



















# Acrónimos

- ADC** coeficiente de difusión aparente. 15, 18, 19, 32, 33, 35, 36, 50, 52, 54
- ADN** ácido desoxirribonucleico. 21
- ARNm** Ácido ribonucleico mensajero. 23
- CNNs** redes neuronales convolucionales. 5, 7, 28, 29, 32, 37, 39, 44, 48, 59
- DL** *deep learning*. 4–7, 15, 25, 31, 32, 36, 37, 39, 56, 59, 60, 63, 64
- DWI** imagen potenciada en difusión. 15, 16, 19, 32, 35, 36, 50, 52, 54
- EDC** químicos interruptores endocrinos. 22
- GAP** global average pooling. 40, 50
- Grad-CAMs** Gradient-Weighted Class Activation Maps. 44, 57, 58, 60
- HBP** hiperplasia benigna de próstata. 22, 23, 32
- IA** inteligencia artificial. 25
- IDE** entorno de desarrollo integrado. 37
- IRM** imagen por resonancia magnética. 5–8, 10, 16, 32, 33, 35, 36, 39, 41–43, 47, 50, 53, 54, 59
- ML** machine learning. 25, 37
- NN** redes neuronales. 6, 7, 25–28
- PCR** reacción en cadena de polimerasa. 23
- PSA** antígeno prostático específico. 4, 22
- ReLU** rectified lineal unit. 30, 45

**RF** radiofrecuencia. 10–12, 14

**RMN** Resonancia magnética nuclear. 7, 13, 16, 35

**RX** rayos-X. 8

**T2W** imagen potenciada en T2. 15

**TAC** tomografía axial computarizada. 7, 8

# Índice general

Resumen	I
Resum	III
Abstract	V
Acrónimos	VIII
Índice general	1
I Memoria	1
1 Introducción	3
1.1 Motivación	3
1.2 <i>Challenge ProstateX-2</i>	5
1.3 Objetivos	5
1.4 Guía de la memoria	6
2 Marco teórico	7
2.1 Imagen por resonancia magnética	7
2.1.1 Introducción a la imagen por resonancia magnética	7
2.1.2 Fundamentos físicos de la resonancia magnética	8
2.1.3 Adquisición de imágenes de resonancia magnética	10
2.1.4 Modalidades de imagen por resonancia magnética	15
2.2 Anatomía de la próstata	19
2.2.1 Cáncer de próstata	21
2.2.2 Escala de Gleason	23
2.3 Deep learning y redes neuronales	25
2.3.1 Introducción a las redes neuronales	25

2.3.2	Algoritmo de retropropagación . . . . .	27
2.3.3	Fundamentos de las redes neuronales convolucionales . . . . .	28
2.4	Estado del arte . . . . .	31
<b>3</b>	<b>Materiales</b>	<b>35</b>
3.1	Base de datos de ProstateX-2. . . . .	35
3.2	Software . . . . .	37
3.3	Hardware . . . . .	38
<b>4</b>	<b>Métodos</b>	<b>39</b>
4.1	Redes Neuronales Multi-Modales . . . . .	39
4.1.1	Combinación tardía . . . . .	41
4.1.2	Combinación temprana . . . . .	42
4.1.3	Redes híbridas . . . . .	43
4.2	Interpretabilidad . . . . .	44
4.2.1	Grad-CAMs. . . . .	44
<b>5</b>	<b>Experimentos y Resultados</b>	<b>47</b>
5.1	Adecuación de los datos y métricas utilizadas . . . . .	47
5.1.1	Pre-procesamiento de la base de datos . . . . .	47
5.1.2	Métricas y evaluación . . . . .	49
5.2	Detección de cáncer de próstata en IRM monomodal . . . . .	50
5.2.1	Arquitectura . . . . .	50
5.2.2	Resultados . . . . .	52
5.2.3	Discusión de resultados con arquitecturas monomodales. . . . .	53
5.3	Aplicación de arquitecturas multi-modales . . . . .	54
5.3.1	Arquitectura . . . . .	54
5.3.2	Resultados . . . . .	55
5.3.3	Discusión de los resultados. . . . .	56
5.4	Interpretabilidad . . . . .	57
<b>6</b>	<b>Conclusiones y líneas futuras</b>	<b>59</b>
6.1	Conclusiones. . . . .	59
6.2	Líneas futuras . . . . .	60
<b>II</b>	<b>Presupuesto</b>	<b>61</b>
<b>7</b>	<b>Presupuestos</b>	<b>63</b>
7.1	Presupuestos parciales . . . . .	63
7.1.1	Coste de mano de obra . . . . .	63
7.1.2	Coste de maquinaria . . . . .	64
7.2	Presupuestos totales . . . . .	65
	<b>Bibliografía</b>	<b>67</b>

Parte I

# Memoria



# Introducción

## 1.1 Motivación

Según la organización mundial de la salud, el cáncer se encuentra entre las diez mayores causas de muerte en el mundo. Esta incidencia aumenta si nos centramos en los países más desarrollados, donde algunos tipos de cáncer se encuentran entre las cinco causas más comunes [1]. Si además se observa la incidencia de los diferentes tipos de cáncer, el cáncer de próstata es el tercero más común en todo el mundo, y está solo por detrás del cáncer de pulmón en el caso de los hombres [2].

Además, según el Observatorio Global del Cáncer de la Organización Mundial de la salud, se espera que la incidencia del cáncer siga subiendo en los próximos años. Por ejemplo, en países como España, como refleja la figura 1.1, esta llegaría a más de 50000 casos al año en 2040 [3], aumentando en más de un 40% el número de casos actual:

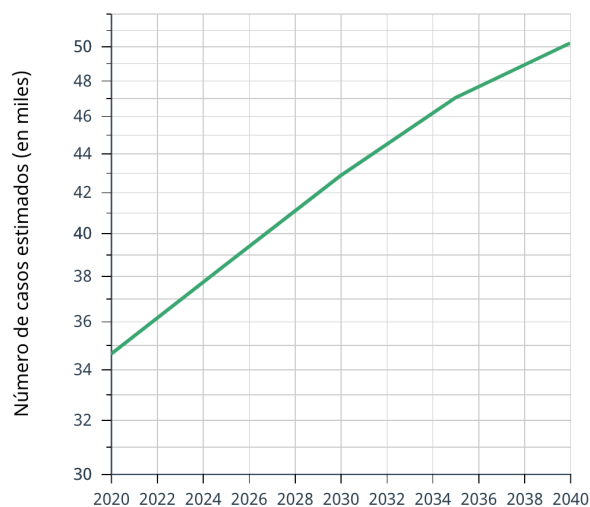
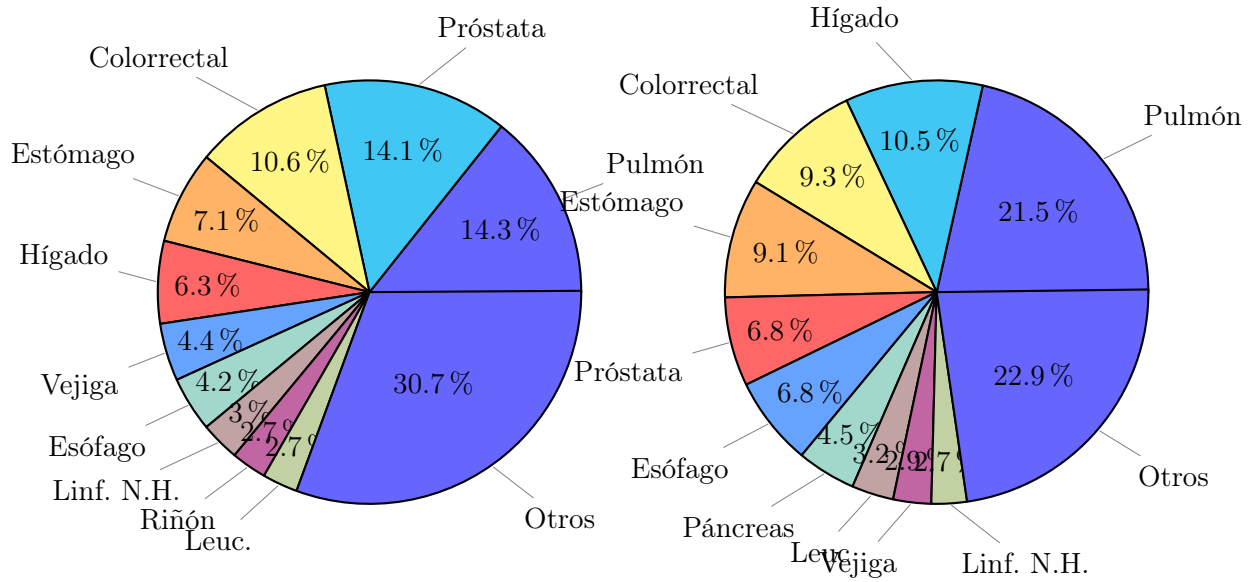


Figura 1.1: Predicción de la incidencia del cáncer de próstata en España [3].

Sin embargo, aunque el cáncer de próstata sea de los más comunes en todo el mundo en número de casos, no tiene tanta incidencia en la mortalidad como otros tipos de cáncer, ya que aunque el 14.1 % de los casos de cáncer son de próstata, solo el 6.8 % de las muertes de cáncer son debidas a este [2], como se refleja en la figura 1.2.



**Figura 1.2:** Porcentaje de casos y muertes por tipo de cáncer [2].

Existen varias razones por las que el cáncer de próstata no es tan mortal como otros, pero un factor crucial para una mayor supervivencia es la detección temprana de la enfermedad, ya que es clave para un correcto diagnóstico y poder así comenzar con el tratamiento adecuado lo antes posible.

En la actualidad, la detección del cáncer de próstata en la práctica clínica se basa en un análisis de sangre en el que se buscan niveles altos de antígeno prostático específico (PSA) y en el caso de que la proteína se encuentre en elevadas concentraciones se confirma el diagnóstico con una biopsia del tejido. Tras la biopsia se clasifica el grado de agresividad del cáncer en la escala de Gleason. El problema que existe con este método de diagnóstico se debe a que las biopsias son altamente invasivas, por ello, existe un gran interés clínico para obtener marcadores mediante técnicas menos invasivas, como la imagen por resonancia magnética. Además, de esta forma el proceso pasaría a ser totalmente indoloro para el paciente, evitando el tiempo de reposo que deben de tomar después de la biopsia y los posibles efectos secundarios que puede tener la extracción del tejido.

Por eso, la motivación del trabajo se encuentra en explorar nuevas formas de diagnóstico del cáncer, como se explicará posteriormente, basadas en la imagen por resonancia magnética, que hagan uso de técnicas de *deep learning* (DL) para dar solución a esta necesidad clínica.



## 1.2 *Challenge ProstateX-2*

El reto *ProstateX-2* [4] fue un *challenge* público llevado a cabo por la Asociación Americana de Físicos en Medicina, junto con la Sociedad Internacional de Óptica y Fotónica. Esta es una asociación que lleva más de 60 años dedicados a la aplicación de la física al campo de la medicina, mejorando la seguridad y precisión de los procedimientos que requieran de radiación, como el campo de la imagen médica o de la radioterapia.

En el año 2017 publicaron este reto, que tenía como objetivo el desarrollo de algoritmos para la obtención de biomarcadores multiparamétricos cuantitativos de imagen por resonancia magnética (IRM) de próstata, para así poder determinar el grupo de grado de Gleason de cáncer de próstata en el caso de que el paciente tenga cáncer de próstata o poder identificar que no lo tiene.

Este *challenge* tenía como novedoso el uso de IRM en el diagnóstico del cáncer de próstata, ya que no es un método que esté muy extendido en la práctica clínica, debido principalmente a que para un médico es difícil diferenciar una hiperplasia benigna de próstata de una maligna. Por ello, este *challenge* pretendía explorar nuevas vías de diagnóstico, aplicando las mejoras en las técnica de *deep learning* (DL), de forma que se pueda crear una herramienta de ayuda al diagnóstico para el médico.

La metodología del *challenge* fue la siguiente: primero se liberó la partición de imágenes de entrenamiento, para que los diferentes grupos de investigación fueran entrenando a sus modelos. Pasado un tiempo les entregaron las imágenes de test sin el *ground truth*, de forma que evaluarían los modelos sin saber la etiqueta verdadera de cada imagen, mandando finalmente los resultados que habían obtenido.

Los resultados del challenge fueron pobres [4], ya que ninguno de los modelos que participó fue capaz de obtener resultados aceptables. Aunque concluyeron que era posible automatizar la detección de los casos de pacientes con cáncer, a la hora de gradar la agresividad de los mismo los modelos fallaban, demostrando la dificultad de esta tarea.

Aunque el *challenge* terminó hace casi 4 años, las bases de datos siguen siendo públicas para que cualquier persona las pueda utilizar, de forma que será la base de datos que se implementará durante el desarrollo de todos los modelos de este trabajo.

## 1.3 Objetivos

El objetivo final de este trabajo es el desarrollo de algoritmos basados en redes neuronales convolucionales (CNNs) que sean capaces de combinar diferentes modalidades de IRM con el objetivo de poder diagnosticar si el paciente padece de cancer o no y en el caso que clasifique la imagen como patológica diferenciar entre cánceres más leves y más graves.

De esta forma, se crearía un sistema de ayuda al diagnóstico que ayudaría al facultativo en el diagnóstico de pacientes, además de habilitar una nueva vía de diagnóstico del cáncer de próstata, ya que a día de hoy en la práctica clínica el diagnóstico del cáncer de próstata se realiza mediante una biopsia del tejido.

Para la consecución de este objetivo, se van a abordar los siguientes puntos:

1. Adecuación de la base de datos del *challenge ProstateX-2* para su posterior tratamiento.
2. Revisar la literatura científica para estudiar las formas en las que se abordó el reto cuando se realizó.
3. A partir de los datos del diagnóstico de los casos proporcionados por los médicos generar el *ground truth* para el posterior entrenamiento.
4. Agrupar los datos en tres grupos distintos, uno en el que se encuentren los de grado más bajo, otro con los de grado más alto y otro donde se encuentren las imágenes sin tumor.
5. Realizar una partición de la base de datos en diferentes subconjuntos, dependiendo de si se utilizarán para el entrenamiento de la red, para su adecuada validación o para testear el sistema.
6. Entrenar diferentes modelos combinando de manera distinta las capas que compongan la red y las modalidades de imagen por resonancia magnética (IRM).
7. Optimizar los hiperparámetros de la red con el objetivo de mejorar los resultados.
8. Evaluar los resultados obtenidos para poder compararlos con los expuestos durante el reto en 2017, para de esta forma poder valorar la calidad de los mismos.
9. Realizar un análisis a partir de los resultados obtenidos y extraer conclusiones de estos, además de proponer posibles mejoras y líneas futuras de investigación en las que poder mejorar el resultado del trabajo.

## 1.4 Guía de la memoria

En este apartado se va a exponer de una manera breve de que van a consistir los siguientes capítulos del trabajo:

1. El capítulo 2 introducirá el marco teórico del trabajo, desarrollando las bases teóricas en las que se asientan la IRM y los modelos de DL y las redes neuronales (NN). También se expondrá cual es el estado del arte en el diagnóstico del cancer de próstata con métodos de aprendizaje profundo.
2. El capítulo 3 desarrollará los materiales utilizados para el desarrollo de este trabajo, tanto la base de datos como el software y el hardware empleados.
3. En el capítulo 4 se presentarán los métodos implementados durante la realización del trabajo.
4. En el capítulo 5 se expondrán los experimentos que se han llevado a cabo y los resultados obtenidos de los diferentes modelos.
5. El capítulo 6 extraerá las conclusiones finales de trabajo, además de proponer diferentes líneas futuras de mejora.

## Capítulo 2

# Marco teórico

En el presente capítulo se presentará el marco teórico del trabajo. Para ello, en primer lugar se explicará el funcionamiento de la imagen por resonancia magnética (IRM) y los fundamentos físicos en los que se asienta. A continuación, se realizará una explicación tanto de los fundamentos básicos de la anatomía de la próstata como de las características del cáncer de próstata, para seguidamente introducir los métodos actuales de su diagnóstico. Posteriormente, se procederá a introducir las técnicas de *deep learning* (DL), centrado especialmente en las redes neuronales (NN) y en las redes neuronales convolucionales (CNNs), y una vez introducidas las técnicas de aprendizaje profundo se enmarcará el estado del arte actual en el diagnóstico del cancer de próstata mediante técnicas de DL

### 2.1 Imagen por resonancia magnética

#### *2.1.1 Introducción a la imagen por resonancia magnética*

La primera propuesta al uso de la propiedad de resonancia magnética de los núcleos se produce en el año 1974, cuando Paul C. Lauterbur y Peter Mansfield publican de manera independiente los artículos en los que postulan el uso de gradientes de campo magnético para la localización espacial de señales de Resonancia magnética nuclear (RMN). Es este trabajo el que posteriormente en el año 2003 les valdrá para obtener el Premio Nobel en Fisiología y Medicina. [5]

Desde que se diseñó el primer equipo de IRM su uso no ha hecho más que aumentar. Entre sus mayores avances se encuentra el de ser la primera tecnología que permitía obtener imágenes tomográficas sin tener que irradiar al paciente, a diferencia de la tomografía axial computarizada (TAC), cuyo primer equipo fue diseñado unos años antes en 1972.

Las mejoras en la IRM se han seguido sucediendo desde ese momento, tanto en la resolución de los cortes, aumentando la calidad de las imágenes, como en el tiempo de adquisición de las mismas, de forma que cada vez requiere menos tiempo la adquisición de las imágenes. Como ejemplo de esto nos encontramos que los primeros equipos tenían una resolución de 4 x 4 x 10

mm [5], mientras que cualquier equipo actual de 1.5 T tiene una resolución de 1.5 x 1.5 x 4 mm [6].

Todo esto no ha hecho más que aumentar su uso hasta convertirse en uno de los procedimientos más comunes en la práctica clínica. Su utilización conlleva ventajas respecto a otras tecnologías como la ecografía, ya que la calidad de la imagen es mucho mejor, o respecto a técnicas como los RX o el TAC ya que, como se ha expuesto anteriormente, su uso no requiere de irradiar al paciente. Sin embargo, entre sus inconvenientes más grandes se encuentra el tiempo de adquisición de las imágenes, que puede requerir de estudios de más de media hora de duración, con todo lo que ello significa para el paciente, tanto por su comodidad como la posibilidad de que en ese tiempo haya variado su posición, artefactando la imagen, aunque actualmente existan técnicas para corregir estos artefactos de movimiento.

A continuación se pretende abordar el funcionamiento de la adquisición de las IRM y los fundamentos teóricos necesarios para comprender su funcionamiento. Si bien una explicación demasiado profunda no es el objetivo de este trabajo, sí que es necesario para abordarlo de una manera efectiva un desarrollo lo suficientemente detallado y riguroso de su funcionamiento.

### 2.1.2 Fundamentos físicos de la resonancia magnética

Para el desarrollo de los fundamentos de la resonancia magnética se han utilizado, a parte de las fuentes citadas, los conocimientos adquiridos sobre el funcionamiento de la resonancia magnética en la asignatura de técnicas de imágenes biomédicas.

Antes de poder indagar en los fundamentos físicos sobre los que se asiente la resonancia magnética, es necesario aclarar algunas propiedades de la materia y del magnetismo. Un campo magnético es el resultado del movimiento de una partícula cargada, ya que este movimiento genera un campo magnético a su alrededor, al que denotaremos como  $\vec{B}$ . Sus unidades se miden en Teslas en el Sistema Internacional, que equivale a  $\frac{\text{Weber}}{\text{m}^2}$ .

Este campo magnético  $\vec{B}$  también se le conoce como inducción magnética, y hay que diferenciarlo de la intensidad de campo magnético  $\vec{H}$ . Esta última la podemos entender como la fuerza magnetizadora aplicada a un cuerpo, mientras que  $\vec{B}$  es el campo magnético que se induce en un cuerpo al someterlo a  $\vec{H}$ .

Ambas son análogas, ya que podemos obtener una a partir de la otra mediante la siguiente relación:  $\vec{B}_{\text{vacío}} = \mu_0 \vec{H}$ , donde  $\mu_0$  es la permeabilidad magnética del vacío. Para otros medios donde las condiciones no sean las del vacío, la expresión queda de la siguiente manera:  $\vec{B} = \mu \mu_0 \vec{H}$ , donde  $\mu$  es la permeabilidad magnética relativa del medio, y es un parámetro adimensional. Podemos deducir de la expresión que si el campo magnético inducido en el interior del material es mayor del que existiría en el vacío,  $\mu$  ha de ser mayor que 1.

