



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA

TRABAJO FIN DE MASTER EN INGENIERÍA BIOMEDICA

**DISEÑO DE UNA SEGMENTACIÓN
AUTOMÁTICA DE GLIOMAS EN IMÁGENES
DE RESONANCIA MAGNÉTICA
MULTIMODAL CON REDES NEURONALES
CONVOLUCIONALES PROFUNDAS**

AUTOR: ALFREDO TORREGROSA LLORET

TUTOR: JUAN MIGUEL GARCÍA GÓMEZ

COTUTOR: JAVIER JUAN ALBARRACÍN

Curso Académico: 2017-18

Agradecimientos

En primer lugar, quiero aprovechar la ocasión para agradecer a Juanmi todo el apoyo y la ayuda que me ha brindado durante los cinco años que he tenido el placer de ser su alumno. Gracias por enseñarme que para formar a un buen profesional primero se debe formar a una persona, y por proporcionarme la oportunidad de llevar a cabo este motivador proyecto. Asimismo, no me podría olvidar de mi cotutor Javi, el cual me introdujo en el apasionante mundo del *deep learning* y cuya ayuda ha sido fundamental para la consecución de este trabajo, así como a todo el equipo del BDSLab que ha velado por mi integración y bienestar en todo momento.

Me gustaría también dar las gracias a toda mi familia, cuyo apoyo ha sido fundamental para impulsarme a adelante en las situaciones desafortunadas que hemos padecido en los últimos tiempos. Del mismo modo, todo esto no hubiese sido posible de no ser por Sofía, el pilar que me sostiene cada día y mi ejemplo a seguir en cada momento.

Finalmente, como no podía ser de otra forma, quiero dedicarle este proyecto a mi tío José. Has sido la motivación que me impulsó a elegir este camino y la razón que me convenció de que mi lugar se encontraba allí donde mi trabajo pudiera ser útil para ayudar a aquella gente que como tú tuvo la mala suerte de obtener el boleto premiado de una lotería en la que nadie quiere participar.

Ojalá este trabajo sirva para evitar algún día que alguien comparta tu cruel destino.

Ojalá este trabajo no tuviese que estar dedicado a ti.

Donde quiera que estés, te quiero tío.

Resumen

Los tumores cerebrales son una de las neoplasias que más afectan a la población a nivel mundial. Estos suponen graves consecuencias para la salud del paciente, así como en el aspecto social y económico. De entre todos los tipos de tumores cerebrales, los gliomas suponen la mayor parte de los mismos y son habitualmente clasificados en Glioma de Alto Grado (HGG) y Glioma de Bajo Grado (LGG). A pesar de los grandes avances médicos y tecnológicos, el diagnóstico y pronóstico de estos tumores continua siendo pobre, especialmente en los HGG, para los que la esperanza de vida es inferior a los dos años y cuyo tratamiento, extremadamente agresivo, puede causar severos problemas al paciente.

El diagnóstico de estas patologías ha mejorado notablemente gracias a la introducción en la práctica clínica de la Imagen por Resonancia Magnética (IRM) junto a algoritmos y sistemas de ayuda al diagnóstico que están demostrando un gran potencial, no solo para mejorar este, sino también el tratamiento de este tipo de enfermedades. Este es el caso de la segmentación de tumores cerebrales, donde las Redes Neuronales Convolucionales (CNN) han demostrado la capacidad de obtener resultados a la altura de los logrados por radiólogos expertos, tarea que anteriormente no podía ser introducida en la práctica clínica debido al tiempo que esta requería, difícilmente disponible por parte del facultativo.

El presente proyecto aborda el desarrollo e implementación de distintas metodologías basadas en el uso de CNN para la automatización de la tarea de segmentación de gliomas y de sus distintas componentes tumorales, clasificadas en Núcleo tumoral Sin Realce de contraste (NSR), Edema Peritumoral (EP) y Núcleo tumoral con Realce de contraste (NR). Para ello se hizo uso del conjunto de datos del *BRAIn Tumor Segmentation* (BRATS) 2017, constituido por un total de 285 estudios de IRM de HGG y LGG, los cuales incluían secuencias potenciadas en T1, previo a la administración de un agente de contraste y posterior a la misma, y T2, con y sin Atenuación de FLuidos por Inversión Recuperación (FLAIR). Los resultados obtenidos fueron analizados mediante la plataforma proporcionada por la organización del BRATS y comparados con los resultados reportados en el estado del arte. De este modo, mediante el uso de una red U-Net se logró un DICE medio de 0.7331, 0.8624 y 0.7267, para el NR, tumor total y núcleo tumoral, respectivamente. Así, la inclusión de este tipo de metodologías favorece tanto al paciente como al facultativo posibilitando la realización de diagnósticos cuantitativos más precisos.

Palabras clave: Neuroimagen, glioma, aprendizaje profundo, segmentación, imagen médica.

Resum

Els tumors cerebrals són una de les neoplàsies que més afecten la població a nivell mundial. Estos suposen greus conseqüències per a la salut del pacient, així com en l'aspecte social i econòmic. D'entre tots els tipus de tumors cerebrals, els gliomes suposen la major part dels mateixos i són habitualment classificats en Glioma d'Alt Grau (HGG) i Glioma de Baix Grau (LGG). A pesar dels grans avanços mèdics i tecnològics, el diagnòstic i pronòstic d'estos tumors continúa sent pobre, especialment en els HGG, per als que l'esperança de vida és inferior als dos anys i el tractament dels quals, extremadament agressiu, pot causar severos problemes al pacient.

El diagnòstic d'estes patologies ha millorat notablement gràcies a la introducció en la pràctica clínica de la Imatge per Ressonància Magnètica (IRM) junt amb algoritmes i sistemes d'ajuda al diagnòstic que estan demostrant un gran potencial, no sols per a millorar este, sinó també el tractament d'este tipus de malalties. Este és el cas de la segmentació de tumors cerebrals, on les Xarxes Neuronals Convolucionales (CNN) han demostrat la capacitat d'obindre resultats a l'altura dels aconseguits per radiòlegs experts, tasca que anteriorment no podia ser introduïda en la pràctica clínica a causa del temps que esta requeria, difícilment disponible per part del facultatiu.

El present projecte aborda el desenrotllament i implementació de distintes metodologies basades en l'ús de CNN per a l'automatització de la tasca de segmentació de gliomes i dels seus distintes components tumorals, classificades en Nucli tumoral Sense Realç de contrast (NSR), Edema Peritumoral (EP) i Nucli tumoral amb Realç de contrast (NR). Per a això es va fer ús del conjunt de dades del BRAIn Tumor Segmentation (BRATS) 2017, constituït per un total de 285 estudis d'IRM de HGG i LGG, els quals inclouen seqüències potenciades en T1, previ a l'administració d'un agent de contrast i posterior a la mateixa, i T2, amb i sense Atenuació de FLuids per Inversió Recuperació (FLAIR). Els resultats obtinguts van ser analitzats per mitjà de la plataforma proporcionada per l'organització del BRATS i comparats amb els resultats reportats en l'estat de l'art. D'esta manera, per mitjà de l'ús d'una xarxa U-Net es va aconseguir un DICE mig de 0.7331, 0.8624 i 0.7267, per al NR, tumor total i nucli tumoral, respectivament. Així, la inclusió d'este tipus de metodologies afavorix tant el pacient com al facultatiu possibilitant la realització de diagnòstics quantitativs més precisos.

Paraules clau: Neuroimagen, glioma, aprenentatge profund, segmentació, imatge mèdica.

Abstract

Brain tumors are one of the neoplasms that most affect the population worldwide. These suppose serious consequences for the health of the patient, as well as in the social and economic aspect. Among all types of brain tumors, gliomas suppose most of them and are usually classified into High Grade Glioma (HGG) and Low Grade Glioma (LGG). Despite the great medical and technological advances, the diagnosis and prognosis of these tumors continues to be poor, especially in HGG, for which the life expectancy is less than two years and whose treatment, extremely aggressive, can cause severe problems to the patient.

The diagnosis of these pathologies has improved markedly thanks to the introduction into clinical practice of Magnetic Resonance Imaging (MRI) along with algorithms and diagnostic aid systems that are demonstrating great potential, not only to improve this, but also the treatment of this type of diseases. This is the case of the segmentation of brain tumors, where the Convolutional Neural Networks (CNN) have demonstrated the ability to obtain results at the height of those achieved by expert radiologists, a task that previously could not be introduced in clinical practice due to the time that this required, hardly available by the doctor.

The present project addresses the development and implementation of different methodologies based on the use of CNN for the automation of the task of segmentation of gliomas and their different tumor components, classified as non-enhancing tumor (NSR), peritumoral edema (EP) and enhancing tumor (NR). To this end, the data set of the BRAin Tumor Segmentation (BRATS) 2017 was used, consisting of a total of 285 MRI studies of HGG and LGG, which included T1-weighted sequences, prior to the administration of a contrast agent and subsequent to it, and T2, with and without Fluid Attenuation by Inversion Recovery (FLAIR). The results obtained were analyzed through the platform provided by the BRATS organization and compared with the results reported in the state of the art. In this way, through the use of a U-Net network, an average DICE of 0.7331, 0.8624 and 0.7267 was achieved for the NR, whole tumor and tumor core, respectively. Thus, the inclusion of this type of methodologies favors both the patient and the physician, making it possible to carry out more precise quantitative diagnoses.

Key words: Neuroimaging, glioma, deep learning, segmentation, medical image.

Índice general

Resumen	III
Resum	VII
Abstract	IX
Índice general	XI
I Memoria	1
1 Introducción	3
1.1 Anatomía del cerebro humano	3
1.2 Tumores cerebrales	5
1.3 Aprendizaje máquina y aprendizaje profundo en radiología	7
1.4 Motivación	11
1.5 Hipótesis	12
1.6 Objetivos	12
1.7 Resumen de los capítulos restantes	12
2 Estado del arte	13
2.1 Algoritmos de segmentación de tumor cerebral	13
2.2 Métodos generativos probabilísticos	14
2.3 Métodos discriminativos	15
3 Materiales	17
3.1 Conjunto de datos <i>BRAIn Tumor Segmentation</i> (BRATS) 2017	17
3.2 Equipamiento <i>hardware</i>	21
3.3 Equipamiento <i>software</i>	21

4 Métodos	23
4.1 Arquitecturas de redes utilizadas	23
4.2 Agregación de modelos	30
4.3 Preparación de los datos	34
4.4 Estrategia de entrenamiento	38
4.5 Experimentos realizados	39
4.6 Evaluación	42
5 Resultados y discusiones	45
5.1 Segmentación de los casos de LGG	45
5.2 Segmentación conjunta de los casos de LGG y HGG	47
5.3 Discusiones	55
6 Conclusiones y líneas futuras	59
6.1 Conclusiones	59
6.2 Líneas futuras	60
Bibliografía	61
Índice alfabético	67
Lista de figuras	67
Lista de tablas	69
Lista de acrónimos	71
II Presupuesto	73
7 Presupuesto desglosado	75
7.1 Objetivo y estructura	75
7.2 Coste de recursos humanos	76
7.3 Coste de equipamiento	78
8 Resumen del presupuesto	79
8.1 Presupuesto de ejecución material	79
8.2 Presupuesto de ejecución por contrata	79
8.3 Presupuesto final del proyecto	80

Parte I

Memoria

Capítulo 1

Introducción

En el presente capítulo se introducirán los conceptos relacionados con la anatomía y fisiología del cerebro humano, así como de la fisiopatología de los tumores cerebrales. Por otro lado se expondrá la motivación que inspiró el trabajo, la hipótesis que trata de comprobar y se resumirá el resto de secciones para facilitar la lectura del documento.

1.1 Anatomía del cerebro humano

El sistema nervioso humano se encuentra anatómicamente y fisiológicamente dividido en Sistema Nervioso Central (SNC) y Sistema Nervioso Periférico (SNP). Más específicamente, el SNC está comprendido por el cerebro y la médula espinal, estructuras que se encuentran protegidas por toda una estructura ósea, compuesta por el cráneo y el canal vertebral, además de las meninges y otros espacios normales relacionados [1]. El cerebro es el órgano encargado de almacenar información, generar pensamientos y determinar las reacciones ante toda clase de estímulos que recibe el individuo, generando señales que se transmiten al resto del organismo para desencadenar las mismas [2].

1.1.1 Sustancia gris y sustancia blanca del cerebro

El cerebro está constituido principalmente por dos tipos de sustancias que deben su nombre a su apariencia visual. Estas se tratan de la sustancia gris y de la sustancia blanca, las cuales presentan ciertas características, además de la visual, como por ejemplo la diferente vascularización del tejido.

La sustancia gris se encuentra situada en la superficie cerebral y cerebelar, así como el en el núcleo basal, situado en la parte superior del tronco encefálico: diencefalo, formado principalmente por el tálamo, el hipotálamo y el epitálamo; núcleo del tronco cerebral y el núcleo cerebelar profundo. Estas regiones se encuentran esquematizadas en la figura 1.1. Asimismo, la materia gris se encuentra compuesta por cuerpos neuronales, células gliales, neuropilo y capilares, presentando así un suministro vascular que cuadruplica el de la sustancia blanca. Por otro lado, la sustancia blanca se encuentra situada principalmente en las regiones cerebrales subcorticales y profundas, compuesta principalmente por axones neuronales mielinizados por los cuales se transmiten señales

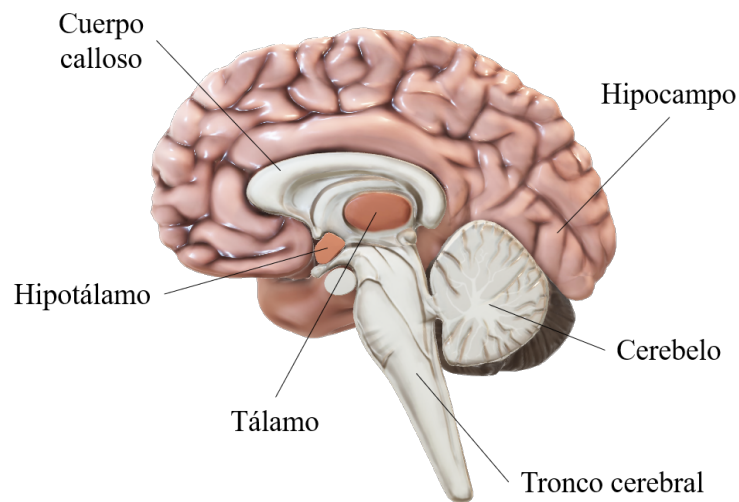


Figura 1.1: Esquema de las distintas regiones cerebrales.

desde y hacia distintas partes de la sustancia gris cerebral [1]. En la figura 1.2 puede observarse como se organiza la sustancia gris y la sustancia blanca en el cerebro. A la izquierda de la figura se presenta un corte transversal de un cerebro de una adquisición de Resonancia Magnética (RM) potenciada en T1, en el centro se encuentra el mismo corte en una adquisición de RM potenciada en T2 y a la derecha se presenta un esquema de la organización de la sustancia blanca y gris en las imágenes.

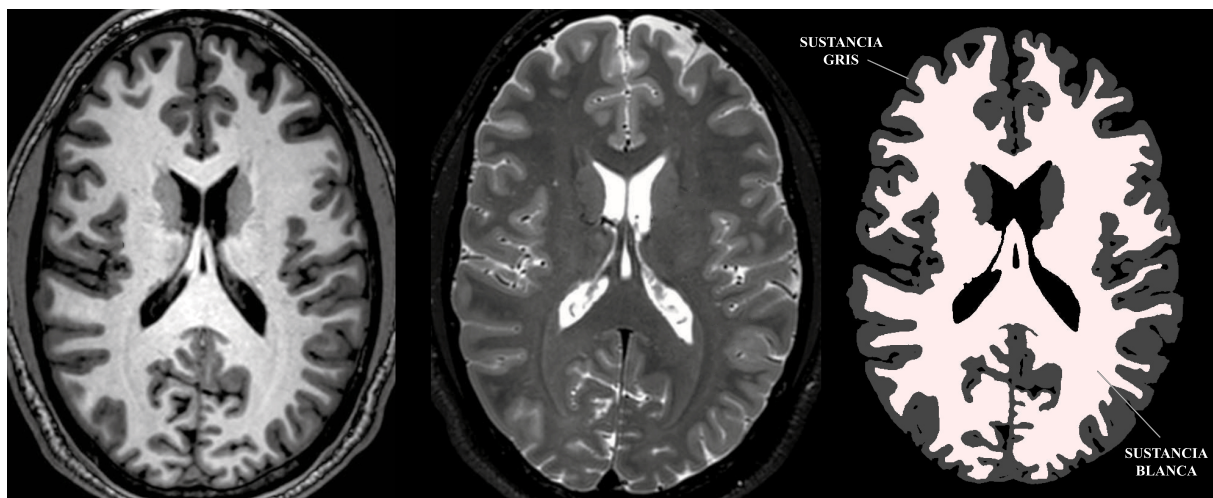


Figura 1.2: Organización de la sustancia blanca y sustancia gris en el cerebro. En la figura se puede observar una imagen potenciada en T1, T2 y un esquema de la distribución de la sustancia gris y blanca en el cerebro, respectivamente. Adaptada de [1].

1.1.2 *Células del sistema nervioso central*

El SNC está compuesto principalmente por neuronas y células gliales, como astrocitos y oligodendrocitos, las cuales cumplen una función de soporte de aquellas. Estas células provienen de las células madre del SNC y han sufrido un proceso de diferenciación, pasando por multitud de tipos celulares progenitores hasta alcanzar su tipo final. Cada tipo se encuentra diferenciado por su morfología, función y expresión de distintos marcadores, como por ejemplo la proteína acética fibrilar glial que es expresada por los astrocitos [3]. Las funciones principales de cada una de estas células son las siguientes:

- **Neuronas.** Se tratan de una células compuestas por dos partes, soma y axón, las cuales son las encargadas de recibir y transmitir señales a distintas partes del organismo.
- **Astrocitos.** Tipo celular más abundante en el conjunto de las células gliales, desempeñan multitud de funciones entre las cuales se encuentra la de transportar nutrientes hacia las neuronas.
- **Oligodendrocitos.** Perteneciente al conjunto de las células de la glía, estas se encargan de la protección de los axones neuronales formando alrededor de ellos una vaina de mielina, la cual aumenta la velocidad de transmisión de las señales eléctricas.

