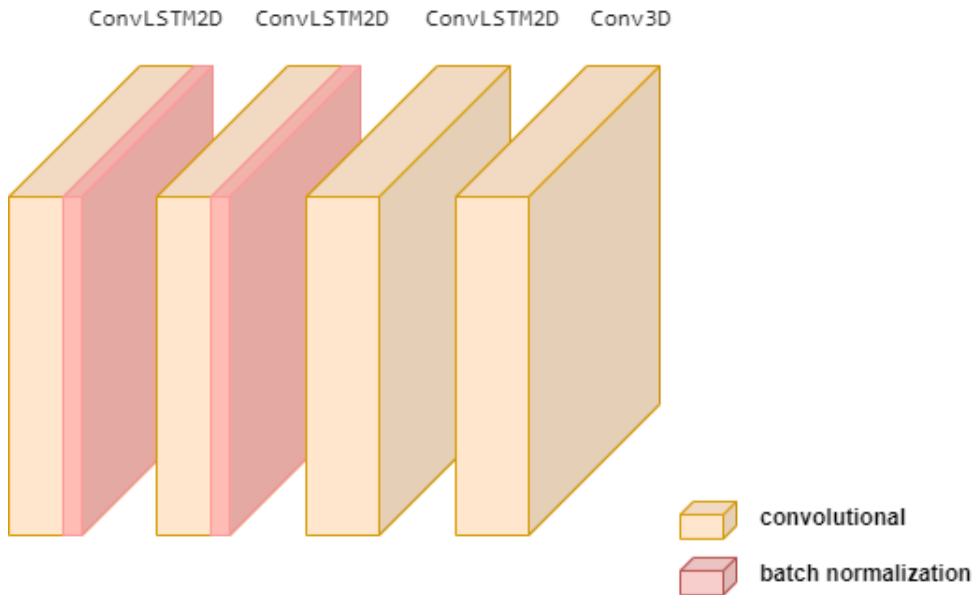


Figura 6.1: Arquitectura de la CNN para la predicción de vídeo de fotogramas siguientes

A pesar de ser el primer paso a realizar, la adquisición de información es una pieza clave debido a que será determinante en la distinción de algunos de los modelos propuestos. Por una parte, es fundamental precisar con qué conjunto de datos se va a trabajar, y por otra con cuántos casos. Podemos poner como ejemplo el caso 1.1 y el caso 1.2, en los que la única diferencia entre ellos es la cantidad de muestras seleccionadas: 100 y 1000 muestras respectivamente. Después de recoger y adaptar los datos a nuestro problema se va a dividir el conjunto entero en dos subconjuntos para realizar el entrenamiento y la validación.

En cuanto a la composición de la arquitectura de este modelo, la cual se representa en la Figura 6.1, se ha hecho uso de varias capas de convolución *ConvLSTM2D* situadas una detrás de otra. Estas capas realizan una tarea similar a las *LSTM* trabajando con operaciones de convolución en lugar de multiplicaciones matriciales [36]. Además, posibilita a nuestro modelo interpretar la información visual en forma de imagen 2D y entender secuencias temporales al mismo tiempo. Debido a que se espera conseguir una imagen como resultado, por asegurar que el conjunto de salida tiene la misma dimensionalidad que la entrada. Por lo tanto, podemos concatenar todos los estados de la red de predicción y los introducimos en una capa convolucional de 1×1 para generar la predicción final.

A cada capa de *Conv2D* le sigue una capa *Batch Normalization* con el fin de acelerar el proceso de entrenamiento y permitir que cada capa aprenda de manera más independiente de otras.

Este modelo termina con una última capa de convolución *Conv3D* empleada para extraer características visuales. En este tipo de capas, a diferencia de las las convoluciones 2D, el filtro se desliza en tres dimensiones (x, y, z) . Un ejemplo de uso es el de las imágenes médicas, en las que se construye un modelo utilizando los cortes de imágenes como bloques de 3D. De la misma manera, los datos basados una secuencia de fotogramas tienen una dimensión temporal adicional, lo que permite que su aplicación sea más efectiva. Además, se hace uso de una función de activación sigmoidea para que la capa de salida produzca el resultado en probabilidad para la clasificación binaria.

6.2. Comportamiento del modelo

El objetivo de este conjunto de experimentos es comparar las diferentes variantes del modelo propuesto en este apartado. Dado que siempre es posible obtener una menor tasa de error, no podemos utilizar esta medida para valorar la calidad de un modelo y se deberán usar otras más representativas [35].

Por una parte, una vez se hayan construido y entrenado los modelos propuestos podemos generar algunas predicciones de fotogramas para cada uno los casos en los que se ha trabajado. Ésto será de gran ayuda para una evaluación rápida y concreta en la que se podrá valorar una determinada secuencia de imágenes. Por otra parte, en el caso de querer estudiar el modelo en su totalidad se ha querido representar gráficamente el rendimiento obtenido mediante los datos de precisión y la pérdida en los conjuntos de entrenamiento y de validación.

6.2.1. Predicción de frames

Las características de cada modelo propuesto es esencial para la obtención de los resultados. Para cada caso, se ha escogido una secuencia de imágenes aleatoria del conjunto de validación y basándose en los diez primeros fotogramas, podemos permitir que cada modelo prediga otros diez nuevos fotogramas. Las figuras en este apartado comparan los resultados que se esperan obtener con los resultados obtenidos para cada caso, colocando cada frame en su respectiva posición en la fila superior e inferior respectivamente.

Las primeras tres figuras muestran los resultados obtenidos para el primer experimento, compuestos por el caso 1.1, el caso 1.2 y el caso 1.3 en los que la característica en común es el conjunto de datos empleado: *MNIST*. La Figura 6.2 muestra los resultados para el caso 1.1, donde se puede determinar claramente en qué zona de la imagen se han querido dibujar los números. Sin embargo, el resultado no deja nada queda claro cuáles son. La escasez de muestras empleadas en el entrenamiento ha hecho que los resultados no sean igual de buenos que los esperados.

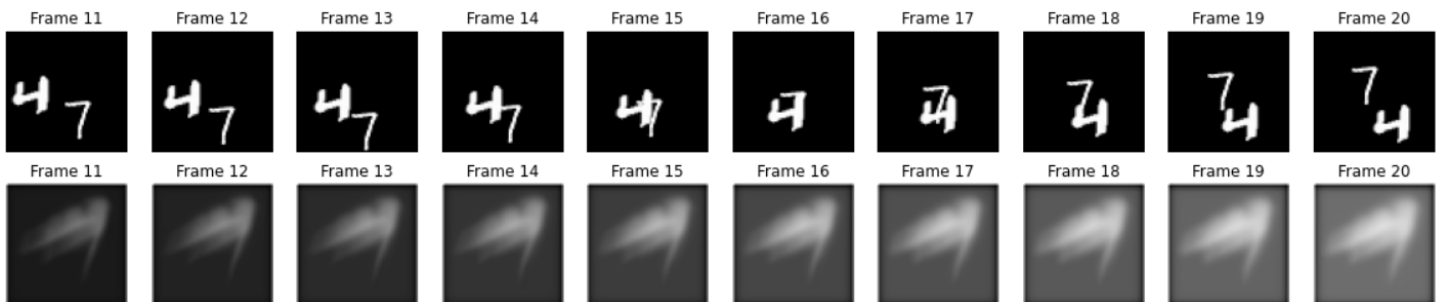


Figura 6.2: Predicción de frames para el caso 1.1

De momento esto no supone ningún problema debido a que esta base de datos dispone de más muestras con las que crear nuevos casos. En consecuencia, se ha propuesto el caso 1.2 con el que se espera obtener resultados más satisfactorios. La mayor diferencia respecto a la propuesta anterior es el número de muestras recogidas de toda la base de datos, pasando de entrenar el modelo con 100 muestras a 1000 muestras.

Con este aumento de muestras utilizadas se espera obtener alguna evolución positiva, y efectivamente se puede observar cómo los resultados han mejorado notablemente. En este caso, en la Figura 6.3 se puede intuir de manera sencilla los números en los frames predichos. Es destacable cómo el Frame 11 muestra claramente el dibujo de los números correspondientes, en este caso el 7 y el 9. Sin embargo, en los siguientes fotogramas los resultados van perdiendo eficacia hasta llegar al último donde casi no se consigue ver nada.

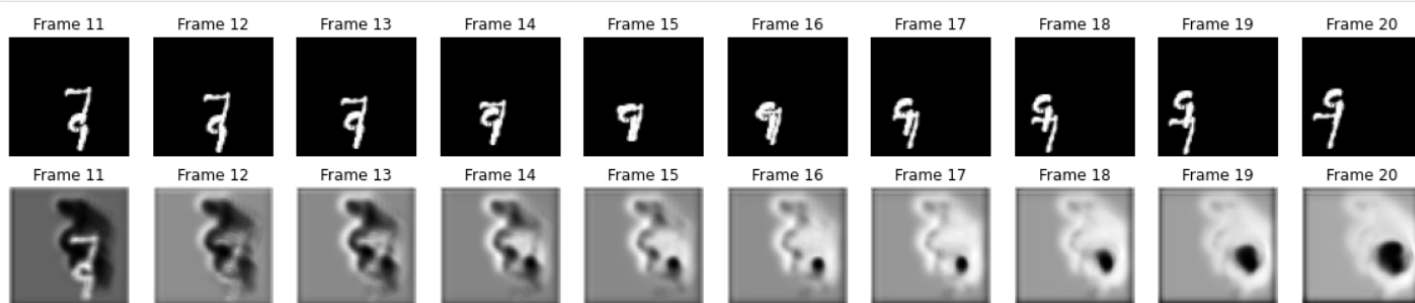


Figura 6.3: Predicción de frames para el caso 1.2

A pesar de conseguir resultados aceptables, se ha decidido elaborar un tercer caso para este experimento. La Figura 6.4 muestra los resultados obtenidos para el caso 1.3 dónde se mantiene la cantidad de muestras recogidas pero se aumenta el número de ciclos en el entrenamiento. En este caso, a diferencia de los dos casos anteriores, en vez de 5 se han realizado 20 ciclos. Por lo tanto, queda demostrado que el aumento de las veces que el modelo analiza cada ejemplo no siempre trae mejores resultados, ya que en general ninguno de los frames predichos en este caso muestra un número claro.