A partir de este punto, para poder introducir el concepto de espín de un átomo se va a hacer una aproximación desde el punto de vista de la mecánica clásica que, sin ser correcto, permitirá comprender mejor posteriormente los conceptos de mecánica cuántica. Desde un punto de vista clásico, una partícula que está orbitando respecto de un punto tiene un momento angular  $\vec{J}$ , con  $J = r \times mv$ , donde  $r$  es el radio de giro de la partícula,  $m$  la masa y  $v$  la velocidad. En el caso de que la partícula además estuviera cargada, tendría un momento magnético asociado, que es proporcional al momento angular de la forma:

$$\vec{\mu} = \gamma \vec{J} \quad (2.1)$$

donde  $\gamma$  es el cociente giromagnético, que dependerá de la masa y de la carga de la partícula en movimiento.

Haciendo un símil con la mecánica clásica, el vector momento angular describe las propiedades mecánicas de una partícula girando sobre sí misma, siendo este movimiento de giro sobre sí mismo el movimiento de *spinning*. Son estas propiedades las que vienen representadas por el vector de espín  $\vec{s}$  y, teniendo en cuenta este símil, se podría reescribir la fórmula de 2.1 de la forma:

$$\vec{\mu} = \gamma \vec{s} \quad (2.2)$$

siendo  $\vec{s}$  el momento angular de espín de la partícula.

De este modo, desde un punto de vista clásico la expresión de la energía de la partícula quedaría de la siguiente manera [7]:

$$E = -\gamma J B_0 \cos \theta \quad (2.3)$$

De forma que, como se puede extraer de 2.3, tanto  $J$  como  $\theta$  pueden tomar cualquier valor y por lo tanto la partícula puede tener cualquier valor de energía potencial. Sin embargo, como se sabe gracias a la mecánica cuántica, esto no es así, ya que los valores de energía de la partícula están discretizados. Para ello, debemos de reescribir la fórmula de la energía con esta nueva expresión:

$$E = -m\gamma\hbar B_0 \quad \text{con} \quad m = -j, -j+1, \dots, j-1, j \quad (2.4)$$

$\hbar$  es la constante reducida de Plank ( $\hbar = h/2\pi$ ). Como se ve, la energía viene determinada como un múltiplo del valor  $m$ , donde  $j$  es el valor del número cuántico de espín de la partícula. De esto extraemos que el valor mínimo de energía (cuanto) es  $\gamma\hbar B_0$ . Si nos centramos en el caso del protón, núcleo del átomo de hidrógeno, se tiene que el número cuántico de espín vale  $j = 1/2$ , de forma que hay dos estados energéticos posibles:

$$E_{\uparrow} = -\frac{1}{2}\gamma\hbar B_0 \quad (2.5a)$$

$$E_{\downarrow} = +\frac{1}{2}\gamma\hbar B_0 \quad (2.5b)$$

De esta forma, se puede comprobar que, un protón en presencia de un campo magnético solo puede estar en dos estados energéticos posibles, que llamaremos *espín up* ( $\uparrow$ ) y *espín down* ( $\downarrow$ ), siendo el estado  $E_{\uparrow}$  el estado de menor energía y por tanto el más estable. Para que un protón en este estado energético pase al estado  $E_{\downarrow}$  debe absorber un fotón con una energía mínima de  $\gamma\hbar B_0$ .

De igual manera, un partícula en el estado *down* pasará al estado *up* liberando energía en forma de un fotón de la misma energía.

Sabiendo que la energía de un fotón depende de su frecuencia de la forma siguiente:  $E = \hbar\omega_{\text{RF}}$  [7]. Si se iguala esta expresión con la del cuanto de energía necesario para cambiar de estado energético el protón ( $\gamma\hbar B_0$ ) se obtiene la frecuencia de resonancia de Larmor:

$$\omega_{\text{RF}} = \gamma B_0 \quad (2.6)$$

Como se puede extraer de 2.6, la frecuencia de resonancia de una partícula depende únicamente del campo  $B_0$  al que esté sometido y a su cociente giromagnético. En el caso de un campo magnético de 1 T, la frecuencia de Larmor de un protón es de 42.6 MHz. Dado que el hidrógeno es una de los componentes más abundantes del cuerpo en todos los tejidos, la IRM se centra en la detección de tejidos con hidrógeno.

### 2.1.3 Adquisición de imágenes de resonancia magnética

Hasta ahora, se ha analizado el comportamiento de los protones en presencia de un campo magnético desde un punto de vista microscópico, para poder comprenderlo en profundidad. Sin embargo, en imagen médica las imágenes no están representadas por todos los protones del tejido, sino por píxeles, o en este caso vóxeles. Estos vóxeles están conformados por miles de millones de protones, de forma que lo que realmente se está midiendo es el comportamiento macroscópico de todos ellos, en lugar de la contribución individual de cada uno.

Por ello, el vector de magnetización que se detecta es el valor de la suma neta de todos los protones que estarán en los dos estados energéticos posibles,  $E_{\uparrow}$  y  $E_{\downarrow}$ :

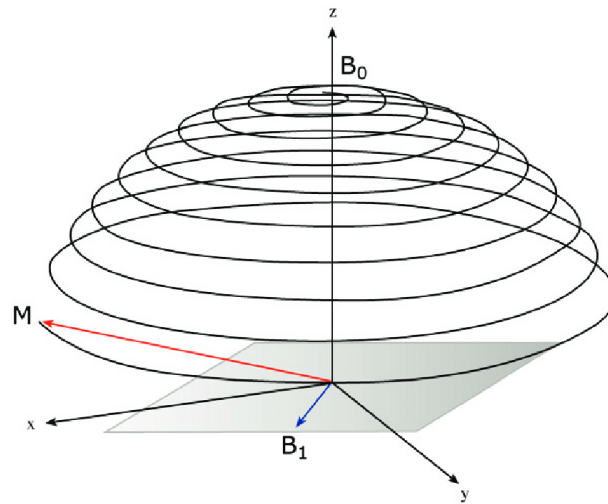
$$\vec{M}_0 = \sum_{i=1}^{n_s} \vec{\mu}_i \quad (2.7)$$

siendo  $n_s$  el numero total de espines que hay en el vóxel. Además, hay que tener en cuenta que, por estadística, la mayoría de ellos tenderán a estar en el estado energético más bajo, ya que de esta forma se encuentran más estables, cuando están con el espín *up*. De esta forma, lo que se adquiere es un vector neto que estadísticamente tenderá a estar hacia arriba cuando los protones estén sometidos a un campo magnético, y a ser nulo cuando no lo estén, debido a que en ese caso las orientaciones de los protones son aleatorias.

Además, de este razonamiento se puede extraer la importancia de contar en las IRM con un  $B_0$  fuerte ya que, a mayor intensidad de campo, mayor será el vector de magnetización neto, que es el que medirá nuestro dispositivo. De esta forma, con un tamaño de vóxel menor podemos obtener señales más fuertes, mejorando así la resolución espacial de la imagen.

Sin embargo, este estado con una magnetización neta *up* puede ser excitado con la aplicación de fotones con la energía adecuada,  $hf_p$ , siendo  $f_p$  la frecuencia de precesión, que como se ha dicho anteriormente, para el hidrógeno en un campo de 1 T es de 42.57 MHz. Estos fotones serían los pulsos de radiofrecuencia (RF). Con estos pulsos los protones pasarían de forma progresiva del estado menos energético *up* al estado más energético *down*, cambiando la orientación del vector neto de magnetización. Este cambio no se produce con un movimiento recto, sino en un movimiento de rotación sobre el eje del campo, conocido como movimiento nutación, como en la figura 2.1.

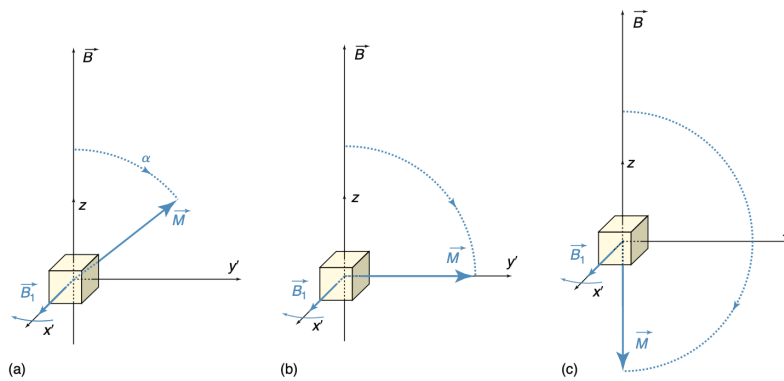
Se ha explicado el movimiento de nutación para conocer de manera correcta la relajación de los protones bajo impulsos de RF, sin embargo, en posteriores representaciones se representará la relajación con los ejes x e y transformados (que se denota de la forma  $x'$  e  $y'$ ). Este nuevo sistema



**Figura 2.1:** [8] Movimiento de nutación de un protón sometido a un impulso de RF

de referencia rotará en sintonía con el movimiento de nutación del protón, simplificando de esta forma la representación de la relajación.

Una vez aplicado el impulso de RF, progresivamente mas protones cambian del momento de espín *up* al *down*, de forma que en algún punto se iguala el número de protones en el estado de espín *up* y los protones en el estado de espín *down*. En ese momento la suma neta de los momentos de espín es cero, y se da lo que se conoce como un impulso de saturación, o de  $90^\circ$ . Sin embargo, si se sigue suministrando energía mediante impulsos de RF, habrá un momento en el que todos los protones pasen a estar en el estado de mayor energía y el vector neto de magnetización pase a tener un ángulo de  $180^\circ$ , lo que se denomina un pulso de inversión. Esto se representa en la figura 2.2.



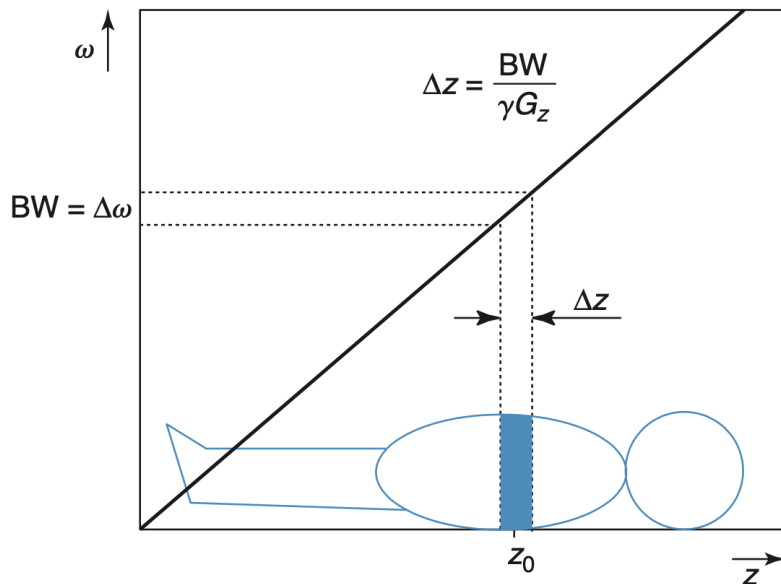
**Figura 2.2:** [7] (a) El vector de magnetización con un cierto ángulo  $\alpha$  sobre la dirección del campo magnético. (b) El vector de magnetización después de ser sometido a un pulso de saturación ( $90^\circ$ ). (c) El vector de magnetización después de ser sometido a un pulso de inversión ( $180^\circ$ ).

De esta forma ya se tiene un método para poder excitar los protones, para posteriormente se puedan obtener las imágenes. Sin embargo, utilizando los pulsos de radiofrecuencia excitemos todos los átomos de hidrógeno del cuerpo, ya que todos tienen la misma frecuencia de precesión.

Por lo tanto, es necesario ser selectivos a la hora de que parte del cuerpo es la que queremos excitar.

Para ello, hay que hallar métodos que permitan que la frecuencia de Larmor sea distinta en diferentes partes del cuerpo. Debido a la ecuación 2.6, la frecuencia de precesión de los protones depende únicamente de dos variables, del coeficiente giromagnético y del campo magnético al que estén sometidos. Dado que el coeficiente giromagnético es dependiente de la naturaleza de la partícula, este siempre será igual para todos los átomos de hidrógeno. Por ello, es necesario que el campo magnético no sea uniforme en todo el cuerpo, sino que este vaya variando. De esta idea es de donde nace la utilización de los gradientes de campo magnético.

El concepto se basa en aplicar un gradiente de campo magnético, haciendo que varíe uniformemente desde una punta del cuerpo del paciente hasta la otra, de manera que en ningún punto el campo magnético es exactamente igual.



**Figura 2.3:** [7] Aplicación de un gradiente de campo magnético al paciente. El corte que se pretende seleccionar es el correspondiente al tramo  $\Delta z$ , que tiene asociados un rango de pulsos de RF que le excitan,  $BW = \Delta\omega$

Como se ve en la figura 2.3, para poder seleccionar un corte del paciente únicamente es necesario excitar los átomos de hidrógeno que caigan dentro del ancho de banda determinado por la intensidad del campo magnético en esa posición. De esta forma, se consigue excitar solo la parte deseada.

Una vez se ha conseguido excitar las partículas de la zona deseada, cuando se deja de suministrar estos pulsos de RF, los átomos que habían pasado al estado más energético *down* vuelven progresivamente a tener el espín *up*. Este proceso, en el que el vector neto de magnetización vuelve a su estado inicial después de haber recibido un pulso de RF se denomina relajación. Y, aunque todas las partículas sean hidrógeno, no todas vuelven a su estado inicial de la misma forma, ya que se ven altamente influenciadas por su entorno histológico.



Este proceso de relajación es el que proporciona la información necesaria para la obtención de la imagen y como depende fuertemente del medio histoquímico en el que se encuentre, nos permitirá diferenciar los diferentes tipos de tejidos atendiendo a su naturaleza.

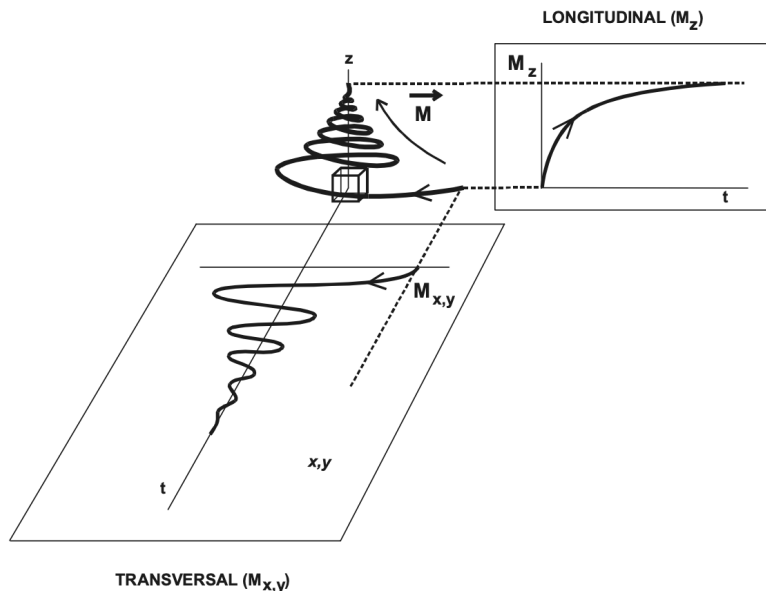
Para captar la señal debida a la relajación se disponen al rededor de la máquina de RMN un conjunto de bobinas que se ven excitadas por las variaciones del vector de magnetización, debida a la inducción electromagnética, induciendo una corriente eléctrica que proporcionará la información necesaria para la formación de la imagen.

Esta relajación puede ser además analizada en diferentes componentes, dependiendo de si la analizamos en la componente longitudinal y en la componente transversal.

La relajación en la componente longitudinal se conoce también como relajación *espín - red*, o se suele usar la denotación de relajación  $T_1$ . Esta es la correspondiente a la componente z del vector de magnetización [9]. Es la resultante del incremento desde  $M_0 \cos \alpha$  hasta  $M_0$  y se debe a la interacción de los espines con el entorno de macromoléculas [7].

Como contrapartida se encuentra la relajación en la componente transversal, llamada también relajación *espín - espín* o relajación en  $T_2$ . Se debe a los cambios del vector de magnetización en los ejes x e y [9]. En el caso de la relajación *espín - espín* hay que tener en cuenta que cuando se relaja el vector de magnetización lo hace con un movimiento de precesión, de forma que induce una corriente alterna en la bobina receptora. La expresión de la señal queda de la siguiente forma [7]:

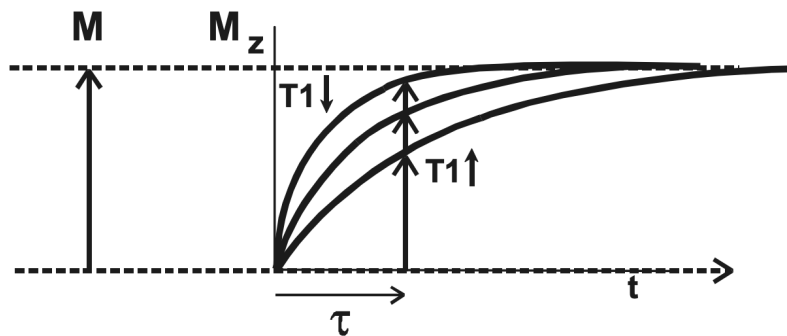
$$s(t) = M_0 e^{-t/T_2} \quad (2.8)$$



**Figura 2.4:** [10] Representación de la relajación en los ejes longitudinal y transversal

Como se ve en la figura 2.4, la evolución en el eje longitudinal tiene un comportamiento exponencial, mientras que en el eje transversal presenta la forma de una sinusoidal amortiguada.

Centrándonos en el comportamiento de la relajación en el eje longitudinal, lo que se obtiene en la representación de la relajación es el tiempo en el que tarda el vector de magnetización en volver a su longitud inicial. Esta longitud vendrá determinada por la densidad protónica, ya que a mayor número de protones en el vóxel mayor será la intensidad del vector de magnetización. Por esta razón, si la información que se utilizara para la formación de la imagen es la altura que llega a alcanzar la función exponencial obtendríamos la imagen de densidad protónica. Sin embargo, se perdería una información muy importante, que es la rapidez con la que llega a ese estado inicial, que viene determinado por lo fácil que sea de captar la energía que están perdiendo esos protones el medio. Otra modalidad de imagen que se puede obtener de esta relajación es cuando se toma la longitud del vector del magnetización después de haber pasado un corto periodo de tiempo ( $\tau$ ) desde que se ha dejado de aplicar el impulso de RF.



**Figura 2.5:** [10] Representación de tres curvas de relajación longitudinal, con la misma densidad protónica pero con diferente velocidad de relajación

De la figura 2.5 se extrae que el intervalo de tiempo  $\tau$  ha de ser corto, debido a que si pasa demasiado tiempo las curvas llegan asintóticamente al mismo punto y sería más difícil de diferenciarlas. Esta es la modalidad que se conoce como  $T_1$ , y permite diferenciar los tejidos por la velocidad de relajación en el eje longitudinal. Como la intensidad de la imagen depende del valor de  $\tau$ , es difícil comparar imágenes en modalidad  $T_1$  con diferente valor de  $\tau$ , sin embargo, se suele considerar como estándar de referencia aquel valor de  $\tau$  que hace que los tejidos con  $T_1$  cortos se representen hiperintensos mientras que los tejidos con  $T_1$  largos se representen hipointensos [10].

A parte de la información proporcionada por la relajación longitudinal, podemos extraer mucha información de la bioquímica del medio con al relajación transversal [10]. Como se ha dicho anteriormente, su forma es la de una sinusoidal amortiguada, cuya amplitud depende de lo coordinadas que estén las fases de los protones en ese momento.

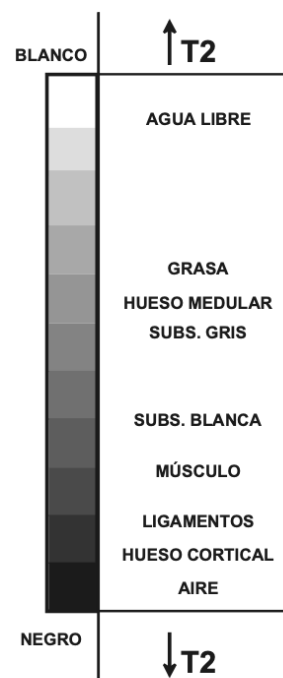
Un tiempo de amortiguación rápido indica una gran asincronismo en la relajación, y por tanto un valor de  $T_2$  bajo.