1.2 Tumores cerebrales

Los gliomas son un tipo de tumores que provienen del interior del SNC y que tienen características similares a las células de la glía, como por ejemplo astrocitos, oligodendrocitos o sus precursores [3], las cuales han sido descritas en subsección 1.1.2. Las primeras personas en acuñar este término fueron Harvey Cushing y Percival Bailey [4], los cuales realizaron un gran esfuerzo en clasificar y describir los distintos tumores cerebrales conocidos hasta la fecha.

Desde las primeras aportaciones de Bailey al campo de la neurociencia y, en concreto, al ámbito de los tumores cerebrales, la clasificación de estos neoplasmas ha estado en continua evolución y se han sugerido distintos sistemas de clasificación de los mismos. Si bien es cierto que los aportes de Bailey supusieron un gran avance en el campo de la neurociencia, actualmente la clasificación que más peso ha cobrado es la de la Organización Mundial de la Salud (OMS) [5], frente a otras clasificaciones como las de Kernohan [6] o St Anne-Mayo [7]. Es por ello que en esta sección estudiaremos los diferentes tipos tumorales en función de la clasificación de la OMS, así como el impacto socioeconómico que estos suponen en la actualidad.

1.2.1 *Clasificación según la OMS*

La gradación de los tumores cerebrales según la OMS está basada en la clasificación de los comportamientos biológicos y aspecto histológico de los distintos tumores cerebrales [8]. De este modo, esta clasificación se trata de una escala de malignidad en la cual encuadrar los diferentes gliomas conocidos hasta la fecha y consta de cuatro grados, recogidos por la gradación realizada por la OMS, los cuales son descritos a continuación:

- **Grado I.** Tumores benignos que presentan un crecimiento lento y están asociados a una alta supervivencia a largo plazo.

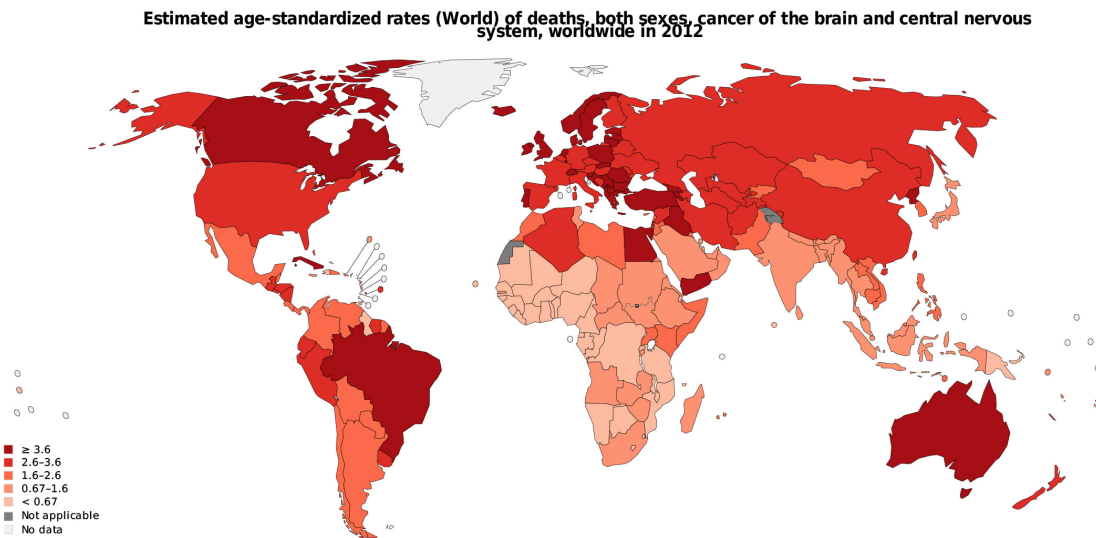


Figura 1.3: Tasas estimadas de casos incidentes cáncer de cerebro y del sistema nervioso central, para ambos sexos, por edad, en todo el mundo en 2012. Fuente [11], disponible en <http://gco.iarc.fr/today/home>.

- **Grado II.** Tumores que presentan un crecimiento relativamente lento y que en ocasiones pueden convertirse en tumores de grado más alto.
- **Grado III.** Tumores malignos y a menudo se convierten en tumores de grado más alto.
- **Grado IV.** Tumores malignos que se reproducen rápidamente y que presentan una elevada agresividad.

Habitualmente se hace referencia a los tumores que presentan un crecimiento lento, es decir, aquellos etiquetados como Grado I o II, son definidos como LGG, mientras que los que presentan una mayor agresividad (Grado III o IV) son considerados HGG. No obstante, en muchas ocasiones la OMS desaconseja esta generalización ya que dentro de estos grupos existen multitud de tumores biológica y estructuralmente distintos [5], tales como meningiomas, tumores astrocíticos o tumores de la región pineal.

1.2.2 Impacto socioeconómico

Los tumores cerebrales se tratan de una neoplasia que afecta a toda la población mundial, con mayor tasa de incidencia en los países desarrollados [9], tal y como se puede apreciar en la figura 1.3. Más concretamente, la mayor parte de los tumores cerebrales se trata de los llamados gliomas, suponiendo alrededor del 70% del total de todos los tumores cerebrales [10].

A pesar de todos los avances médicos y los esfuerzos en investigación realizados, el diagnóstico de estos pacientes continua siendo pobre, con una esperanza de vida de dos años o menos para HGG, los cuales requieren de un tratamiento agresivo instantáneo, el cual no está exento de problemas asociados, y algo mayor en los casos de LGG, en los cuales el tratamiento agresivo es retrasado todo lo posible [12]. Este tratamiento consiste en la extirpación quirúrgica de la lesión, tratando de no afectar las zonas elocuentes, seguida de quimioterapia [13].

Asimismo, los costes asociados a la enfermedad son elevados, según se recoge en un estudio realizado entre los años 1990 y 2010, en el cual se estima que el coste por caso es de alrededor

a 20 538€ en España [14]. Otros estudios aproximan que los costes indirectos¹ asociados a los casos de glioblastoma, neoplasia englobada dentro de los HGG, alcanzan cifras alrededor de los 111 926.90€ [13].

Además, los pacientes diagnosticados con un tumor cerebral ven considerablemente mermada su calidad de vida debido a la aparición de fenómenos de ansiedad y depresión, llegando a disminuir bruscamente en aquellos pacientes que sufren una recidiva del tumor tras el tratamiento [15]. A pesar de que no se ha demostrado que el ánimo del paciente se correlacione con el éxito de la terapia, este es un aspecto importante a tener en cuenta a la hora de tratar al paciente y dirigir la terapia de la forma que menor detrimento cause al mismo.

1.3 Aprendizaje máquina y aprendizaje profundo en radiología

En medicina y, más concretamente, en radiología ha surgido una gran necesidad de incluir nuevas técnicas, no solo para liberar carga de trabajo al profesional sanitario sino también para mejorar el seguimiento del paciente buscando un diagnóstico temprano atendiendo al pronóstico y la respuesta al tratamiento. Además, dados los avances en las tecnologías médicas utilizadas hoy en día en la práctica clínica diaria, actualmente se dispone de multitud de parámetros y variables obtenidos directa e indirectamente del paciente, los cuales son, en la mayoría de los casos, desaprovechados por la inexistencia de un algoritmo adecuado para la explotación de este tipo de información.

En este contexto surgen las técnicas de aprendizaje máquina, de las que el aprendizaje profundo es el estado del arte actual. El aprendizaje máquina se trata de una rama de la llamada Inteligencia Artificial (IA), la cual comprende multitud de técnicas que permiten a un sistema informático o algoritmo aprender de su entorno, adaptarse a los cambios en el mismo y proporcionar soluciones ante un determinado problema. Por ello, se puede definir el aprendizaje máquina como el proceso de optimización de un criterio de desempeño usando datos o experiencia pasada, para finalmente obtener un modelo para ser utilizado para solventar el problema para el cual fue diseñado [16]. La mayoría de los algoritmos de aprendizaje máquina requieren de una fase previa de extracción de características con las cuales alimentar el modelo que se trata de diseñar. Este hecho convierte a los modelos extremadamente dependientes de esta fase, por lo que es necesario un buen procesamiento previo de los datos. Del mismo modo, existe un tipo de métodos que no requieren de esta etapa previa, los cuáles reciben el nombre de aprendizaje por representación, donde el algoritmo utiliza las mejores características para clasificar los datos proporcionados [17].

De este modo, el aprendizaje profundo no es más que una especialización del llamado aprendizaje por representación, compuesto principalmente por el uso de redes neuronales diseñadas con multitud de capas junto con una serie de filtros convolucionales con parámetros libres. Así, el algoritmo es capaz de aprender el conjunto de características que reflejan una jerarquía de estructuras en los datos que permiten expresar representaciones complejas mediante representaciones más simples [17]. Tal y como se puede observar en figura 1.4, las Redes Neuronales Convolucionales (CNN) diseñadas para una determinada tarea de clasificación son capaces de extraer en las capas más superficiales características sencillas, tales como intensidades o bordes, y finalmente reconocer formas y objetos más complejos en sus capas más profundas. En la sección

¹Asociados a la pérdida de productividad del paciente.

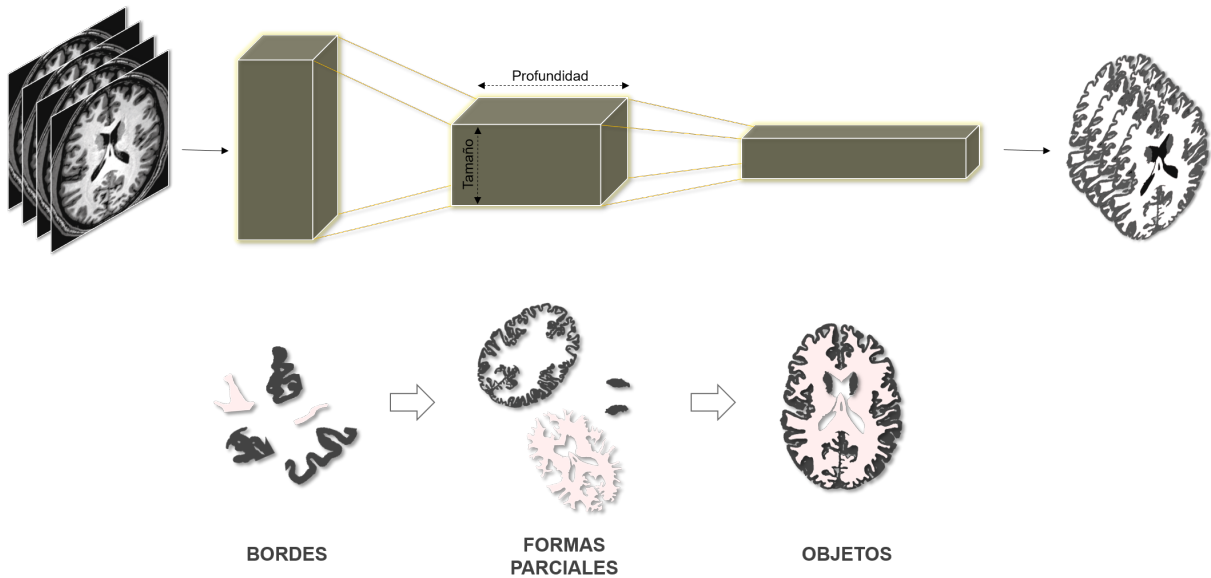


Figura 1.4: Esquema de un ejemplo de CNN para segmentación de sustancia gris y sustancia blanca a partir de una imagen T1 postcontraste.

subsección 1.3.1 se introducirá brevemente el funcionamiento de esas redes. En los últimos años el aprendizaje profundo ha sufrido un gran aumento en cuanto a número de adeptos y aplicaciones se refiere debido principalmente a las nuevas posibilidades de computación en paralelo que existen hoy en día, a la existencia de nuevos conjuntos de datos etiquetados de libre acceso y la mejora en técnicas de entrenamiento y optimización [17].

1.3.1 Introducción a las CNN.

En la base de las CNN se encuentran las redes neuronales artificiales, las cuales tienen como pieza principal las llamadas neuronas artificiales o nodos. Cada una de estas neuronas es un clasificador logístico, el cual es entrenado para encontrar aquellos pesos (W) y sesgos (b) que mejor predicen el problema por el cual se diseña la red. De este modo, la neurona recibe una entrada (X) y aplica una transformación lineal a esta, la cual se rige por la Ecuación 1.1, para obtener unas determinadas predicciones (y).

$$W \cdot X + b = y \quad (1.1)$$

No obstante, estas predicciones no son más que pesos que indican la probabilidad de que la entrada pertenezca a una determinada clase. Es por ello que es necesario convertir estos pesos a probabilidades de pertenencia a una determinada clase en la última capa de una red de clasificación. Para ello se utiliza la función *softmax*, la cual se encuentra definida en la Ecuación 1.2, donde $S(y_i)$ hace referencia a la probabilidad de pertenencia a una determinada clase i .

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (1.2)$$

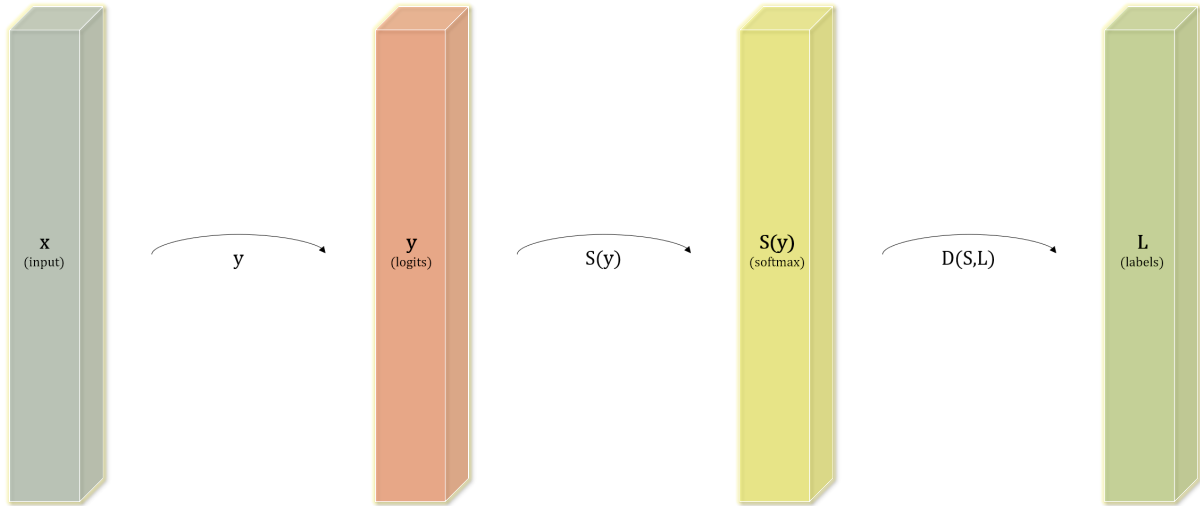


Figura 1.5: Resumen del proceso de obtención de la segmentación mediante el uso de un modelo lineal.

Una vez obtenidas las probabilidades de cada una de las entradas, estas son codificadas de forma binaria a través del llamado *one hot encoding*, asignando el número 1 a aquella clase con mayor probabilidad y 0 al resto de clases. Finalmente, se mide el desempeño del modelo, es decir, lo bien que el modelo ha predicho las clases a partir de las entradas de entrenamiento. La forma más común de realizar esta tarea en clasificación es mediante la llamada *Cross-Entropy*, la cual mide la distancia entre las etiquetas predichas y las etiquetas reales mediante la Ecuación 1.3. Todo el proceso desde que se reciben la entrada (X) hasta que se obtienen las etiquetas predichas por el modelo (L) se encuentra resumido en la figura 1.5.

$$D(S, L) = - \sum_i L_i \log(S_i) \quad (1.3)$$

De este modo, el objetivo de la etapa de entrenamiento de cualquier modelo se trata del de minimizar la *Cross-Entropy*. En la práctica se trata de minimizar el promedio de todas estas distancias ($D(S(y_i), L_i)$). Este promedio recibe el nombre de función de coste o *Loss* (\mathcal{L}) y se encuentra descrito en Ecuación 1.4.

$$\mathcal{L} = \frac{1}{N} \sum_i D(S(y_i), L_i) \quad (1.4)$$

En este proceso de optimización de la *Loss* es habitual el uso del gradiente descendiente estocástico, basado en el método de gradiente descendiente, como optimizador, el cual se diferencia de este en que divide los datos en pequeñas piezas aleatorias para computar la *Loss*. Así, el proceso de optimización conlleva más iteraciones para conseguir el mismo resultado pero en mucho menos tiempo, lo cual ha supuesto un gran avance y un gran aliciente para el uso de este tipo de redes.