Una mala predicción de frames puede ser causada por diversas razones, pero sin duda cada caso en particular también influye en el resultado. En esta ocasión partimos de un frame en el que los dos números se colocan uno encima del otro. Por lo tanto, debemos tener en cuenta que este modelo tendrá más difícil diferenciar de qué números se trata. De hecho, desde un principio no se puede reconocer que los números a predecir son el 2 y el 3.

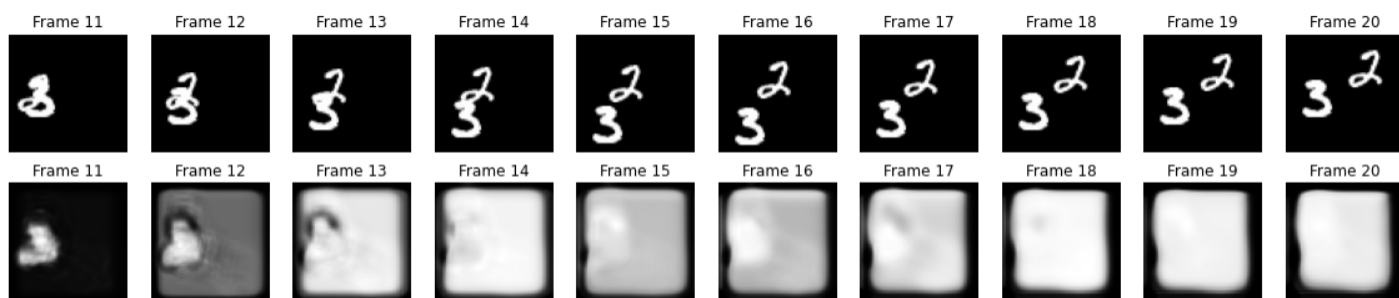


Figura 6.4: Predicción de frames para el caso 1.3

Una vez analizado el primer experimento en el que se han hecho las pruebas pertinentes, es hora de estudiar uno nuevo con una nueva base de datos en la que su contenido es esencial en nuestro proyecto. Se trata del caso 2.1 y el caso 2.2 en los que se trabaja con la base de datos *Lung-Pet-Ct-Dx*. Este experimento consta de dos casos y la única diferencia entre ambos casos en el número de ciclos establecido en el entrenamiento, 5 y 20 respectivamente.

Dentro de este experimento, nos encontramos ante la Figura 6.5 en la que podemos observar los frames predichos para el caso 2.1. Comparando los frames que se pretenden obtener con los que se han obtenido, se puede intuir de manera clara la forma del pulmón. Sin embargo, la calidad y la precisión de los frames que se esperan obtener es bastante mayor que los predichos. Además, podemos destacar que a medida que se predicen los frames, el resultado es cada vez más oscuro, por lo que el último frame no muestra demasiada información relevante.

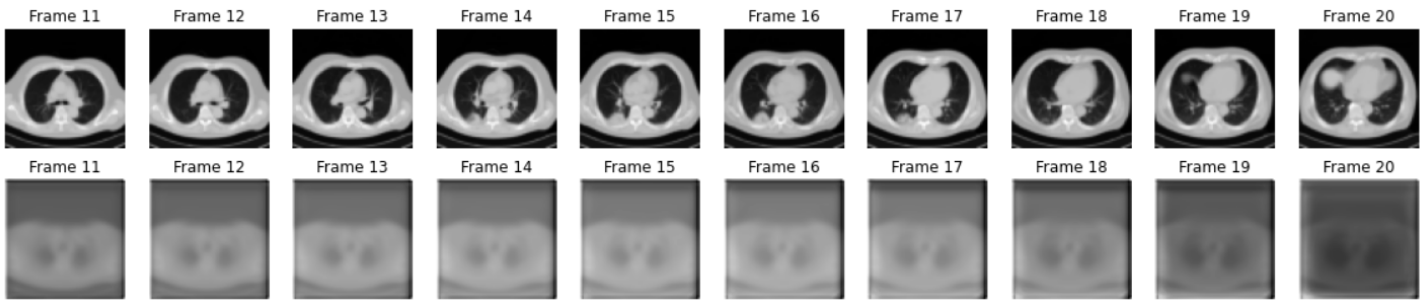


Figura 6.5: Predicción de frames para el caso 2.1

Para finalizar con este experimento y la visualización de frames, se ha llevado a cabo un último caso para esta técnica de predicción. Se trata del caso 2.2 en el que se ha aumentado el número de ciclos de entrenamiento, y que se espera que este pequeño cambio consiga mejorar los resultados.

En esta ocasión, podemos observar claramente una ligera mejora, tanto en los resultados obtenidos independientes como en la evolución que tiene la predicción de los últimos frames. No obstante, la predicción del frame número 20 para este caso, recogido en la Figura 6.6, no es del todo exacta en cuanto al frame esperado. De todos modos, se puede destacar el gran progreso que ha tenido nuestro modelo comparándolo respecto al caso anterior.

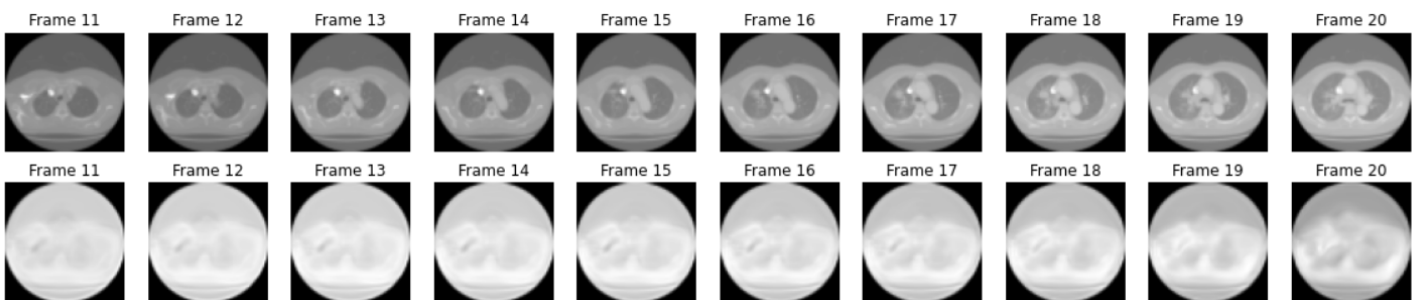


Figura 6.6: Predicción de frames para el caso 2.2

Una vez estudiados la predicción de frames en los dos primeros experimentos podemos sacar en claro varias cuestiones. Por una parte, la cantidad de ejemplos utilizados en el entrenamiento influye de manera significativa en la obtención de buenos resultados. Por el contrario, también ha quedado demostrado que la cantidad de ciclos no es del todo determinante en un modelo bien construido.

6.2.2. Rendimiento del modelo

Los modelos que implementan soluciones de aprendizaje automático para resolver problemas médicos necesitan saber cómo cuantificar el rendimiento del modelo. En el Aprendizaje Automático, los modelos son tan útiles como la calidad de sus predicciones [37]; y generalmente el objetivo no solo se centra en crear modelos, sino crear modelos de alta calidad. Sin embargo, ésta no es la finalidad principal del proyecto: Siempre conviene construir modelos diferentes para analizar diversas estrategias y así poder alcanzar la máxima precisión a la hora de contruir el modelo definitivo. Por consiguiente, en este subapartado se analiza el rendimiento que ha tenido cada uno de los casos propuestos mediante gráficas que muestran la precisión y la pérdida en los conjuntos de entrenamiento y validación.

Una curva de aprendizaje es un gráfico que muestra el tiempo en el eje de las abscisas X y el aprendizaje en el eje de las ordenadas Y . Durante el entrenamiento de un modelo, estas curvas pueden dar una idea de lo bien que está aprendiendo el modelo [38]. También puede evaluarse en un conjunto de datos de validación para tener una idea de lo bien que está generalizando el modelo. La forma de una curva de aprendizaje puede utilizarse para diagnosticar el comportamiento de un modelo y, a su vez, sugerir el tipo de cambios de configuración que pueden realizarse para mejorar el rendimiento.

Es fundamental analizar el rendimiento de los experimentos propuestos en esta capítulo para determinar si los modelos se comportan de manera adecuada. En la Figura 6.7a se recogen los gráficos de precisión y pérdida para el caso 1.1 en los que podemos destacar el aumento de la precisión en los dos primeros ciclos. Ésto indica que la red de este modelo está aprendiendo rápidamente. Sin embargo, a partir del tercer ciclo la curva se aplana, lo que indica que no se necesitan demasiadas iteraciones para seguir entrenando este modelo.

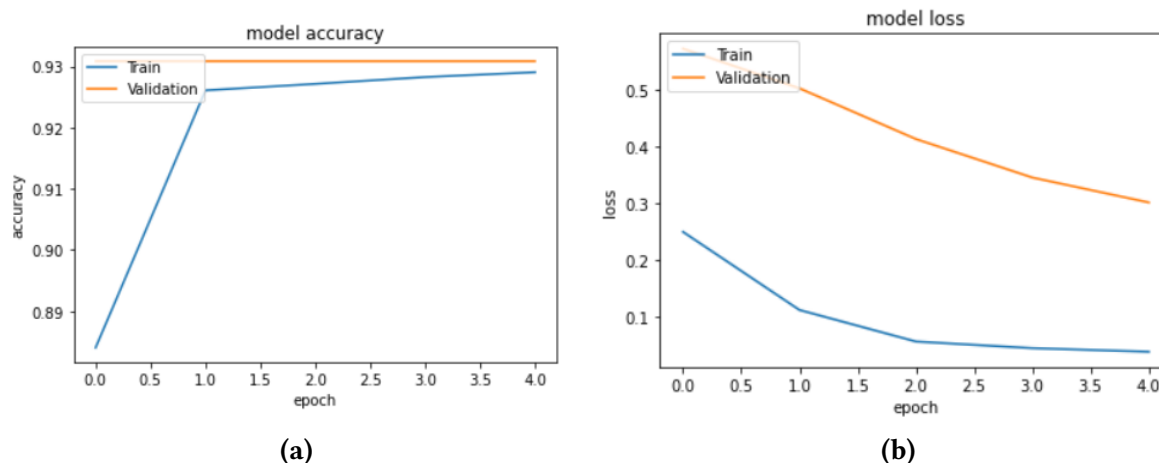


Figura 6.7: Rendimiento del caso 1.1

Un modelo infraajustado (*underfitting*) se produce cuando la red neuronal no puede captar las características de los conjuntos de entrenamiento y no puede ajustarse bien a los objetivos [39]. También puede identificarse por una pérdida de entrenamiento que disminuye y sigue disminuyendo al final del gráfico. Esto indica que el modelo es capaz de seguir aprendiendo y de realizar posibles mejoras adicionales o que el proceso de entrenamiento se detuvo prematuramente, como en la Figura 6.7b.