#### 2.1.4 Modalidades de imagen por resonancia magnética

Una vez explicados los métodos de adquisición de imágenes por resonancia magnética, queda exponer los diferentes tipos de modalidades de imagen que se utilizan para el diagnóstico clínico. Como abordar todos los tipos de modalidades de imagen escapa del propósito de este trabajo, se expondrá la forma de adquisición y características principales de las modalidades que se usarán posteriormente en los modelos de *deep learning* (DL): imagen potenciada en T2 (T2W), imagen potenciada en difusión (DWI) y coeficiente de difusión aparente (ADC).

##### *Imagen potenciada en T2 (T2W)*

Como se ha explicado anteriormente, las imágenes potenciadas en T2 son las obtenidas a partir del tiempo de relajación transversal posterior a un impulso de radiofrecuencia, por lo que no se volverá a explicar su forma de adquisición y se pasará a explicar las características diferenciales de este tipo de imágenes.



**Figura 2.6:** [10] Escala de intensidades de las imágenes potenciadas en T2

En las imágenes T2W, como se dijo anteriormente, se mide el sincronismo de la fase de las partículas en la relajación. En este tipo de imágenes los líquidos se ven hiperintensos, por lo que siempre que se tenga una imagen donde los líquidos se vean de color blanco se esta ante una imagen potenciada en T2. Además, se diferencia de las imágenes T2\* ya que estas últimas no tienen las inhomogeneidades de campo magnético corregidas ni los efectos de variaciones de susceptibilidades fijas [10].

Como se ve en la figura 2.6, después del aire el tejido con mayor intensidad es la grasa, seguido del hueso medular y la sustancia gris, que se diferencia de la sustancia blanca ya que esta última es menos intensa. Por último los músculos y ligamentos se ven más oscuros, hasta llegar al hueso cortical y al aire que son prácticamente negros.

Un ejemplo de imagen potenciada en T2 es la figura 2.7, donde se puede comprobar las intensidades de este tipo de imágenes. Por ejemplo, el músculo en esta imagen es de un color muy oscuro, a diferencia del hueso medular que hay en los lados de la imagen, o de la grasa. Si observamos las equinas superiores de la imagen, donde ya no se ha recogido el cuerpo del paciente sino el aire, estas zonas son totalmente oscuras, lo que está en consonancia con la figura 2.6

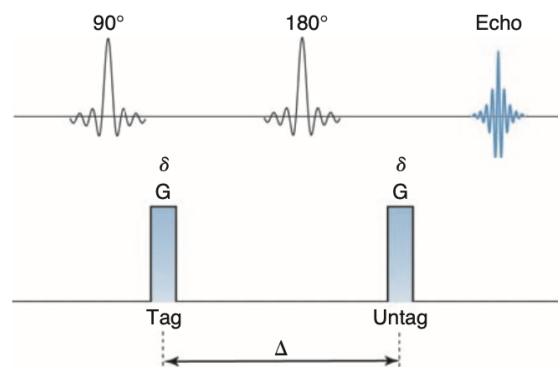


**Figura 2.7:** Imagen potenciada en T2 de la próstata, del *Challenge ProstateX-2*

### *Imágen potenciada en difusión (DWI)*

La segunda modalidad de IRM que se va a implementar en los modelos es la imagen potenciada en difusión (DWI). Esta técnica sirve para representar cuál es el grado de libertad de movimiento de las partículas en el tejido que se está visualizando. Además, la RMN es la única técnica que permite tener este tipo de información de un tejido *in vivo* [10].

Como resultado de la agitación térmica, las partículas de un líquido se mueven en un movimiento aleatorio, lo que normalmente se conoce como movimiento Browniano. Es este movimiento el que se consigue medir mediante la DWI. La visualización de la difusión es posible mediante una secuencia específica de tipo *espín-espín EPI*, conocida como *gradiente pulsado espín-eco* o también secuencia de Stejskal - Tanner, representada en la figura 2.8.



**Figura 2.8:** [7] Secuencia EPI para la obtención de imágenes potenciadas en difusión.

La señal correspondiente a una señal de difusión es la siguiente [7]:

$$S(b) = S_0 e^{-bD} \quad (2.9a)$$

$$b = \gamma^2 \delta^2 \left( \Delta - \frac{\delta}{3} \right) G^2 \quad (2.9b)$$

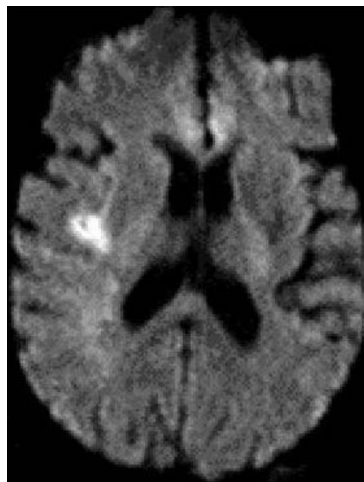
Donde  $S(b)$  es la señal de difusión obtenida,  $S_0$  representa la señal si no hubiera difusión,  $\delta$  representa la duración de los gradientes,  $\Delta$  es intervalo de tiempo que hay entre los dos gradientes,  $D$  es el coeficiente de difusión del líquido que se está midiendo y  $G$  es la amplitud del gradiente.

Es necesario tener en cuenta la forma del tejido en el que se produce la difusión. Si este tejido permite la difusión sin restricción en ninguna de sus direcciones, es decir, que es igual de probable que una partícula difunda en cualquier dirección, se dice que hay una difusión isotrópica. Sin embargo, existen tejidos como pueden ser las fibras nerviosas, donde hay claramente una dirección predominante, en este caso, la probabilidad de que las partículas difundan en la dirección de la fibra es mucho mayor de que lo hagan perpendicularmente, esto se conoce como difusión anisotrópica.

Cuando se mide la difusión, esta medición se realiza únicamente en un eje, por lo que para conocer la difusión en los demás ejes es necesario volver a calcular la imagen variando la dirección de los gradientes. Una vez calculadas las difusiones en todos los ejes se puede calcular la imagen potenciada en difusión isotrópica:

$$ISO = \sqrt[3]{DWIx + DWIy + DWIz} \quad (2.10)$$

De esta forma, ya se tiene una imagen en la que se puede apreciar la difusión en todas las direcciones del espacio, siendo las zonas hipointensas las zonas donde mayor difusión hay, mientras que las zonas hiperintensas corresponden a zonas donde existen restricciones a la difusión. La figura 2.9 es un ejemplo de imagen potenciada en difusión.



**Figura 2.9:** [10] Imagen potenciada en difusión de una accidente vascular agudo, donde la zona hiperintensa del lóbulo temporal derecho demuestra una oclusión que no está permitiendo el paso de la sangre.

### *Coefficiente de difusión aparente (ADC)*

Finalmente, la última modalidad que se va a implementar es el coeficiente de difusión aparente (ADC). Estas imágenes se adquieren a partir de las imágenes potenciadas en difusión explicadas anteriormente. Las imágenes de ADC sirven para dar información cuantitativa de la difusión en una dirección determinada ( $mm^2/s$ ).

Para calcular el ADC es necesario obtener una imagen potenciada en T2, y otra imagen potenciada en difusión. Una vez se tienen estas dos imágenes únicamente es necesario aplicar la siguiente fórmula:

$$ADC = \frac{\ln\left(\frac{DWI}{T_2}\right)}{-b} \quad (2.11)$$

Donde DWI es la imagen potenciada en difusión,  $T_2$  la misma imagen potenciada en  $T_2$ , y  $b$  es la expresión de la fórmula de 2.9

El valor del mapa del coeficiente de difusión aparente se calcula para cada eje, por lo que es necesaria la construcción de un tensor de difusión con todos los mapas de ADC posibles en cada una de las direcciones, y como estamos midiendo en un volumen de tres dimensiones hay un total de 9 ejes de medida distintos:

$$\begin{pmatrix} ADC_{xx} & ADC_{xy} & ADC_{xz} \\ ADC_{yx} & ADC_{yy} & ADC_{yz} \\ ADC_{zx} & ADC_{zy} & ADC_{zz} \end{pmatrix} \quad (2.12)$$

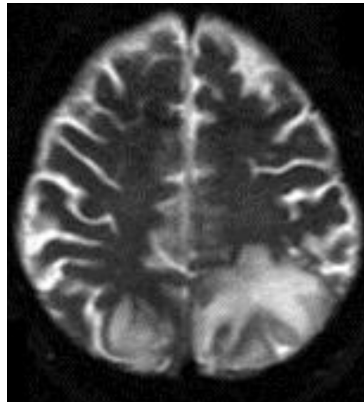
Sin embargo, como este tensor es simétrico, realmente solo son necesarias seis imágenes para su construcción. Una vez se tiene el tensor construido, se debe descomponer en sus autovalores y autovectores. Por último, con los autovalores y autovectores calculados se puede calcular el ADC promedio en todas las direcciones:

$$ADC = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (2.13)$$

Donde ADC es el mapa promedio, y  $\lambda_1$ ,  $\lambda_2$  y  $\lambda_3$  son los tres autovectores asociados a los tres autovalores principales del tensor, que representan sus ejes principales.

En las imágenes de ADC las zonas hiperintensas indican zonas de alta difusión en esa dirección, mientras que las zonas de baja intensidad indican baja difusión. Un ejemplo de imagen ADC es la figura 2.10.

Para representar mejor las diferencias entre las diferentes tipos de imagen de resonancia magnética, en la figura 2.11 se muestra el mismo corte de un paciente pero en diferentes modalidades.



**Figura 2.10:** [11] Imágen ADC del cerebro después de sufrir un síndrome por descompresión

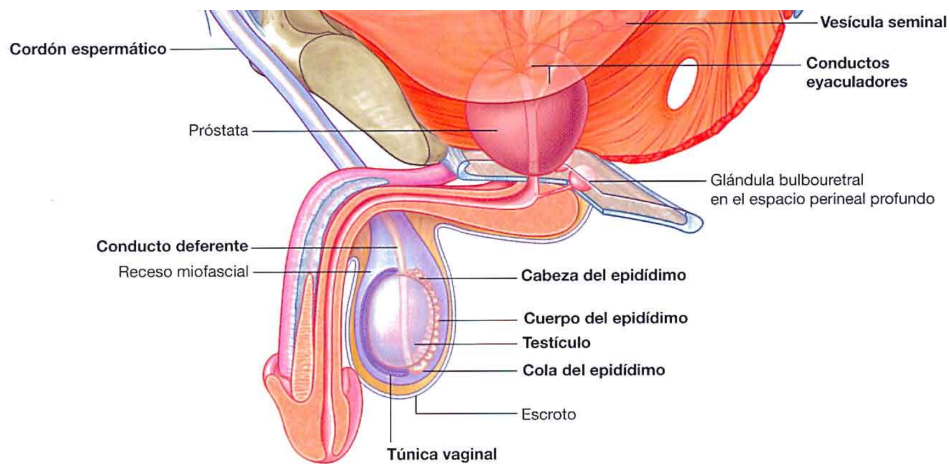


**Figura 2.11:** Comparación del mismo corte en diferentes modalidades de resonancia magnética, de izquierda a derecha: T2, DWI y ADC.

## 2.2 Anatomía de la próstata

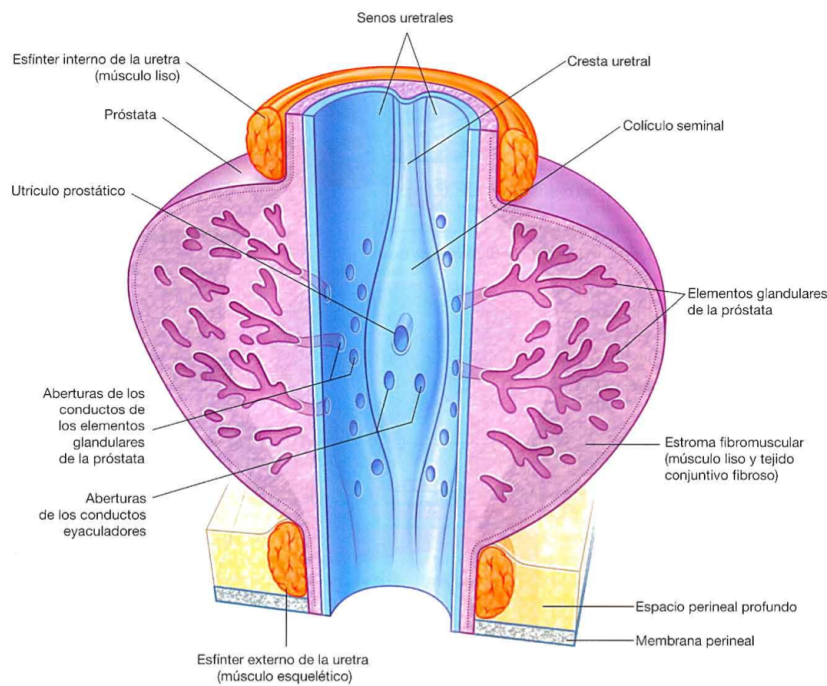
Como se ha dicho en la introducción, este trabajo se basa en el desarrollo de un modelo de gradación de cáncer de próstata. Para ello es necesario conocer la anatomía de la misma, ya que resulta de relevancia para comprender los diferentes métodos de diagnóstico y las dificultades que se afrontan. Con este fin se han usado los conocimientos adquiridos en la asignatura de anatomía, además de los atlas anatómicos citados.

La próstata es una estructura del aparato reproductor masculino. Tiene forma de cono redondeado invertido con la base de mayor tamaño. Rodea la uretra en la cavidad pélvica, situado inferior a la vejiga, posterior a la sínfisis del pubis y anterior al recto. Continúa por encima del cuello de la vejiga y tiene un vértice más estrecho apoyado sobre el suelo pélvico. [12]. Esto se puede ver en la figura 2.12.



**Figura 2.12:** [12] Vista general de la posición de la próstata en el aparato reproductor masculino

La próstata está formada por alrededor de 30-40 glándulas complejas individuales, que crecen desde el epitelio de la uretra hasta la pared que la rodea. Se puede ver como que este conjunto de glándulas aumentan la pared de la uretra hasta dar forma a lo que se conoce como próstata, sin embargo, cada una de estas glándulas es individual y desembocan en los senos prostáticos independientemente de las demás, como se ve en la figura 2.13.



**Figura 2.13:** [12] Porción prostática de la uretra masculina

La función principal de la próstata es la de secretar un fluido rico en calcio, iones citrato y fosfato, enzimas de coagulación y profibrinolisisina. Este líquido es ligeramente alcalino, necesario para la supervivencia de los espermatozoides y puedan realizar la fecundación, debido a que las secreciones de la vagina son ácidas (en un rango de 3.5 a 4.0 pH), y el rango de movilidad óptimo de los espermatozoides se encuentra entre 6.0 y 6.5 pH. De esta forma, el líquido prostático sirve para



neutralizar la acidez del medio y aumentar la supervivencia y fertilidad de los espermatozoides [13].

### ***2.2.1 Cáncer de próstata***

Aunque cáncer es un término muy general para referirnos a un conjunto de enfermedades, todas se basan en la misma premisa, una replicación celular muy rápida y descontrolada, que puede estar localizada en un órgano específico o extenderse a otras zonas del cuerpo, lo que se conoce como metástasis.

Esta división descontrolada produce en ocasiones la pérdida de funcionalidad del órgano donde se está produciendo el tumor, que es la razón por la que el cáncer causa la muerte del paciente. Las razones por las que se produce esta división descontrolada de las células son demasiado variadas como para poder abarcarlas todas en este trabajo, pero en general suelen estar producidas por mutaciones en genes que se encargan de controlar el ciclo celular, que es el que debe decidir cuándo las células deben dividirse, o provocar la muerte de la célula cuando esta está dañada, proceso denominado apoptosis.

Un ejemplo de las mutaciones más frecuentes en la formación de tumores es la mutación del gen TP53, que codifica el factor de transcripción p53. Entre sus muchas funciones, una es la de inducir la apoptosis de las células cuando detecta que se ha producido una lesión en ella que es imposible de reparar y que puede producir un mal funcionamiento de la misma [14], por eso también se le conoce como gen supresor tumoral. Estas mutaciones se producen por daños en el material genético de la célula, que cambian la secuencia de nucleótidos, lo que provoca que la transcripción de las proteínas no sea la correcta. Existen diferentes factores que pueden provocar variaciones en el ácido desoxirribonucleico (ADN), pero podemos agruparlos en tres grupos diferenciados: los factores físicos, debidos a la exposición a radiación ionizante; los factores químicos, debido a exposición a elementos carcinógenos como son el formaldehído o el bromuro de etidio; y por último los biológicos, debidos a infecciones por virus o bacterias, como es el caso del virus del papiloma humano.

En cuanto al cáncer de próstata, existen diferentes factores de riesgo que predisponen al paciente de padecerlo. Para ello, se puede diferenciar entre dos grupos, los factores endógenos y los factores exógenos. Los factores endógenos son aquellos que son inherente al paciente, mientras que los factores exógenos se deben principalmente a la interacción del paciente con el entorno. Algunos ejemplos de esto son [15]:

- Factores endógenos

1. Historial familiar, ya que el cáncer de próstata se da más en personas cuyos familiares también lo han padecido.
2. Hormonales, debido a que altos niveles de andrógenos se asocian a mayor probabilidad de cáncer de próstata.
3. La edad, siendo el grupo de personas mayores donde mayor incidencia de cáncer de próstata hay.

- Factores exógenos
  1. La dieta, que es un factor crucial, ya que un alto consumo de grasas poliinsaturadas aumenta el riesgo de padecer cáncer de próstata
  2. Agentes ambientales, especialmente merecen atención los químicos interruptores endocrinos (EDC), que se han relacionado con un aumento de la incidencia del cáncer de próstata

La próstata permanece con un tamaño pequeño durante toda la niñez hasta el comienzo de la pubertad, que empieza a desarrollarse hasta alcanzar su máximo tamaño sobre los 20 años debido al efecto de la testosterona. A partir de ese momento su tamaño se estabiliza durante toda la época adulta hasta los 50 años. En ese momento en algunos hombres la próstata decrece de tamaño por la disminución de los niveles de testosterona. Sin embargo, en otros hombres de avanzada edad lo que ocurre es que la próstata aumenta de tamaño, conocido como fibroadenoma benigno prostático, o también hiperplasia benigna de próstata (HBP) [13].

La HBP afecta a más del 60 % de la población de más de 60 años, y este porcentaje aumenta conforme aumenta la edad. Entre todos los síntomas de la HBP los más comunes son [16]:

- Micción frecuente
- Sensación de vaciado incompleto de la vejiga
- Interrupción en el flujo de la micción
- Flujo débil

Por último, en algunos casos en hombres de esa edad en lugar de tener una HBP lo que acaban desarrollando es un cáncer maligno. Una vez comienza el desarrollo tumoral en la próstata, el crecimiento de las células cancerígenas se ve incrementado debido a la acción de la testosterona [13].

Entre los cánceres de próstata, el tipo de cáncer más común es el adenocarcinoma de tipo acinar. Es un carcinoma invasivo de la próstata con células neoplásicas epiteliales con diferenciación secretora, con una gran variedad de patrones histológicos. Además, normalmente están ausentes las células basales. A parte del adenocarcinoma de tipo acinar común, hay otras variantes de adenocarcinoma de próstata: atrófico, pseudohiperplásico, microcístico, de glándula espumosa, mucinoso, células en anillo de sello, células gigantes pleomórficas y variante sarcomatoide [17].

La sintomatología que causa para el paciente es muy similar a la causada por la HBP, por lo que sin más pruebas es difícil de diferenciarlas. Con el fin de poder diferenciarla de una HBP y poder diagnosticarla correctamente existen diferentes pruebas [17]:

1. Examen rectal digital, donde el médico palpa la próstata a través del recto para detectar cambios en la rigidez del tejido de la próstata o zona anormales. Sin embargo, no es específico para el cáncer de próstata y puede dejar de diagnosticar entre un 25 % y un 50 % de casos que podrían ser detectados mediante otras pruebas.
2. Análisis de antígeno prostático específico (PSA) (de sus siglas en inglés *prostate-specific antigen*). Esta prueba es una de las que más ampliamente extendidas se encuentra en el diagnóstico del cáncer de próstata, ya que altos valores de PSA están estrechamente ligados

a padecer cáncer de próstata. Sin embargo, algunos pacientes pueden estar en fases poco avanzadas del cáncer de próstata y es posible que no se detecte mediante esta prueba.