Así, el modelo tan solo será capaz de representar relaciones lineales entre las entradas y las salidas. No obstante, si se aumenta la complejidad de la red y se añaden más capas y entre medio de ellas se introduce una función no lineal, por ejemplo una *Rectified Linear Units* (ReLU), la red adquiere la capacidad de interpretar relaciones no lineales. La función ReLU, descrita en la

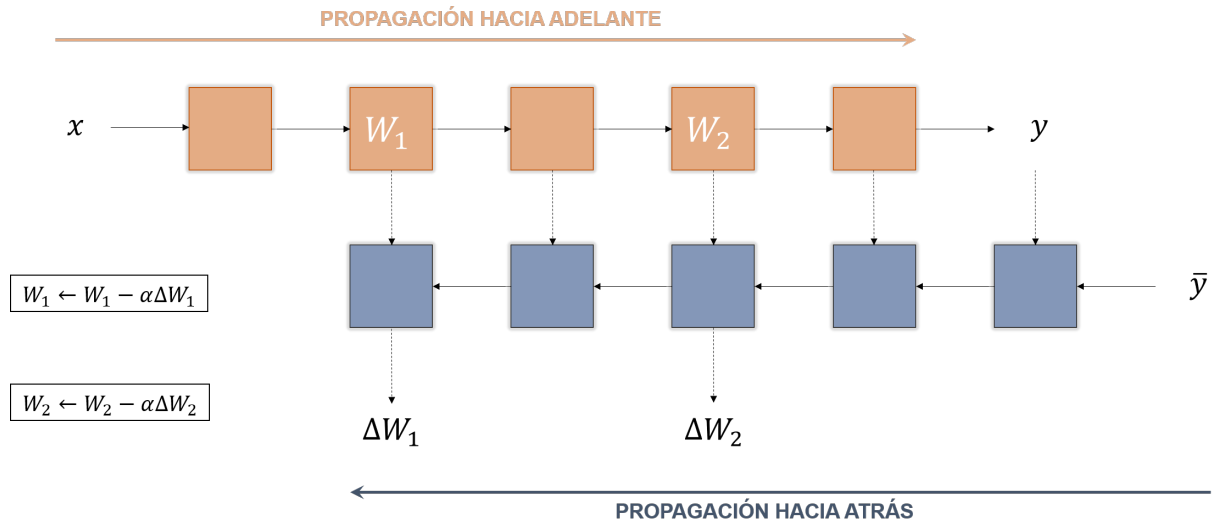


Figura 1.6: Resumen del proceso de optimización de una red neuronal. En rojo se muestra el grafo diseñado formado por un conjunto de operaciones y en azul se muestra el grafo derivado de la utilización de la regla de la cadena, obtenido para el cálculo de los gradientes de los pesos.

Ecuación 1.5, se trata de una función lineal a trozos, la cual es lineal si la entrada es mayor que 0 y es 0 en cualquier otro caso.

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x > 0 \end{cases} \quad (1.5)$$

Cuando se tiene una red neuronal con varias capas, este modelo está compuesto por un conjunto de operaciones que poseen parámetros que deben ser ajustados, como por ejemplo las transformaciones lineales, y en otros casos no, como la función ReLU. Durante el proceso de optimización se tratará de actualizar estos pesos para garantizar la correcta clasificación de las entradas del modelo. Utilizando el *stochastic gradient descent* como optimizador, cada *batch*² del conjunto de entrenamiento va a pasar por dos procesos: el proceso de propagación hacia adelante y el proceso de propagación hacia atrás. De esta forma, durante la propagación hacia adelante las entradas sufrirán un conjunto de operaciones para obtener la clasificación final. Posteriormente un nuevo grafo será creado a partir de las operaciones utilizadas en la propagación hacia adelante, de modo que los datos se propagan hacia atrás a través de la red, se combinan gracias a la regla de la cadena y se obtienen gradientes de una forma muy eficiente. Este nuevo grafo puede ser completamente obtenido a partir de las operaciones realizadas en la red, y lo proporciona de forma automática la mayoría de sistemas de aprendizaje profundo. Este proceso recibe el nombre de propagación hacia atrás, y es fundamental para la actualización de los pesos de la red. Así, durante la propagación hacia atrás la clasificación obtenida será comparada con la clasificación verdadera y se calcularán los gradientes de los pesos que junto a un parámetro denominado ratio de aprendizaje (α) permitirá actualizar los pesos para la siguiente iteración. Este proceso se encuentra resumido en figura 1.6.

²Subgrupo de muestras de entrenamiento utilizado para actualizar los pesos de la red durante el proceso de optimización basándose en los resultados de la clasificación de todo el subgrupo en vez de en una sola muestra.

No obstante, cuando se trabaja con imágenes u otro tipo de información estructurada existe mucha información que no se encuentra únicamente representada por el valor de la intensidad de cada uno de los píxeles que la compone, sino que tiene especial importancia las relaciones que estos valores puedan tener en todo el espacio que representa la imagen. Es por ello que en estos casos no es suficiente con tener un conjunto de matrices multiplicando cada una de las capas y es necesario añadir otro tipo de operaciones, tales como las convoluciones, que permitan compartir los parámetros de la red a lo largo del espacio. Estas convoluciones van a comprimir las dimensiones espaciales de estas imágenes mientras aumentan la profundidad de las mismas, lo que corresponde aproximadamente a la complejidad semántica de su representación. Estas redes son las llamadas CNN y serán explicadas con mayor detalle en la Capítulo 4.

1.4 Motivación

Tras esta introducción, tenemos los elementos básicos para establecer la motivación que impulsó la realización del presente proyecto. Así, los gliomas son neoplasias que afectan a gran parte de la población mundial cuyas consecuencias suelen ser graves tanto a nivel de la salud del paciente, como en el aspecto social y económico. El diagnóstico de las enfermedades que afectan al cerebro y al sistema nervioso en general ha sufrido una mejora considerable gracias a la aparición del la IRM, la cual ha permitido el estudio de regiones cerebrales que no era posible anteriormente. Actualmente, el estudio neurológico es, en la mayoría de casos, realizado mediante IRM [18].

El glioma se trata de la neoplasia más común entre todos los tumores cerebrales presentando diferentes grados de malignidad los cuales se encuentran directamente relacionados con el pronóstico del paciente y las distintas regiones histológicas [19]. A pesar de su alta incidencia, el pronóstico de los pacientes que padecen esta enfermedad sigue siendo pobre e insuficiente, dando lugar a una esperanza de vida mediana de dos años [12], [18], en el caso de HGG, y que se incrementa en el caso de LGG [10], [12].

Actualmente el diagnóstico de estas patologías es realizado mediante criterios cualitativos, como por ejemplo el estudio de las intensidades observables en las imágenes del estudio multimodal de IRM, o mediante criterios manuales cuantitativos sencillos tales como el *Response Evaluation Criteria in Solid Tumors* (RECIST), consistente en la medida del diámetro más largo medible en el interior de la lesión en adquisiciones transversales [12], [20], [21]. Asimismo, uno de los principales hándicaps que presenta la tarea de segmentación manual de estas lesiones se trata del tiempo que requiere y de la dificultad de la misma debido a la gran heterogeneidad que las caracteriza. Según lo reportado en [12], cada estudio multimodal de IRM de un caso de glioma requiere alrededor de 60 minutos. Todos estos criterios se encuentran sujetos a una gran variabilidad inter e intraobservador, lo que impide la obtención de métricas y conclusiones robustas.

Por todo lo anteriormente expuesto, surge la necesidad de utilización de métodos capaces de realizar la tarea de segmentación de estas lesiones de forma automática y que no se encuentren sujetos a la variabilidad intrínseca al observador. De este modo, se logrará liberar al radiólogo y será posible la caracterización biológica de la lesión, tanto a nivel morfológico como a nivel funcional mediante el estudio de biomarcadores de imagen. Gracias a ello el diagnóstico y el tratamiento será más preciso y apropiado para cada uno de los pacientes.

1.5 Hipótesis

La presente trabajo final de máster tiene como base la siguiente hipótesis:

- La tarea de segmentación de gliomas puede ser automatizada mediante la elaboración de metodologías basadas en CNN, logrando resultados equiparables a la segmentación de un experto.

1.6 Objetivos

El objetivo principal de la presente tesis es el de **diseñar, elaborar y validar una metodología de segmentación automática de gliomas mediante la utilización de CNN**. Para ello será necesario la consecución de una serie de objetivos especificados a continuación:

- Investigación bibliográfica del estado del arte en segmentación de tumores cerebrales.
- Diseño de diferentes arquitecturas basadas en CNN para ejecutar la tarea de segmentación de gliomas a partir de un estudio multimodal de cerebro.
- Entrenamiento de las redes anteriormente dichas con un conjunto etiquetado de casos de HGG y LGG.
- Comparación de los resultados obtenidos con el estado del arte y obtención de conclusiones.

1.7 Resumen de los capítulos restantes

- **Capítulo 2.** Estudio del estado del arte de la segmentación de tumores cerebrales, principalmente compuesto por métodos generativos probabilísticos y métodos discriminatorios.
- **Capítulo 3.** Descripción de los materiales utilizados para la ejecución del proyecto, compuestos por la base de datos de casos de gliomas y el equipamiento *hardware* y *software*.
- **Capítulo 4.** Descripción de la metodología empleada para la preparación de los datos, diseño e implementación de arquitecturas, obtención de resultados y evaluación de los mismos.
- **Capítulo 5.** Presentación y discusión de los resultados obtenidos en el proyecto de acuerdo a las distintas metodologías aplicadas.
- **Capítulo 6.** Conclusiones extraídas de los resultados obtenidos y futuras líneas de trabajo en el campo de la segmentación de tumores cerebrales.

Capítulo 2

Estado del arte

En este capítulo se realizará una revisión bibliográfica sobre los métodos de segmentación existentes en el estado del arte así como las diferentes aplicaciones clínicas de los mismos.

2.1 Algoritmos de segmentación de tumor cerebral

La tarea de segmentación de tumores cerebrales, de entre los cuales se encuentran los gliomas, no es sencilla y conlleva multitud de objetivos a conseguir. Entre ellos se encuentra la necesidad de que esta segmentación sea lo más rápida posible, en tanto que no se entorpezca el flujo de trabajo en la práctica clínica diaria. Asimismo, esta debe ser reproducible y precisa, con el fin de garantizar la mejor atención sanitaria posible al paciente, objetivo último de cualquier tarea relacionada con la ingeniería biomédica. De este modo será posible la realización de un primer diagnóstico del tipo de tumor, la planificación de una terapia acorde con la patología y la evaluación de distintos aspectos relacionados con ella, como por ejemplo la respuesta al tratamiento.

La dificultad de esta tarea subyace a la multitud de estructuras a delimitar y la heterogeneidad intrínseca a este tipo de tumores, donde coexisten varios tipos de tejidos. Tanto es así que ni entre expertos se logra alcanzar, en muchas ocasiones, un consenso de la extensión de la lesión. A todo esto se suma la falta de consenso en los protocolos de IRM utilizados en los distintos hospitales y los posibles artefactos presentes en las adquisiciones. Es por todo ello por lo que la tarea de segmentación automática de gliomas no es fácilmente alcanzable [12].

Tal y como se describe en el trabajo realizado por Menze et al. [12], actualmente las metodologías de segmentación de este tipo de tumores pueden ser clasificadas en dos grandes familias, los métodos generativos probabilísticos y los métodos discriminativos, los cuales serán comentados a continuación.

2.2 Métodos generativos probabilísticos

Los métodos generativos probabilísticos se tratan de aquellos métodos que modelan la distribución conjunta de los datos y las clases a las que pertenecen. Más concretamente, estos son algoritmos que generan un modelo estadístico de como se distribuye cada una de las clases en el conjunto de datos proporcionados con el fin de obtener la segmentación final.

Estos métodos están basados principalmente en la extracción de características de imagen que permitan la caracterización del tejido tumoral, dado que algunas de sus características son conocidas, tales como el comportamiento angiogénico o el aumento de la celularidad, presentes en los llamados *Hallmarks of Cancer* [22]. En el trabajo realizado por Juan-Albarracín et al. [23], [24] se presenta un algoritmo de segmentación automático del edema tumoral y del tumor con realce basado en un modelo de mixturas finito llamado *Directional Class Adaptive Spatially Varying Finite Mixture Model* (DCA-SVFMM). Este consiste en un algoritmo de agrupación, también conocido como algoritmo de *clustering*, que combina el modelado tumoral mediante mixtura de Gaussianas con *continuous Markov Random Fields* para sacar partido de la similitud y redundancia local de las imágenes. Este método hace uso de las imágenes potenciadas en T1 y potenciadas en T1 postcontraste, imágenes potenciadas en T2 y adquisiciones FLAIR, secuencias típicas de un estudio cerebral de IRM, así como información obtenida de una segmentación previa basada en atlas del tejido sano cerebral. Posteriormente, tal y como se presenta en [24], los autores fueron capaces de distinguir cuatro hábitats tumorales distintos en cada lesión. Esta distinción fue realizada mediante la adición de una nueva etapa de segmentación basada en un algoritmo de *clustering* de dos clases utilizando únicamente la información proporcionada por los mapas de dos biomarcadores de imagen. Dichos biomarcadores fueron previamente obtenidos de las secuencias llamadas *Dynamic Susceptibility Contrast* (DSC), y son los siguientes: *relative Cerebral Blood Volume* (rCBV) y *relative Cerebral Blood Flow* (rCBF). Así, los cuatro hábitats detectados fueron tumor con realce altamente angiogénico, tumor con realce poco angiogénico, edema periférico potencialmente infiltrado por tumor y edema periférico vasogénico.

Por otro lado, se ha trabajado en la combinación de la IRM junto con la espectroscópica por RM, como el estudio realizado por Jan et al. [25]. En este se crean imágenes nosológicas, definidas como una imagen creada a partir de otras y que aporta información útil para el estudio de una determinada enfermedad [26]. Además, estas imágenes permiten la clasificación de los distintos comportamientos patológicos presentes mediante la utilización de métodos de procesamiento de imagen y de reconocimiento de patrones. Así, este método consistía en dos partes: la primera realizaba una segmentación no supervisada basada en atlas a partir de imágenes potenciadas en T2 y en la detección del edema peritumoral basado en los valores de intensidad de estas imágenes, mientras la segunda hacía uso de técnicas de aprendizaje supervisado utilizando las imágenes de espectroscopía mencionadas para obtener la segmentación final.

Otras aproximaciones a la tarea de segmentación es mediante la utilización de atlas de imágenes. Un atlas de imágenes consiste en una base de datos de casos, por ejemplo de estudios cerebrales de IRM, etiquetados por un profesional. De este modo, una aproximación común de segmentación de tejidos cerebrales sanos se realiza mediante la modelización de estos usando mixturas de gaussianas empleando el algoritmo de *Expectation-Maximization* (EM). Este es el caso del estudio realizado por Menze et al. [27], donde los autores introducen un modelo generativo probabilístico a partir de imágenes multidimensionales que generaliza los modelos de segmentación basados en

EM. Este método extrae una distribución previa de un atlas probabilístico de tejidos cerebrales y distribuciones posteriores de la lesión conjuntamente de los datos de imagen.

2.3 Métodos discriminativos

Los métodos discriminativos son aquellos que modelan la distribución *a posteriori* de las segmentaciones dados los datos. Este tipo de métodos no proporcionan un modelo para cada una de las clases si no que aprenden la frontera de decisión entre las mismas.

Actualmente, el estado del arte de la segmentación de tumores cerebrales se encuentra en su mayoría copado por las CNN, las cuales son clasificadores discriminativos. Estas únicamente aprenden a relacionar unos datos de entrada a un espacio de salida sin aprender un modelo explícito de cada una de las clases. Así, las CNN han mostrado un futuro prometedor en este campo por el hecho de ser capaces de alcanzar un desempeño de análisis semejante al de un experto. Este hecho es notable cuando se observan los resultados del reconocido reto BRATS impulsado por la sociedad *Medical Image Computing and Computer Assisted Intervention* (MICCAI), en el cual participan algunas de las mejores universidades del mundo en este ámbito para tratar de lograr un delineamiento de estos tumores lo más preciso posible. Es así que viendo los grupos mejor posicionados en este reto, todos ellos hacen uso de CNN para realizar esta tarea. Por ello, lo que en esta sección se presentan son algunos de los algoritmos que mejores resultados han proporcionado a la comunidad científica en cuanto a segmentación de gliomas se refiere.

El método que mejores resultados obtuvo en el BRATS 2017 es el llevado a cabo por el grupo del *Imperial College of London*. En este trabajo presentado por Kamnitsas et al. [28] se introdujo un modelo discriminativo de segmentación de gliomas basado en la conjunción de diversos modelos basados en CNN. Las arquitecturas entrenadas fueron totalmente dispares y entrenadas de forma distinta con el fin de introducir una gran varianza entre ellas, para poder así posteriormente construir el modelo mediante la agregación de todas ellas, también conocido como *ensemble*. Los autores denominaron a este modelo como *Ensemble of Multiple Models and Architectures* (EMMA) y alcanzaron valores de coeficiente DICE¹ del 0.73, 0.89 y 0.79, en la segmentación del tejido tumoral con realce, del tumor total y del núcleo tumoral, respectivamente. Cabe destacar que los autores hicieron uso de algunas de las arquitecturas de CNN más contemporáneas para realizar su modelo. Estas fueron DeepMedic [29], FCN [30] y U-Net [31]. De este modo, estas redes fueron entrenadas completamente por separado y posteriormente fueron puestos en común en el *ensemble*.

En segundo lugar se encuentra el grupo de la *University College of London*, representado por Guotai Wang. En este caso, Wang et al. [32] presentaron un método de segmentación basado en una “cascada” de CNNs para segmentar las subregiones secuencialmente, es decir, separaron el problema inicial más complejo de segmentar cuatro clases distintas, incluyendo la clase de tejido sano, en tres subtarefas más sencillas. De este modo lograron construir del problema inicial, tres subtarefas de clasificación binarias, y sacaron partido de la estructura jerárquica del tumor, reduciendo así el número de falsos positivos. Los resultados obtenidos por este grupo fueron semejantes a los alcanzados por el grupo anterior, con un DICE del 0.78, 0.87 y 0.77, en la

¹Estadístico utilizado para estudiar la similitud entre dos muestras. Es ampliamente utilizado para estudiar el desempeño de un modelo de segmentación comparando los resultados de este con las muestras originales correctamente delineadas.

segmentación del tejido tumoral con realce, del tumor total y del núcleo tumoral, respectivamente, en la última etapa del reto.

Finalmente, en tercer lugar se encuentra el trabajo realizado por Isensee et al. [33], en el cual se hizo uso de una arquitectura basada en la conocida U-Net, la cual fue modificada respecto a su concepción inicial con el fin de maximizar el desempeño en la tarea de segmentación de gliomas. Con este modelo, los autores lograron resultados tan prometedores como los anteriormente expuestos, alcanzando un DICE de 0.86 para el tumor total, 0.78 para el núcleo tumoral y 0.65 para el tumor con realce.