Por esta misma razón, en el caso 1.2 se ha aumentado el número de ejemplos de entrenamiento para poder estudiar cómo se comporta el modelo. La Figura 6.8a muestra que la tendencia de ambos conjuntos se mantiene en una alta precisión. A pesar de que se aprecia un pico a la baja, no es importante ya que la bajada no es muy notable en cuanto a la escala total.

En lo que respecta a las curva de pérdida para el caso 1.1 y el caso 1.2, ilustrados en la Figura 6.7b y la Figura 6.8b respectivamente, podemos observar una notable bajada para ambos conjuntos. Sin embargo, podemos destacar una diferencia considerable en cuanto a los valores que recogen estas funciones. Generalmente, debido al aumento de muestras empleadas en el caso 1.2, los valores son inferiores que para el caso 1.1.

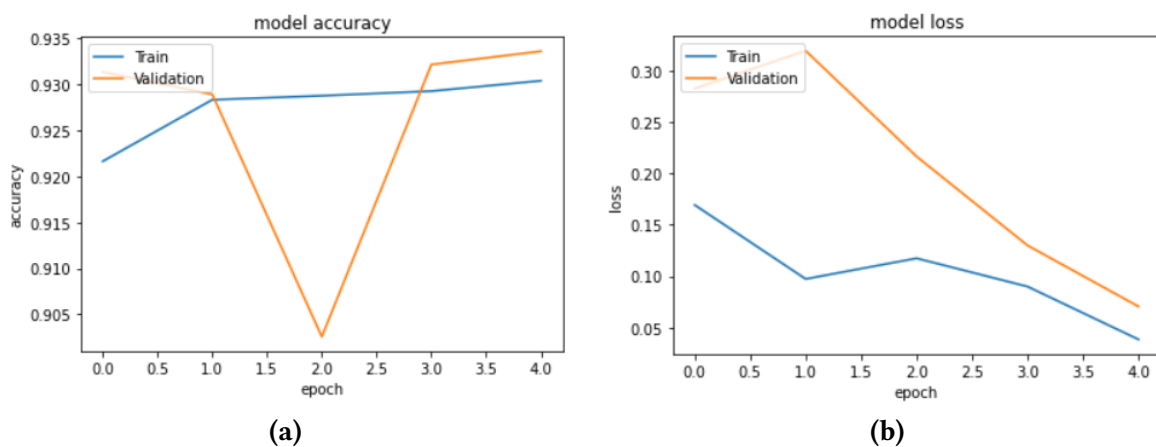


Figura 6.8: Rendimiento del caso 1.2

Dentro del mismo experimento, se ha aumentado el número de ciclos de 5 a 20 con el fin de corroborar esta tendencia y efectivamente se cumple la evolución esperada. Al igual que para los casos anteriores, el gráfico de la Figura 6.9a para el caso 1.3 muestra una tendencia a la alta en la curva de precisión y luego se aplana en valores superiores al 90%. Por otra parte, en relación a la pérdida en el conjunto de entrenamiento ilustrado en la Figura 6.9b, sigue una tendencia a la baja y una vez el modelo ha aprendido lo necesario se mantiene estable en un punto. Además, la pérdida de validación disminuye hasta un punto de estabilidad y tiene una pequeña diferencia con la pérdida de entrenamiento. Ésto es señal de un modelo ajustado adecuadamente.

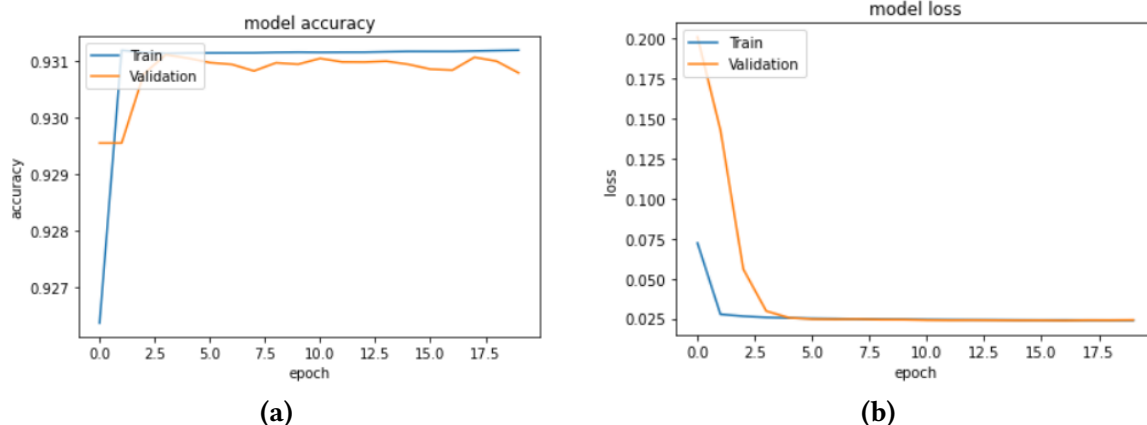


Figura 6.9: Rendimiento del caso 1.3

En relación con el segundo experimento, se van a analizar los dos casos que han obtenido resultados muy similares. Por una parte, los gráficos de precisión en la Figura 6.10a y Figura 6.11a recogen resultados bastante inferiores a los del experimento anterior que no llegan al 10 % de acierto. Podemos destacar que el conjunto de validación es superior que al del entrenamiento, por lo que es posible que el conjunto de validación esté formado por ejemplos más sencillos que el conjunto de entrenamiento. Por consiguiente, se recomienda probar a hacer una validación cruzada del modelo o simplemente volver a entrenar con una mezcla diferente de conjuntos de entrenamiento y validación para ver si la tendencia persiste.

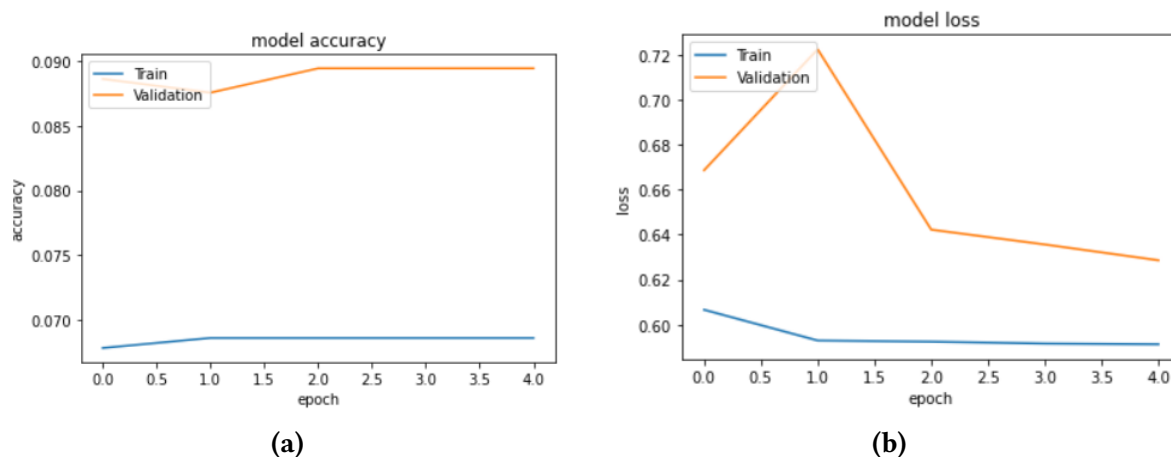


Figura 6.10: Rendimiento del caso 2.1

Referente a las gráficas de pérdida (Figura 6.10b y Figura 6.11b) muestran, ambas representaciones recogen valores muy altos, en consecuencia de la baja precisión obtenida. De todos modos, podemos destacar el comportamiento de las curvas que va disminuyendo a medida que se entrena. Al igual que en el experimento anterior, son necesarias más de 5 ciclos para conseguir una estabilidad al final del entrenamiento.

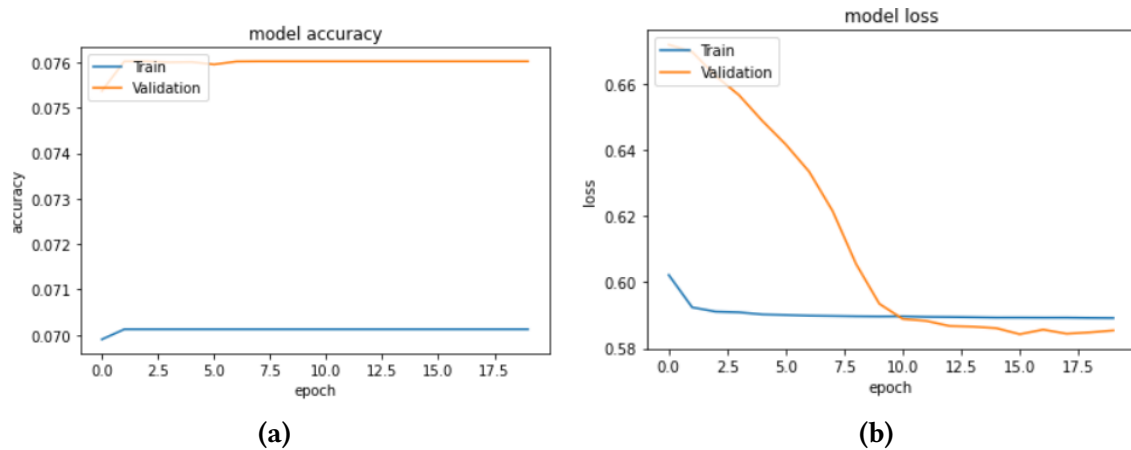


Figura 6.11: Rendimiento del caso 2.2

6.3. Técnicas de mejora

La visualización de datos es una de las mejores formas de humanizar los datos para facilitar su comprensión y obtener las tendencias relevantes de los mismos. Esta actividad puede ser crucial cuando todavía se está tratando de optimizar el modelo. Una de las principales dificultades de cualquier enfoque de aprendizaje automático es hacer que el modelo sea generalizado y obtener resultados razonables [40]. Visualizar la precisión de entrenamiento frente a la precisión de validación a lo largo de un número de ciclos es una buena manera de determinar si el modelo ha sido suficientemente entrenado. Esto es importante para que el modelo no esté subentrenado ni sobreentrenado. Los hiperparámetros, como el número de nodos por capa de la red neuronal y el número de capas en la red, pueden tener un impacto significativo en el rendimiento del modelo.