3. Niveles en orina de PCA3 (por las siglas en inglés de *prostate cancer gene 3*). Es un biomarcador no codificante de Ácido ribonucleico mensajero (ARNm) que se detecta en la orina y puede ser medido mediante amplificación por reacción en cadena de polimerasa (PCR). Se encuentra altamente sobreexpresado en casos de adenocarcinoma de próstata, a diferencia de casos de HBP.
4. Otras variaciones genéticas como son mutaciones en la línea germinal de genes como BRCA1 y BRCA2 entre otros.
5. Biopsia de tejido. Esta prueba sirve para confirmar definitivamente la presencia de cáncer en la próstata aunque, debido a que es una técnica invasiva se suele dejar como prueba para confirmar el diagnóstico en el caso de que otra de las pruebas anteriormente expuestas diese positivo.

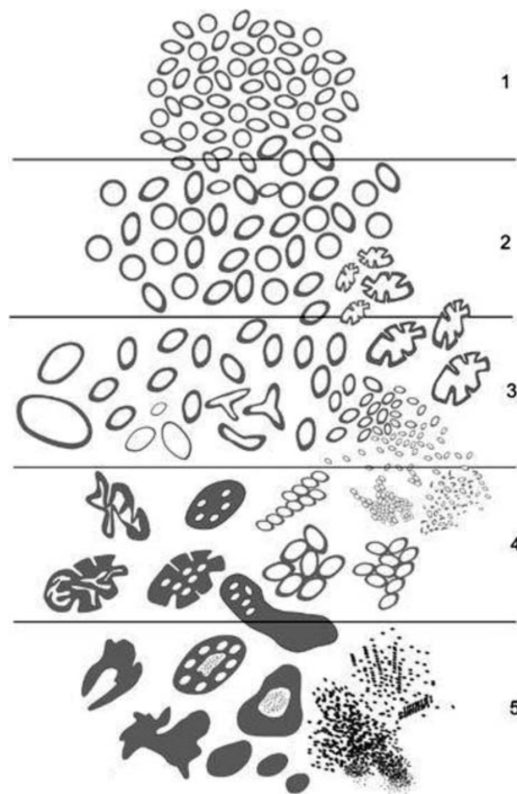
### 2.2.2 *Escala de Gleason*

Como se ha dicho anteriormente, la prueba que se usa para confirmar la presencia de cáncer de próstata es la biopsia. Esta permite el estudio de la arquitectura histológica del tejido. Para la extracción de la muestra se realiza una punción en la próstata con una aguja. Posteriormente ese tejido es fijado con formaldehído y se procesa con parafina para que pueda ser laminada al microtomo. Finalmente la muestra se tinte con hematoxilina-eosina para la visualización de las estructuras al microscopio [18].

Los médicos, para diagnosticar el tejido como cancerígeno, se centran en el estudio del tejido tanto a bajos aumentos (20x - 100x) como a altos aumentos (200x). A bajos aumentos la estructura del tejido no es un continuo sino que suele estar interrumpida con un patrón infiltrante que crece alrededor de las glándulas benignas. A altos aumentos los objetivos para el diagnóstico son la morfología de las células, donde las células cancerígenas son de mayor tamaño, poseen núcleos más prominentes, y el citoplasma de las células es más basofílico que el del tejido sano [17]. Existen otras características histológicas, pero las anteriores son las de mayor relevancia para el diagnóstico.

El sistema que está más extendido en el diagnóstico del cáncer de próstata es la escala de Gleason. Es un sistema desarrollado en los años ochenta, que está basado únicamente en la estructura del tejido de carcinomas que no han sido sujetos de radioterapias, ya que entonces su estructura no sería la misma. En este sistema las características citológicas como el tamaño de los núcleos o polimorfismos nucleares no son relevantes para el diagnóstico.

La escala de Gleason se basa en la formación de distintos patrones en el tejido. En el se han diferenciado cinco patrones distintos, y la gradación total será la suma de los dos patrones más comunes en el tejido. De esta forma, un tejido donde los patrones más comunes sean el 4 y el 3 tendría una gradación en la escala de Gleason de 7 [17]. Estos patrones están representados en la figura 2.14.



**Figura 2.14:** [17] Diagrama esquemático con los patrones de los diferentes grados de adenocarcinoma de próstata

Actualmente, en la escala de Gleason los patrones del 1 al 3 corresponden a lo que se denomina glándulas bien formadas que consisten en agrupaciones de células. En el grupo 3 el carcinoma se infiltra entre las glándulas sanas, a diferencia de los patrones 1 y 2, donde las células forman agregados. En cuanto a los patrones 4 y 5 las estructuras pueden ser muy variadas. Dentro del grupo 4 se diferencian cuatro formas [17]:

- Glándulas pobremente formadas
- Glándulas fusionadas
- Glándulas cribriformes
- Estructuras glomeruloides

Por último, en el grupo cinco las células se encuentran muy poco diferenciadas, con nula similitud con las células sanas. Los patrones que forman pueden aparecer de las siguientes formas [17]:

- Células individuales
- Capas de células prácticamente planas sin evidencia de formación glandular
- Comedocarcinoma

## 2.3 Deep learning y redes neuronales

### 2.3.1 Introducción a las redes neuronales

El desarrollo de los métodos de *deep learning* y las redes neuronales se nutre de los artículos citados y de los conocimientos adquiridos en la asignatura de imágenes biomédicas.

La inteligencia artificial (IA) como término es un concepto amplio y aportar una sola definición es complicado debido a la amplia variedad de campos donde se puede aplicar la IA. Sin embargo, una posible definición sería que es un área de estudio de la informática, donde los ordenadores son capaces de desarrollar un comportamiento que sea similar al pensamiento humano y que le permita sobre una tarea aprender, razonar y corregirse a si misma [19].

La IA es un campo que engloba muchos términos, entre ellos el de *machine learning* (ML). Se puede definir un algoritmo de ML como un proceso computacional que recibe la entrada de datos, y que es capaz de dar una respuesta a esos datos sin haber sido programado literalmente para producir una salida en particular [20]. Los algoritmos de ML convencionalmente trabajan con datos que han sido previamente manipulados por un ser humano, donde se extraen un conjunto de características que pueden ser de utilidad para el trabajo del algoritmo de ML. De esta forma, aunque imiten el comportamiento humano, no son completamente autónomos, ya que necesitan de este preprocesado de los datos para obtener resultados satisfactorios.

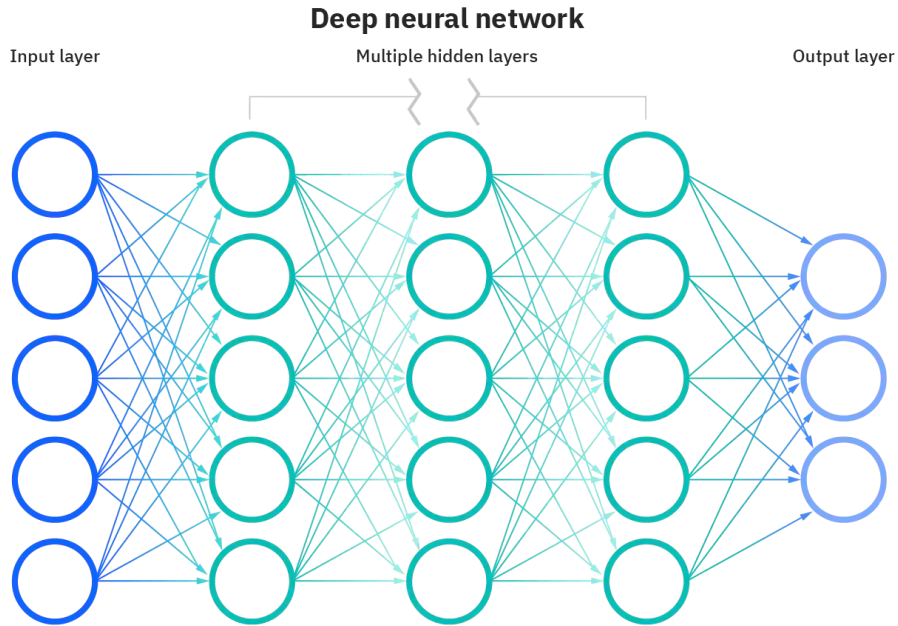
Dentro del ML se puede englobar un conjunto de técnicas de aprendizaje profundo, o también conocidas como de *deep learning* (DL). El DL se puede clasificar como un subconjunto del ML. Las técnicas de DL tratan de simular el comportamiento del cerebro humano, tratando de aprender el conocimiento en diferentes niveles de representación y abstracción, extrayendo patrones y jerarquías de los datos, en lugar de aplicar simplemente algoritmos específicos [21].

Como se ha dicho, en el campo del DL los modelos puede aprender de los datos, y este aprendizaje puede ser de tres maneras distintas:

1. Supervisado: donde junto a los datos se tiene la etiqueta con la salida correcta
2. Débilmente supervisado: donde parte de la información de las etiquetas no esta presente
3. No supervisado: cuando al algoritmo no se le proporciona ninguna etiqueta de los datos

Es dentro del DL donde se enmarca el uso de la redes neuronales profundas. Estas son redes neuronales artificiales con un gran número de capas que le otorgan esa profundidad, y permiten asemejarse más al comportamiento humano. Están inspiradas en la transmisión de los impulsos eléctricos del cerebro, donde una neurona conecta con muchas otras compartiendo la información a través de estos impulsos. Una posible esquematización de su estructura es la de la figura 2.15:

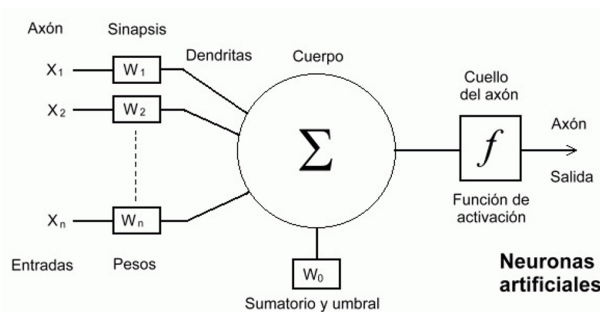
Son estas capas profundas las que permiten a las redes neuronales (NN) comportarse como lo haría el cerebro humano, simulando los impulsos nerviosos y las conexiones entre ellas. Están compuestas por una capa de entrada en la que se introducen los datos, la capa de salida donde la NN realiza su predicción y un conjunto de capas ocultas. Y aunque el uso de redes neuronales se haya popularizado en los últimos años no son algo nuevo. Ya en 1960 Rosenblatt propuso el uso del perceptrón simple para la clasificación lineal [23]. Sin embargo, su uso no se ha popularizado hasta recientemente, debido principalmente a un problema tecnológico, la tremenda carga de cómputo



**Figura 2.15:** [22] Estructura de una red neuronal profunda, con entradas (*inputs*), salidas (*outputs*) y las capas ocultas (*hidden layers*)

que suponía calcular estas NN, pero con el desarrollo de programas para la paralelización del cálculo y la popularización de las tarjetas gráficas, el tiempo de entrenamiento de NN se acortó enormemente.

Como se ha dicho anteriormente, tras la idea de la creación de redes neuronales estaba la intención de asemejarse al funcionamiento del cerebro humano. Para ello, sus unidades principales son muy similares a las neuronas del cerebro, donde cada unidad o nodo tiene muchas entradas que le conecta al resto de neuronas, luego combina estas entradas para producir una salida que se comunicará a otras neuronas. Estas entradas serían como las dendritas y la salida recuerda al axón de la neurona. Un esquema de la neurona de una NN es el de la figura 2.16:



**Figura 2.16:** [24] Representación esquemática del modelo computacional de una neurona

Las diferentes  $x_n$  son todas las entradas que le llegan a la neurona, siendo  $w_n$  los pesos por los que la neurona pondera cada entrada. Posteriormente está  $w_0$ , que se le conoce como valor de bias. Este es un valor que se suma a la ponderación de las entradas por los pesos. Finalmente, la

predicción de la neurona pasa por una función de activación antes de convertirse en la salida final de la neurona. Se podría esquematizar el funcionamiento de la neurona de la siguiente manera:

$$z = w_0 + \sum_n^N x_n w_n \quad (2.14a)$$

$$Z = f(z) \quad (2.14b)$$

Donde  $z$  es la predicción de la neurona y  $f(z)$  es la función de activación, siendo  $Z$  la salida final de la neurona.

### 2.3.2 Algoritmo de retropropagación

En el espíritu de la concepción de las NN está el intento de parecerse a la forma de funcionar del razonamiento humano, por lo que las NN son capaces de aprender de los datos que se le proporcionan. Este aprendizaje lo hacen mediante la modificación de los valores de los pesos y del valor de bias de forma que minimicen una función de error. Para ello Rumelhart formuló en 1986 [25] el algoritmo de retropropagación (o *back-propagation* en inglés).

Este algoritmo está compuesto de dos pasos. La primera etapa, también conocida como *feed forward* realiza una primera predicción con los pesos iniciales de la red. Una vez realizadas estas predicciones calcula el error mediante una función de coste. El objetivo final del entrenamiento de la red será minimizar el error que calcule la función de coste. Durante esta primera fase del algoritmo de retropropagación la red recibe como entrada datos de ejemplo con los que realiza esas predicciones.

Tras llevar a cabo el paso de *forward propagation* y haber comparado la predicción del modelo con las etiquetas, se pasa al segundo paso del algoritmo, la retropropagación. Para la implementación de la retropropagación existen diferentes optimizadores para obtener el resultado, a modo de ejemplo se explicará el funcionamiento del descenso de gradiente estocástico (SGD por sus siglas en inglés), uno de los más sencillos pero más ampliamente extendidos.

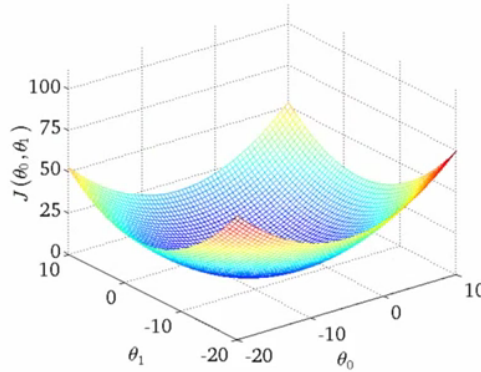
En este algoritmo los pesos se recalculan siguiendo la siguiente expresión [26]:

$$w = w - \eta \cdot \nabla_w J(w; x^{(i)}; y^{(i)}) \quad (2.15)$$

Donde  $w$  es el valor de los pesos,  $\eta$  es la tasa de aprendizaje y  $J$  es la función de coste con la que se calcula el error. De esta forma, a cada aplicación del algoritmo de retropropagación se actualizan los valores de los pesos de forma que se ajusten mejor a los datos de entrenamiento, consiguiendo así que la red aprenda a realizar mejores predicciones.

De la expresión de 2.15 se puede extraer que el proceso de aprendizaje de una NN se basa en la minimización de una función de coste, variando los valores de los pesos y del bias de forma que minimicen esta función. El valor de la tasa de aprendizaje es el que marcará la velocidad a la que se llega a este mínimo local de la función de coste. Un valor muy alto hará que los valores de los pesos se actualicen cambiando de valor muy rápido, lo que puede favorecer llegar antes a ese mínimo local, sin embargo, mediante valores demasiado altos puede suceder que nunca se llegue a un mínimo local, ya que las actualizaciones serían demasiado bruscas. Por el contrario, si el

valor de la tasa de aprendizaje fuese demasiado pequeño, los cambios en los valores de los pesos no serían suficiente para llegar a un mínimo local en un tiempo razonable. En la figura 2.17 se representa el plano de la función de coste con un mínimo local.



**Figura 2.17:** [27] Representación en un plano de la función de coste

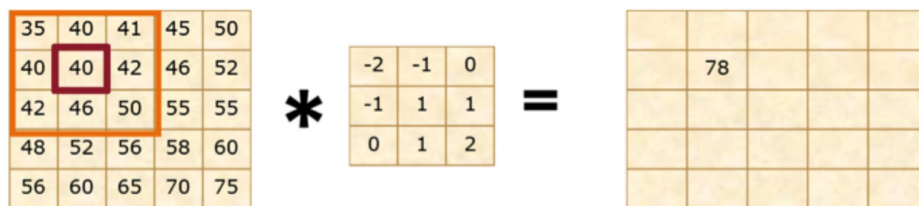
### 2.3.3 Fundamentos de las redes neuronales convolucionales

Dentro de la diversidad de arquitecturas de las NN, una destaca en su uso en el análisis de imágenes, las redes neuronales convolucionales (CNNs). La estructura de estas redes se basa en la forma en la que el cortex del cerebro procesa y reconoce las imágenes.

Una CNNs debe estar compuesta como mínimo por una capa convolucional. Para comprender mejor el funcionamiento de las CNNs se explicará a continuación el proceso de convolución de dos señales, que es en lo que se fundamenta este tipo de redes y se puede expresar de la siguiente forma:

$$x[n] \otimes h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n - k] \quad (2.16)$$

Donde  $x[n]$  es la señal que se pretende convolucionar y  $h[n]$  corresponde al kernel por el que se convoluciona la señal. Un ejemplo gráfico para comprender el funcionamiento de la convolución en matrices es el de la figura 2.18, donde se centra el kernel sobre un punto de la matriz asemejando el funcionamiento de una ventana deslizante. Posteriormente multiplica cada valor del kernel con el valor correspondiente en la matriz, y la suma de todos los valores será el resultado, que corresponderá al mismo punto sobre el que se ha centrado el kernel ( $35 \cdot -2 + 40 \cdot -1 + 0 \cdot 41 + -1 \cdot 40 + 40 \cdot 1 + 42 \cdot 1 + 42 \cdot 0 + 46 \cdot 1 + 50 \cdot 2 = 78$ ).



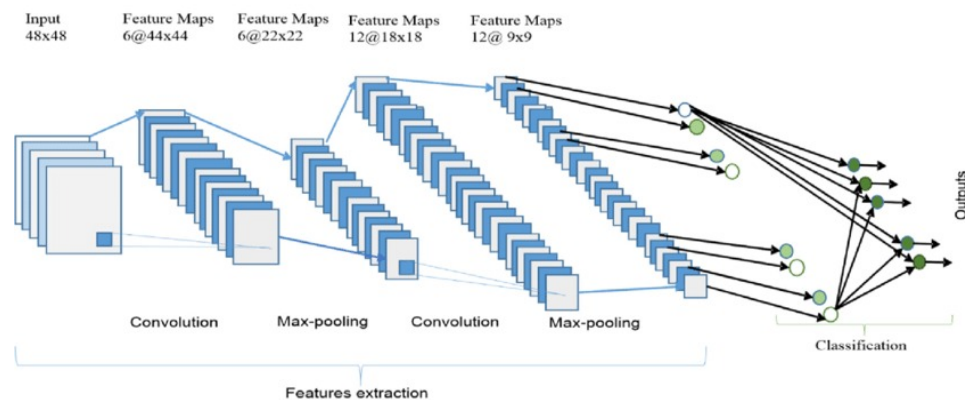
**Figura 2.18:** [28] Esquemización gráfica del funcionamiento de convolución en matrices



Como se ha dicho, el funcionamiento de la convolución será el de una ventana deslizante, donde se vayan convolucionando parches de la matriz original del tamaño del kernel. En este punto hay varias características del kernel a tener en cuenta: el tamaño, el número de puntos que se desplaza cada vez que se desliza por la imagen (lo que se denomina *stride*) y el funcionamiento del mismo cuando llega al borde y como rellena los píxeles que faltan, lo que se conoce como *padding*. Una posibilidad sería que cuando llegue al borde las zonas donde no haya píxeles las rellene como ceros, lo que se conoce como *zero-padding*.

Las CNNs se pueden entender como un extractor de características, donde las diferentes capas convolucionales buscan características distintas en las imágenes. Para entender el funcionamiento de una capa convolucional se va a dividir en las partes que la componen, que se representan en la figura 2.19:

1. Convolución
2. Función de activación
3. *Pooling*



**Figura 2.19:** [29] Arquitectura de una red neuronal convolucional (CNNs)

La parte de la convolución funciona como se ha explicado anteriormente, donde en los bloques convolucionales se extraen patrones de la imagen a diferentes niveles de complejidad, por eso las capas convolucionales se les llama extractores de características. Son los valores del kernel los que se actualizan en cada paso del algoritmo de retropropagación para minimizar la función de coste.