Capítulo 3

Materiales

En este capítulo se expondrán los materiales utilizados para la elaboración del presente de proyecto, comenzado por la base de datos utilizada y el tipo de imágenes que dispone así como las herramientas necesarias para la ejecución del mismo.

3.1 Conjunto de datos *BRAIn Tumor Segmentation* (BRATS) 2017

El conjunto de datos utilizado tanto para el entrenamiento del algoritmo como para la evaluación de los resultados del mismo corresponde a la base de datos facilitada por la iniciativa BRATS 2017 [12], [34]. A continuación se describirá este conjunto de datos junto con las técnicas utilizadas para el preproceso y la anotación del mismo. Más información sobre el conjunto de datos y la iniciativa BRATS 2017 pueden ser encontrada en <https://www.med.upenn.edu/sbia/brats2017/data.html>.

3.1.1 Descripción del conjunto de datos de entrenamiento

La base de datos utilizada en este trabajo para el entrenamiento del sistema de segmentación automático está compuesta por un total de 285 estudios de RM, los cuales se encuentran divididos en 75 y 210 casos de LGG y HGG, respectivamente.

Cada uno de los casos del conjunto de datos incluye cuatro tipo de estudios de RM distintos, correspondientes a adquisiciones potenciadas en T1, adquisiciones potenciadas en T1 tras la administración de un agente de contraste al paciente (T1c), adquisiciones potenciadas en T2 y adquisiciones potenciadas en T2 con FLAIR. Información detallada de las imágenes utilizadas en el presente trabajo se encuentra en la tabla 3.1. Adicionalmente, para cada caso de la base de datos, se proporciona la máscara de segmentación de los tejidos tumorales. La descripción de este archivo, del conjunto de clases que los algoritmos van a tratar de clasificar, así como la metodología seguida para la anotación del mismo se encuentra detallado en la subsección 3.1.2.

Tal y como se describe en [12] todas las imágenes de este conjunto de datos fueron preprocesadas para permitir la correcta anotación, clasificación y desempeño de los algoritmos. Para ello todas las imágenes fueron corregistradas rígidamente al la adquisición T1c, dado que esta presentaba la mayor resolución espacial en la mayor parte de los casos. Adicionalmente, todas las imágenes

Tabla 3.1: Descripción de las imágenes del conjunto de datos de entrenamiento y validación.





Imagen	Potenciación	Agente de contraste	Modo de adquisición	Orientación	Espesor de corte (mm)
T1	T1	No	2D	Transversal, sagital	1-6
T1c	T1	Sí	3D	-	1
T2	T2	No	2D	Transversal	2-6
FLAIR	T2	No	2D	Transversal, sagital, coronal	2-6

fueron interpoladas mediante un interpolador lineal para presentar siempre el mismo tamaño de vóxel, el cual se decidió fijar en 1 mm de resolución isotrópica. En la figura 3.1 puede observarse un ejemplo de un caso de HGG mientras que en la figura 3.2 se encuentra un caso de ejemplo de LGG, ambas con sus correspondientes anotaciones mostradas en la parte derecha de las figuras.

3.1.2 Anotación del conjunto de datos de entrenamiento

Todas las imágenes del conjunto de datos fueron anotadas por un equipo de radiólogos expertos de Berna, Debrecen y Boston. Para ello se siguió una metodología, detallada en [12], para asociar las etiquetas a cada uno de los vóxeles de las imágenes del conjunto de datos.

De este modo, se definieron cuatro clases con las cuales etiquetar los distintos tipos de componentes del tumor, además de definir una clase de fondo para el resto de estructuras. Las etiquetas definidas fueron las siguientes:

- **Etiqueta 0**  - Tejido Sano (TS) y fondo
- **Etiqueta 1**  - Núcleo tumoral Sin Realce de contraste (NSR) y Núcleo Necrótico (NN)
- **Etiqueta 2**  - Edema Peritumoral (EP)
- **Etiqueta 4**  - Núcleo tumoral con Realce de contraste (NR)

Cabe destacar que el NR fue definido como etiqueta 4 en esta base de datos. Con el fin de evitar posible confusiones durante el proceso de validación y de preparación de los datos, se decidió reanotar este componente como etiqueta 3.

Una vez definidas las etiquetas en las cuales se segmentarían las imágenes, el protocolo utilizado para la anotación de las mismas se resume en los siguientes puntos:

- La zona de edema fue segmentada en primer lugar a partir de las imágenes potenciadas en T2 y posteriormente verificada y corregida mediante las imágenes FLAIR.
- El núcleo tumoral grueso fue segmentado mediante la evaluación de las hiperintensidades en los casos de HGG en las imágenes T1 postcontraste, junto con las componentes heterogéneas de la lesión hiperintensa visible en las imágenes T1 y las regiones hipointensas en T1.
- El núcleo con realce fue segmentado mediante la umbralización de la hiperintensidades de la imágenes T1 postcontraste del núcleo tumoral grueso.

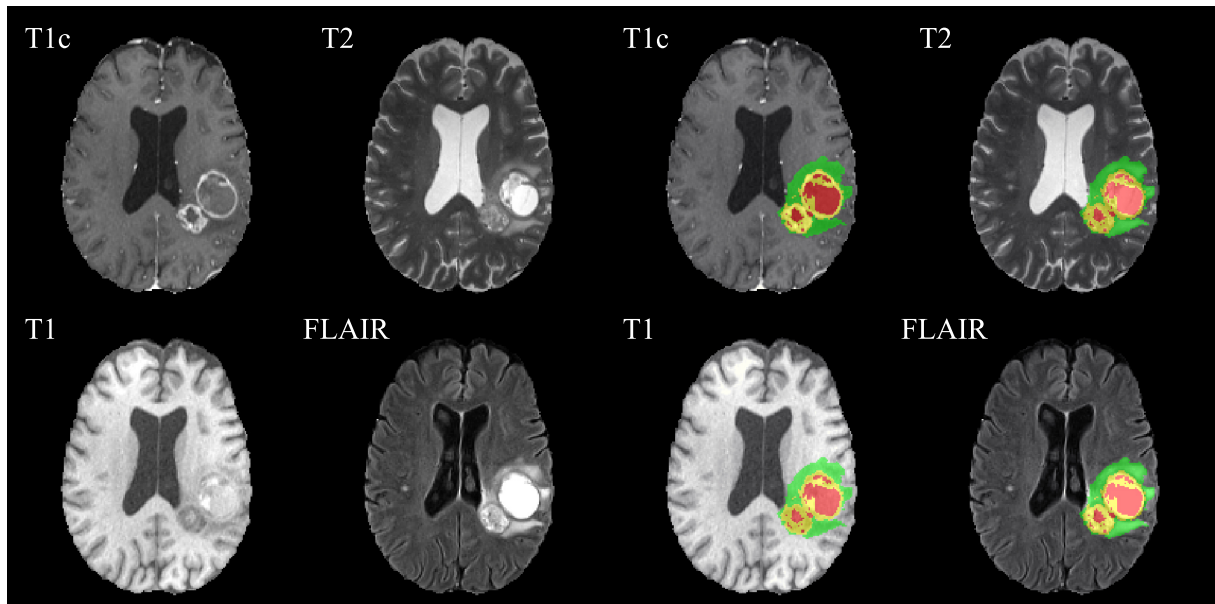


Figura 3.1: Ejemplo de imágenes correspondientes a un caso de HGG.

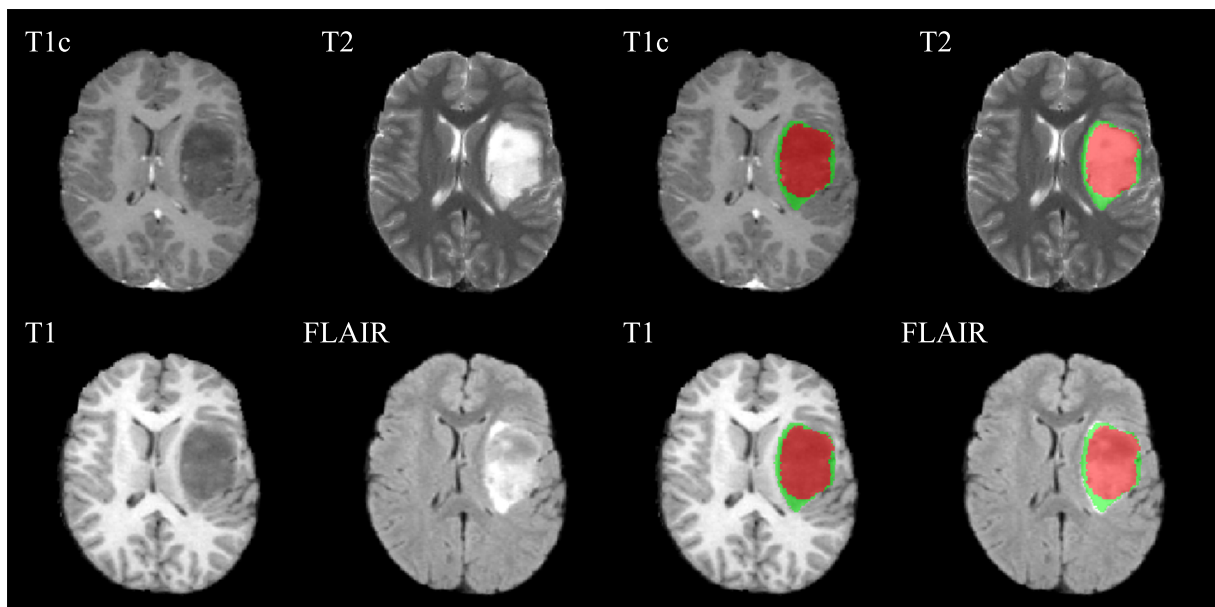


Figura 3.2: Ejemplo de imágenes correspondientes a un caso de LGG.

- El núcleo necrótico fue segmentado a partir de las imágenes T1 postcontraste definiendo aquellas zonas tortuosas de baja intensidad en el interior del anillo con realce de la zona tumoral.
- El núcleo tumoral que no presentaba realce fue definido como toda aquella zona restante del grueso tumoral segmentado.

Debido a la dificultad de la tarea a resolver y que todo el proceso de anotación fue realizado por radiólogos expertos, fue necesario el establecimiento de un algoritmo para alcanzar un consenso para obtener las segmentaciones finales ante una posible discrepancia entre observadores. De este modo, se estableció como algoritmo de consenso el voto mayoritario jerárquico, el cual se encuentra descrito en el algoritmo 1, donde n_{Total} hace referencia al número de radiólogos que desempeñaron la tarea de anotación, y $n_{componente}$ hace referencia al número de votos para cada una de las componentes tumorales descritas. Posteriormente, las clases NSR y NN fueron fusionadas en una sola clase que fue anotada con el número uno, tal y como se comentó anteriormente y se define en <https://www.med.upenn.edu/sbia/brats2017/data.html>. Es por ello que, dadas las características de estas dos clases, en el resto de este proyecto se referirá a la clase uno como NSR.

Algoritmo 1 Voto mayoritario jerárquico. Adaptado de [12].

```

etiqueta ← "TS" {Tejido sano}
si ( $n_{EP} + n_{NSR} + n_{NN} + n_{NR}$ ) ≥  $n_{Total}/2$  entonces
    etiqueta ← "EP" {Edema peritumoral}
si ( $n_{NSR} + n_{NN} + n_{NR}$ ) ≥  $n_{Total}/2$  entonces
    etiqueta ← "NSR" {Núcleo tumoral sin realce de contraste}
    si ( $n_{NN} + n_{NR}$ ) ≥  $n_{Total}/2$  entonces
        etiqueta ← "NN" {Núcleo necrótico}
        si  $n_{NR} ≥ n_{Total}/2$  entonces
            etiqueta ← "NR" {Núcleo tumoral con realce de contraste}
        fin si
    fin si
fin si
fin si

```

3.1.3 Descripción del conjunto de datos de validación

El conjunto de datos de validación proporcionado por la organización del BRATS 2017 cuenta con 46 casos, en los cuales se encuentran casos de LGG y casos de HGG. Cada caso está constituido por el mismo tipo de imágenes que el conjunto de entrenamiento, las cuales fueron preprocesadas del mismo modo. El tipo de glioma de cada uno de los casos no ha sido revelado por la organización por motivos de equidad en la evaluación del *challenge*. Es por ello por lo que este conjunto de validación tan solo será utilizado para la evaluación de aquellas arquitecturas preparadas para segmentar indistintamente casos de LGG y HGG. Para el resto, se realizará una partición de los datos de entrenamiento, la cual será detallada en la sección 4.4.

Tabla 3.2: Especificaciones técnicas de la GPU NVIDIA® GTX Titan X.

	Especificaciones
Memoria	12 GB
Frecuencia de la memoria	7010 MHz
Bus de memoria	384 bit
<i>Boost clock</i>	100/1075 MHz
Tecnología de procesamiento	28 nm

3.2 Equipamiento *hardware*

Para el entrenamiento de todas las CNN profundas se hizo uso de una Unidad de Procesamiento Gráfico (GPU) NVIDIA® GTX Titan X, cuyas características se encuentran especificadas en la tabla 3.2. Asimismo, para la realización del resto de cálculos y tareas se utilizó un ordenador MSI GP62 6QF *Leopard Pro* (Core i7, 16Gb de RAM).

3.3 Equipamiento *software*

En el presente trabajo todos el proceso de creación de las diferentes arquitecturas fue realizado utilizando Python como lenguaje de programación. Para ello, la versión utilizada fue la 3.6.

Asimismo, para la construcción de las CNN se utilizó la librería *open source* TensorFlow™[35], diseñada para computación numérica de altas prestaciones y desarrollada por los investigadores e ingenieros del equipo de Google Brain dentro de la organización de IA de Google. Más información, incluyendo los enlaces de descarga de esta librería, puede ser encontrada en <https://www.tensorflow.org/>.

Capítulo 4

Métodos

En este capítulo se expondrán los métodos utilizados para la ejecución del presente proyecto, centrados en todas las etapas necesarias para la elaboración de un sistema de segmentación mediante CNN.

4.1 Arquitecturas de redes utilizadas

Tal y como se ha comentado en los capítulos anteriores, el presente proyecto consiste en el diseño de metodologías de segmentación de gliomas mediante el uso de CNN. Por ello, en primer lugar es necesario especificar las distintas arquitecturas utilizadas en este proyecto. De este modo, las CNN están compuestas principalmente por las siguientes componentes:

- **Capas convolucionales.** En el ámbito del procesamiento de imagen, la operación convolución se define como una operación matemática entre un filtro de dimensiones $N \times M \times O$ y una imagen de dimensiones $I \times J \times K$. La operación consiste en la multiplicación del filtro por cada una de las regiones de la imágenes especificando un determinado tamaño de paso. En estas capas es necesario la especificación de algunas características, las cuales se encuentran descritas a continuación:
 - *Padding.* Dada la definición previa de convolución, es fácil observar que en los extremos de la imagen no es posible aplicar el filtro. En esta situación es necesario realizar una operación de *padding*, que puede ser principalmente de dos tipos distintos. El primero de ellos corresponde al llamado *padding same*, consistente en añadir ceros (*zero padding*) alrededor de la imagen con el fin de poder operar y asignar nuevos valores a los bordes de la imagen. En este caso el tamaño de la imagen resultante coincide con el de la entrada. La otra posibilidad consiste en no realizar ningún método de *padding* a la imagen, lo que se conoce como *padding valid* y proporciona una imagen de tamaño inferior a la imagen de entrada. En todas las redes creadas en este trabajo se utilizará *padding same*.
 - *Stride.* Se corresponde con el tamaño de paso con el que el filtro se desplaza a través de la imagen. De este modo, un *stride* mayor que uno siempre dejará píxeles de la imagen sin ser procesados y, consecuentemente, una imagen resultante de menor tamaño, independientemente del tipo de *padding* que se utilice.

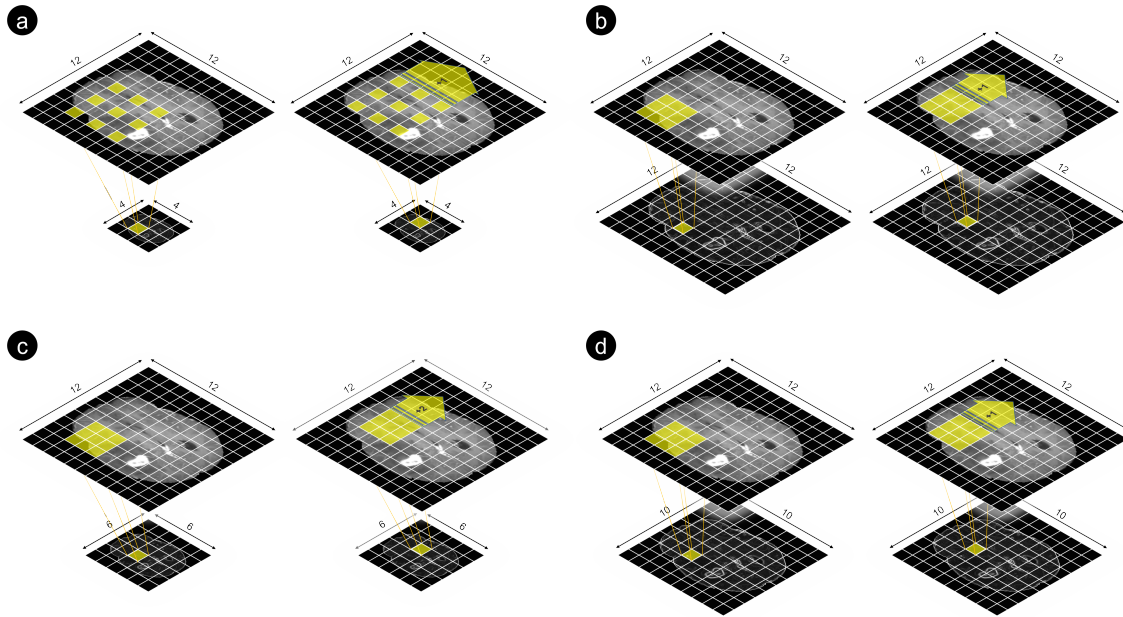


Figura 4.1: Características de las capas convolucionales.