Con el fin de reducir el efecto del sobreajuste [41], se han utilizado múltiples soluciones basadas en diferentes estrategias.

- **Early-stopping.** Las arquitecturas de redes neuronales casi siempre tienen un número de parámetros demasiado grande, y esto conlleva a un posible sobreajuste. Para su posible solución se propone la reducción de parámetros y la detención temprana, la cual se basa en el juicio humano, o en modelo de carácter automatizado.
- Simplificación del modelo.
- Ampliación del conjunto de entrenamiento.
- Técnicas de regularización.

Modelos de clasificación

La clasificación es la técnica de extracción de datos más comúnmente aplicada y se utiliza un conjunto de ejemplos preclasificados para desarrollar un modelo que sea capaz de clasificar otro conjunto no etiquetado. La representación de imágenes para la tarea de clasificación utiliza a menudo métodos de extracción de características [42] que han demostrado ser eficaces para diferentes tareas de reconocimiento visual. *Deep Learning* reduce la tarea de desarrollar un nuevo extractor de características, automatizando la fase de extracción y aprendizaje de componentes.

El proceso de clasificación de datos implica dos fases, aprendizaje y clasificación. En la fase de aprendizaje, el algoritmo de clasificación analiza los datos de entrenamiento y en la fase de clasificación, se utilizan los datos de prueba para estimar la precisión de las reglas de clasificación. Además, involucra un conjunto de entrenamiento de datos con etiquetas de clase conocidas [43] el cual se emplea para construir un modelo que posteriormente se aplica a un conjunto de prueba que posee datos con etiquetas de clase desconocidas.

La importancia de la imagen médica digital ha hecho que el análisis de imágenes médicas desempeñe un papel indispensable en la terapia clínica [44]. Este capítulo tiene como objetivo distinguir las imágenes médicas según un determinado criterio, como las patologías clínicas o las modalidades de imagen. Un sistema fiable de clasificación de imágenes médicas es capaz de ayudar a médicos en la interpretación rápida y precisa de las imágenes.

7.1. Clasificación de imágenes 3D

La clasificación de imágenes 3D puede considerarse como una de las técnicas más comunes dentro de todos los métodos de clasificación de *Deep Learning*. La aplicación por excelencia es el campo médico, el cual se emplea para predecir la presencia de anomalías en imágenes de tipo TC. Este subapartado muestra los pasos necesarios para construir una CNN en 3D y a diferencia de las CNN en 2D [45], pueden tomar una secuencia de fotogramas en dos dimensiones para así aprender representaciones para datos volumétricos.

Con el fin de observar un comportamiento en los resultados totalmente distinto a los anteriores, se ha llevado a cabo un nuevo experimento, el cual se emplea un método especialmente creado para estudiar secuencias de fotogramas, *3D image classification*, a dos bases de datos, *MosMedData*, y *DB_SBRT*.

Experimento	Caso	DB	Method	Epochs
3	3.1	MosMedData	3D	100
	3.2	DB_SBRT	3D	100

Tabla 7.1: Características de los modelos de clasificación de imágenes 3D

En la Figura 7.1 se puede observar la representación gráfica de la red neuronal propuesta para este método, la cual está formada por un total de 17 capas. Las cuatro capas convolucionales 3D (*Conv3D*) siguen el mismo funcionamiento que las mencionadas en el Capítulo 6. Sin embargo, existe una pequeña diferencia: Dos de estas capas están formadas por 64 filtros, seguida de una de 128 y otra de 256. De todos modos, el tamaño del *kernel* que se aplica se mantiene en una estructura de $3 \times 3 \times 3$. A cada una de estas capas le sigue una capa de *Max-pooling* con un filtro de tamaño 2, seguida de otra capa de *Batch Normalization*.

Como los datos se almacenan en tensores de rango 3 de forma, se procede a añadir una dimensión de tamaño 1 en el cuarto eje para poder realizar convoluciones 3D en los datos. La nueva estructura es la siguiente:

$$(muestras, altura, anchura, profundidad, 1)$$

La extracción de características concluye mediante el uso de una capa de *Global Average Pooling* en la que los mapas de características de la última capa convolucional se vectorizan [18] y se introducen en una capa totalmente conectada formada por un total de 512 neuronas. Después de ignorar el 30 % de los casos mediante la capa *Dropout* se vuelve a hacer uso de otra capa totalmente conectada, la cual estará formada por una sola neurona.

Teniendo en cuenta que una serie de TAC tiene muchos cortes, el modelo va a emplear el conjunto entero como una imagen en 3D. En la Figura 7.2 se visualiza un montaje de cortes elegido al azar, compuesto por 40 en imágenes.

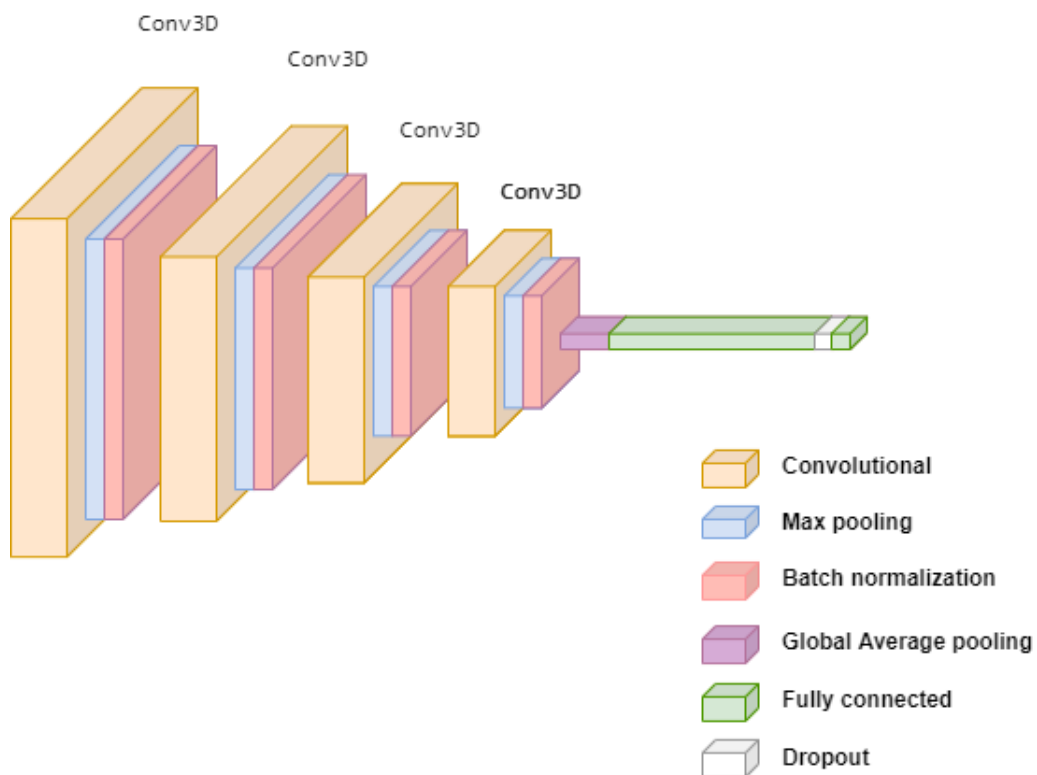


Figura 7.1: Arquitectura CNN para la clasificación de imágenes 3D

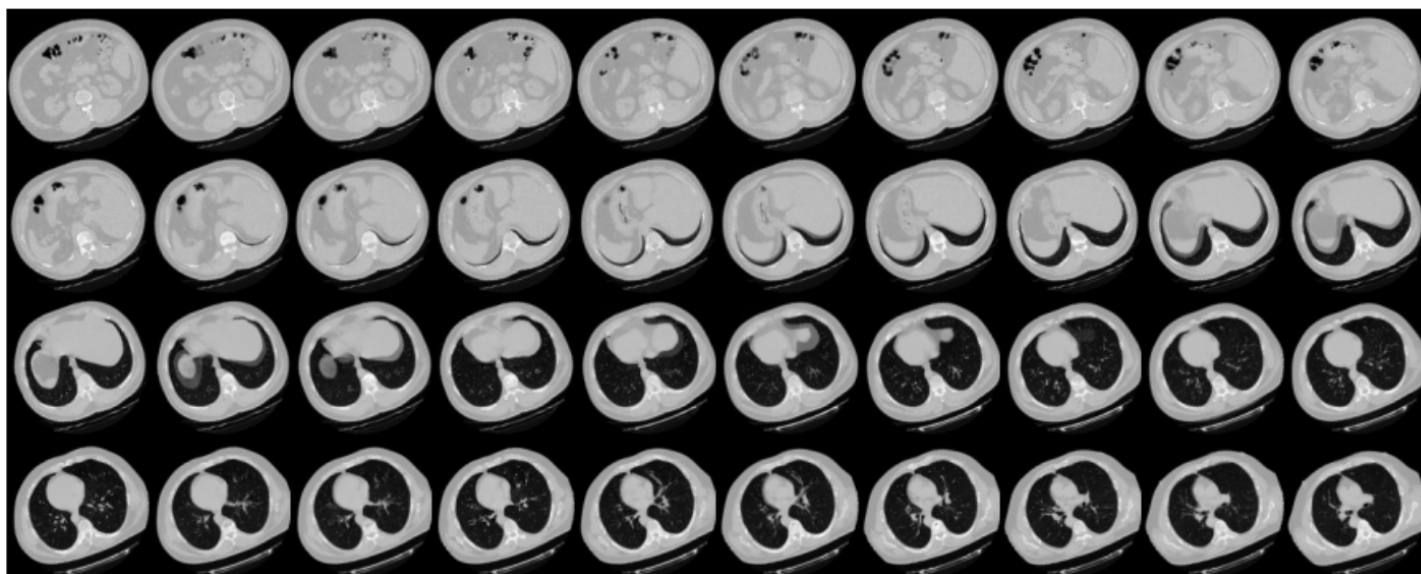


Figura 7.2: Montaje de los primeros 40 los cortes de un TAC

7.2. Clasificación de imágenes 2D

En este capítulo también se tiene como objetivo identificar las mejores variables que puedan servir para generar un modelo de clasificación confiable. Por lo tanto, después de trabajar con volúmenes en 3D se ha decidido implementar lo propio para imágenes en 2D. A continuación, en la Tabla 7.2, se presentan dos experimentos para esta fase del proyecto.