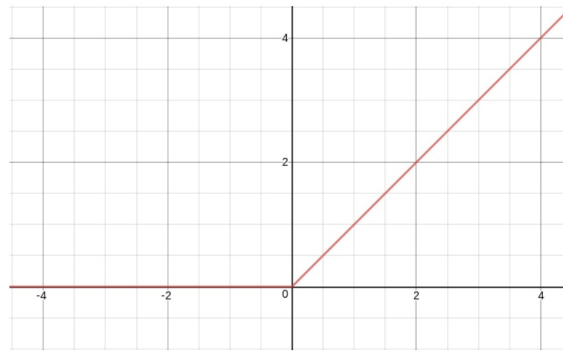
Un factor que hay que tener en cuenta es que, aunque para simplificar su representación se han representado los kernels de forma bidimensional, en realidad en CNNs con imágenes los kernels suelen ser tridimensionales. La profundidad del kernel debe ser igual a la de la entrada que recibe. Por ejemplo, en el caso de que la entrada fuera una imagen de  $256 \times 256 \times 3$ , un posible kernel sería  $5 \times 5 \times 3$ , de forma que el resultado de la convolución sería de  $256 \times 256 \times 1$ . Si en esa capa se usaran 10 kernels distintos, se obtendrían 10 mapas de características distintos, que se amontonarían uno después del otro, formando un volumen de  $256 \times 256 \times 10$ . De esta forma, conforme se avanza en la estructura de la red aumenta la profundidad de las salidas de las capas.

Cabe destacar que, normalmente, tras cada bloque convolucional se aplica una función de activación, que permitirá obtener relaciones no lineales en los patrones obtenidos. Una de las funciones

de activación más comunes en capas convolucionales es la *rectified lineal unit (ReLU)*. Esta función se puede expresar de la siguiente manera:

$$f(z) = \max(0, z) \quad (2.17)$$

Y su comportamiento se puede representar gráficamente como en la figura 2.20:



**Figura 2.20:** [30] Función ReLU

De esta forma, se consigue que solo se activen las neuronas cuando su salida sea positiva. Sin embargo, existen multitud de funciones de activación distintas según el propósito que se les quiera dar. Una de las más utilizadas para la capa de salida de una red es la función *softmax*, muy utilizada para la realización de predicciones. La función se puede expresar de la siguiente manera:

$$f(z) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad \text{para } j = 1, \dots, K \quad (2.18)$$

Entre sus propiedades se encuentra que es diferenciable y además proporciona una probabilidad de la predicción que realiza la red y se puede usar para clasificadores que no sean binarios, siendo  $K$  las diferentes etiquetas en las que la red puede clasificar los datos.

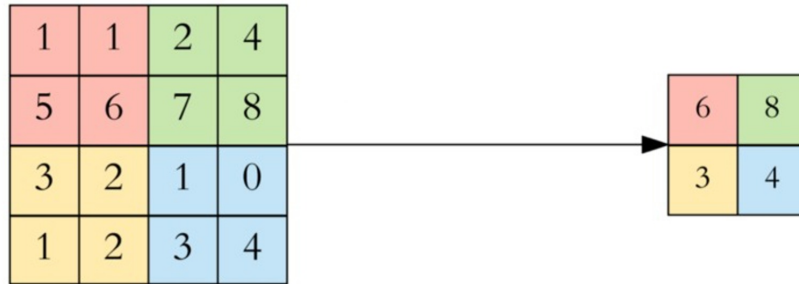
A continuación se pasará a explicar algunos tipos de capas muy comunes que serán útiles posteriormente para la explicación de los diferentes modelos implementados.

### **Capa *max pooling***

Con el objetivo de reducir la dimensionalidad de las siguientes capas se suele implementar un método de *pooling*, donde de alguna manera se reduce las dimensiones de las salidas de una capa antes de entrar en la siguiente. Por ejemplo, en la figura 2.19 se puede comprobar como a la salida de la primera capa convolucional se tiene unas dimensiones de  $44 \times 44 \times 6$ , y la entrada de la siguiente capa es de tamaño  $22 \times 22 \times 6$ . Hay diferentes formas de conseguir este resultado, pero el más común es la aplicación de un proceso denominado *max pooling*.

En el *max pooling* el proceso que se sigue para reducir las dimensiones de las capas es el de desplazar una ventana deslizante por la imagen del filtro, y de todos los píxeles que ocupe esa ventana solo quedarse con el de mayor valor. Por ejemplo, con un tamaño de ventana de  $2 \times 2$

y un *stride* de 2, las dimensiones de la imagen se reducen a la mitad, como se puede ver en la figura 2.21.

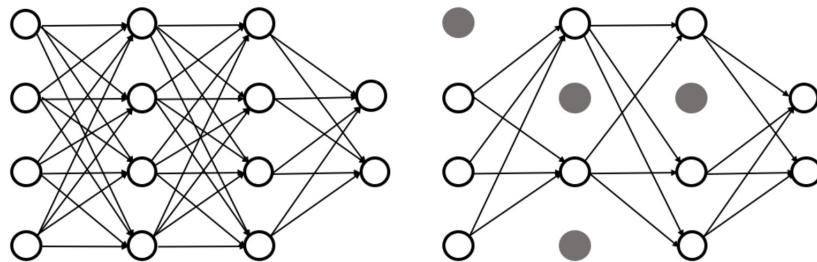


**Figura 2.21:** [31] Representación de la reducción de la dimensionalidad mediante *max pooling* con un tamaño de ventana de 2x2 y un *stride* de 2

### Capa dropout

Uno de los mayores problemas en la creación de modelos de DL es que durante el entrenamiento la red se ajuste demasiado a los datos de entrenamiento, y en lugar de aprender información de los datos lo que realice sea memorizar los resultados. Esto se conoce como sobreajuste o *overfitting*. Una de las formas más comunes de tratar de minimizar este *overfitting* es mediante las capas *dropout*. Estas capas lo que hacen es, de manera aleatoria, desconectar un porcentaje de los nodos de esa capa, de esta forma la red no podrá sobreajustarse a los datos con los que se entrena. Su funcionamiento se representa en la figura 2.22.

Un parámetro importante a configurar en estas capas es el porcentaje de nodos que desactiva. Si es demasiado pequeño tal vez no solucione el problema del sobreajuste, sin embargo, si se desconectaran un gran número de neuronas puede que la red no consiguiera aprender.



**Figura 2.22:** [32] Representación de una red neuronal sin *dropout* y otra con *dropout*

## 2.4 Estado del arte

Actualmente en la práctica clínica el proceso de diagnóstico del cáncer de próstata sigue realizándose de manera manual por los médicos. En general el proceso suele consistir en, tras una prueba cuyos resultados hagan sospechar de la presencia de cáncer en la próstata, realizar una biopsia del tejido. Posteriormente estas muestras serán primero procesadas para poder ser tratables al microscopio y a continuación analizadas una a una por un patólogo para diagnosticar finalmente si el paciente padece cáncer de próstata y en que estadio se encuentra.

El proceso del diagnóstico de cáncer de próstata es una labor tediosa y que conlleva mucho tiempo, tanto para la preparación de la muestra como en el diagnóstico del facultativo. Por ello en los últimos años se han desarrollado diferentes técnicas de ayuda al diagnóstico, con el fin de facilitar esta tarea al médico.

Entre las razones por las que los últimos años los sistemas de ayuda al diagnóstico han sufrido un auge se encuentra el crecimiento de las tecnologías de DL. Además, la concienciación de los facultativos con las ventajas de estos sistemas ha provocado la adecuación de grandes bases de datos de datos clínicos. Estos datos son la base para el desarrollo de sistemas de ayuda al diagnóstico, y en concreto de algoritmos de DL.

Con el desarrollo de la arquitectura de las CNNs se creó el caldo de cultivo perfecto para el auge de estos métodos de diagnóstico basados en imágenes. Actualmente gran parte de los modelos de DL con CNNs se enfocan en el estudio de imágenes histológicas. Se han llegado a realizar modelos que fueron entrenados con cerca de 7000 casos de pacientes a los que se les había realizado una biopsia de la próstata [33].

Sin embargo, aunque estos modelos hayan conseguido proporcionar una herramienta a los médicos que les facilite el trabajo del diagnóstico de pacientes con cáncer de próstata, siguen teniendo un problema fundamental, y es que todos estos modelos requieren de la realización de una biopsia al paciente, con los problemas que para este conlleva, además de que el tiempo de preparación de las muestras sigue ralentizando considerablemente el proceso de diagnóstico. Actualmente, esta es una limitación no solucionada en la práctica clínica y existe un gran interés por obtener marcadores no invasivos mediante el análisis de imagen de resonancia magnética.

La razón por la que se ha demorado tanto tiempo la creación de modelos que utilicen esta modalidad de imagen médica se debe a que en la práctica clínica no está muy extendida su utilización con este fin, debido a que como se ha expuesto anteriormente en el trabajo, resulta difícil diferenciar en algunas ocasiones las HBP de un adenocarcinoma en estas imágenes.

Debido a esto han surgido *challenges* como el de *ProstateX-2*, donde se proporciona a investigadores una base de datos con imágenes de resonancias magnéticas de próstata, para fomentar la investigación y creación de modelos que utilicen estas modalidades de imagen, con el fin de obtener biomarcadores que permitan diagnosticar el cáncer de próstata mediante esta modalidad de imagen.

Dentro de este *challenge* que fue realizado ya hace cuatro años se crearon diferentes modelos que utilizaban las IRM de diferente manera para obtener el diagnóstico del cáncer de próstata. Como además este TFG está enfocado dentro de este *challenge*, se ha realizado un estudio de los diferentes trabajos que fueron publicados utilizando esta base de datos, para comparar diferentes maneras de abordar el proyecto y cuales fueron sus resultados.

De todos los estudios se puede realizar una separación en dos grandes grupos, aquellos que realizaron modelos que trabajaban con los volúmenes de la resonancia magnética, y los que utilizaban las imágenes de manera separada. Como ejemplo del primero se encuentra el trabajo de Chian, E. [34]. En él se trata de clasificar las imágenes utilizando como entrada de la CNNs volúmenes con todas las imágenes del paciente, con estos recortados al rededor de la próstata con un radio que cubra todo el órgano, de forma que así se facilita a la red que se centre en las zonas de interés para el diagnóstico. Para ello utiliza los volúmenes potenciados en  $T_2$ , las DWI, los ADC

y las imágenes KTrans. Se toman todos estos volúmenes y se conforman de forma que la entrada de la red tiene cuatro dimensiones, siendo las tres primeras las correspondientes a los volúmenes de IRM, y la última para las diferentes modalidades implementadas. Con el entrenamiento de este modelo consigue obtener una precisión del 73% en la clasificación entre pacientes que no tienen cáncer y los que si tienen, sin embargo, no es capaz de estimar la gradación del cáncer.

Como ejemplo del segundo grupo, donde se utilizan las imágenes de resonancia magnética en dos dimensiones, se encuentra el trabajo de Yuan et al. [35]. En el utilizan también un modelo multiparamétrico como el anterior, donde combinan las imágenes potenciadas en  $T_2$  tanto sagital como transaxial, además de las imágenes de ADC. Al igual que en el trabajo anterior, segmentan la imagen para crear una región de interés, con el fin de facilitar al modelo la extracción de características que sean significativas para la clasificación.

Sin embargo, los modelos a la hora de obtener una correcta gradación de los tumores no consiguen obtener buenos resultados, como se publicó en el artículo del *challenge* [4], por ello, gran parte de la investigación actual se centra en modelos que sean capaces de determinar de una manera precisa el grado del cáncer de próstata en imágenes de resonancia magnética.



# Materiales

### 3.1 Base de datos de ProstateX-2

Como se ha indicado en el capítulo 2 del presente trabajo, para la elaboración del mismo se va a hacer uso de la base de datos de imagen por resonancia magnética (IRM) del *challenge ProstateX-2*. Desde la finalización del *challenge* en 2017 la base de datos está a libre disposición del todo el mundo que quiera trabajar y realizar modelos con ella.

Para ello se proporcionaba una base de datos de 161 casos de IRM, cada caso perteneciente a un paciente distinto. En cada uno de los casos se proporcionaban diferentes modalidades de IRM:

- Potenciada en  $T_2$ , tanto sagital como transaxial.
- Ktrans transaxial
- DWI transaxial
- ADC transaxial

Todas las imágenes fueron tomadas en dos tipos distintos de escáneres de RMN de 3T de Siemens, el MAGNETOM Trio y el Skyra. Para la adquisición de las imágenes potenciadas en  $T_2$  se utilizó una secuencia espín-eco turbo con una resolución de plano de 0.5 mm, y un grosor de corte de 3.6 mm. Las imagen potenciada en difusión (DWI) fueron tomadas con una secuencia eco-planar con una resolución de plano de 2 mm y un grosor de corte de 3.6 mm. Los valores  $b$  de las imágenes de difusión fueron 50, 400 y 800. De estos valores se pudieron calcular los mapas de ADC.

Todas las imágenes, a excepción de las imágenes KTrans, se proporcionan en formato DICOM. DICOM surge en el año 1993 [36] como respuesta a todos los diferentes formatos que habían ido apareciendo durante los últimos años para el almacenamiento de las imágenes radiológicas. De esta forma, se consigue crear un formato estandarizado en el que almacenar gran parte de las imágenes radiológicas, facilitando así su posterior almacenamiento y tratamiento.

Los archivos en formato DICOM tienen una cabecera de datos donde se encuentra toda la información relevante del estudio, como por ejemplo la resolución de la imagen, que máquina a

adquirido la imagen, o los datos del paciente. De esta forma no solamente se almacena la imagen sino también toda la información que posteriormente puede resultar de utilidad para el futuro diagnóstico.

Para la realización de los modelos de DL en este trabajo se utilizarán las imágenes potenciadas en  $T_2$  transaxial, además de las imágenes DWI y los ADC. Se descartan las imágenes KTrans debido a que no proporcionaban las etiquetas necesarias para el propósito del trabajo.

Las imágenes que se proporcionan potenciadas en  $T_2$  tienen en su gran mayoría una resolución de 384x384. Sin embargo, en algunos pacientes las imágenes proporcionadas son de mayor tamaño. En el caso de las imágenes de DWI y ADC, todas las imágenes tienen la misma resolución, 128x84. Los archivos en formato DICOM de las imágenes  $T_2$  ocupan un tamaño promedio de 300 KB, mientras que las imágenes de DWI y ADC ocupan un espacio menor (debido a su menor tamaño), con una media de 30 KB.

Junto a las imágenes proporcionadas se adjunta en la base de datos un fichero *csv* con el corte donde se encuentra la lesión, el punto donde se encuentra y el grupo al que pertenece (a excepción de las imágenes Ktrans, que no proporcionaban esta información). Mediante este fichero se realiza el *ground truth* para el entrenamiento de la red neuronal, como se explicará posteriormente. Toda esta base de datos fue realizada a partir del criterio de médicos que diagnosticaron a los pacientes uno a uno para poder disponer de estas etiquetas, además de marcar el punto de la lesión en los casos de los pacientes que padecían de cáncer. Para poder diagnosticarlos todos los casos fueron clasificados a través de biopsias guiadas de la próstata.

Además, cada lesión tiene asignado diferentes grupos dependiendo de su grado en la escala de Gleason que se distribuyeron de la siguiente manera:

Grupo 1	Escala de Gleason $\leq 6$
Grupo 2	Escala de Gleason 3+4
Grupo 3	Escala de Gleason 4+3
Grupo 4	Escala de Gleason 8
Grupo 5	Escala de Gleason 9-10

**Tabla 3.1:** Grupos de grados de lesión cancerígena [37].

Como se ve en la tabla 3.1 se han juntado en 5 grupos los diferentes grados de la escala de Gleason, siendo los primeros grupos los más leves y los últimos los grupos con los estadios de cáncer de próstata más avanzados.

De los 161 casos disponibles en el *dataset* facilitado, se utilizarán en el modelo final 160, debido a que en unos de los casos no se dispone de todas las modalidades de IRM que se utilizan. De esta forma, los grupos se reparten el número como se expone en la tabla 3.2:



Grupo:	Número de casos
No cáncer	77
Grupo 1	28
Grupo 2	34
Grupo 3	15
Grupo 4	4
Grupo 5	2

**Tabla 3.2:** Número de casos por cada grupo del conjunto de datos

## 3.2 Software

Durante el desarrollo de este TFG se ha hecho uso de diferentes herramientas informáticas, ya sea para el preprocesado de los datos como para la creación de los modelos de DL. Para la etapa del preprocesado se ha utilizado la plataforma MATLAB®, en su versión v.R2020b. Es propiedad de la empresa The Mathworks Inc. En el programa se combinan tanto un lenguaje de programación propio como un entorno de desarrollo. Es un lenguaje ampliamente utilizado especialmente en sectores científicos de investigación, en ingeniería y en finanzas. Esto se debe a las múltiples *tool-boxes* con las que proporciona programas específicos para diferentes ámbitos, como pueden ser el ML, la visión artificial, el procesamiento de señales o el procesamiento de imágenes, entre otras. Para la realización de este TFG se ha hecho uso de la *Image processing toolbox*.

Para la creación de las redes neuronales convolucionales (CNNs) el lenguaje de programación que se ha utilizado ha sido Python. La razón por la que se ha utilizado Python y no MATLAB para la creación de los modelos de DL es que en la actualidad el programa más extendido para la creación de este tipo de modelos es Python, debido a que es gratuito y la gran cantidad de librerías de ciencia de datos y procesado que hay a disposición de todo el mundo. Por ello se consideró que la mejor forma de desarrollar este trabajo sería utilizando Python.

La versión de Python utilizada en este trabajo es la 3.8.5, y como programa para la edición de código y su ejecución se ha utilizado el entorno de desarrollo integrado (IDE) PyCharm Professional 2020.3.3 gracias a las múltiples herramientas que proporciona, entre ellas la de realizar una conexión SSH con el servidor donde se ejecutaban remotamente los cálculos.

Para el desarrollo de los modelos con Python se han utilizado múltiples librerías durante todo el proceso, como *NumPy* para el cálculo con matrices, *Pandas* para la lectura de archivos *csv* o *scikit-learn* entre otros, pero se quieren destacar dos por su importancia para el desarrollo del proyecto: *Keras* y *Tensorflow*.

*Keras* es una API de código abierto para el desarrollo de redes neuronales con Python. Fue diseñado por el ingeniero de Google François Chollet como una librería de alto nivel que facilitara la creación de modelos de DL en Python. *Tensorflow* es el *backend* sobre el que trabaja Keras. Fue desarrollado por Google y proporciona herramientas para la realización de algoritmos de aprendizaje profundo y para la paralelización de los cálculos, de forma que pueda aprovechar los numerosos núcleos de las tarjetas gráficas para el procesamiento de los cálculos de una manera más eficiente.

### 3.3 Hardware

Es importante conocer el hardware que se ha utilizado durante la elaboración del trabajo, ya que de sus características vienen determinadas los tiempos de cálculo de los diferentes modelos y el rendimiento al ejecutarlos. Para ello se han utilizado dos equipos distintos, por un lado un portátil MacBook Pro con las características de la tabla 3.3:

Sistema operativo:	Manjaro Linux x86_64
Procesador:	Intel® Core i5-4278U
Frecuencia:	2.60 GHz
Memoria:	8 GB de RAM
Almacenamiento:	512 GB SSD
Tarjeta gráfica:	Intel Iris

**Tabla 3.3:** Características del equipo empleado

Sin embargo, para la realización de todos los cálculos necesarios para el diseño de un modelo de aprendizaje profundo las características del equipo anteriormente mencionadas no serían suficientes. Es por ello que se trabaja en remoto con un servidor con el hardware apropiado para este propósito, proporcionado por el CVBLab. Las características principales del servidor son las expuestas en la tabla 3.4:

Sistema operativo:	Ubuntu 20.04.1 LTS
Procesador:	Intel® Core i7-7700K
Frecuencia:	4.20 GHz
Memoria:	32 GB de RAM
Tarjeta gráfica:	Nvidia Titan Xp

**Tabla 3.4:** Características del servidor proporcionado por el CVBLab

De las características de la tabla 3.4 destacar principalmente la tarjeta gráfica de la que dispone, una Nvidia Titan Xp, ya que es la que permitirá la paralelización del cálculo de la red neuronal de forma que el tiempo de entrenamiento se acorte considerablemente.

## Capítulo 4

# Métodos

En este capítulo se van a exponer los métodos utilizados para el desarrollo de los diferentes modelos de aprendizaje profundo. Para ello se explicaran las diferentes arquitecturas que se han implementado durante el transcurso del trabajo, explicando las características de cada una. Posteriormente se profundizará en el desarrollo de los Grad-CAMs y cómo se pueden utilizar para estudiar la interpretabilidad de los resultados y del comportamiento de la red.