- *Número de filtros.* Se trata del parámetro que define el número de mapas de características que se van a extraer de las imágenes. Por ejemplo, una capa convolucional que recibe como entrada una imagen tridimensional RGB de tamaño $H \times W \times D \times 3$, dará como resultado una imagen $H \times W \times D \times F$, donde F hace referencia al número de filtros utilizados en la capa.
- *Tamaño de filtro.* Parámetro que describe el tamaño de *kernel* utilizado en las capas convolucionales. así, se decidió utilizar un filtro isotrópico de tamaño $3 \times 3 \times 3$, de modo que durante el entrenamiento no exista una dirección preferente en la cual encontrar las características de la imagen. Además, se decidió utilizar un tamaño de $3 \times 3 \times 3$ para disminuir el coste computacional mientras se mantenía un buen desempeño de la red.
- *Ratio de dilatación.* Se trata de la separación entre los diferentes píxeles del filtro utilizado con el fin de ampliar información contextual de la imagen. En el presente trabajo todas las convoluciones serán aplicadas con un ratio de dilatación de uno.

Estas características se encuentran resumidas en la figura 4.1. En ella se puede observar el efecto que tiene aplicar la operación convolución sobre una imagen de tamaño 12×12 mediante un filtro de tamaño 3×3 tras aplicar: **a** *stride* igual a uno, ratio de dilatación igual a dos y *padding valid*; **b** *stride* igual a uno, ratio de dilatación igual a uno y *padding same*; **c** *stride* igual a dos, ratio de dilatación igual a uno y *padding same*; y **d** *stride* igual a uno, ratio de dilatación igual a uno y *padding valid*.

- **Capas *pooling*.** Capas utilizadas para disminuir la dimensionalidad de las imágenes de entrada, a menudo necesario para aumentar la profundidad de la red con un consumo computacional sostenible, mientras se mantiene información de cada uno de los píxeles de la imagen, a diferencia de lo que ocurre cuando se especifica un *stride* superior a uno. La operación de *pooling* más utilizada en la práctica y utilizada igualmente en este trabajo se

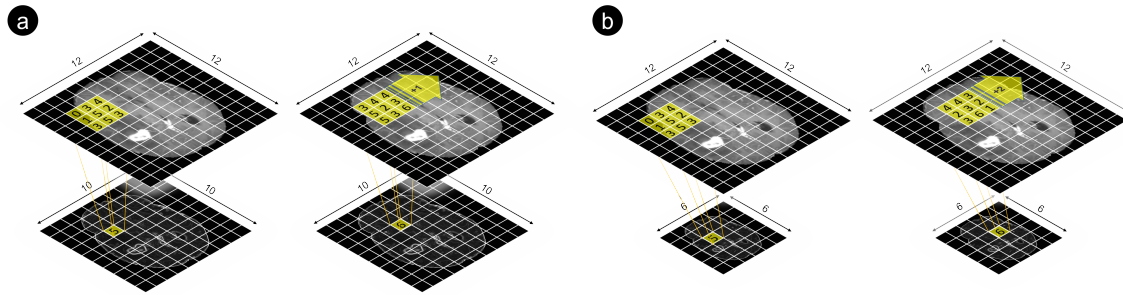


Figura 4.2: Ejemplos de la operación *max pooling*.

trata del *max pooling*. Esta consiste en calcular el máximo de un determinado vecindario de la imagen para cada uno de los píxeles de la imagen de entrada y se caracteriza por presentar una mayor precisión que una operación convolución. Asimismo, es necesario especificar dos hiperparámetros adicionales: el *pooling size* y el *stride*. En la figura 4.2 se presentan dos ejemplos de *max pooling* con un *pooling size* de 3x3. El primero de ellos (a) muestra un caso con *stride* igual a uno y *padding valid*, mientras que el segundo ejemplo (b) es el resultado de aplicar *stride* igual a dos y *padding same*.

- **Capas densas o *fully connected*.** Se tratan de perceptrones multicapa en los cuales cada neurona de una capa se encuentra conectada a cada una de las neuronas de la capa siguiente.
- **Capas *dropout*.** Esta capa tiene como finalidad proporcionar una mayor robustez a la red y evitar sobreajuste. Estas consisten en “ignorar” aleatoriamente algunas de las neuronas de las capas *fully-connected* durante el proceso de entrenamiento, lo que obliga a la red a aprender representaciones redundantes de los datos, forzando a que todas las neuronas aprendan y sean utilizadas durante el proceso de clasificación. El resultado de ello equivale a la creación de muchas redes de pequeño tamaño que posteriormente son agregadas y tomado un consenso de sus resultados.

A continuación, se presentarán las distintas arquitecturas de redes realizadas en este trabajo, especificando las el conjunto de características utilizadas en cada caso.

4.1.1 U-NET

Las U-NET son redes que están siendo ampliamente utilizadas en tareas de segmentación debido a los buenos resultados que presentan. Estas redes se caracterizan por presentar dos fases diferenciables. La primera de ellas, conocida como *contracting path*, consiste en un conjunto de capas convolucionales y de capas de *pooling* que van disminuyendo la dimensionalidad de las imágenes de entrada al mismo tiempo que aumenta la profundidad de la red debido al incremento gradual de filtros en cada una de las capas. Esta primera etapa puede ser considerada como la fase de extracción de características, ya que por cada filtro utilizado se tendrá un mapa de características y estos filtros irán aprendiendo a obtener aquellas características que mejor definan el problema que se trata de solucionar a medida que avanza el entrenamiento de la red.

La segunda de estas fases, conocida como *expanding path* puede ser considerada como la etapa de clasificación propiamente dicha, en la cual se va reduciendo gradualmente el número de filtros en cada capa y se sustituyen las capas de *pooling* por operaciones de aumento de la dimensiona-

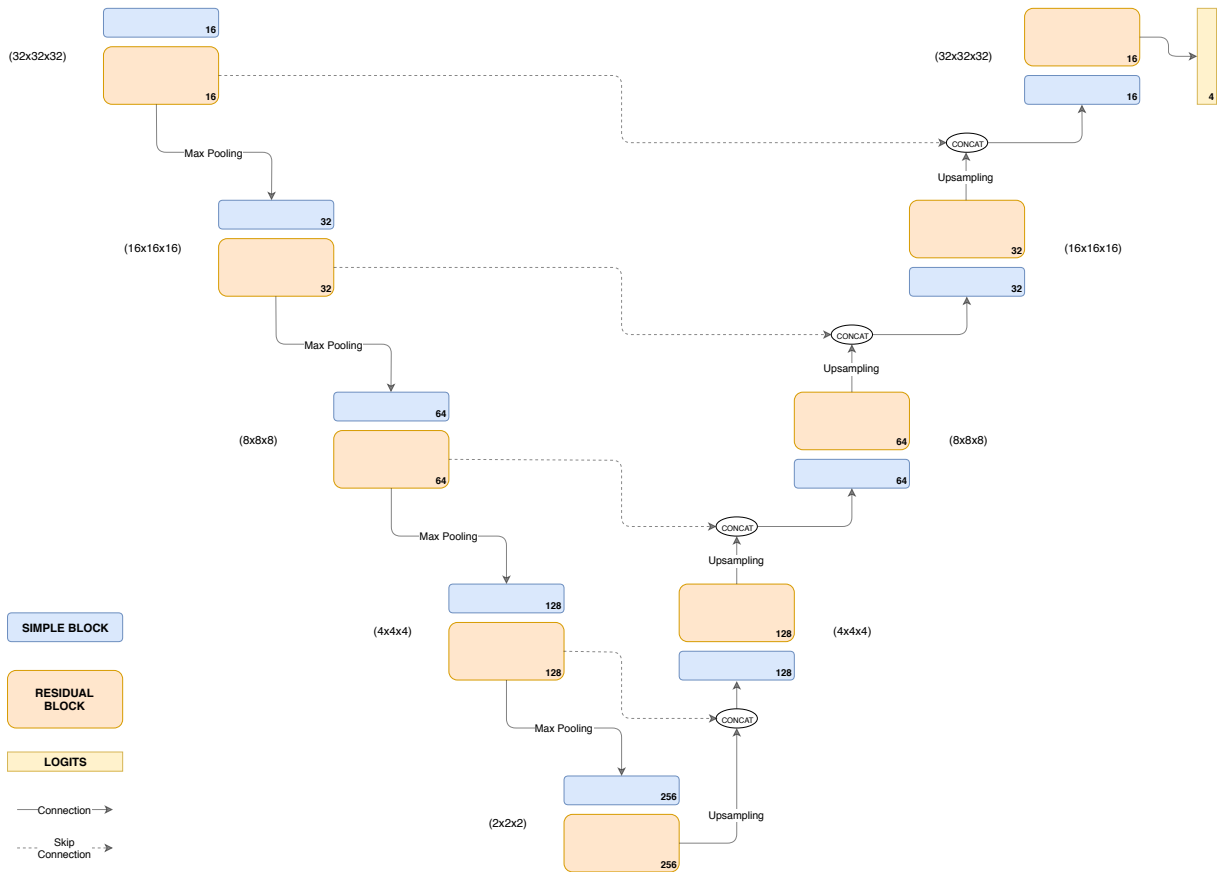


Figura 4.3: Arquitectura U-NET con cinco niveles de profundidad.

lidad. En todos los experimentos realizados se utilizó la operación convolución traspuesta como operación de aumento de la dimensionalidad. De este modo, los mapas de características irán aumentando de dimensionalidad hasta alcanzar las dimensiones originales. Así, la red trata de condensar toda la información extraída en la primera etapa hasta quedarse con aquella realmente importante y no redundante. A la salida de esta etapa es común encontrar una o varias capas densas. En el presente trabajo se tratarán redes con y sin capas densas para comprobar el efecto de las mismas sobre el desempeño de la red. Algunos ejemplos de redes U-NET están ilustrados en las figuras 4.3-4.4, donde se puede observar dos redes con cinco y cuatro niveles de profundidad, respectivamente.

Así mismo, estas redes se encuentran organizadas por bloques con una determinada estructura. Los bloques utilizados fueron definidos como *Simple Block* (SB) y *Residual Block* (RB). Los SB, representado en la figura 4.6, se tratan de los bloques más sencillos compuestos únicamente por una convolución 3D, un bloque de normalización de la salida de la convolución para evitar posibles sesgos en cuanto al rango del mapa de características y una función ReLU, definida en la Ecuación 1.5. Por otro lado, los RB, representados en la figura 4.5 e introducidos por He et al. [36], se tratan de bloques más complejos, los cuales están formados por dos convoluciones 3D, seguidas respectivamente por dos bloques de normalización de los *batches* y por operaciones ReLU, con la salvedad de que se añade una conexión residual entre el inicio y el final del bloque. La adición de este tipo de bloque fuerza a que cada capa aprenda alguna característica adicional diferente a las que se encuentran codificadas en los mapas de características de entrada ya que estos están

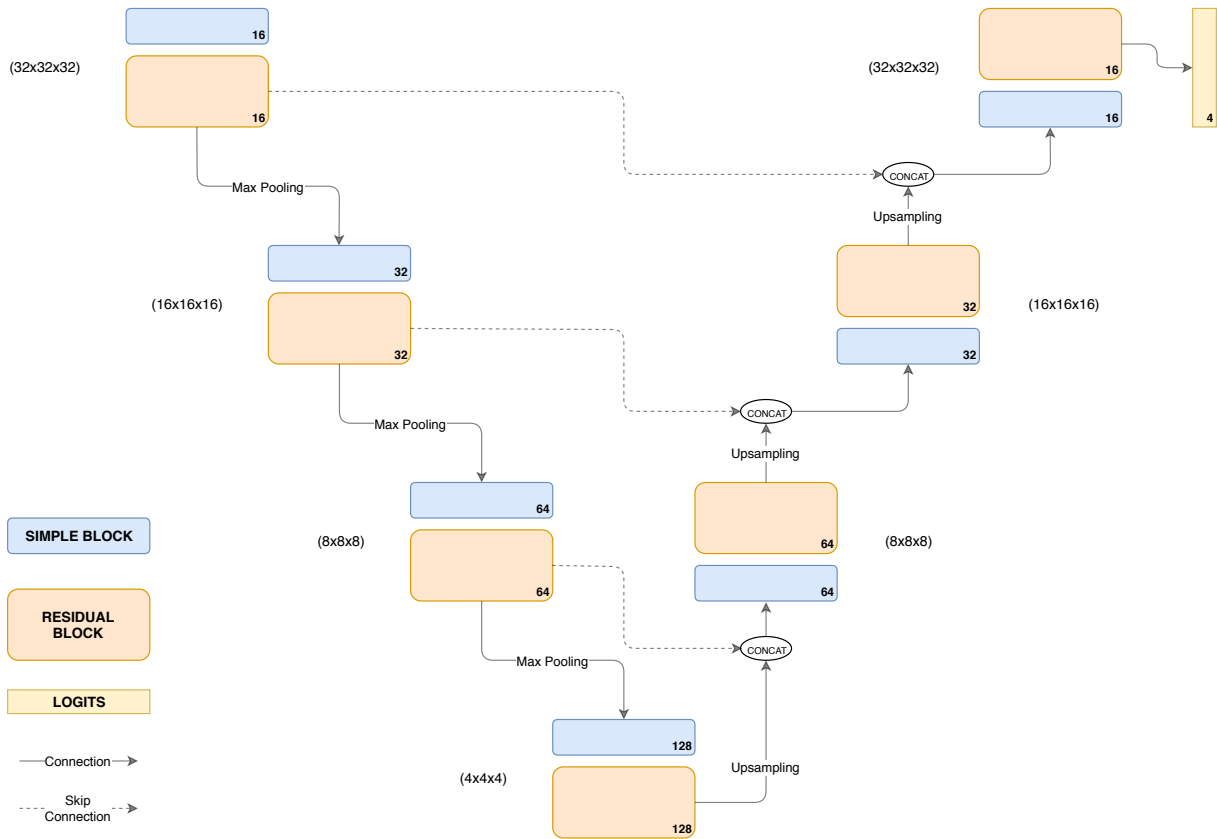


Figura 4.4: Arquitectura U-NET con cuatro niveles de profundidad.

siendo añadidos al final del bloque. Además de estos bloques, existen otro tipo de conexiones residuales en estas redes, tal y como se puede apreciar en las Figuras 4.3-4.4, en la cual la salida de cada conjunto de SB y RB se encuentra conectado mediante una concatenación a la entrada del conjunto de bloques, homólogos respecto a dimensiones de los mapas de características, de la etapa de clasificación. Finalmente, existe un bloque denominado *logits*, el cual precede a las operaciones *softmax* definida en la Ecuación 1.2, que se trata de una convolución con número de filtros igual al número de clases que se trata de estimar. Esta capa tiene como finalidad realizar la estimación final de las segmentaciones de cada una de las clases a estimar, previas a la transformación de sus valores a valores de probabilidad mediante la operación *softmax*.

4.1.2 W-NET

Otro tipo de redes utilizadas consistieron en la unión de dos redes U-NET, de modo que se presentaban dos etapas en las cuales se extraían características mediante la disminución de la dimensionalidad de la imagen y aumento de la profundidad de la misma mediante capas *max pooling* y dos etapas de *upsampling* de los mapas de características. Estas redes fueron denominadas como W-NETs, dada la topología que presentan. Un ejemplo de red W-NET se encuentra representado en la figura 4.8.

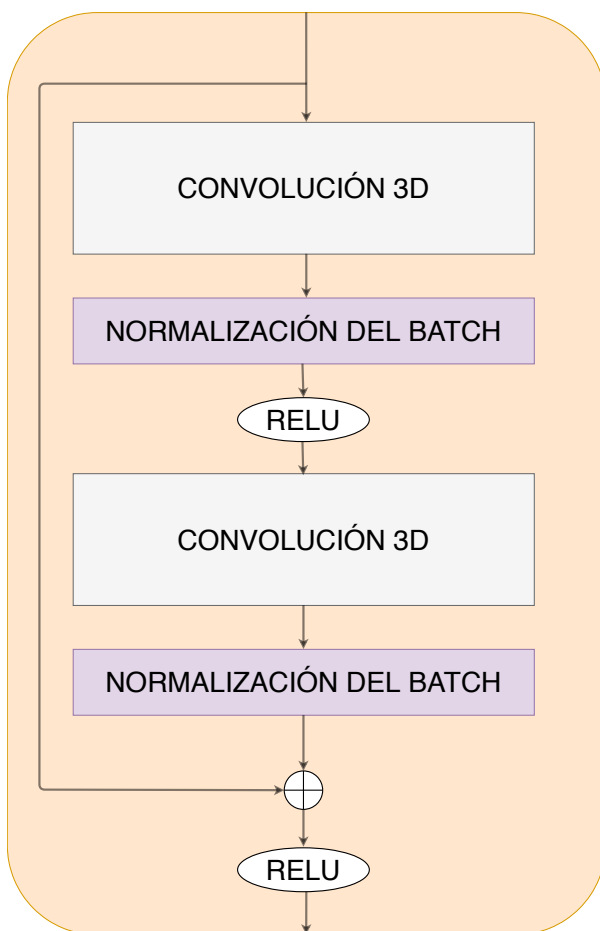


Figura 4.5: Estructura de un *Residual Block*.