Ambos experimentos trabajan con la base de datos *DB_SBRT* en los que el objetivo de clasificación es diferente. Por una parte, en el experimento 4 se pretende construir un modelo capaz de distinguir en un conjunto de datos a qué parte del cuerpo corresponde las tomografías disponibles. Se trata de una clasificación binaria, en la que nos interesa saber si las imágenes muestran las partes 'ABDOMEN', 'CHEST' o 'TORAX'. En caso positivo la imagen será clasificada como 'ACT' y en caso negativo como 'Other'. Por otra parte, en el experimento 5 se pretende realizar la distinción en cuanto al tipo de imagen, dónde nos interesa saber si el contenido es de tipo 'AXIAL' o no ('Other').

Experimento	Caso	Clases	NN	Salto de imagen
4	4.1	ACT, Other	NN1	100
	4.2	ACT, Other	NN1	10
5	5.1	AXIAL, Other	NN1	100
	5.2	AXIAL, Other	NN1	10

Tabla 7.2: Características de los modelos de clasificación

La base de datos *DB_SBRT* es muy amplia y trabajar con ella en su totalidad supone un gran coste computacional. De la misma manera, teniendo en cuenta que las series imágenes pueden considerarse secuencias temporales, se ha decidido prescindir de algunas. Por consiguiente, a la hora de recoger las imágenes, en el caso 4.1 y en el caso 5.1 se ha hecho un salto de 100 imágenes y en el caso 4.2 y en el caso 5.2 un salto de 10.

En cuanto a los resultados obtenidos para el experimento 4 y 5, recogidos en la Tabla 7.4, podemos asegurar que nos encontramos ante varios casos de *overfitting*. Por muy preciso que sea el modelo, es sospechoso que la tasa de acierto sea tan elevada. En este caso se ofrece demasiada flexibilidad a los parámetros y se acaba generando un modelo que encaja perfectamente con los datos que ha sido entrenados pero que no es capaz de realizar la función básica de un modelo estadístico: ser capaz de generalizar a nueva información.

El primer paso para evitar este problema es reducir la complejidad del modelo. En redes neuronales podemos disminuir el número de capas o neuronas o utilizar técnicas de regularización como el *dropout* o el *early stopping*. Se han probado varias alternativas pero en ningún caso se ha conseguido mejorar el modelo de manera notable, y por esta razón se ha optado por llevar a cabo un nuevo experimento con una red neuronal totalmente nueva.

Experimento	Caso	Clases	NN	Salto de imagen
6	6.1	ACT, Other	NN2	10
	6.2	AXIAL, Other	NN2	10

Tabla 7.3: Características de los modelos de clasificación

En la Figura 7.3 se puede observar la representación gráfica de la red neuronal propuesta para este experimento, formada por un total de 7 capas. Cada una de las dos capas convolucionales 2D (*Conv2D*) crean un núcleo de convolución que se asocia con la entrada y ayuda a producir un tensor de salidas. La primera de ellas está formada por 128 filtros, y la segunda por 64. A cada una de estas capas le sigue una capa *Pooling* para simplificar la salida al realizar un submuestreo no lineal, lo que reduce el número de parámetros que la red necesita aprender.

Siguiendo el aprendizaje de características, la arquitectura de una red neuronal convolucional cambia a la clasificación. La penúltima capa, totalmente conectada, genera un vector de K dimensiones, donde K es el número de clases que la red será capaz de predecir. Este vector contiene las probabilidades de una imagen perteneciente a cada una de las clases, este caso $K = 2$. La última capa de la arquitectura de la red neuronal convolucional utiliza una capa de clasificación como *softmax* para generar la salida de clasificación.

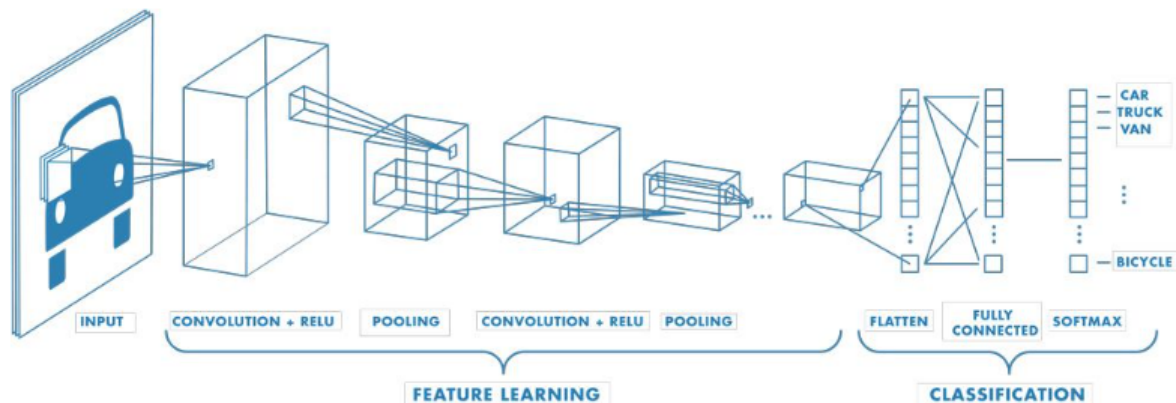


Figura 7.3: Arquitectura CNN para la clasificación de imágenes 2D

7.3. Resultados de los experimentos

Es importante estudiar bien los resultados de los experimentos y analizar su rendimiento. En general las redes neurales han mostrado una capacidad clasificatoria igual o superior que las técnicas estadísticas convencionales. A diferencia de los dos experimentos realizados hasta ahora, todos los siguientes se ha empleado como herramienta de clasificación. En cuanto al experimento 3, debido a su largo proceso de 100 ciclos, se ha decidido mostrar el resultado mediante los gráficos de la Figura 7.4.

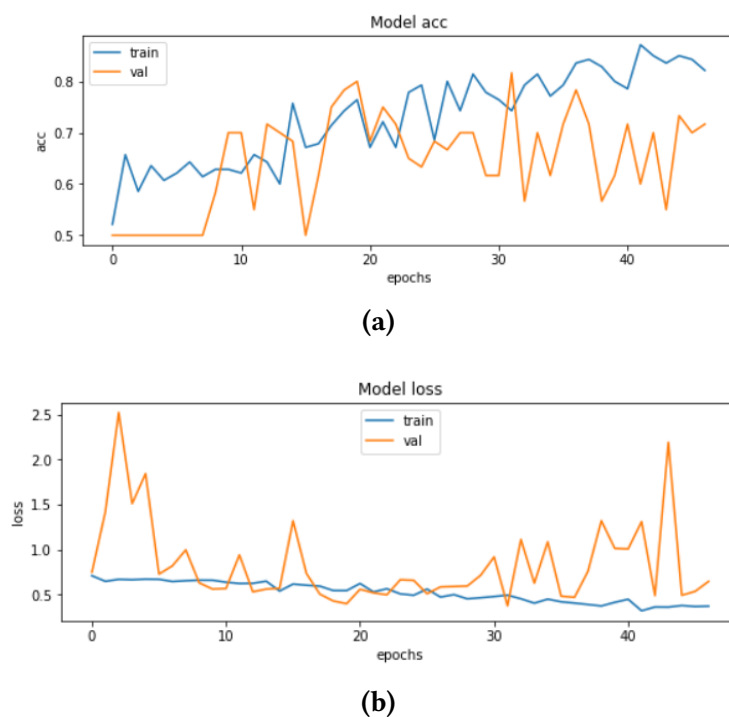


Figura 7.4: Rendimiento del caso 3.1

En cuanto al gráfico de precisión (Figura 7.4a) podemos observar una evolución positiva, representada mediante el color azul, en el conjunto de datos de entrenamiento: Cuantos más ciclos realizados el modelo obtiene un mejor resultado. Con lo que respecta a los datos de pérdida (Figura 7.4b), los valores van disminuyendo de forma regular en el conjunto de entrenamiento, por lo que podemos asegurar la efectividad del modelo.

La clasificación de imágenes en 2D permite estudiar cada tomografía independientemente de la serie secuencial a la que pertenece. A continuación se muestran varios bloques de ejemplos de clasificación para el experimento 4 y el experimento 5. El primer bloque (Figura 7.5) consta de 10 ejemplos seleccionados de manera aleatoria, en los que cada caso se ha clasificado de manera correcta: Todos los casos pertenecen a la clase 'ACT' y efectivamente esa ha sido la predicción.

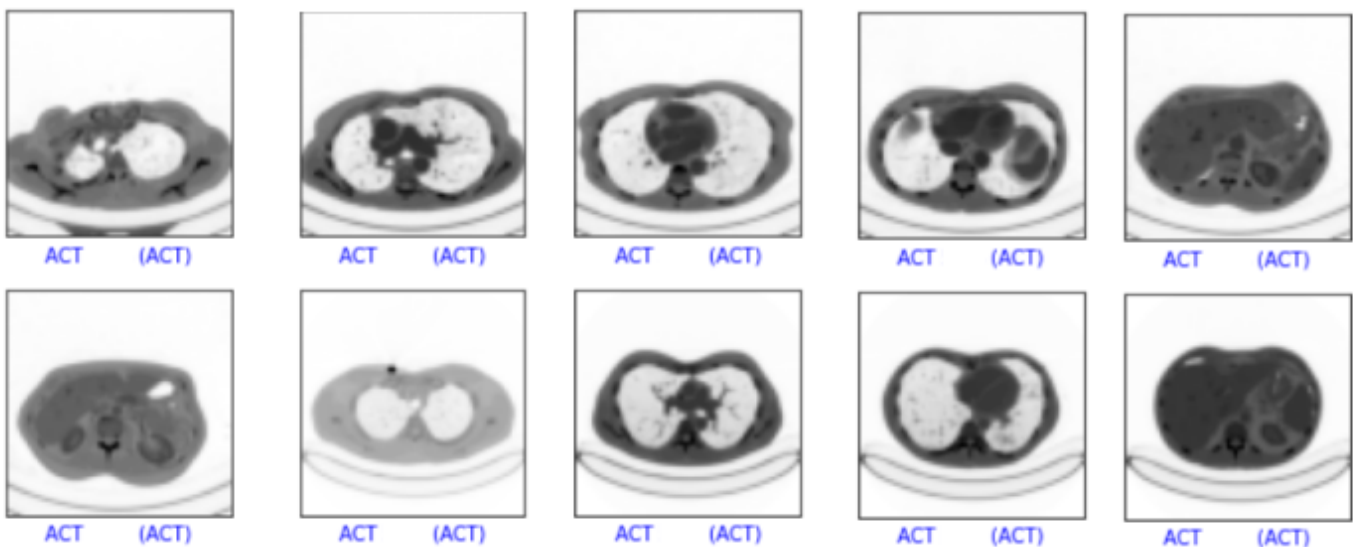


Figura 7.5: Primer bloque de ejemplos de clasificación (Experimento 4)

El segundo bloque de imágenes (Figura 7.6) recoge cinco ejemplos también pertenecientes a la clase 'ACT'. En esta ocasión, nos encontramos antes un caso mal clasificado, el segundo, el cual ha sido clasificado como 'Other' y realmente no lo es. A pesar de que el primer y el segundo caso tienen bastantes similitudes, la predicción de clases no concuerda. No obstante, el resto de los casos en este bloque son muy distintos al primero y aun así coinciden en la clasificación.