### 4.1 Redes Neuronales Multi-Modales

Las redes neuronales convolucionales (CNNs) supusieron un gran avance en el uso de modelos de DL para la clasificación de imágenes. Como se ha explicado en anteriores capítulos, las CNNs simulan el comportamiento del cerebro cuando analiza una imagen, siendo capaces de extraer características más o menos abstractas que podrá utilizar para realizar una predicción sobre la imagen.

Sin embargo, cuando se habla de redes neuronales convolucionales (CNNs) no existe una arquitectura única a implementar. Algunas de las arquitecturas más comunes en la clasificación de imágenes son: *AlexNet*, *GoogleNet*, *ResNet*, *VGG16* y *VGG19*.

Para la realización de este trabajo se ha obtenido por el uso del *transfer learning*, un método de DL que se basa en utilizar modelos preentrenados para, a partir de esa base, reentrenarlos con el objetivo que se pretenda, de forma que la inicialización de los pesos no será aleatoria sino que ya ha sido entrenada para la clasificación de imágenes.

En este trabajo se va a trabajar tanto con el modelo VGG16 como el modelo VGG19, ya que ambos están disponibles en la API de *Keras* con los pesos inicializados al entrenamiento con la base de datos *Imagenet*. Se entrenarán ambos modelos en las diferentes arquitecturas que se definirán posteriormente, de forma que se quedará el modelo que mejor resultado proporcione en cada una de las diferentes combinaciones, siempre partiendo de los mismos pesos. Como el objetivo del trabajo es combinar las diferentes modalidades de IRM, se probaran diferentes

formas de combinar todos los tipos de imágenes, con el fin de corroborar cual otorga el mejor resultado.

La arquitectura del modelo VGG16 se muestra en la figura 4.1.

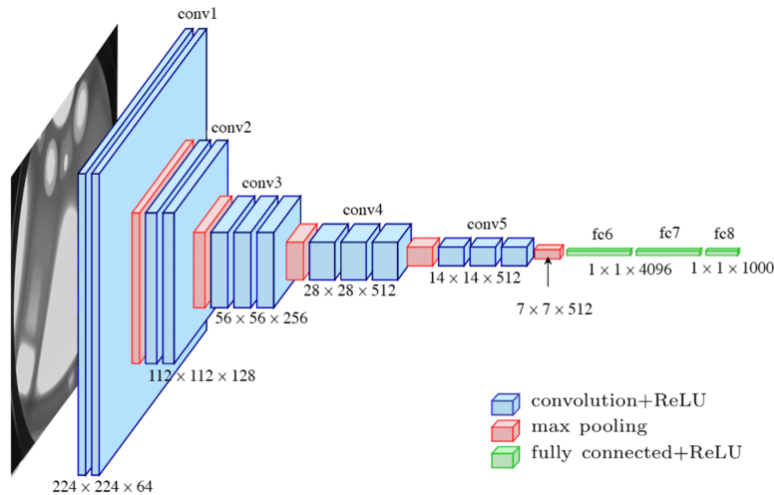


Figura 4.1: [38] Arquitectura del modelo VGG16

Se puede ver en la figura 4.1 que la arquitectura del modelo VGG16 tiene dos partes diferenciadas. Una primera parte compuesta por cinco bloques de capas convolucionales y *max pooling* que realizará la extracción de características de la imagen y una segunda parte de capas *Flatten* y densamente conectas, que se encargarán de la realización de la predicción. La estructura del modelo VGG19 es la misma, la única variación es que en los 3 últimos bloques convolucionales hay 4 capas convolucionales y no 3 como en la VGG16.

Las capas *Flatten* procesan la entrada que reciban, de forma que si reciben una matriz devolverán a la salida los mismos valores pero en un vector. Las capas densamente conectas son capas donde todos los nodos están conectados a la capa anterior. Para los intereses de este trabajo la parte del modelo que se utilizará será únicamente el extractor de características, es decir, los bloques convolucionales con *max pooling*. Se desechan las capas *Flatten* y densamente conectadas ya que actualmente el estado del arte es la utilización de capas *global average pooling (GAP)*, ya que disminuyen el número de parámetros y ayudan a prevenir el sobre ajuste. Funciona realizando un promediado de todos los canales a la salida del último bloque convolucional, de forma que la salida de la capa GAP sea un vector de características unidimensional, representativo de cada uno de los mapas de activación en la imagen.

Esta será la entrada de una capa de *dropout* para finalmente entrar en la capa de predicción que tendrá como función de activación la función *softmax*, cuya expresión matemática es la de la fórmula 2.18.

#### 4.1.1 Combinación tardía

El mayor reto a la hora de afrontar la arquitectura del modelo fue hallar la forma óptima en la que combinar las diferentes modalidades de IRM. En un primer momento se optó por utilizar tres modelos distintos donde cada uno recibiera como entrada una modalidad de imagen. Cada modelo actualiza sus pesos independientemente de los demás, y posteriormente se unen con una capa de concatenación. Su esquema se muestra en la figura 4.2.

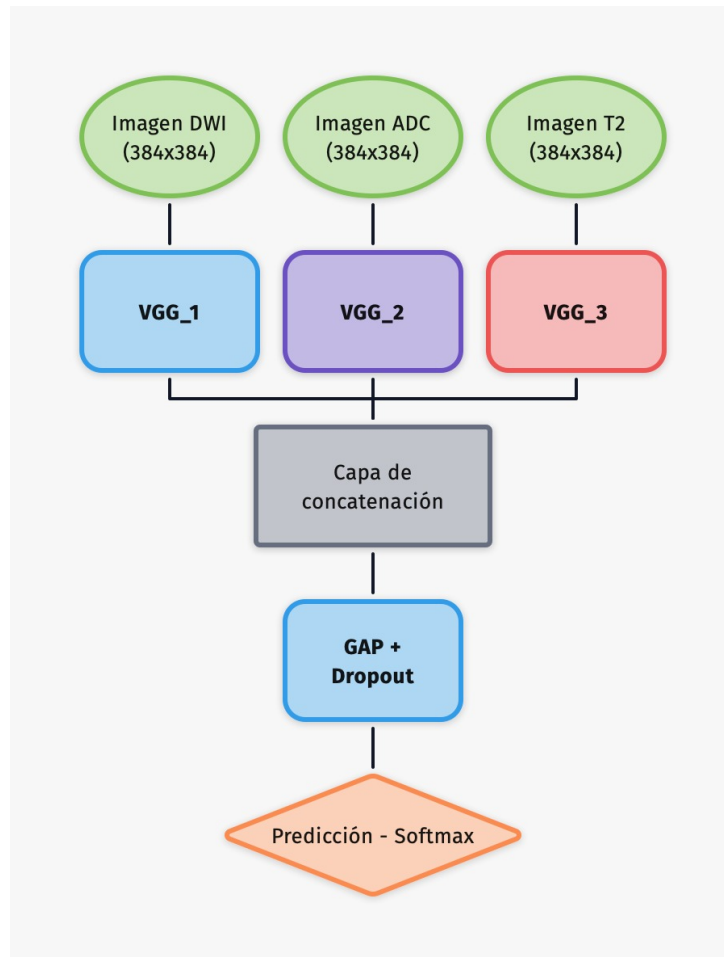


Figura 4.2: Arquitectura de combinación tardía multimodelo

Se ha denominado red de combinación tardía ya que cada modalidad de IRM es analizada utilizando todos los bloques convolucionales del modelo y es a la salida de este último que se han combinado las tres modalidades.

Llegados a este punto es razonable considerar una variación de esta arquitectura, ya que para cada modalidad de IRM se está utilizando un modelo independiente de las otras modalidades, por lo que tendrán distintos pesos. De este razonamiento surge la idea de, también con una combinación tardía, utilizar el mismo modelo con los mismos pesos para cada una de las modalidades, quedando esta propuesta de arquitectura de la forma de la figura 4.3.

De esta forma, los pesos se adaptarían a trabajar con las tres imágenes en conjunto, aunque estas entren de manera independiente a la red. El resto de la arquitectura, como se ve en la figura 4.3 se mantiene de la misma manera

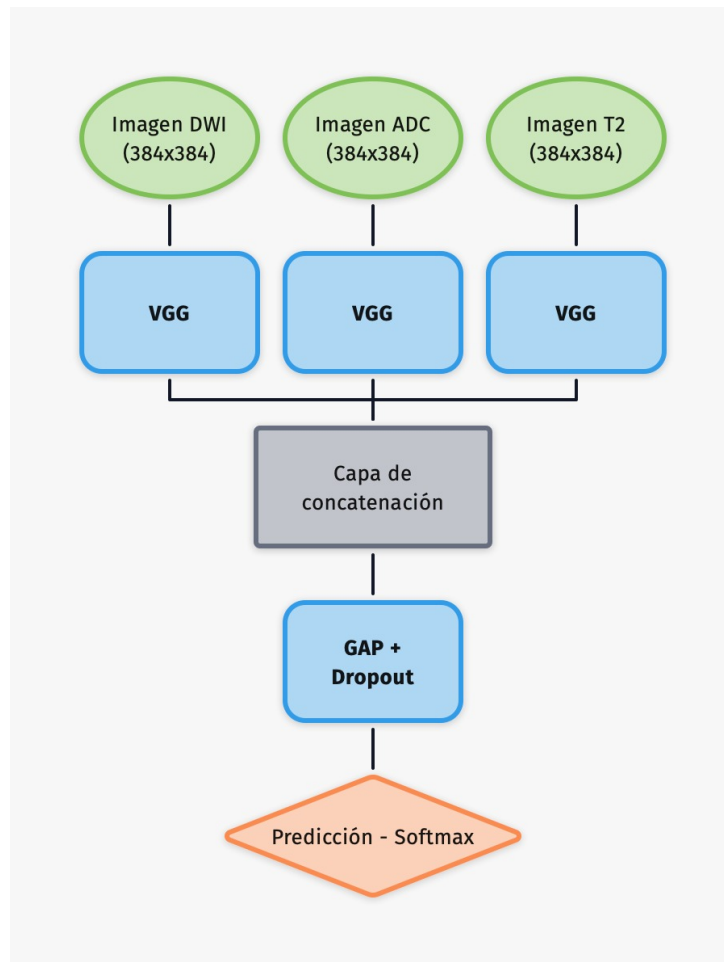


Figura 4.3: Arquitectura de combinación tardía monomodelo

#### 4.1.2 Combinación temprana

Hasta ahora la combinación de las diferentes modalidades de IRM se ha realizado posteriormente a la extracción de características. Sin embargo existen otras formas en las que se podrían combinar las diferentes imágenes. Una de ellas es asemejar su comportamiento al de una imagen en formato *rgb*, donde la imagen está compuesta por tres canales y cada uno de ellos tiene una información distinta, en el caso del *rgb* un canal aporta la información del color rojo en la imagen, otra del color verde y otra del color azul.

De este razonamiento surge la idea de combinar las tres modalidades en una única imagen con tres canales, donde cada canal aporta la información de una modalidad distinta de IRM, y sería la red la encargada de elegir como extraer las características de cada uno de los canales. Su arquitectura queda de la forma de la figura 4.4.

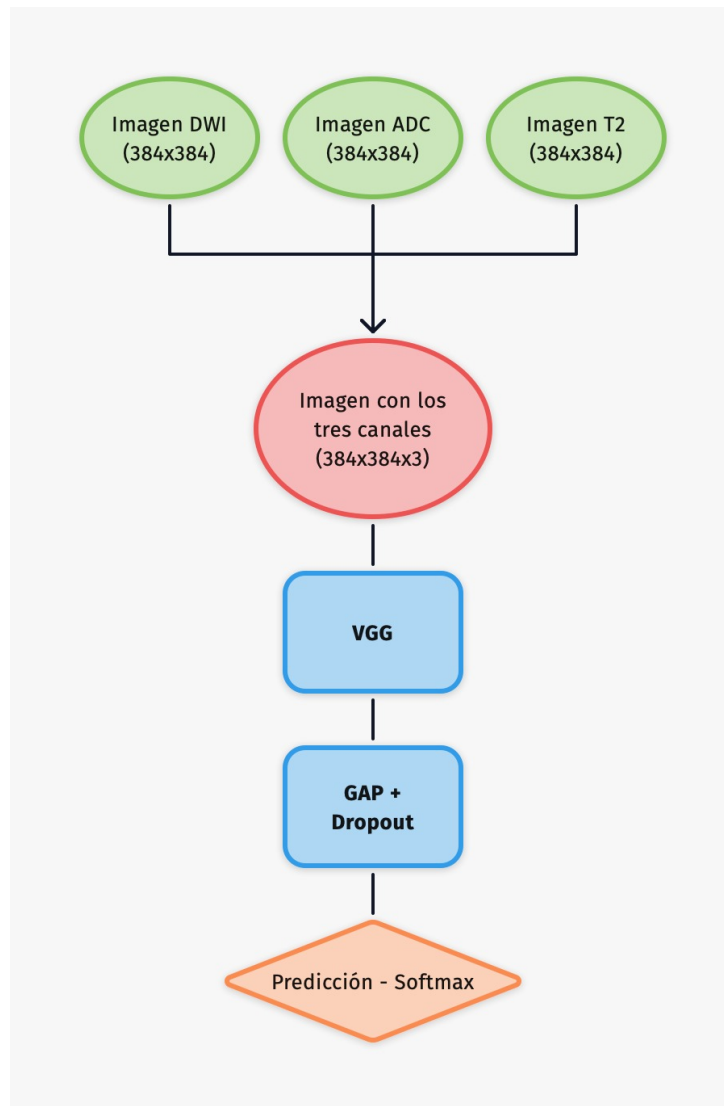


Figura 4.4: Arquitectura de combinación temprana

#### 4.1.3 Redes híbridas

La última arquitectura que utilizará en este trabajo se basa en la idea de combinar las características de las arquitecturas anteriores. Utilizará una primera parte multimodelo donde cada imagen es analizada por separado, como en la combinación tardía, pero en vez de concatenar los mapas de características después de haber pasado por todos los bloques convolucionales, se realizará la unión a un nivel intermedio de la red.

Como se ha explicado anteriormente, se están utilizando los cinco bloques convolucionales de los modelos VGG16 y VGG19, así que para esta arquitectura la combinación se realizará a la altura del tercer bloque, de forma que los tres primeros bloques estarán destinados a la extracción de características individuales de cada modalidad de IRM, y se dejará a los dos últimos bloques para que saquen características conjuntas de las tres imágenes. Su esquema se puede ver en la figura 4.5.

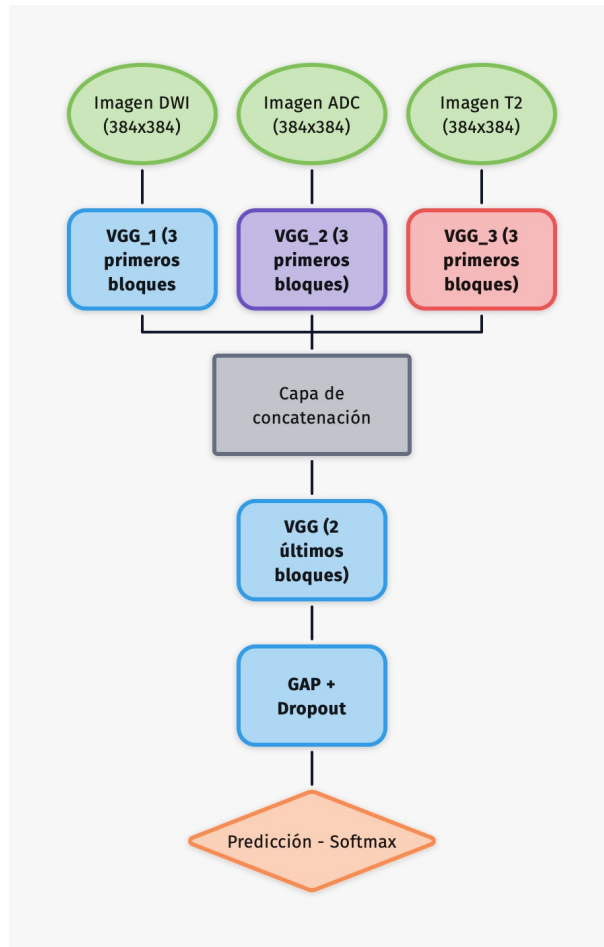


Figura 4.5: Arquitectura de combinación híbrida

## 4.2 Interpretabilidad

### 4.2.1 Grad-CAMs

Con el objetivo de estudiar el comportamiento de la CNNs y poder interpretar los resultados de la misma, se hará uso de Gradient-Weighted Class Activation Maps (Grad-CAMs), que es un método débilmente supervisado de segmentación semántica [39].

El proceso para la obtención de los Grad-CAMs consta de dos pasos. En el primero se realiza una retropropagación parcial a partir de la predicción de la red hasta la salida de la activación de la última capa convolucional. De aquí se realiza un promediado en dos dimensiones para obtener el mapa de características con las zonas donde más se ha fijado la red a la hora de realizar la predicción. Este mapa se puede expresar de la siguiente forma [39]:

$$\alpha_{c,d_L} \leftarrow \frac{1}{N_L^2} \sum_{i=1}^{N_L} \sum_{j=1}^{N_L} \frac{\partial y_c}{\partial \hat{A}dL^L(i,j)} \quad (4.1a)$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left( \sum_k \alpha_k^c A^k \right) \quad (4.1b)$$



Siendo  $\alpha_{c,d_L}$  los gradientes,  $y_c$  es la predicción de la red para esa imagen y  $\hat{A}dL^L$  es el mapa de activación de la última capa convolucional. Posteriormente se hace una suma ponderada de los mapas de activación  $A^k$  y los pesos, pasando finalmente los resultados por la función de activación ReLU, para mantener aquellos patrones con una influencia positiva sobre la predicción.



# Experimentos y Resultados

## 5.1 Adecuación de los datos y métricas utilizadas

### 5.1.1 *Pre-procesamiento de la base de datos*

El preprocesado de la base de datos es una parte fundamental del trabajo, ya que es crucial la forma en la que se adecúan los datos para el modelo que se vaya a implementar. En este caso, la base de datos está formada por una carpeta por cada paciente, y cada paciente tiene tantas carpetas como modalidades de IRM se le han tomado.

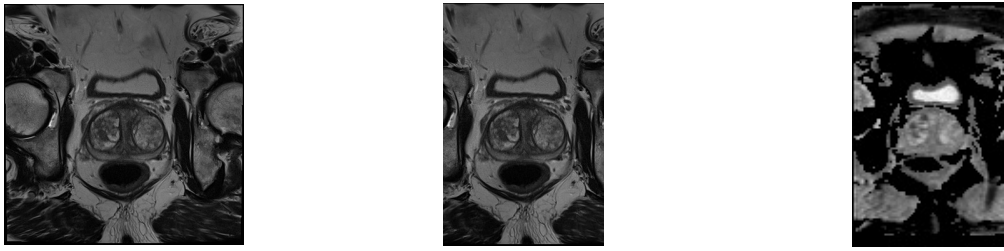
Dentro de la carpeta de cada modalidad se disponen los archivos en formato *dicom*, teniendo un archivo por cada corte de esa modalidad de IRM. El primer paso para la adecuación del formato de los datos es convertir todos los cortes en formato *dicom* de cada una de las modalidades que se quieran utilizar, de todos los pacientes, a volúmenes en formato *.mat*, para poder trabajarlo posteriormente con MATLAB. Con este fin se crea un *script* que recorre todas las carpetas, lee todas las imágenes en formato *dicom* y crea con ella una matriz tridimensional.

Posteriormente, se deben extraer los cortes que se vayan a utilizar en el futuro entrenamiento de la red, ya que no se utiliza todo el volumen. Para ello se crea otro *script* que selecciona que corte elige de cada uno de los volúmenes, dependiendo de si el paciente está sano o no. En el caso de que el paciente no presente cáncer de próstata se cogen cinco cortes centrales (el corte del medio del volumen, dos por delante y dos por detrás). Se hace de esta forma para poder aportar más información a la red de la anatomía de una próstata sana.

Si por el contrario el paciente tiene cáncer el script selecciona el corte donde el médico ha señalado que se encuentra la lesión. Estos cortes se exportan en formato *.png* para convertirlos a imagen y que la red pueda leerlos posteriormente.

Además, el formato en el que se encuentran las imágenes potenciadas en T2 tiene un ratio 1:1 e incluye más parte del cuerpo que las demás modalidades utilizadas. Por ello, se recortan las

imágenes potenciadas en T2 para que se ajusten a las dimensiones de las demás imágenes. El resultado de este recorte se puede ver en la figura 5.1.