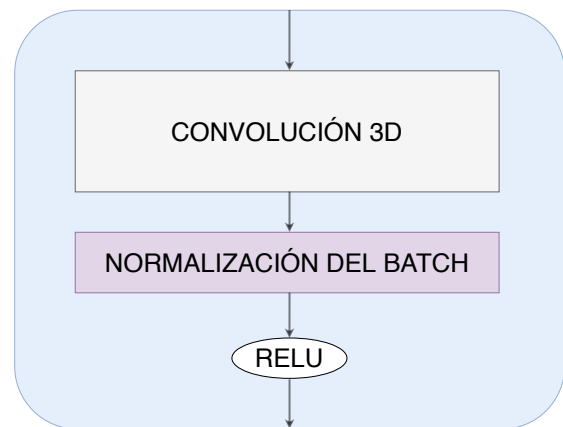


Figura 4.6: Estructura de un *Simple Block*.

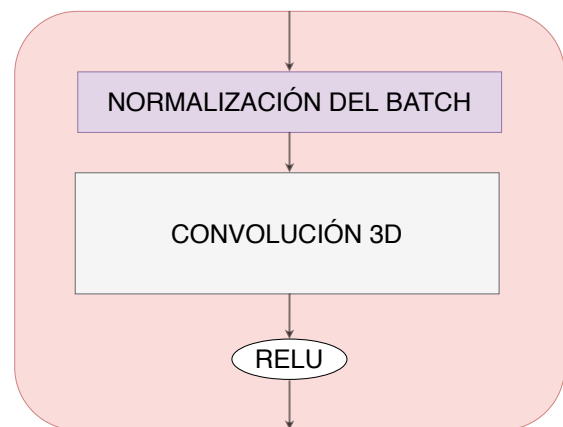


Figura 4.7: Estructura de un *Inverted Simple Block*.

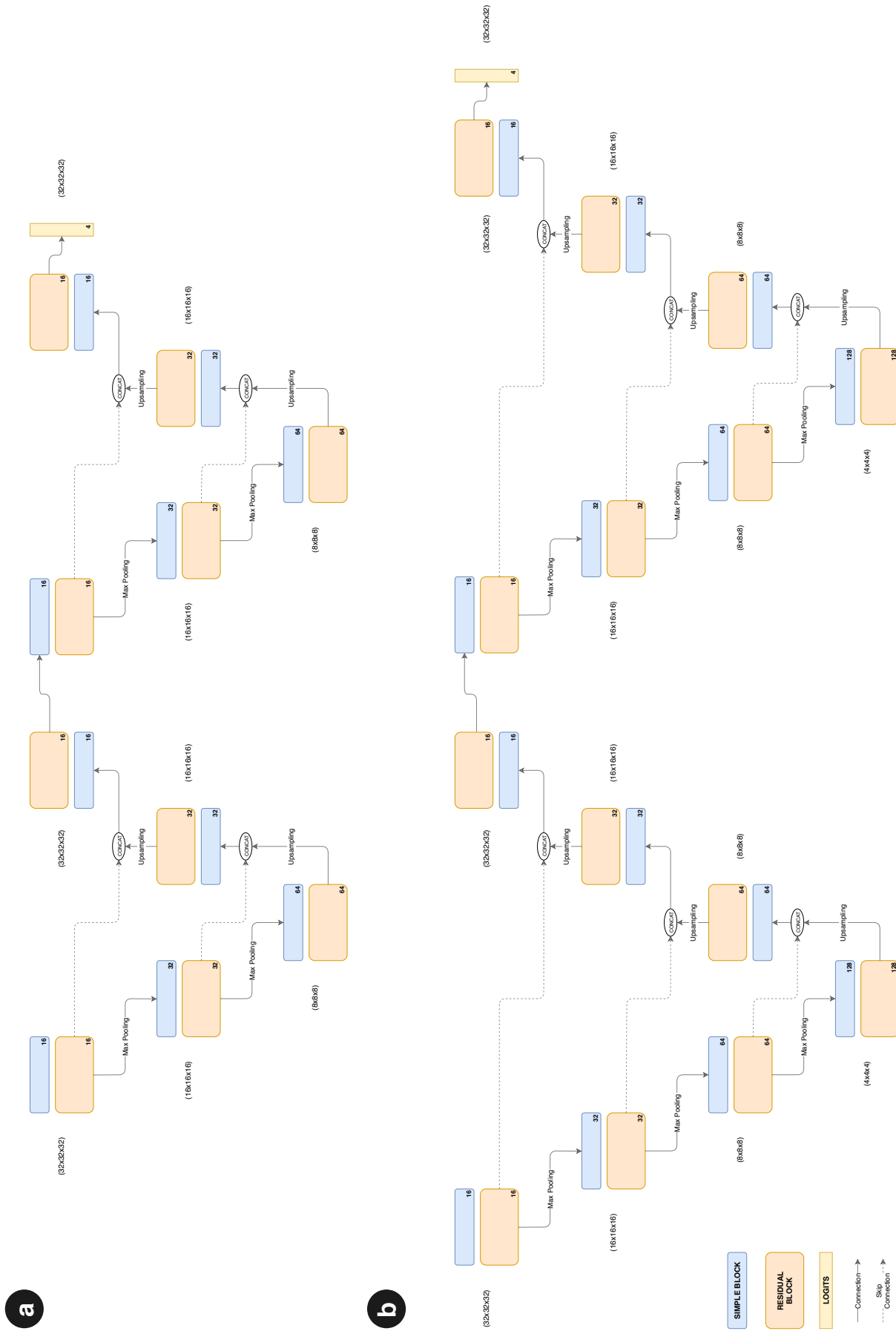


Figura 4.8: Arquitectura W-NET con tres y cuatro niveles de profundidad, representados en las redes **a** y **b**, respectivamente.

4.1.3 Red tubular

Finalmente, la última arquitectura de red utilizada en este proyecto es la llamada red en forma de tubo o red tubular. Esta red no consta de capas *pooling* por lo que la dimensionalidad de los mapas de características no se reduce. En cambio los mapas de características continúan con las mismas dimensiones (32x32x32) a lo largo de toda la red y van aumentando su profundidad dependiendo del número de filtros de salida especificados en cada uno de los bloques definidos.

En la red representada en la figura 4.9 se puede apreciar una ligera diferencia con respecto a las redes introducidas anteriormente. Esta reside en la utilización de distintos bloques, los cuales han sido definidos como *Inverted Simple Block* (ISB) y están descritos en la figura 4.7. Estos bloques presentan una gran similitud a los SB con la salvedad que en aquellos la normalización del *batch* es introducida previa a la convolución 3D y en este es posicionada después.

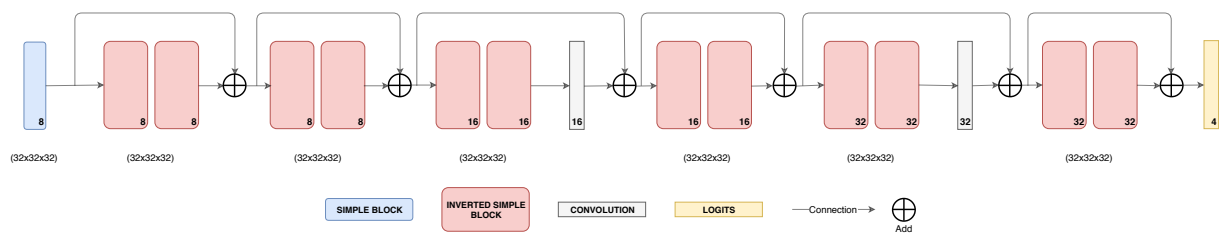


Figura 4.9: Arquitectura de red tubular.

4.2 Agregación de modelos

En esta sección se abordará el problema desde un enfoque distinto al que se ha seguido en las secciones previas. En este caso no se utilizó una única red para la resolución del problema, sino que se diseñaron varias redes con el fin de dividir el problema en pequeñas tareas más fácilmente abarcables. Con esta idea se siguieron dos estrategias distintas, las cuales serán explicadas a continuación.

4.2.1 Bosque de redes

El bosque de redes tiene como objetivo la segmentación individual de cada una de las componentes tumorales y su posterior agregación de acuerdo a un conjunto de reglas. Las diferentes componentes a segmentar se encuentran resumidas en la figura 4.10 y son las siguientes: segmentación de la totalidad del tumor, EP, NSR y NR.

Una vez todas las segmentaciones han sido obtenidas independientemente es necesario establecer unas premisas a la hora de asignar las etiquetas finales a cada píxel, las cuales se encuentran descritas en la tabla 4.1. De este modo, la primera de las redes de segmentación de la totalidad del tumor sería aquella que establecería la extensión del tumor, de forma que todo aquel píxel que no sea etiquetado como tumor por esta red directamente será etiquetado como tejido sano. Por otro lado, se decidió dar prioridad a la parte activa del tumor y etiquetar un píxel como NR siempre que la red de NR haya detectado ese píxel. Respecto a las dos zonas restantes, se decidió dar prioridad al EP frente al NSR.

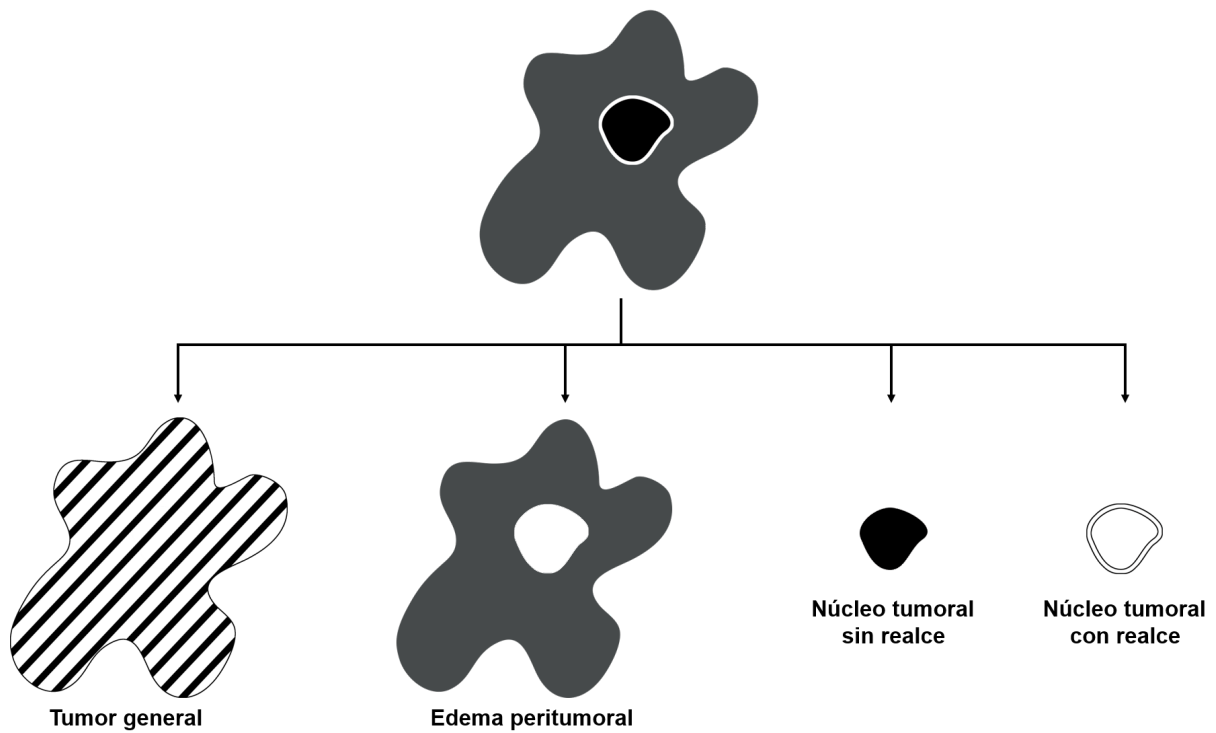


Figura 4.10: Tareas de segmentación a resolver por el bosque de redes.

Tabla 4.1: Tabla de verdad resumiendo la metodología de agregación de las distintas redes del bosque de redes.

Tumor total	NR	NSR	EP	Etiqueta
1	1	1	1	3
1	1	1	0	3
1	1	0	1	3
1	1	0	0	3
1	0	1	1	2
1	0	1	0	1
1	0	0	1	2
1	0	0	0	0
0	0

Así, todas las redes utilizados en este bosque fueron diseñadas con la misma arquitectura, la cual se presenta en la figura 4.12. Esta se trata de una arquitectura U-NET como las mostradas anteriormente, con cuatro niveles de profundidad. La principal diferencia con las redes anteriores reside en el hecho de que se utilizaron distintos tipos de bloques. En este caso se utilizaron únicamente convoluciones y unos bloques denominados *Dense Block* (DB), los cuales son la conjunción de tres SB cuyas salidas se encuentran concatenadas con todas las entradas a estos bloques anteriores, tal y como se puede apreciar en la figura 4.11. Asimismo, todas las redes diseñadas recibieron como datos de entrada las imágenes T1, T2 y FLAIR de cada uno de los casos y mediante la metodología descrita en sección 4.3.

4.2.2 Cascada de redes

Con el mismo objetivo que anteriormente, se experimentó con un diseño en “cascada” consistente en abarcar la segmentación de cada uno de los tejidos uno a uno y permitir que cada red diseñada se beneficie de los resultados obtenidos de las redes anteriores. Así, la primera de las redes de la cascada corresponde a aquella diseñada para la segmentación total del tumor. Esta primera red recibe como datos de entrada las imágenes potenciadas en T1 postcontraste, T2 y FLAIR de cada uno de los casos y proporciona como salidas la segmentación del tumor total, así como las mapas de probabilidad de pertenencia al tumor, las cuales reciben el nombre de *posteriors* (salida de la capa *softmax*).

La segunda de las redes es la encargada de realizar la segmentación del núcleo tumoral formado por la unión de la clase NSR y NR. Esta red recibe como datos entrada las mismas imágenes que la red anterior junto con las *posteriors* del tumor total y, del mismo modo que anteriormente, proporciona como salida la segmentación del núcleo tumoral y las *posteriors* del núcleo tumoral. Asimismo, la segmentación proporcionada por esta red se interseca con el mapa de segmentación del tumor total proporcionado por la primera red, con el fin de eliminar falsos positivos y limitar los resultados de esta red a la región previamente detectada como tumor.

Finalmente, la última red será la encargada de segmentar la componente tumoral con realce, es decir, la llamada NR, a partir del mismo tipo de imágenes que anteriormente, junto con las *posteriors* del núcleo tumoral. Del mismo modo que para la segmentación del núcleo tumoral,

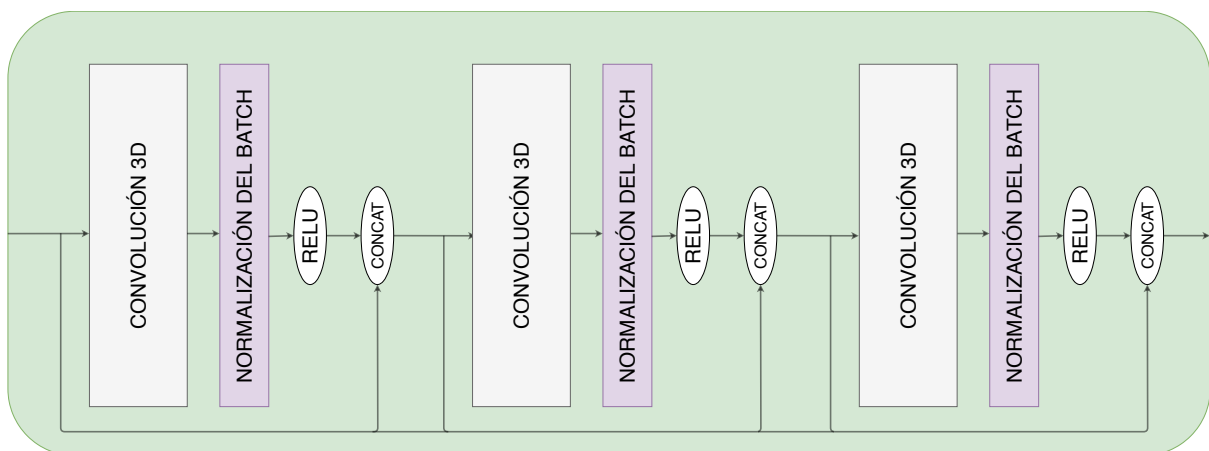


Figura 4.11: Estructura de un *Dense Block*.

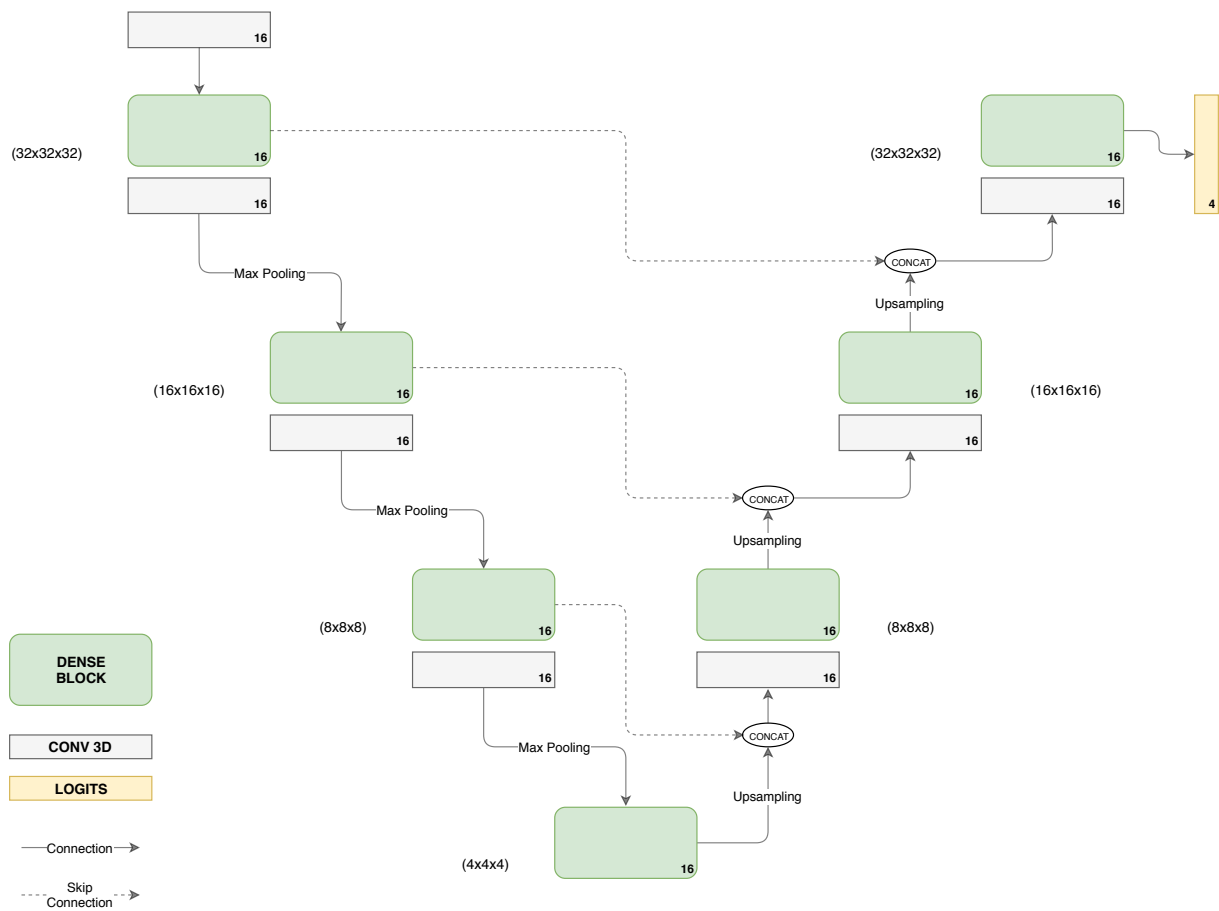


Figura 4.12: Arquitectura de red utilizada para cada una de las CNN diseñadas en el bosque de redes.

la segmentación del NR fue intersectada por está por motivos de delimitación y precisión del resultado final.