Figura 7.6: Segundo bloque de ejemplos de clasificación (Experimento 4)

A pesar de que las imágenes pertenecientes a las clase 'Other' son muy reducidas, se ha buscado un ejemplo con el fin de mostrar casos con esta condición. Podemos observar como el último caso de la Figura 7.7 incluye una imagen de esta clase pero clasificada como 'ACT'. De la misma manera, encontramos el caso opuesto en el que una imagen de la clase 'ACT' se ha clasificado como 'Other'.

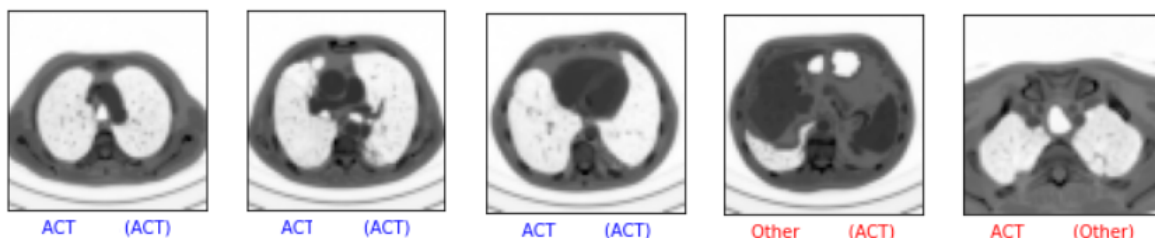


Figura 7.7: Tercer bloque de ejemplos de clasificación (Experimento 4)

Todos los ejemplos mostrados hasta ahora, independientemente de su etiqueta, presentan tomografías claramente pertenecientes a la parte torácica, es decir, se puede intuir de manera clara que pertenecen a la clase 'ACT'. Por una parte, los tres primeros casos que se recogen en la Figura 7.8 representan casos que claramente no pertenecen a esta clase. En los dos primeros, nuestro modelo ha decidido clasificarlos como 'ACT', por lo que la clasificación se ha efectuado de manera incorrecta. Por otra parte, el cuarto caso de este bloque no muestra información, y aún así el clasificador no ha acertado con la clase real.

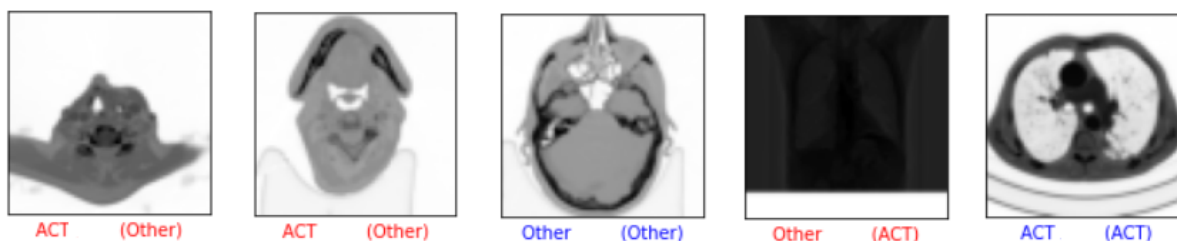


Figura 7.8: Cuarto bloque de ejemplos de clasificación (Experimento 4)

A pesar de que las técnicas empleadas son muy similares y que el proceso es idéntico, en el experimento 5 se ha querido llevar a cabo una clasificación de dos clases diferentes. En este caso vamos a trabajar con los planos tomográficos y la clase de interés corresponde a la 'AXIAL'. La Figura 7.9 recoge el primer bloque de este experimento compuesto por 8 muestras. Con el fin de entrar en contexto se han escogido estos casos en los que nuestro modelo asegura que todas las clases pertenecen a esta clase.

Por el contrario, en la Figura 7.10 se muestran imágenes pertenecientes a la otra clase. En esta ocasión, los ejemplos están previamente etiquetados como 'Other' y después de los proceso de entramiento y validación, nuestro clasificador ha considerado etiquetar esas imágenes como tal. A simple vista es más que evidente que las muestras expuestas en este bloque pertenecen a las tomografías en su plano coronal. La cuestión del modelo es considerar si para él también le resulta tan sencillo hacer esta distinción.

Las aplicaciones de *Deep Learning* se han convertido en una valiosa herramienta para resolver problemas complejos y es muy importante proporcionar fiabilidad a los resultados obtenidos [46]. Debemos tener en cuenta que por mucho que obtengamos resultados si éstos no son del todo decisivos, el modelo no es muy efectivo debido a que la decisión tomada no es muy sólida.

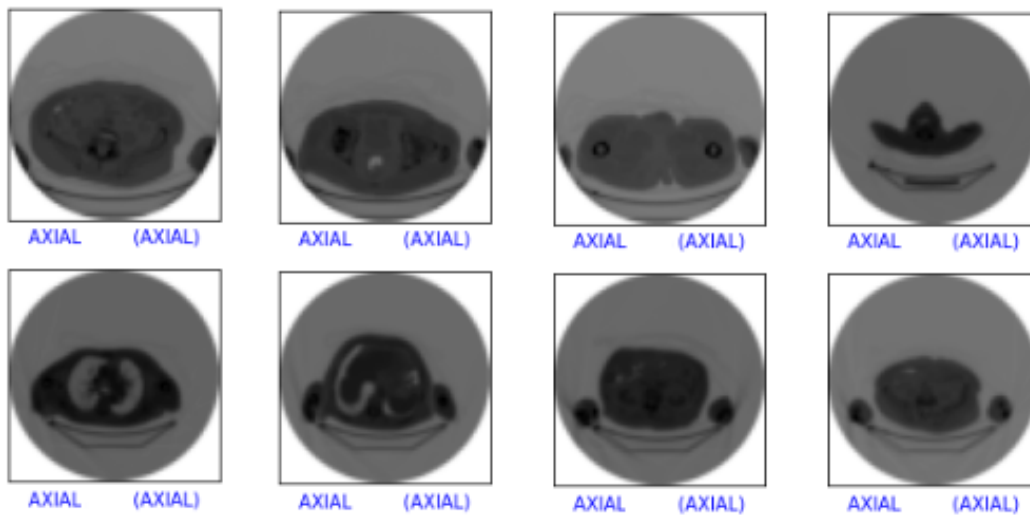


Figura 7.9: Quinto bloque de ejemplos de clasificación (Experimento 5)

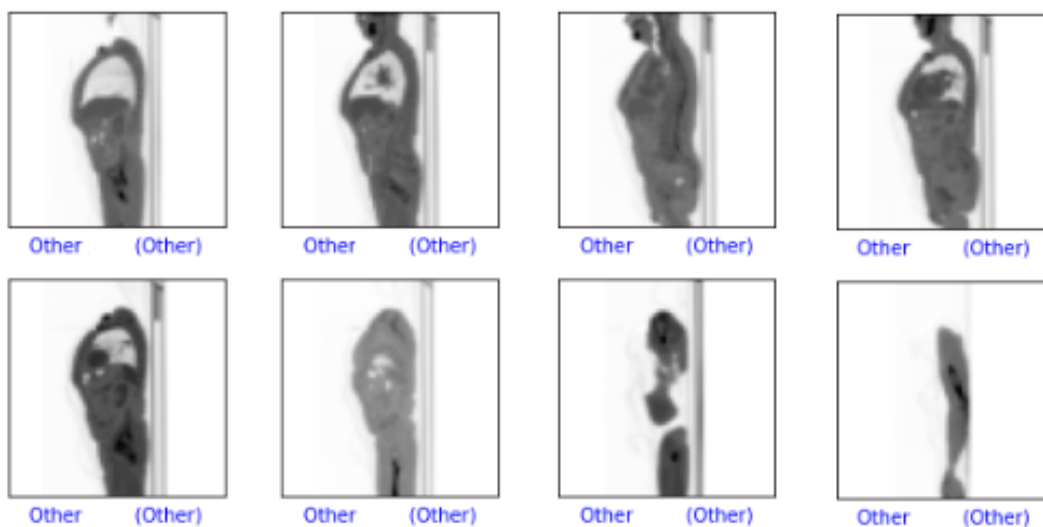


Figura 7.10: Sexto bloque de ejemplos de clasificación (Experimento 5)

Una de las opciones que vamos a estudiar aquí, es asumir la incertidumbre en el proceso, es decir, no buscar una distinción radical de clases [47], sino incluir también la posibilidad de error para cada una de las predicciones del modelo.

Por esta misma razón, en los bloques que vienen a continuación nos vamos a fijar en la tasa de fiabilidad que los modelos han establecido para cada resultado. En los dos primeros casos de la Figura 7.11 nos encontramos con predicciones con una tasa del 51 %. Por lo tanto, podemos asegurar que el modelo con el que se ha trabajado ha decidido clasificar estas dos imágenes como 'Other' pero sin mucha fiabilidad. Dicho de otra manera, también se considera que esos casos pertenecen en un 49 % a la clase 'AXIAL'.