**Figura 5.1:** Representación del mismo corte en T2, esa imagen recortada y la misma imagen en ADC

Para poder trabajar con estas imágenes se han de almacenar en un directorio ordenado, donde se guarden todas las modalidades y separado en diferentes grupos. Para ello, se crea otro *script* que lee que grado de cáncer de próstata tiene cada paciente y los ordena según su gravedad, separando a los pacientes en tres grupos distintos:

1. Grado Bajo (GB): corresponde a los pacientes con un grado de lesión bajo, en este caso los grupos 1 y 2 de la base de datos de *ProstateX-2*.
2. Grado alto (GA): se asigna a los grupos con estadíos de cáncer más avanzados, los correspondientes a los grupos 3, 4 y 5 de la base de datos *ProstateX-2*.
3. No cancer (NC): en este último grupo estan todas las imágenes en las que no haya ninguna lesión

El número total de imágenes en cada grupo queda como se presenta en la tabla 5.1:

Grupo	Nº de imágenes
GB	62
GA	21
NC	385

**Tabla 5.1:** Número de imágenes que componen el dataset

Como se observa en la tabla 5.1, el número de imágenes de pacientes sin cáncer es mucho mayor que del resto de imágenes. Esto se debe a un problema de desbalanceo en la base de datos, donde la mayoría de los pacientes son pacientes sanos y hay pocos casos de pacientes con estadíos altos de cáncer. Sin embargo, se aplicarán medidas posteriormente para tratar de paliar ese desbalanceo de la base de datos.

A continuación se realiza la partición del *dataset*, ya que el entrenamiento de la CNNs seguirá el método de *hold out*, en el que se separan dos subconjuntos para validar y testear los modelos desarrollados. En este caso, se destina el 70% de los datos al entrenamiento del modelo, el 20% a su evaluación y un 10% para la validación durante el entrenamiento. En esta separación se ha tenido en cuenta también el número de imágenes de cada clase, de forma que en cada uno de los grupos se mantienen las proporciones de imágenes de cada clase.

Para sintetizar el proceso de trabajo que se ha seguido durante todo el pre-procesamiento de la base de datos, se ha creado el siguiente esquema:

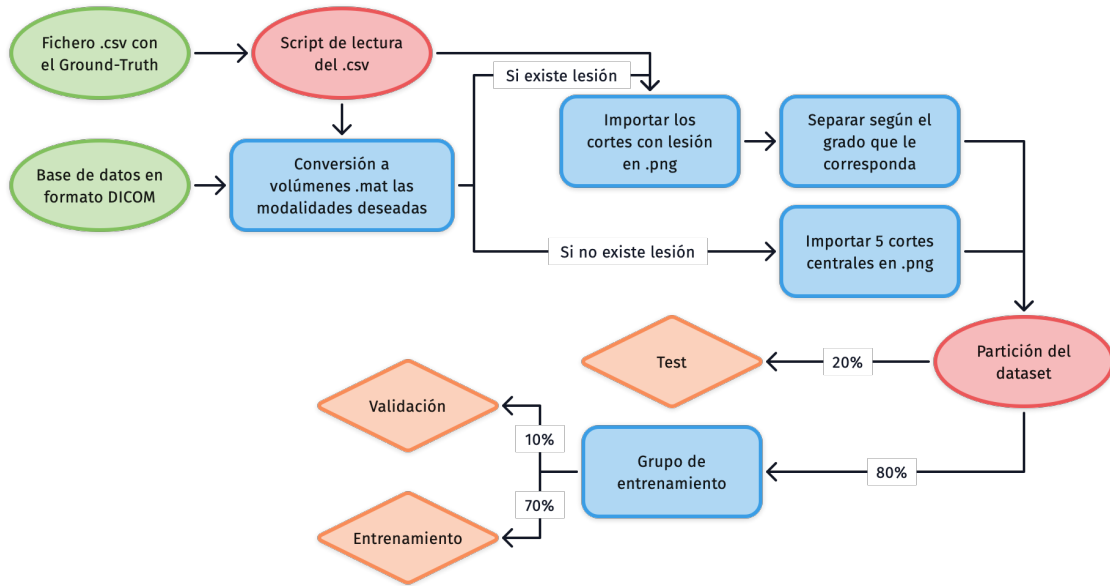


Figura 5.2: Esquema del flujo de trabajo implementado durante el pre-procesamiento de los datos

### 5.1.2 Métricas y evaluación

Para medir las pérdidas del modelo se utiliza como función de coste la *categorical crossentropy* (entropía cruzada categórica). Es un tipo de función de pérdidas muy extendido en clasificadores multiclase, donde cada ejemplo puede permanecer únicamente a una clase. Esta función de pérdidas viene definida mediante de la siguiente manera:

$$Loss = - \sum_{i=1}^I y_i \cdot \log \hat{y}_i \quad (5.1)$$

Donde  $\hat{y}_i$  es la predicción que realiza la red e  $y_i$  es el *ground truth* real para esa imagen, estando ambos en formato *one-hot-encoding*. Por último,  $I$  corresponde al número de categorías.

También es necesario definir algunas métricas que se utilizarán posteriormente con el grupo de test para evaluar el rendimiento de los diferentes modelos. Se medirá tanto la precisión como el *F1-score* y el coeficiente *Kappa*. La precisión se calcula como el porcentaje de imágenes que han sido clasificadas de un grupo que realmente pertenecen a ese grupo, lo que se expresa de la siguiente manera:

$$P = \frac{VP}{VP + FP} \quad (5.2)$$

Donde VP hace referencia a los verdaderos positivos y FP a los falsos positivos.

Por otro lado, el valor de *F1-score* se calcula como la media armónica de los valores de precisión y de exhaustividad [40] (que también suele llamarse *recall*). La precisión ya se ha definido anteriormente, y la exhaustividad (R) se mide de la siguiente forma:

$$R = \frac{VP}{VP + FN} \quad (5.3)$$

Siendo FN el número de falsos negativos.

Teniendo en cuenta ambas expresiones, el valor de *F1-score* resulta de la siguiente manera:

$$F_1 = \frac{2PR}{P + R} \quad (5.4)$$

Por último, el coeficiente *Kappa* es un valor que sirve para ajustar el efecto que puede tener el azar en los resultados obtenidos en variables categóricas [41]. En este caso se utiliza el valor *kappa* cuadrático, que es el mismo que utilizaron en la evaluación del *challenge*.

## 5.2 Detección de cáncer de próstata en IRM monomodal

### 5.2.1 Arquitectura

Antes de exponer los resultados de las arquitecturas multi-modales, también se entrenan modelos que únicamente utilizan una de las modalidades, para comprobar si la aplicación de la multimodalidad supone una mejora respecto a este tipo de modelos.

Se entrenan tres modelos distintos, cada uno utilizando una modalidad distinta de IRM y la arquitectura que se implementa fue la de la VGG19, empleando una capa global average pooling (GAP) y una *dropout* al final del modelo antes de entrar a la capa de salida. En el caso de las imágenes potenciadas en T2 se utilizan las imágenes a una resolución de 384x384 (en este caso no se utilizan las imágenes potenciadas en T2 recortadas ya que no se iban a utilizar en conjunto con el resto de modalidades). Como existen algunas pocas imágenes que no están con ese tamaño se reescalan estas para que tengan las mismas dimensiones. De esta forma, la estructura del modelo queda de la forma de la tabla 5.2.

Por otro lado se encuentran las imágenes DWI y los ADC. Estas son de menor resolución, 128x84. También se implementa la arquitectura de VGG19 como en el caso de las imágenes potenciadas en T2. Se utilizan las imágenes en su resolución nativa sin tener que reescalar ninguna ya que todas tienen las mismas dimensiones, a diferencia de lo que ocurre con las imágenes potenciadas en T2. Como ambas modalidades de imagen tienen las mismas dimensiones utilizan la misma arquitectura, la desarrollada en la tabla 5.3

Además los hiperparámetros de los modelos se configuran de la siguiente forma:

- *Batch size* = 16
- *Learning rate* = 0.00001
- *Dropout* = 0.25
- Optimizador = SGD - Nesterov
- Número de épocas = 400

Como el número de imágenes del que se dispone es limitado, se ha aplicado *Data augmentation* al conjunto de entrenamiento. *Data augmentation* consiste en crear más imágenes a partir de

las que ya se dispone, realizando ciertas modificaciones sobre ellas. De esta forma se consigue aumentar artificialmente el tamaño de nuestra base de datos. En este caso se han aplicado:

- Rotaciones ( $\pm 5^\circ$ )
- Desplazamiento en ambos ejes (máximo un 10 %)
- *Zoom* de la imagen (máximo un 10 %)

También se tiene en cuenta el desbalance entre las diferentes clases, que queda patente en la tabla 5.1. Por ello durante el entrenamiento se aplican diferentes pesos a las clases. Estos pesos hacen que el error de la función de pérdidas tenga más en cuenta las pérdidas provocadas por los grupos con menos imágenes (en este caso GA) que las provocadas por el grupo por más imágenes (en este caso NC). Estos pesos se distribuyen según el número de imágenes de cada etiqueta, de forma que el error se reparta de forma balanceada. Para la obtención de los pesos se aplica la expresión 5.5.

Capa	Tipo	Dimensiones	Parámetros
input_1	InputLayer	(384,384,1)	0
block1_conv1	Conv2D	(384,384,64)	1792
block1_conv2	Conv2D	(384,384,64)	36928
block1_pool	MaxPooling2D	(192,192,64)	0
block2_conv1	Conv2D	(192,192,128)	73856
block2_conv2	Conv2D	(192,192,128)	147584
block2_pool	MaxPooling2D	(96,96,128)	0
block3_conv1	Conv2D	(96,96,256)	295168
block3_conv2	Conv2D	(96,96,256)	590080
block3_conv3	Conv2D	(96,96,256)	590080
block3_conv4	Conv2D	(96,96,256)	590080
block3_pool	MaxPooling2D	(48,48,256)	0
block4_conv1	Conv2D	(48,48,512)	1180160
block4_conv2	Conv2D	(48,48,512)	2359808
block4_conv3	Conv2D	(48,48,512)	2359808
block4_conv4	Conv2D	(48,48,512)	2359808
block4_pool	MaxPooling2D	(24,24,512)	0
block5_conv1	Conv2D	(24,24,512)	2359808
block5_conv2	Conv2D	(24,24,512)	2359808
block5_conv3	Conv2D	(24,24,512)	2359808
block5_conv4	Conv2D	(24,24,512)	2359808
block5_pool	MaxPooling2D	(12,12,512)	0
global_average_pooling	GAP2D	(512)	0
dropout	Dropout	(512)	0
predictions	Dense	(3)	0

**Tabla 5.2:** Arquitectura del modelo monomodal para imágenes potenciadas en T2

Capa	Tipo	Dimensiones	Parámetros
input_1	InputLayer	(128,84,1)	0
block1_conv1	Conv2D	(128,84,64)	1792
block1_conv2	Conv2D	(128,84,64)	36928
block1_pool	MaxPooling2D	(64,42,64)	0
block2_conv1	Conv2D	(64,42,128)	73856
block2_conv2	Conv2D	(64,42,128)	147584
block2_pool	MaxPooling2D	(32,21,128)	0
block3_conv1	Conv2D	(32,21,256)	295168
block3_conv2	Conv2D	(32,21,256)	590080
block3_conv3	Conv2D	(32,21,256)	590080
block3_conv4	Conv2D	(32,21,256)	590080
block3_pool	MaxPooling2D	(16,10,256)	0
block4_conv1	Conv2D	(16,10,512)	1180160
block4_conv2	Conv2D	(16,10,512)	2359808
block4_conv3	Conv2D	(16,10,512)	2359808
block4_conv4	Conv2D	(16,10,512)	2359808
block4_pool	MaxPooling2D	(8,5,512)	0
block5_conv1	Conv2D	(8,5,512)	2359808
block5_conv2	Conv2D	(8,5,512)	2359808
block5_conv3	Conv2D	(8,5,512)	2359808
block5_conv4	Conv2D	(8,5,512)	2359808
block5_pool	MaxPooling2D	(4,2,512)	0
global_average_pooling	GAP2D	(512)	0
dropout	Dropout	(512)	0
predictions	Dense	(3)	0

**Tabla 5.3:** Arquitectura del modelo monomodal para imágenes DWI y imágenes ADC

$$\frac{\text{número de muestras}}{\text{número de clases} \cdot \text{número de muestras por clase}} \quad (5.5)$$

### 5.2.2 Resultados

A continuación se presenta en la tabla 5.4 los resultados obtenidos en la modalidad T2 para el subconjunto de test. En la tabla 5.5 se exponen los resultados obtenidos en la modalidad DWI y en la tabla 5.6 los resultados para la modalidad ADC.



<i>Monomodal - T2</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.18	0.21	12
Grado Alto	0	0	4
No cancerígeno	0.87	0.85	95
Valores medios ponderados:	0.76	0.75	
Valor Kappa:	0.089		

**Tabla 5.4:** Resumen de resultados de la arquitectura monomodal T2

<i>Monomodal - DWI</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.10	0.12	12
Grado Alto	0	0	4
No cancerígeno	0.86	0.80	95
Valores medios ponderados:	0.74	0.70	
Valor Kappa:	-0.079		

**Tabla 5.5:** Resumen de resultados de la arquitectura monomodal DWI

<i>Monomodal - ADC</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.09	0.09	12
Grado Alto	0	0	4
No cancerígeno	0.85	0.87	95
Valores medios ponderados:	0.74	0.75	
Valor Kappa:	-0.042		

**Tabla 5.6:** Resumen de resultados de la arquitectura de combinación tardía con tres modelos

### 5.2.3 *Discusión de resultados con arquitecturas monomodales*

A la vista de los resultados expuestos en el apartado anterior se observa que el uso de una única modalidad de IRM no es suficiente para poder detectar el cáncer de próstata. Estos resultados entran dentro de lo esperable, ya que normalmente el diagnóstico por parte de expertos radiólogos no se apoya únicamente en una modalidad, sino que combina varias extrayendo información distinta de cada una para poder llegar a un diagnóstico. Sin embargo, resulta de interés para compararlo posteriormente con los resultados obtenidos en multimodalidad.

Estos resultados tampoco mejoran los presentado en el *Challenge ProstateX-2*, donde los mejores modelos obtuvieron un coeficiente *kappa* de 0.27 [4], mientras que el mejor de los modelos monomodales solo ha obtenido un valor de 0.089.

En general se puede observar de los resultados que los modelos no son capaces de detectar correctamente las imágenes de pacientes con cáncer, y las únicas en las que aciertan son en las imágenes sanas, debido principalmente a que son la mayoría del conjunto de datos.

## 5.3 Aplicación de arquitecturas multi-modales

### 5.3.1 Arquitectura

Una vez visto que los resultados que se obtienen mediante arquitecturas monomodales no son satisfactorios, se realizan diferentes experimentos en los que se combinan las modalidades de IRM de diferentes maneras, como ya se expuso en el capítulo 4. Las cuatro diferentes arquitecturas que se prueban son las siguientes:

1. Combinación tardía con tres modelos
2. Combinación tardía con un modelo
3. Combinación temprana
4. Combinación híbrida

El esqueleto de la arquitectura para la extracción de características se selecciona de forma empírica para cada configuración entre la VGG16 y la VGG19, utilizando el modelo que mejor resultados proporcione.

En cuanto a los datos, es necesario tener en cuenta los diferentes tamaños de las imágenes, ya que las imágenes T2 recortadas tienen un tamaño de 384x253, mientras que las imágenes de ADC y DWI tienen un tamaño de 128x84. Como para poder combinar las diferentes modalidades es necesario que tengan el mismo tamaño, las imágenes de ADC y DWI se reescalan al tamaño de las imágenes de T2.

Al igual que en los modelos monomodales, se emplea un proceso de *data augmentation* a las imágenes para aumentar de manera artificial el tamaño del *dataset*. Se aplican las siguientes transformaciones:

- Rotaciones ( $\pm 5^\circ$ )
- Desplazamiento en ambos ejes (máximo un 10 %)
- *Zoom* de la imagen (máximo un 10 %)

También se asignan pesos a las diferentes clases para equilibrar la diferencia en el número de imágenes que hay de cada una (tabla 5.1).

Por último, se lista la configuración de los hiperparámetros de entrenamiento de las redes:

- *Batch size* = 16
- *Learning rate* = 0.00001
- *Dropout* = 0.25
- Optimizador = SGD - Nesterov
- Número de épocas = 400

### 5.3.2 Resultados

En esta sección se detallan los resultados obtenidos por las diferentes arquitecturas. En la tabla 5.7 se exponen los resultados para el conjunto de test de la arquitectura de combinación tardía. En la tabla 5.8 se encuentran los resultados para la arquitectura de combinación tardía monomodelo. La tabla 5.9 muestra los resultados obtenidos en la arquitectura de combinación intermedia y por último la tabla 5.10 detalla los resultados de la arquitectura de combinación híbrida.

<i>Combinación tardía multimodelo</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.50	0.57	12
Grado Alto	0	0	4
No cancerígeno	0.95	0.94	95
Valores medios ponderados:	0.86	0.87	
Valor Kappa:	0.591		
Mejor modelo:	VGG16		

**Tabla 5.7:** Resumen de resultados de la arquitectura de combinación tardía multimodelo

<i>Combinación tardía monomodelo</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.64	0.61	12
Grado Alto	0	0	4
No cancerígeno	0.93	0.92	95
Valores medios ponderados:	0.87	0.85	
Valor Kappa:	0.594		
Mejor modelo:	VGG19		

**Tabla 5.8:** Resumen de resultados de la arquitectura de combinación tardía monomodelo

<i>Combinación temprana</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.53	0.59	12
Grado Alto	0	0	4
No cancerígeno	0.94	0.93	95
Valores medios ponderados:	0.86	0.86	
Valor Kappa:	0.571		
Mejor modelo:	VGG19		

**Tabla 5.9:** Resumen de resultados de la arquitectura de combinación temprana

<i>Combinación intermedia</i>	Precisión	F1-score	Nº imágenes
Grado Bajo	0.35	0.50	12
Grado Alto	0	0	4
No cancerígeno	0.93	0.87	95
Valores medios ponderados:	0.83	0.79	
Valor Kappa:	0.387		
Mejor modelo:	VGG19		

**Tabla 5.10:** Resumen de resultados de la arquitectura de combinación intermedia

### 5.3.3 *Discusión de los resultados*

Se observa que todos los modelos que combinan distintas modalidades en el aprendizaje superan el uso de una única modalidad. Cabe destacar la alta precisión a la hora de detectar las imágenes sanas (en el mejor de los casos de 95 %) y que con estas arquitecturas se obtienen precisiones aceptables en los casos de pacientes con cáncer de próstata en estadios menos avanzados (en el mejor de los casos llegando a un 64 %).

Si bien es cierto que ningún modelo es capaz de detectar las imágenes de pacientes con cánceres de más alto grado, hay que tener en cuenta el bajo número de imágenes de este tipo que hay disponibles en la base de datos del *challenge*. Una posible estrategia para mitigar este problema es utilizar el método *leave-one-out*, que sin embargo no se utiliza en este trabajo debido a los altos tiempos de entrenamiento que requiere.

Tres de las cuatro arquitecturas funcionan mejor utilizando como modelo base VGG19, a excepción de la combinación tardía multimodal, donde se consiguen resultados levemente mejores utilizando VGG16.

Como estadístico para comparar con los resultados obtenidos en el *challenge* se usa el valor *Kappa* cuadrático ponderado, ya que es el mismo que se utiliza en el reto para medir el rendimiento de los modelos. Como se dijo anteriormente, el modelo que mejor resultado obtuvo en el *challenge* obtuvo un valor kappa de 0.27. De los experimentos llevados a cabo en este trabajo, en la mejor de las arquitecturas se obtiene un valor de kappa igual a 0.594, que es más del doble que el resultado obtenido en el *challenge*. Este resultado se obtiene en la combinación tardía de las modalidades empleando un único modelo.

Incluso en el *challenge* trataron de combinar los mejores resultados que habían obtenido, pero aún así el valor kappa solo llegó a 0.40, aún por debajo del obtenido en este trabajo, si bien es cierto que en el *challenge* se trató de diferenciar entre los cinco grupos de la escala de Gleason, mientras que este trabajo busca diferenciar únicamente entre grados bajos y altos.

Los resultados, aunque mejoran los obtenidos en el *challenge* y a la implementación de una sola modalidad, siguen lejos de poder proporcionar una herramienta de ayuda al diagnóstico que sea fiable para los médicos. El desafío de crear modelos de DL que utilicen las imágenes de resonancia magnética para el diagnóstico del cáncer de próstata es una tarea que sigue sin estar resuelta en la literatura, como se puede ver en los resultados del *challenge* [4]. Todavía no se han creado modelos que consigan diagnosticar de manera fiable el cáncer de próstata de este modo, por lo que sigue siendo necesaria la investigación en este campo.