Una vez obtenidas las tres segmentaciones anteriores, el resto de componentes, es decir, el EP y el NSR, pueden ser obtenidos directamente de la resta de la máscara del tumor total con la del núcleo tumoral y de la máscara del núcleo tumoral con la del NR, respectivamente. Este postproceso se ilustra en la figura 4.13. Cabe destacar que todas las *posteriors* obtenidas por las distintas redes fueron postprocesadas filtrándolas por componentes conexas y eliminando aquellas componentes que poseen menos de 50 píxeles, ya que fueron consideradas ruido producido durante el proceso de clasificación o falsos positivos que físicamente no pueden ser tumor debido a su tamaño.

La arquitectura de redes utilizadas en la cascada no discierne mucho de las introducidas en las secciones anteriores. La red de segmentación del tumor total fue diseñada por una red U-NET equivalente a la presentada en la figura 4.3. Del mismo modo, la última red de segmentación de el NR fue construida con una arquitectura U-NET de cuatro niveles de profundidad, como la red plasmada en la figura 4.4. Finalmente, la red de segmentación del núcleo tumoral es la única de ellas que presenta alguna diferencia respecto a las anteriores. Esta red fue el resultado de la adición de tres capas *fully-connected*, con 16 neuronas las dos primeras y dos neuronas la última, junto con dos capas *dropout* [37] que preceden a las dos últimas capas.

4.3 Preparación de los datos

La preparación de los datos es una de las etapas críticas en la tarea de entrenamiento de un algoritmo de aprendizaje automático. En esta etapa se preparan los casos que componen el conjunto de datos y serán preprocesados con el fin de garantizar la compatibilidad con TensorFlow, así como un mejor desempeño en la tarea.

Todos los clasificadores desarrollados en este proyecto se basan en una metodología *patch-wise*. Así, cada red recibe como datos de entrada pequeños fragmentos de las imágenes originales en vez de la imagen completa. Esto es así debido a que se desea obtener filtros, también llamado *kernels*, capaces de analizar pequeñas piezas de la imagen para obtener características, tales como bordes o texturas. Mediante esta metodología la red presenta mejores propiedades de regularización ya que cada filtro debe ser bueno para las distintas regiones presentes en una imagen, así como en otras regiones que puedan aparecer en otras imágenes del conjunto de datos de entrenamiento.

Posteriormente, estos *patches* son agrupados en pequeños conjuntos denominados *batches*. La razón de realizar esta agrupación es que los algoritmos de optimización comúnmente utilizados en este ámbito, como los algoritmos de *stochastic gradient descent*, permiten calcular la suma de gradientes con respecto a un conjunto de muestras para posteriormente actualizar los pesos de la red de forma más consistente utilizando el gradiente acumulativo. El tamaño de *batch* fue establecido en 64, lo que significa que en cada iteración la red observa un total de 64 muestras de todas las clases a clasificar gracias al balanceo de la muestras, el cual se encuentra detallado en la sección 4.3. Mientras la situación ideal es que la red fuera capaz de observar todas las muestras en cada iteración, en la práctica esta situación es imposible dadas las altas prestaciones de computación que ello requeriría. Así, el número establecido de 64 *patches* por *batch* presentaba un compromiso adecuado entre desempeño de la red y tiempo de computación.

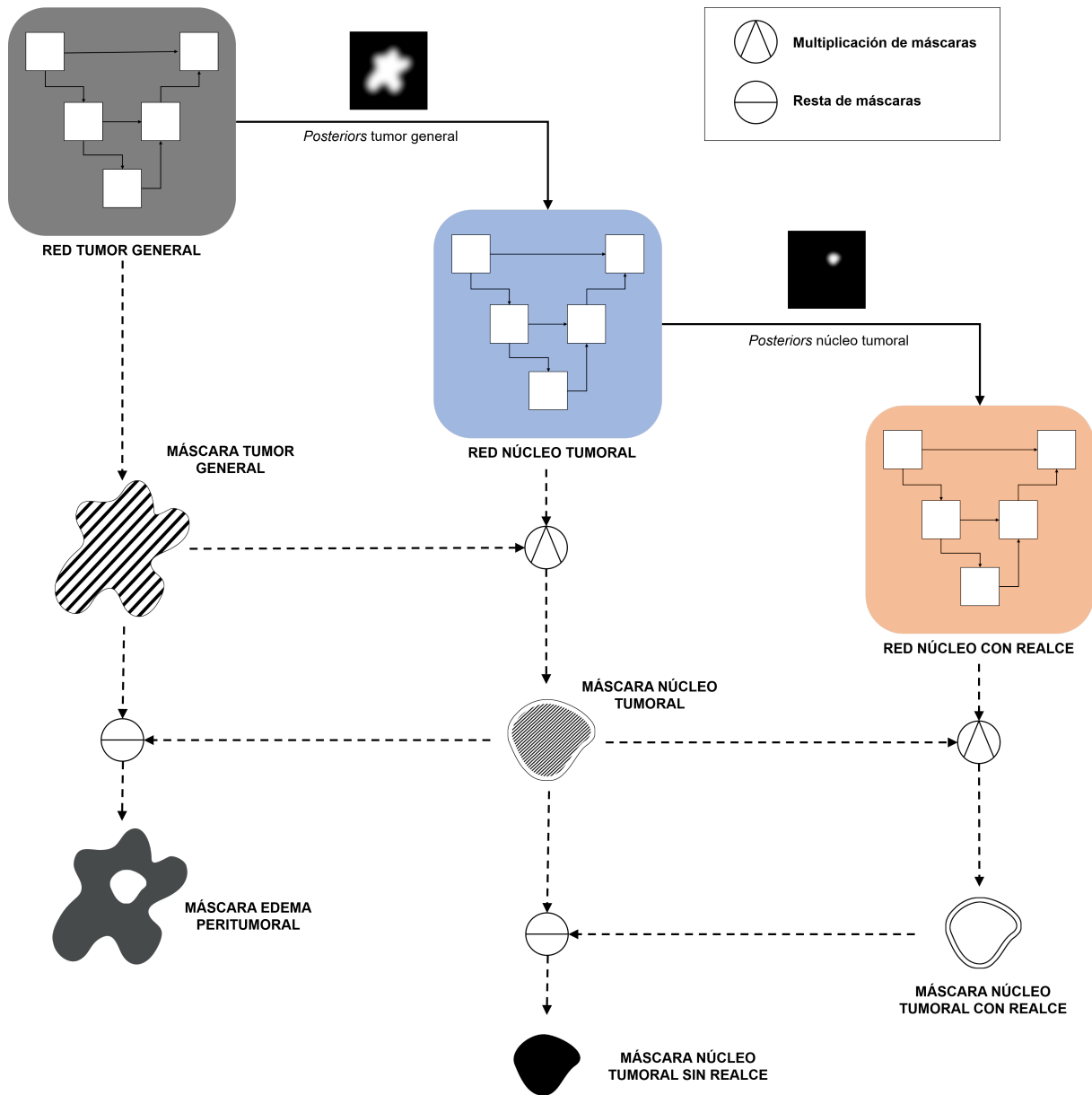


Figura 4.13: Estrategia en cascada para la segmentación de gliomas.

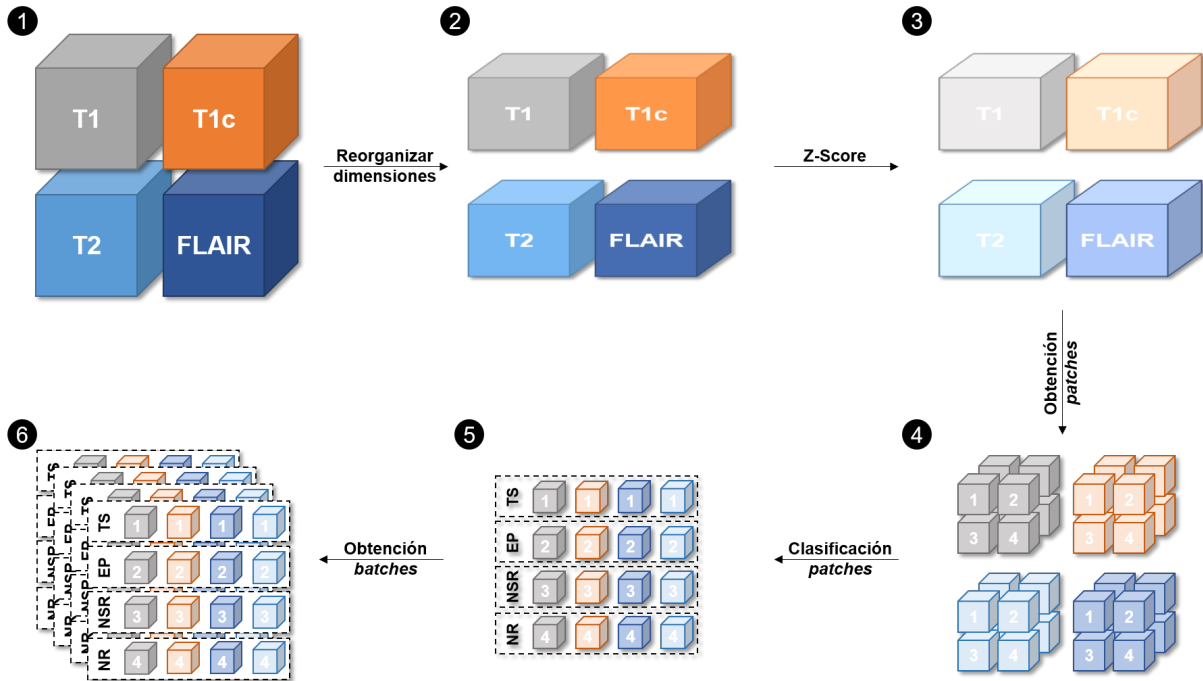


Figura 4.14: Proceso de preparación de los datos y obtención de los *batches* de entrenamiento.

De este modo, el proceso de preparación de los datos se divide en seis pasos, los cuales se resumen en la figura 4.14, y enumerados a continuación:

1. **Carga de los casos.** Durante esta fase se carga en memoria las diferentes modalidades de imagen de RM para cada paciente, así como las etiquetas manuales para el entrenamiento.
2. **Reorganización de las dimensiones.** Con el fin de adaptar los datos a los requerimientos de Tensorflow, se realizó una reorganización de las dimensiones.
3. **Normalización de los datos.** Una vez reorganizadas las dimensiones de los datos, todas las imágenes son normalizadas mediante el *Z-Score*. Esta operación normaliza los valores de la intensidad de cada imagen a un valor medio de cero y una desviación típica de uno. Esta fase es de vital importancia ya que los distintos rangos numéricos de los datos de cada una de las modalidades podría sesgar considerablemente el modelo.
4. **Obtención de los *patches*.** Tras la normalización de los datos, las imágenes son seccionadas uniformemente en pequeños *patches* de $32 \times 32 \times 32$. Esta operación se realiza de forma automática mediante un algoritmo que calcula el solapamiento mínimo necesario para cubrir la extensión total de la imagen con el tamaño de *patch* indicado. Esta operación se encuentra ejemplificada en figura 4.15. Cabe destacar que en la imagen se ejemplifica un caso en 2D, el cual puede ser extrapolable al caso 3D, tal y como se realiza en este proyecto.
5. **Clasificación de los *patches*.** Cada uno de los *patches* obtenidos se clasifican según la nomenclatura descrita en subsección 3.1.2. Esta clasificación se realiza de acuerdo al contenido de cada *patch*. De este modo, para cada *patch* del conjunto de datos se calcula el ratio ($ratio_{patch}$) de tejido patológico sobre tejido sano y, posteriormente, se clasifica según el algoritmo 2.

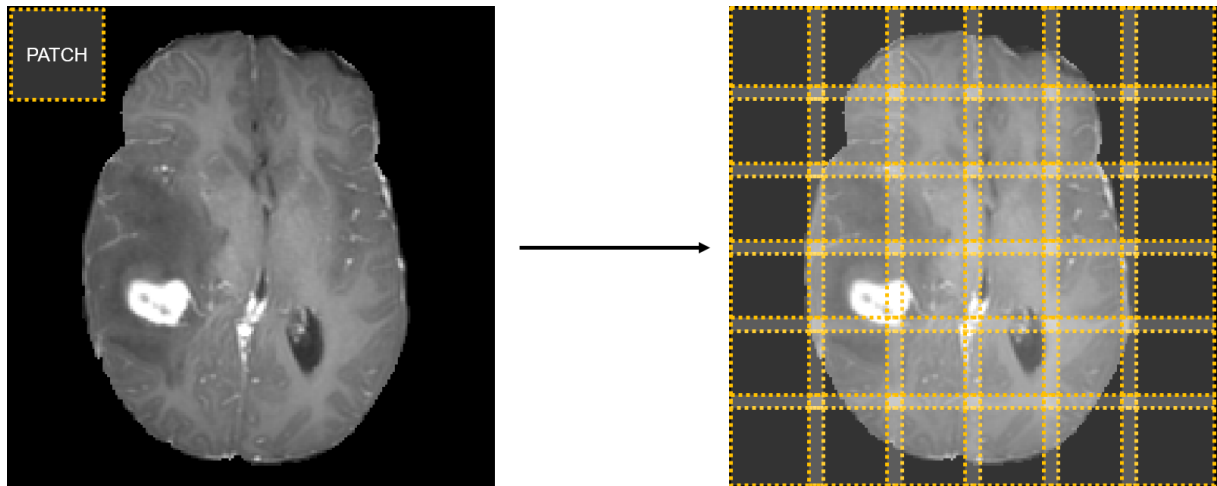


Figura 4.15: Proceso obtención de los *patches* de una imagen.

Algoritmo 2 Algoritmo de clasificación de un *patch*.

```
si  $ratio_{patch} > 0,1$  entonces
    clase predominante =  $\arg \max_{clase} (\#GT_{clase})$ 
si no
    Patch de la clase 0 {Tejido sano}
fin si
```

Donde $\#GT_{clase}$ hace referencia al número de píxeles correspondiente a una clase determinada.

6. **Creación de *batches* balanceados.** Por último, gracias a la clasificación previa de los *patches*, es posible crear *batches* balanceados de modo que cada *batch* contenga una representación equitativa de individuos de cada clase. El motivo de ello es el corregir el problema causado por el desbalanceo de clases y no sesgar el clasificador hacia la clase más prevalente.

4.4 Estrategia de entrenamiento

El conjunto de entrenamiento consta de un total de 285 casos, incluyendo HGG y LGG. Para la etapa de entrenamiento se decidió realizar una partición de los datos en base a un porcentaje del total, estrategia conocida con el nombre de *hold-out*. De este modo, el 80% de los casos fueron utilizados en la etapa de entrenamiento, mientras que el 20% restante fue utilizado como casos de validación para analizar el desempeño del algoritmo. Esta partición es necesaria para elegir de una forma objetiva el modelo que mejores prestaciones presenta frente al problema que se trata de solucionar para posteriormente obtener el desempeño final del modelo con el conjunto de datos de validación, introducido en la subsección 3.1.3, liberado por la organización del BRATS.

Una vez realizada la partición de los datos se probaron distintas metodologías y arquitecturas de CNN, las cuales fueron entrenadas con la partición del 80% del conjunto de datos de entrenamiento y posteriormente fue evaluado el desempeño de cada una de ellas mediante el 20% restante de las muestras. De este modo, se obtuvo una estimación del desempeño de las distintas redes y aquellas que presentaron unos resultados mejores fueron elegidas y reentrenadas con el 100% de los datos para ser evaluadas con el conjunto de validación. Esta estrategia se encuentra resumida en la figura 4.16. La decisión de evaluar tan solo aquellas redes con mejor desempeño fue debido a que este proceso era realizado en línea a través de la plataforma del BRATS, por lo que se trataba de un proceso lento y dependiente de la organización.