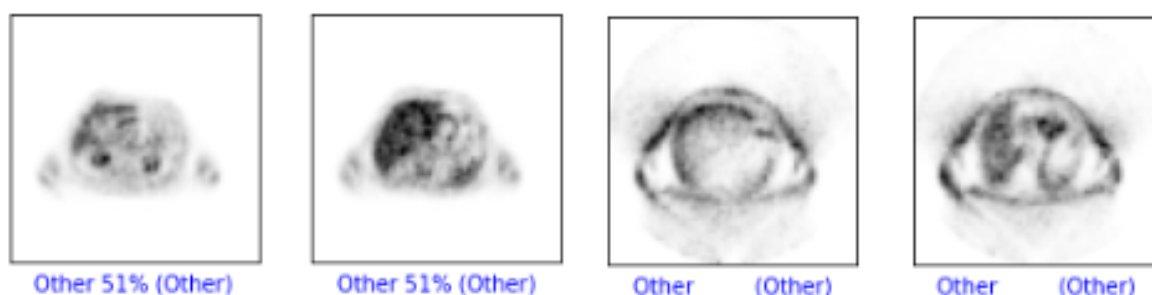


Figura 7.11: Séptimo bloque de ejemplos de clasificación (Experimento 5)

En el caso contrario, se ha considerado oportuno incluir el octavo bloque (Figura 7.12 de ejemplos para estudiar otra circunstancia destacable. Podemos determinar que a pesar de que la primera predicción no se ha realizado de manera correcta, tampoco es un resultado tan malo debido a la baja tasa de fiabilidad. Lo mismo ocurre con el segundo ejemplo, en el que el modelo establece que la clasificación no concuerda con la clase real. Aun así, trabajando con dos clases, la tasa del 50 % indica que el modelo no tiene mucha idea de cómo afrontar este caso.

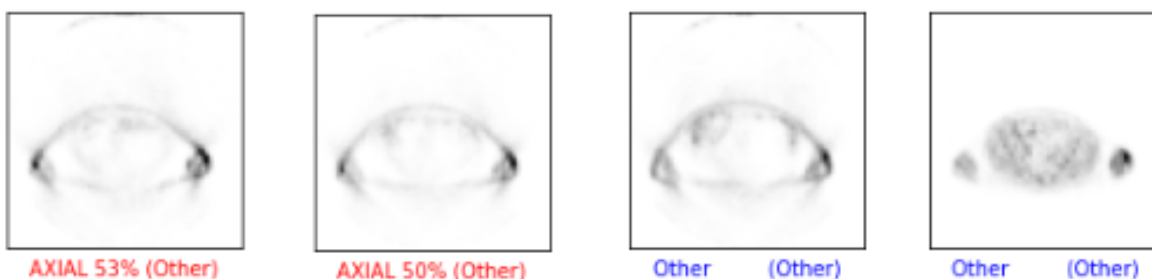


Figura 7.12: Octavo bloque de ejemplos de clasificación (Experimento 6)

Los modelos de Deep Learning tienen capacidad limitada para aprender de grandes cantidades de datos. Esto hace que las predicciones que realizan no sean del todo certeras cuando se exponen a situaciones desconocidas o distintos tipos de incertidumbre.

Para terminar con el estudio de ejemplos en los modelos de clasificación, se ha incluido un último bloque (Figura 7.13) correspondiente a otro caso particular. Los 16 ejemplos que se muestran a continuación corresponden a la clase 'AXIAL' y el modelo de este último experimento así lo ha considerado. La cuestión que se pretende tratar también gira en torno a la tasa de fiabilidad. En esta ocasión nuestro modelo es capaz de determinar que las presentes imágenes pertenecen a una clase con la fiabilidad del 100 %, es decir, sin contemplar lo más mínimo la posibilidad de fallo.

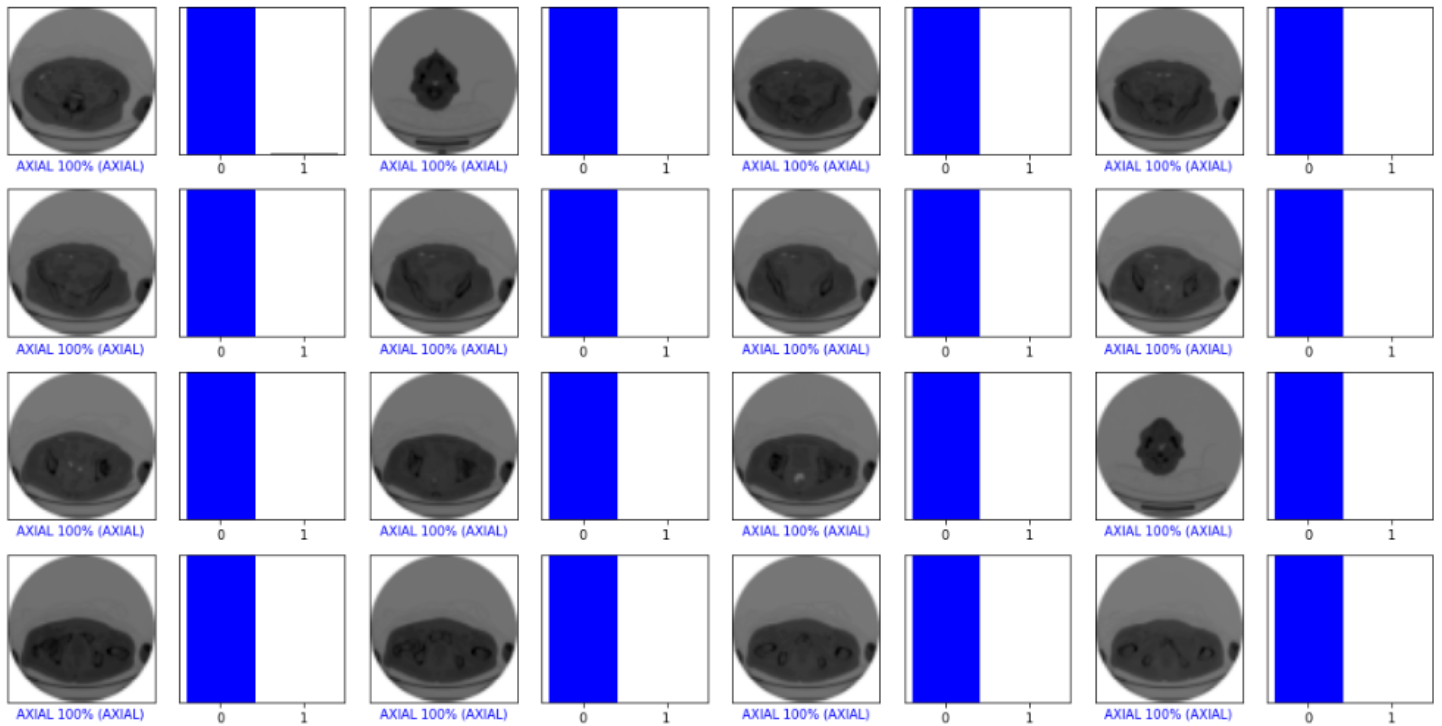


Figura 7.13: *Noveno bloque de ejemplos de clasificación (Experimento 5)*

Estudiar los resultados obtenidos de manera independiente ayuda a identificar los casos más atípicos y precisar una posible solución. De todas maneras, en el caso de querer valorar la efectividad de un modelo debemos ceñirnos a la tasa de acierto de una forma más global. Estos valores se recogen en la Tabla 7.4.

El experimento 3 ha sido en general el que mejor resultados ha obtenido en lo que respecta a su rendimiento. En cuanto a la tasa de acierto también podemos determinar que se ha conseguido valores aceptables en ambos casos. En lo que respecta al experimento 4 y al experimento 5, debido a que el segundo caso es una avance del primero, vemos reflejado la mejora en la tasa de acierto. De todos modos, ésto no quiere decir que se considere un progreso debido al sobreajuste que se ha creado.

Experimento	Caso	Resultado
3	3.1	76,44 %
	3.2	62,79 %
4	4.1	89,14 %
	4.2	94,32 %
5	5.1	54,02 %
	5.2	97,36 %
6	6.1	66,73 %
	6.2	60,89 %

Tabla 7.4: Resultados de los modelos de clasificación

Conclusiones y trabajo futuro

En este trabajo de fin de máster, se han empleado numerosas herramientas que, dadas sus características, desde un principio se esperaba obtener resultados muy diferentes entre sí. Entre ellos podemos destacar aquellos que se representan de manera visual, en forma de imagen, y los que ofrecen un valor numérico. Del mismo modo, estudiar el rendimiento de las técnicas empleadas también puede considerarse otro tipo de resultado.

No es tarea sencilla interpretar un resultado que depende de múltiples factores y que al mismo tiempo consta de varias fases en su desarrollo. Dentro de el campo de *Deep Learning* se han podido estudiar varios comportamientos basados en redes neuronales debido a que se han planteado diferentes experimentos. Por una parte, de manera general, podemos destacar que cada base datos y su composición ha sido determinante a la hora de trabajar en este proyecto. En los modelos de clasificación se han obtenido resultados muy dispares debido a la estructura de las redes entrenadas. A pesar de que todas las bases de datos empleadas en este fase contienen imágenes médicas y que éstas sean a primera vista similares, la estructura interna puede ser totalmente distinta.

Exploración previa. Un estudio previo del contexto en el que ha trabajado ha sido de gran utilidad para comprender el marco del problema. La recopilación de los numerosos artículos disponibles sobre la radiocirugía estereotáctica han contribuido al buen entendimiento de la meta que se ha querido alcanzar. De la misma manera, gracias a la fase preliminar en la que se han explorado imágenes tumorales se ha podido construir un proyecto en el que poder sacar el máximo rendimiento.

Especificidad. Uno de los factores más importantes a tener en cuenta a la hora de trabajar con redes neuronales podría ser la necesidad de un previo estudio de las mismas. Es esencial establecer de manera clara los objetivos para así poder construir el modelo más adecuado a nuestras necesidades. Uno de los problemas causados en este proyecto ha sido la generalización de modelos para cualquiera de los experimentos; y que en consecuencia, no se ha encontrado un modelo capaz de satisfacer los objetivos de cualquier conjunto.

Formación. Otro de los puntos claves en este proyecto ha sido la ejecución gradual en cuanto a la dificultad de la metodología. La posibilidad de trabajar con diferentes bases de datos, de distinto contenido y estructura, ha sido fundamental para realizar un seguimiento durante esta propuesta de investigación. Del mismo modo, en cuanto a las técnicas también han sufrido un aumento de dificultad a medida que se han obtenido resultados. Esto se ha visto reflejado en las estructuras de las redes neuronales, las cuales algunas son muy sencillas y otras más complejas. Sin embargo, este factor no ha sido determinante en la obtención de buenos resultados.