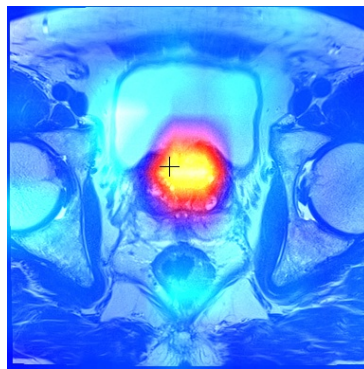
Si bien puede que no sea todavía implementable para diagnosticar algunos estadios del cáncer, si que ha sido capaz en algunos casos de tener buena precisión en grados bajos de cáncer. Teniendo en cuenta esto y con futuras mejoras, podría considerarse su proyección como método de cribado para la detección temprana del cáncer. De esta forma, aunque no se haya creado una herramienta de ayuda al diagnóstico, si que podría ayudar a detectar el cáncer en personas que todavía se encuentran en estadios tempranos de la enfermedad y esta detección temprana es crucial para el índice de supervivencia de los enfermos.

## 5.4 Interpretabilidad

Una vez entrenados los modelos, resulta de interés conocer en que zonas se fijan más para clasificar una imagen con la etiqueta de que tiene cáncer. De esta forma se interpreta mejor el comportamiento de las redes. Es con este fin que se implementan los Gradient-Weighted Class Activation Maps (Grad-CAMs), de los que se realiza una formulación detallada en el capítulo 4.

Los mapas de calor se obtienen utilizando la activación para las imágenes de resonancia magnética. Las zonas de mayor intensidad serán las zonas de mayor importancia para la red a la hora de determinar la etiqueta. Además se marca la zona que el médico determinó como la zona donde se encontraba el tumor en la próstata, para saber si la red realmente consigue delimitar la zona de tejido tumoral.

Para la implementación de los Grad-CAMs se utilizan imágenes del grupo de test de grado bajo que la red consiguiera diagnosticar de manera correcta.

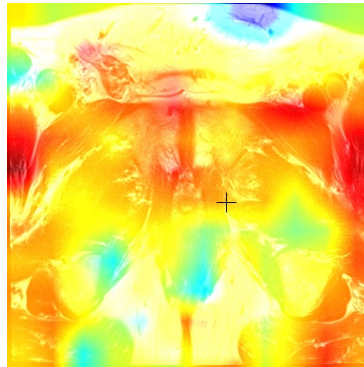


**Figura 5.3:** Ejemplo de Grad-CAMs donde no se detecta correctamente la zona de tejido canceroso

La cruz negra indica la posición que el médico diagnosticó como la zona donde se encontraba el cáncer en la imagen. De la imagen de la figura 5.3 se observa como la red consigue detectar de una manera acertada la posición de la lesión. Esto se debe a que se producen más activaciones en esas zonas.

A diferencia de la figura 5.3, en el Grad-CAM de la figura 5.4 la red no observa la zona de la lesión para asignar la etiqueta de tejido canceroso, sino que observa otros tejidos que no forman parte de la próstata.

En casos como los de la figura 5.3 los Grad-CAMs podrían llegar a ser utilizados como una herramienta de localización de tumores, sin embargo, no siempre funcionan correctamente ya



**Figura 5.4:** Ejemplo de Grad-CAMs donde no se detecta correctamente la zona de tejido canceroso

que, aunque la red acierte con la etiqueta de la imagen, no en todas las ocasiones es capaz de detectar la zona de la lesión, como es el caso de la figura 5.4.

# Conclusiones y líneas futuras

## 6.1 Conclusiones

El objetivo principal de este TFG es el desarrollo de métodos de *deep learning* (DL) capaces de clasificar imágenes de resonancia magnética según el grado de cáncer que padezcan. Para la realización de estos modelos se han propuesto diferentes arquitecturas de redes neuronales convolucionales (CNNs) que extraigan las características necesarias de las imágenes en sus distintas modalidades para poder realizar un diagnóstico.

Para este trabajo, se ha tratado de predecir si el paciente tiene cáncer o no y en el caso de tenerlo poder diferenciar entre estadios de cáncer poco avanzados y aquellos que estaban en grados más altos.

Se han propuesto en el trabajo tanto arquitecturas que utilizaban una única modalidad como arquitecturas que combinaban de diferentes maneras las modalidades de IRM adquiridas de los pacientes. Los modelos que hacían uso de una única modalidad de IRM se han mostrado ineficaces en la tarea de la detección del cáncer de próstata, con un coeficiente kappa peor que el obtenido en el *challenge ProstateX-2*.

Sin embargo, las arquitecturas de multimodalidad han obtenido resultados más precisos en la detección del cáncer, sobretodo con altas precisiones en las clasificaciones de las imágenes que no tenían cáncer y las que estaban en estadios poco avanzados. Pero al contrario que con estas categorías, para la clasificación de cánceres más avanzados no se ha mostrado eficaces.

El coeficiente kappa de la arquitectura que mejor resultado ha mostrado ha sido de 0.594, mientras que el modelo que mejor se mostró durante el *challenge* fue de 0.27, por lo que el modelo de este trabajo lo supera en rendimiento, aunque hay que tener en cuenta que en el *challenge* se trataba de diferenciar entre los cinco grupos de la escala de Gleason, mientras que en este TFG se habían agrupado en solo dos.

Aunque este modelo todavía no es lo suficientemente preciso como herramienta de ayuda al diagnóstico para el paciente, si se consigue suficiente precisión en los casos de cáncer menos avanzados

podría servir como técnica de cribado para la detección temprana del cáncer de próstata, que se ha demostrado como uno de los factores claves de supervivencia.

Además, este es un problema que actualmente no está resuelto en la literatura científica, debido a que ni los propios médicos ni los modelos de DL desarrollados hasta el momento son capaces de utilizar de una manera efectiva las imágenes de resonancia magnética para el diagnóstico del cáncer de próstata.

## 6.2 Líneas futuras

En esta sección se proponen diferentes líneas en las que poder avanzar este proyecto en el futuro.

1. Realización de modelos con una base de datos con más ejemplos. Como se ha visto, uno de los mayores problemas a la hora de afrontar este trabajo ha sido los pocos casos que se proporcionaban en la base de datos, especialmente de los tipos de cáncer más avanzados, que son donde menor precisión se ha obtenido. Por ello resultaría importante en un futuro poder reentrenar estos modelos con más datos, para comprobar si esta baja precisión se debe a la escasez de datos. Además, puede que una base de datos con más modalidades de resonancia magnética pudiera aportar más información a los modelos que fuera útil a la hora de clasificar las imágenes.
2. Segmentación de la próstata. Una de las posibles maneras en las que se podría mejorar la precisión de los modelos sería segmentando las próstatas. De esta manera se les facilitaría el trabajo ya que no habría otros tipos de tejidos que podrían confundir a los modelos, como ha ocurrido en algunos ejemplos que se han expuesto durante el trabajo. Además, si al segmentar la próstata prosiguieran los problemas de baja precisión serviría como demostración de que hay un problema inherente a la modalidad de resonancia magnética, que tal vez resulte que no es la más adecuada para el diagnóstico del cáncer de próstata.
3. Localización del cáncer en el tejido. Ya durante el trabajo se ha realizado una aproximación a posibles técnicas para la localización del cáncer en las imágenes de resonancia magnética con el uso de Gradient-Weighted Class Activation Maps (Grad-CAMs). Resultaría de interés en un futuro, si la precisión de los modelos mejora, comprobar si es posible utilizarlos con este fin.



Parte II

# Presupuesto



## Capítulo 7

# Presupuestos

El objetivo de este capítulo es el desglose de la valoración económica del proyecto. Para ello, se detallan las diferentes secciones que componen el presupuesto del TFG para el desarrollo de un modelo de DL para la gradación multimodal de cáncer de próstata.

### 7.1 Presupuestos parciales

Se desarrolla el presupuesto separándolos en concepto de mano de obra y en coste de maquinaria.

#### 7.1.1 Coste de mano de obra

Se considera costes de mano de obra a los recursos humanos necesarios para llevar este proyecto a cabo. En el caso de este trabajo la mano de obra la componen: D<sup>a</sup>. Valery Naranjo Orneo (como tutora del proyecto), D. Julio José Silva Rodríguez (como cotutor del proyecto) y D. Juan Daniel Saborido Moral (como estudiante del Grado en Ingeniería Biomédica y autor del proyecto). Estos se exponen en la figura 7.1.

Descripción	Uds.	Cantidades	Precio Unitario (€)	Total (€)
Tutora (Catedrática)	h	12	29,50	354,00
Cotutor (Doctorando)	h	36	17,20	619,20
Autor (Estudiante GIB)	h	600	12,50	7.500,00
			<b>Total</b>	8.473,30

**Tabla 7.1:** Cuadro de precios de la mano de obra

### 7.1.2 Coste de maquinaria

En cuanto al coste de maquinaria, se desglosan de manera independiente el *hardware* y el *software* empleados en el desarrollo de este proyecto.

En cuanto al *hardware*, se ha utilizado tanto un ordenador personal para la realización del proyecto como un servidor del CVBLab para el entrenamiento de los modelos de DL. Para el ordenador personal utilizado se considerará un periodo de amortización de 8 años, mientras que para el servidor del CVBLab este periodo será de 4 años. Los costes desarrollados se exponen en la tabla 7.2.

Descripción	Uds.	Cantidades	Precio Unitario (€)	Periodo de amortización (meses)	Intervalo amortizado (meses)	Total (€)
MacBook Pro retina 13"	u	1	1829,00	96	6	114,31
Intel Core i7-7700K	u	1	410,00	48	6	51,25
Nvidia Titan Xp	u	1	1349,00	48	6	168,63
<b>Total</b>						<b>334,19</b>

**Tabla 7.2:** Cuadro de precios del coste del *hardware*

En relación al software implementado en la realización del proyecto se expone a continuación el coste relacionado a ellos. Entre los diferentes programas utilizados se encuentran *Matlab* con las diferentes *toolboxes*. Para la redacción de este TFG se ha utilizado el editor de texto *Overleaf*. Para el desarrollo de los modelos de DL se han utilizado las librerías *Keras* y *TensorFlow*. En el apartado de *software* se ha considerado un periodo de amortización de 1 año y los costes se representan en la tabla 7.3.

Descripción	Uds.	Cantidades	Precio Unitario (€)	Periodo de amortización (meses)	Intervalo amortizado (meses)	Total (€)
Matlab	u	1	800	12	6	400,00
Overleaf	u	1	0,00	12	6	0,00
Image Processing Toolbox	u	1	400,00	12	6	200,00
Python 3.8.5.	u	1	0,00	12	6	0,00
Keras	u	1	0,00	12	6	0,00
Tensorflow	u	1	0,00	12	6	0,00
<b>Total</b>						<b>600,00</b>

**Tabla 7.3:** Cuadro de precios del coste del *software*

## 7.2 Presupuestos totales

Para el cálculo del precio total del proyecto se tendrá en cuenta los precios parciales expuestos anteriormente. Además, se añadirá un 13% en concepto de gastos generales y un 6% como beneficio industrial. Por último, sobre el precio bruto se aplicará el 21% de IVA. El presupuesto total se desglosa en la tabla 7.4.

Capítulos	Importe (€)
1. Coste de mano de obra	8.473,30
2. Coste de maquinaria ( <i>hardware</i> )	334,19
3. Coste de maquinaria ( <i>software</i> )	600,00
<b>Presupuesto de ejecución material</b>	<b>9.407,49</b>
13% de gastos generales	1.222,97
6% de beneficio industrial	637,83
<b>Presupuesto de ejecución por contrata</b>	<b>11.268,29</b>
<b>21% de IVA</b>	<b>2.366,34</b>
<b>Presupuesto total</b>	<b>13.634,63</b>

**Tabla 7.4:** Presupuesto total



# Bibliografía

- [1] *The top 10 causes of death* (vid. pág. 3).
- [2] Hyuna Sung y col. “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries”. En: *CA: A Cancer Journal for Clinicians* 71.3 (mayo de 2021), págs. 209-249. ISSN: 0007-9235. DOI: 10.3322/caac.21660 (vid. págs. 3, 4).
- [3] OMS. *Cancer Tomorrow* (vid. pág. 3).
- [4] Samuel G Armato y col. “PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images "PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images””. En: *J. Med. Imag* 5.4 (2018), pág. 44501. DOI: 10.1117/1.JMI.5.4.044501 (vid. págs. 5, 33, 53, 56).
- [5] Tal Geva. “Magnetic resonance imaging: Historical perspective”. En: *Journal of Cardiovascular Magnetic Resonance* 8.4 (2006), págs. 573-580. ISSN: 10976647. DOI: 10.1080/10976640600755302 (vid. págs. 7, 8).
- [6] Eric Van Reeth y col. “Super-resolution in magnetic resonance imaging: A review”. En: *Concepts in Magnetic Resonance Part A: Bridging Education and Research* 40 A.6 (2012), págs. 306-325. ISSN: 15466086. DOI: 10.1002/cmr.a.21249 (vid. pág. 8).
- [7] Paul Suetens. *Fundamentals of Medical Imaging*. 2.<sup>a</sup> ed. Cambridge: Cambridge University Press, 2009, págs. 64-104. ISBN: 9780511596803. DOI: 10.1017/CB09780511596803 (vid. págs. 9, 11-13, 16, 17).
- [8] Henning Holle. “The Comprehension of Co-speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies”. En: 2007.March (2007) (vid. pág. 11).

- [9] Nadine Barrie Smith y Andrew Webb. *Introduction to Medical Imaging*. 2010. DOI: 10.1017/cbo9780511760976 (vid. pág. 13).
- [10] Jaume Gili. “Introduccion biofisica a la Resonancia Magnetica en Neuroimagen”. En: *Books Medicos 03-2* (2016), pág. 229 (vid. págs. 13-17).
- [11] Ryu Matsuo y col. “Magnetic resonance imaging in breath-hold divers with cerebral decompression sickness”. En: *Case Reports in Neurology* 6.1 (2014), págs. 23-27. ISSN: 1662680X. DOI: 10.1159/000357169 (vid. pág. 19).
- [12] Richard Drake, Wayne Vogl y Adam Mitchell. *Gray Anatomía para estudiantes*. 2005 (vid. págs. 19, 20).
- [13] John E. Hall y Arthur C. Guyton. *Guyton y Hall Tratado de Fisiología médica - John E. Hall - 13° ed. 2016*. 2016 (vid. págs. 21, 22).
- [14] Noa Rivlin y col. *Mutations in the p53 tumor suppressor gene: Important milestones at the various steps of tumorigenesis*. Abr. de 2011. DOI: 10.1177/1947601911408889 (vid. pág. 21).
- [15] David G Bostwick y col. “Human Prostate Cancer Risk Factors”. En: *Cancer* 101.10 (2004), págs. 2371-490. DOI: 10.1002/cncr.20408 (vid. pág. 21).
- [16] J Kellogg Parsons. “Benign Prostatic Hyperplasia and Male Lower Urinary Tract Symptoms: Epidemiology and Risk Factors”. En: (2010). DOI: 10.1007/s11884-010-0067-2 (vid. pág. 22).
- [17] Paolo Boffetta y Pierre Hainaut. *Encyclopedia of cancer*. 2018. DOI: 10.5860/choice.40-1906 (vid. págs. 22-24).
- [18] Aman Rana y col. *High Accuracy Tumor Diagnoses and Benchmarking of Hematoxylin and Eosin Stained Prostate Core Biopsy Images Generated by Explainable Deep Neural Networks*. Inf. téc. 2019. arXiv: 1908.01593v1 (vid. pág. 23).
- [19] Joost N Kok y col. “Artificial intelligence: definition, trends, techniques, and cases”. En: *Artificial intelligence* 1 (2009), págs. 270-299 (vid. pág. 25).
- [20] Issam El Naqa y Martin J Murphy. “Machine Learning in Radiation Oncology”. En: *Machine Learning in Radiation Oncology* (2015), págs. 3-11. DOI: 10.1007/978-3-319-18305-3 (vid. pág. 25).
- [21] W. J. Zhang y col. “On Definition of Deep Learning”. En: *World Automation Congress Proceedings 2018-June* (2018), págs. 232-236. ISSN: 21544832. DOI: 10.23919/WAC.2018.8430387 (vid. pág. 25).
- [22] IBM Cloud Education. *Neural Networks*. 2020 (vid. pág. 26).



- [23] Bernard Widrow y Michael A. Lehr. “30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation”. En: *Proceedings of the IEEE* 78.9 (1990), págs. 1415-1442. ISSN: 15582256. DOI: 10.1109/5.58323 (vid. pág. 25).
- [24] Fernando Sancho Caparrini. *Redes Neuronales: una visión superficial*. 2019 (vid. pág. 26).
- [25] David E Rumelhart, Geoffrey E Hinton y Ronald J Williams. *Learning internal representations by error propagation*. Inf. téc. California Univ San Diego La Jolla Inst for Cognitive Science, 1985 (vid. pág. 27).
- [26] Sebastian Ruder. *An overview of gradient descent optimization algorithms \**. Inf. téc. arXiv: 1609.04747v2 (vid. pág. 27).
- [27] Ayoosh Kathuria. “Intro to optimization in deep learning: Gradient Descent”. En: (2018) (vid. pág. 28).
- [28] Fernando Giménez Palomares, Juan Antonio Monsoriu Serrá y Elena Alemany Martínez. “Aplicación de la convolución de matrices al filtrado de imágenes”. En: *Modelling in Science Education and Learning* 9.1 (2016). ISSN: 1988-3145. DOI: 10.4995/mse1.2016.4524 (vid. pág. 28).
- [29] Md Zahangir Alom y col. “A state-of-the-art survey on deep learning theory and architectures”. En: *Electronics (Switzerland)* 8.3 (2019), págs. 1-67. ISSN: 20799292. DOI: 10.3390/electronics8030292 (vid. pág. 29).
- [30] Abien M Fred Agarap. *Deep Learning using Rectified Linear Units (ReLU)*. Inf. téc. arXiv: 1803.08375v2 (vid. pág. 30).
- [31] Arden Dertat. *Applied Deep Learning - Part 4: Convolutional Neural Networks / by Arden Dertat / Towards Data Science*. 2017 (vid. pág. 31).
- [32] Siyue Wang y col. “Defensive Dropout for Hardening Deep Neural Networks under Adversarial Attacks”. En: (2018). DOI: 10.1145/3240765.3264699. arXiv: 1809.05165v1 (vid. pág. 31).
- [33] Octavian Sabin Tătaru y col. “diagnostics Artificial Intelligence and Machine Learning in Prostate Cancer Patient Management-Current Trends and Future Perspectives and Technology from Târgu Mures”. En: 11 (2021), pág. 354. DOI: 10.3390/diagnostics11020354 (vid. pág. 32).
- [34] Eugene Chian. *3D CNN Classification of Prostate Cancer on PROSTATEx-2 / by Hugesene / Towards Data Science*. 2019 (vid. pág. 32).
- [35] Yixuan Yuan y col. “Prostate cancer classification with multiparametric MRI transfer learning model”. En: (2018). DOI: 10.1002/mp.13367 (vid. pág. 33).

- [36] R N J Graham, R W Perriss y A F Scarsbrook. “DICOM demystified: A review of digital file formats and their use in radiological practice”. En: (). DOI: 10.1016/j.crad.2005.07.003 (vid. pág. 35).
- [37] Asociación Americana de Físicos en Medicina. *PROSTATEx-2 Challenge* (vid. pág. 36).
- [38] Max Ferguson y col. “Automatic localization of casting defects with convolutional neural networks”. En: December (2018), págs. 1726-1735. DOI: 10.1109/bigdata.2017.8258115 (vid. pág. 40).
- [39] Lyndon Chan y col. “HistoSegNet: Semantic segmentation of histological tissue type in whole slide images”. En: *Proceedings of the IEEE International Conference on Computer Vision 2019-October.Iccv* (2019), págs. 10661-10670. ISSN: 15505499. DOI: 10.1109/ICCV.2019.01076 (vid. pág. 44).
- [40] Steven M Beitzel. *ON UNDERSTANDING AND CLASSIFYING WEB QUERIES*. Inf. téc. 2006 (vid. pág. 49).
- [41] J CARLETTA. “Assessing Agreement on Classification Tasks: The Kappa Statistic”. En: *Squibs and Discussions: Assessing Agreement on Classification Tasks: The Kappa Statistic* 22.2 (1996). ISSN: 0891-2017 (vid. pág. 50).