Finalmente, como continuación de este proyecto de fin de máster, existen varias líneas de investigación que quedan abiertas y en las que es posible seguir trabajando. La mayoría de ellas están relacionadas con este proyecto y pueden considerarse una ampliación del proceso. La idea de la propuesta consta de tres fases diferentes. La primera, y la más importante se centra en la detección de nódulos o demás anomalías dentro del área pulmonar. Para ello, se usarán técnicas de segmentación para poder localizar el área exacta afectada. La siguiente fase dará comienzo a partir de las imágenes en las que se hayan detectado nódulos, y tendrá como objetivo clasificar esas irregularidades como malignas o benignas. Para finalizar, en la última fase se pretende analizar la evolución de los tumores malignos con el paso del tiempo y de esta forma determinar la fase de desarrollo en la que se encuentra. Todo ello permitirá acelerar el proceso de detección, evaluación y tratamiento del tumor en caso de ser maligno.

Bibliografía

- [1] Lei Cai, Jingyang Gao, and Di Zhao. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*, 8(11), 2020. Ver página 1.
- [2] Federica Vernuccio, Roberto Cannella, Albert Comelli, Giuseppe Salvaggio, Roberto Lagalla, and Massimo Midiri. Radiomica e intelligenza artificiale: Nuove frontiere in medicina. *Recenti Progressi in Medicina*, 111(3):130–135, 2020. Ver página 2.
- [3] Issam El Naqa and Martin J. Murphy. *What Is Machine Learning?*, pages 3–11. Springer International Publishing, Cham, 2015. Ver página 2.
- [4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. Ver página 2.
- [5] H.H. Aung, A. Sivakumar, S.K. Gholami, S.P. Venkateswaran, Bapi Gorain, and Shadab. Chapter 1 - an overview of the anatomy and physiology of the lung. In Prashant Kesharwani, editor, *Nanotechnology-Based Targeted Drug Delivery Systems for Lung Cancer*, pages 1–20. Academic Press, 2019. Ver páginas 3, 4.
- [6] American Cancer Society. Acerca del cáncer de pulmón. <https://www.cancer.org/es/cancer/cancer-de-pulmon/>, 2022. Ver página 4.
- [7] Golpe al cáncer de pulmón: aclaran el origen de la enfermedad en los no fumadores. https://www.lespanol.com/ciencia/20210907/golpe-cancer-pulmon-aclaran-enfermedad-no-fumadores/610190002_0.html, 2021. Ver página 4.
- [8] I Torres Espallardo. Pet/tac: bases físicas, instrumentación y avances. *Radiologia*, 59(5):431–445, 2017. Ver página 5.
- [9] JR García Garzón, A Rodríguez, and A Cabrera. Tomografía por emisión de positrones de cuerpo completo (pet/tac) con 18f-fluorodesoxiglucosa. *Revista Española de Medicina Nuclear*, 28(2):85–89, 2009. Ver página 6.

- [10] Chia-Hung Chen, Chih-Kun Chang, Chih-Yen Tu, Wei-Chih Liao, Bing-Ru Wu, Kuei-Ting Chou, Yu-Rou Chiou, Shih-Neng Yang, Geoffrey Zhang, and Tzung-Chi Huang. Radiomic features analysis in computed tomography images of lung nodule classification. *PLoS one*, 13(2):e0192002, 2018. Ver página 7.
- [11] Subba R Digumarthy, Atul M Padole, Shivam Rastogi, Melissa Price, Meghan J Mooradian, Lecia V Sequist, and Mannudeep K Kalra. Predicting malignant potential of subsolid nodules: can radiomics preempt longitudinal follow up ct? *Cancer Imaging*, 19(1):1–8, 2019. Ver página 7.
- [12] Jayashree Kalpathy-Cramer, Artem Mamomov, Binsheng Zhao, Lin Lu, Dmitry Cherezov, Sandy Napel, Sebastian Echegaray, Daniel Rubin, Michael McNitt-Gray, Pechin Lo, et al. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. *Tomography*, 2(4):430–437, 2016. Ver página 7.
- [13] Pathmind Chris Nicholson. Neural networks. <https://www.ibm.com/cloud/learn/neural-networks>. Ver página 9.
- [14] Towards Data Science Gavril Ognjanovski. Everything you need to know about neural networks and backpropagation – machine learning easy and fun. <https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-mad>. Ver página 9.
- [15] Towards Data Science Dasaradh S K. A gentle introduction to math behind neural networks. <https://www.ibm.com/cloud/learn/neural-networks>. Ver página 10.
- [16] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks. *CoRR*, abs/1511.08458, 2015. Ver páginas 11, 12.
- [17] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018. Ver página 11.
- [18] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013. Ver páginas 12, 47.
- [19] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017. Ver página 13.
- [20] Mercedes Marqués-Andrés et al. *Bases de datos*. Universitat Jaume I, 2011. Ver página 15.
- [21] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. Emnist: Extending mnist to handwritten letters. In *2017 international joint conference on neural networks (IJCNN)*, pages 2921–2926. IEEE, 2017. Ver página 16.
- [22] Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6):141–142, 2012. Ver página 16.
- [23] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *ATT Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist>, 2, 2010. Ver páginas 16, 34.

-
- [24] Sergey P Morozov, AE Andreychenko, NA Pavlov, AV Vladzmyrskyy, NV Ledikhova, VA Gombolevskiy, Ivan A Blokhin, PB Gelezhe, AV Gonchar, and V Yu Chernina. Mosmeddata: Chest ct scans with covid-19 related findings dataset. *arXiv preprint arXiv:2005.06465*, 2020. Ver página 18.
- [25] Nikolay A Pavlov, Anna E Andreychenko, Anton V Vladzmyrskyy, Anush A Revazyan, Yury S Kirpichev, and Sergey P Morozov. Reference medical datasets (mosmeddata) for independent external evaluation of algorithms based on artificial intelligence in diagnostics. *Digital Diagnostics*, 2(1):49–66, 2021. Ver página 18.
- [26] José Ahas López Portillo. Predicción de errores en producción industrial de piezas mediante clasificación supervisada con desbalanceo de clases. *Universitat Oberta de Catalunya (UOC)*, 2019. Ver página 21.
- [27] Ursula Alcañas Martínez. Resonancia magnética, los protocolos y su práctica. <https://protocolosresonanciamagnetica.org/2015/01/25/planos-anatomicos/>, 2015. Ver página 22.
- [28] Francisco J Galdames, Fabrice Jaillet, and Claudio A Perez. An accurate skull stripping method based on simplex meshes and histogram analysis for magnetic resonance images. *Journal of neuroscience methods*, 206(2):103–119, 2012. Ver página 25.
- [29] Image Analyst. How to do skull stripping. <https://es.mathworks.com/matlabcentral/answers/310584-how-to-do-skull-stripping-to-a-dcm-image-in-matlab>. Ver página 25.
- [30] P. Li, S. Wang, T. Li, J. Lu, Y. HuangFu, and D. Wang. A large-scale ct and pet/ct dataset for lung cancer diagnosis [data set]. the cancer imaging archive. <https://doi.org/10.7937/TCIA.2020.NNC2-0461>, 2020. Ver página 27.
- [31] Amogh Joshi. Next-frame video prediction with convolutional lstms. https://keras.io/examples/vision/conv_lstm/, 2021. Ver página 34.
- [32] Matin Hosseini, Anthony S Maida, Majid Hosseini, and Gottumukkala Raju. Inception-inspired lstm for next-frame video prediction. *arXiv preprint arXiv:1909.05622*, 2019. Ver página 34.
- [33] Jefferson Ryan Medel and Andreas Savakis. Anomaly detection in video using predictive convolutional long short-term memory networks. *arXiv preprint arXiv:1612.00390*, 2016. Ver páginas 34, 35.
- [34] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015. Ver página 35.
- [35] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In *International conference on machine learning*, pages 843–852. PMLR, 2015. Ver página 36.
- [36] Tarun Papparaju. Video frame prediction with keras. <https://medium.com/machine-learning-basics/video-frame-prediction-with-keras-f74dd4743a1f>, 2018. Ver página 36.

- [37] Marianella Álvarez Vega, Laura María Quirós Mora, Mónica Valeria Cortés Badilla, et al. Inteligencia artificial y aprendizaje automático en medicina. *Revista médica sinergia*, 5(8):e557–e557, 2020. Ver página 40.
- [38] Jason Brownlee. How to use learning curves to diagnose machine learning model performance. <https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/>, 2019. Ver página 40.
- [39] Haotian Zhang, Lin Zhang, and Yuan Jiang. Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2019. Ver página 41.
- [40] Vaibhav Sharma. Data visualization for deep learning model using matplotlib. <https://www.pluralsight.com/guides/data-visualization-pluralsightdeep-learning-model-using-matplotlib>, 2019. Ver página 43.
- [41] Xue Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019. Ver página 43.
- [42] Shiqi Yu, Sen Jia, and Chunyan Xu. Convolutional neural networks for hyperspectral image classification. *Neurocomputing*, 219:88–98, 2017. Ver página 45.
- [43] Joshua Zárate-Valderrama, Norka Bedregal-Alpaca, and Víctor Cornejo-Aparicio. Modelos de clasificación para reconocer patrones de deserción en estudiantes universitarios. *Ingeniare. Revista chilena de ingeniería*, 29(1):168–177, 2021. Ver página 45.
- [44] Jianpeng Zhang, Yutong Xie, Qi Wu, and Yong Xia. Medical image classification using synergic deep learning. *Medical image analysis*, 54:10–19, 2019. Ver página 45.
- [45] Hasib Zunair. 3d image classification from ct scans. https://keras.io/examples/vision/3D_image_classification/, 2020. Ver página 46.
- [46] Bogdan Nicolae, Esteban Meneses Rojas, and Elvis Rojas. Evaluando la resiliencia de modelos de deep learning. *Tecnología en Marcha*, 33(1):25–30, 2020. Ver página 53.
- [47] Yolanda de la Hoz. Del laboratorio al mundo real. ¿son fiables las predicciones de deep learning? <https://www.bbvanexttechnologies.com/pills/del-laboratorio-al-mundo-real-son-fiables-las-predicciones-de-deep-learning/>, 2021. Ver página 54.