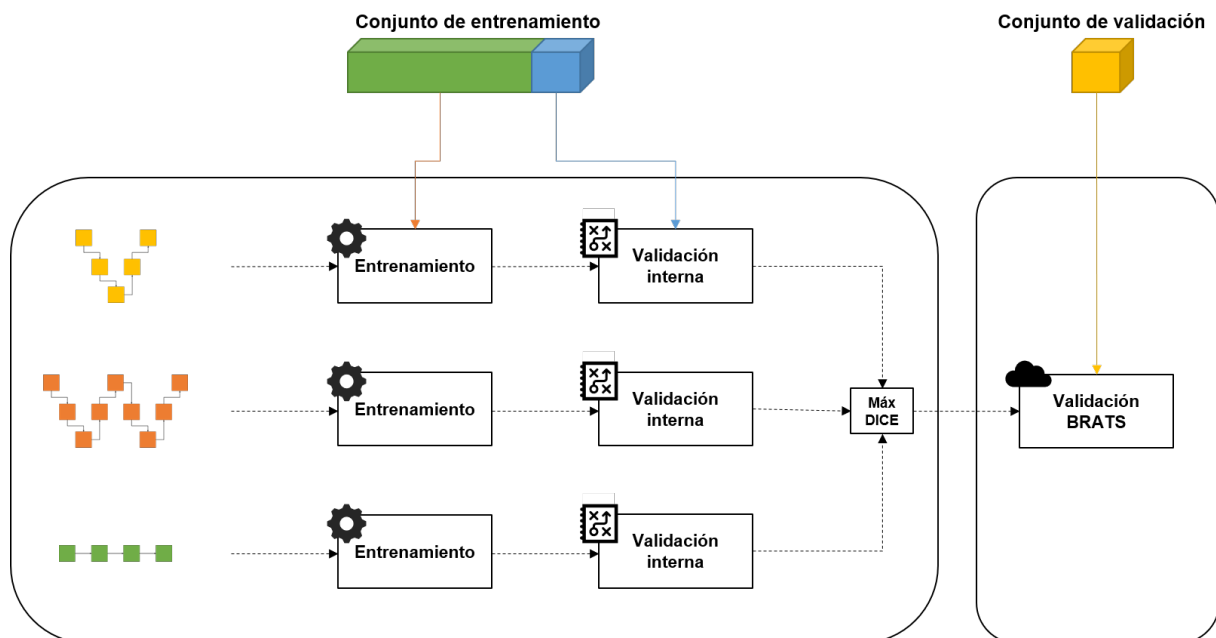


Figura 4.16: Estrategia de entrenamiento utilizada para entrenar las diferentes CNNs para segmentación de tumor cerebral.

Asimismo, es necesario el establecimiento de un método de optimización de la red. Este proceso establece el modo en el que la red va a actualizar sus pesos dados los errores cometidos en cada iteración del entrenamiento, tal y como se ha introducido en la subsección 1.3.1. Para ello el optimizador utilizado es el llamado *Adam* [38], un método para una optimización estocástica caracterizado por su bajo consumo de memoria. Alguna de las bondades de este optimizador es la capacidad de actualizar automáticamente el ritmo de aprendizaje a medida que se avanza en el entrenamiento, de manera que el usuario no tendrá que ajustar la actualización de este hiperparámetro sino que tan solo será necesario especificar el valor inicial del mismo, el cual se decidió establecer en 0.001. Así, durante el entrenamiento de la red, el optimizador actualizará los pesos con el fin de reducir la *loss*, definida en la Ecuación 1.4. El criterio utilizado para el cálculo de la *loss* es la *cross-entropy* de la Ecuación 1.3.

Adicionalmente se decidió utilizar un método de regularización para garantizar la robustez de los modelos diseñados. Así, el proceso de regularización consiste en aplicar barreras artificiales a la red que reducen implícitamente el número de parámetros libres mientras complica el modelo para poder ser optimizado. La regularización es una de las herramientas indispensables para prevenir el llamado sobreajuste u *overfitting*, que consiste en una sobre-especialización de los pesos de la red a los casos del conjunto de entrenamiento, y que conlleva una pérdida de capacidad de generalización del clasificador.

Una de las formas de regularización más típica es la regularización L2, la cual será utilizada en el entrenamiento de todas las redes de este proyecto. Esta técnica consiste en la penalización de la función de pérdida mediante un término adicional, multiplicado por una constante (β), que regula cambios abruptos en la actualización de los pesos. La expresión de la *loss* con la norma L2 está definida en la Ecuación 4.1. Durante el proceso de entrenamiento de todas las redes presentadas en este proyecto se estableció el valor de β en 0.001.

$$\mathcal{L}_{L2} = \mathcal{L} + \beta \frac{1}{2} \|w\|_2^2 \quad (4.1)$$

4.5 Experimentos realizados

En el presente trabajo se decidió separar los experimentos realizados de acuerdo a una serie de criterios, los cuales se introducen a continuación:

- **Gradación tumoral.** Si el tumor se clasifica como HGG o LGG.
- **Tipo de red.** Arquitectura de la CNN utilizada. Por ejemplo, si la red se trata de una U-Net, una red tubular u otro tipo de red.
- **Niveles de profundidad de la red.** Distintos niveles de profundidad y, por lo tanto, distintos número de filtros entrenados para la tarea de clasificación.
- **Canales utilizados.** Diferente combinación de canales utilizados para la tarea de clasificación en base a los tipo de imágenes disponibles (T1, T1c, T2, FLAIR).
- **Data augmentation.** Si se ha utilizado *data augmentation* durante la etapa de entrenamiento o no.

- **Agregación de modelos.** Tipo de estrategia utilizada para la resolución de la tarea de clasificación en base a si se ha utilizado una única red, un conjunto de redes (*ensemble*) u otro tipo de estrategia.

Los primeras redes diseñadas tuvieron el objetivo de segmentar únicamente los casos de LGG. Esta estrategia se basaba en la hipótesis de que al tratarse de grados tumorales distintos, las redes se verían beneficiadas si los casos de LGG y HGG se trataban independientemente. Así, se realizaron una serie de experimentos, los cuales se encuentran resumidos en la tabla 4.2. Para ellos se probaron distintas arquitecturas U-NET variando los niveles de profundidad de la red, los canales, el tipo de regularización utilizada, así como otras características adicionales como el uso de *data augmentation*, para resolver la tarea de segmentación de LGG.

Tabla 4.2: Experimentos realizados con U-NET para la segmentación de LGG.

Experimento	Niveles de profundidad	Canales	Regularización	<i>Data augmentation</i>
LGG001	5	T1c, T2, FLAIR	L2	No
LGG002	5	T1c, T2, FLAIR	L2	Sí
LGG003	4	T1c, T2, FLAIR	L2	No
LGG004	4	T1c, T2, FLAIR	L2	Sí

En los dos primeros experimentos, correspondientes a las redes nombradas LGG001 y LGG002, respectivamente, se utilizaron redes U-NET con cinco niveles de profundidad, tal y como se muestra en la figura 4.3. En esta figura se puede apreciar el número de canales de salida de cada bloque en la parte inferior derecha de cada uno de ellos, además de cuatro capas *max pooling* para reducción de la dimensionalidad de los mapas de características. Cabe destacar que la operación de reescalado de los mapas de características, es decir, la operación inversa a las capas *max pooling*, fue realizado mediante convoluciones traspuestas con *stride* igual a dos. La diferencia entre estas dos redes reside en el uso de estrategias de *data augmentation* con el fin de suplir la falta de casos de entrenamiento de LGG, tal y como se muestra en la subsección 3.1.1. De este modo, en la red LGG002 se aplicaron aleatoriamente rotaciones de 90° y espejados de los *patches* de entrenamiento. Así, la probabilidad de que a un *patch* se le aplique *data augmentation* fue establecida en 0.75.

A continuación, en las dos redes restantes, nombradas LGG003 y LGG004, se decidió reducir la profundidad de la red y, de este modo, su complejidad, con el fin de evitar el posible sobreajuste producido por el reducido conjunto de datos que se poseía. De este modo, se utilizaron redes semejantes a la mostrada en la figura 4.3 pero eliminando un nivel de profundidad y presentando un total de cuatro, tal y como se muestra en la figura 4.4. Asimismo, se realizaron los mismos experimentos utilizando *data augmentation* y sin utilizarlo en los experimentos LGG004 y LGG003, respectivamente.

Tras la experimentación realizada únicamente con el subconjunto de los casos de LGG, se decidió utilizar todo el conjunto de datos para estudiar el efecto que el aumento de la cantidad de muestras

junto con la información que proporcionaba los casos de HGG lograba mejorar el desempeño de estas redes. Para ello, se realizaron dos experimentos más, entrenados con todo el conjunto de datos definido en la subsección 3.1.1, utilizando dos redes idénticas a la de la figura 4.3. Asimismo, se estudio también el efecto de utilizar la estrategia de *data augmentation* introducida anteriormente. Finalmente, se realizó otro experimento, denominado HGGLGG003, en el cual se utilizó misma arquitectura que en el experimento HGGLGG001 pero los datos de entrada fueron preprocesados para eliminar el campo de inhomogeneidades que algunos presentaban mediante el conocido filtro N4 [39]. Los experimentos realizados para el conjunto de datos completo de HGG y LGG se encuentran resumidos en la tabla 4.3.

Tabla 4.3: Experimentos realizados con U-NET para la segmentación de LGG y HGG.

Experimento	Niveles de profundidad	Canales	Regularización	<i>Data augmentation</i>
HGGLGG001	5	T1c, T2, FLAIR	L2	No
HGGLGG002	4	T1c, T2, FLAIR	L2	No
HGGLGG003	5	T1c, T2, FLAIR	L2	No

Del mismo modo, se realizaron un total de tres experimentos utilizando W-NET como arquitectura. Estos experimentos están resumidos en la tabla 4.4 y fueron diseñados para la segmentación conjunta de casos de LGG y HGG. De este modo, en el primero de estos experimentos, denominado HGGLGG004, se utilizó una W-NET con tres niveles de profundidad, mientras que en el HGGLGG005 se utilizó una red con cuatro niveles de profundidad, las cuales se encuentran descritas en las redes a y b de la figura 4.8, respectivamente. Finalmente, en el último de los experimentos se preprocesó lo datos del mismo modo que para el experimento HGGLGG003.

Tabla 4.4: Experimentos realizados con W-NET para la segmentación de LGG y HGG.

Experimento	Niveles de profundidad	Canales	Regularización	<i>Data augmentation</i>
HGGLGG004	5	T1c, T2, FLAIR	L2	No
HGGLGG005	4	T1c, T2, FLAIR	L2	No
HGGLGG006	5	T1c, T2, FLAIR	L2	No

Adicionalmente, se realizó un experimento utilizando una red tubular, la cual se encuentra representada en la figura 4.9. Este experimento, denominado LGG009, fue realizado únicamente para los casos de LGG y sus características se encuentran resumidas en la tabla 4.5. .

Finalmente, se realizaron dos nuevos experimentos utilizando las dos estrategias de agregación de modelos comentadas en la sección 4.2. Estos dos experimentos fueron diseñados para la segmentación conjunta de casos de LGG y HGG.

Tabla 4.5: Experimento realizado con redes tubulares para la segmentación de LGG.

Experimento	Niveles de profundidad	Canales	Regularización	<i>Data augmentation</i>
LGG009	1	T1c, T2, FLAIR	L2	No

Cabe destacar que en todos los experimentos anteriores se utilizó regularización L2, dado que la inclusión de otro tipo de término adicional, como por ejemplo la norma L1, penalizaba mucho el desempeño de la red. Asimismo, se decidió también utilizar todos los canales disponibles excepto las imágenes potenciadas en T1, ya que toda la información visible en estas se encuentra también presentada en las imágenes T1c. Por último, en la tabla 4.6 se resumen los principales parámetros comunes a todos los experimentos.

Tabla 4.6: Principales parámetros establecidos para la etapa de entrenamiento.

Parámetro	Valor
Tamaño de <i>patch</i>	32x32x32
Tamaño de <i>batch</i>	64
Tamaño de filtro	3x3x3
Penalización L2	0.001
Ritmo de entrenamiento	0.001

Una vez realizados todos los experimentos, se seleccionaron aquellas redes que mejores resultados obtuvieron durante el proceso de evaluación interna para ser reentrenadas y evaluados mediante la plataforma del BRATS, tal y como se ha introducido en la sección 4.4. Asimismo, para cada una de las redes seleccionadas, las máscaras fueron filtradas mediante componentes conexas para evitar que en un mismo caso existan más de dos focos tumorales, eliminando los casos con menos número de píxeles. Se decidió realizar este filtrado ya que la probabilidad de encontrar un caso con más de dos focos es muy pequeña. Para evitar eliminar aquellas zonas segmentadas que si sean un foco tumoral, se permitió que si la zona segmentada eliminada con mayor número de píxeles presenta todas las etiquetas definidas (NSR, EP y NR), esta también será incluida en la segmentación final. De este modo, tras este postprocesado, cada caso podrá tener un máximo de tres focos, en caso de que se cumpla la condición anterior, reduciendo segmentaciones espurias que hayan podido ocurrir.

4.6 Evaluación

4.6.1 Figuras de mérito

La evaluación de los resultados, tal y como se ha comentado en la sección 4.4, es realizada de forma *on-line* por medio de la plataforma proporcionada por el BRATS. Esta evaluación, al igual que la evaluación interna, consiste en el cálculo del llamado coeficiente DICE, el cual se trata de un estadístico dirigido a evaluar la similitud entre dos segmentaciones. Este estadístico obtiene la proporción de píxeles coincidentes entre las segmentaciones realizadas por un experto

(A), consideradas como *groundtruth*, y las segmentaciones realizadas por la red que se trata de evaluar (B) del total de píxeles de las dos segmentaciones. El coeficiente DICE se encuentra definido en la Ecuación 4.2.

$$DICE = \frac{2|A \cap B|}{|A| + |B|} \quad (4.2)$$

De este modo, los casos del conjunto de validación proporcionados por el BRATS fueron segmentados de modo *off-line* y posteriormente cargados en la plataforma de la organización, la cual proporcionaba un archivo *csv* con el coeficiente DICE para cada uno de los casos.

4.6.2 Análisis estadístico

Con el fin de medir si los cambios en las arquitecturas y estrategias de clasificación tienen un impacto significativo en los resultados, se realizaron diferentes test de contraste de hipótesis para los resultados de cada modelo. Por ello, se decidió utilizar un test Wilcoxon de contraste de hipótesis. Este consiste en un test no paramétrico para comparar dos poblaciones con muestras independientes, equivalente al test de Mann-Whitney. Así, la hipótesis nula del test es que las medias de ambas poblaciones son iguales y se estudiará si esta hipótesis es aceptada con una confianza del 95 %. De este modo, este test será utilizado para la comparación de los resultados obtenidos por los distintos modelos durante la evaluación interna y también para el estudio del efecto del postprocesado en la validación externa. Cabe destacar que, dado que el primer caso expuesto conlleva la realización de múltiples comparaciones, se decidió utilizar como técnica de corrección el *False Discovery Rate* (FDR). Esto es así ya que en situaciones en las que se realizan varias comparaciones con el mismo conjunto de datos, la probabilidad de cometer error de tipo I aumenta con cada comparación que se realiza. Por ello es necesario reajustar el p -valor para evitar aumentar artificialmente la probabilidad de rechazar la hipótesis nula por el mero hecho de incrementar el número de comparaciones. Dado que no se realizarían un gran número de comparaciones, el FDR fue calculado mediante el procedimiento de Benjamini-Hochberg [40], el cual presenta una mayor robustez en la situación de pocos casos a comparar. Cabe destacar que en los contraste de hipótesis realizados para el estudio del efecto del postprocesado de las máscaras obtenidas en la validación externa, este tipo de correcciones no era necesaria ya que la comparación se realizaría únicamente entre dos grupos.

Resultados y discusiones

En este capítulo se expondrán los resultados obtenidos durante la ejecución del presente proyecto relativos a las diferentes estructuras de CNN utilizadas para el cometido de la segmentación de tumores cerebrales. Asimismo, se discutirán los diferentes resultados obtenidos de acuerdo a su desempeño y a su relevancia clínica.

5.1 Segmentación de los casos de LGG

Las primeras redes en ser probadas en el presente proyecto fueron aquellas que tenían como finalidad de la segmentación únicamente de los casos de LGG, los cuales se caracterizaban a efectos macroscópicos por presentar una menor cantidad de la componente NR, entre muchas más características. La razón de tratar de evaluar solamente los casos de LGG residía en la idea de afrontar los problemas que aparentemente provienen de naturalezas distintas de forma separada, con el fin de aumentar la especificidad de las redes para determinadas tareas.

No obstante, el principal inconveniente que se presentaba para este cometido fue el hecho de contar con un pequeño conjunto de datos de entrenamiento de casos de LGG, el cual contenía un total de 75 casos, completamente inferior al conjunto de datos de casos de HGG, representado por 210 casos. A este aspecto se añadía el inconveniente de que el conjunto de validación liberado por el BRATS no estaba separado en casos de LGG y HGG, lo que dificultaba la evaluación del desempeño de las redes sin utilizar previamente un medio de clasificación de tumores por sus gradación tumoral, añadiendo así un problema nuevo y mayor complejidad al que ya se presentaba.

Los resultados obtenidos para todas las redes diseñadas y entrenadas para el conjunto de casos de LGG se encuentran especificados en la tabla 5.1 y representados en los diagramas de caja y bigotes de la figura 5.1. Tal y como se puede observar, el DICE medio de todas las redes entrenadas para los casos de LGG es semejante y no logra superar un DICE de 0.7, lo que indica pobre desempeño de estas. A grandes rasgos, este dato indica que del total de píxeles segmentados, el ratio de coincidencia de estos no alcanza el 70 % en ninguna de las redes diseñadas. Asimismo, ninguna de las redes expuestas en la presente sección mostraron una diferencia significativa en

Tabla 5.1: Coeficiente DICE obtenido en el proceso de validación interna para los casos de LGG.

Experimento	Media	Desviación estándar	Mediana	Máximo	Mínimo
LGG001	0.6513	0.1598	0.6975	0.8605	0.3283
LGG002	0.5995	0.1548	0.6099	0.8251	0.3468
LGG003	0.6593	0.1547	0.6829	0.8575	0.3508
LGG004	0.6381	0.1597	0.6831	0.8591	0.2955
LGG005	0.6125	0.1567	0.6312	0.8639	0.3115

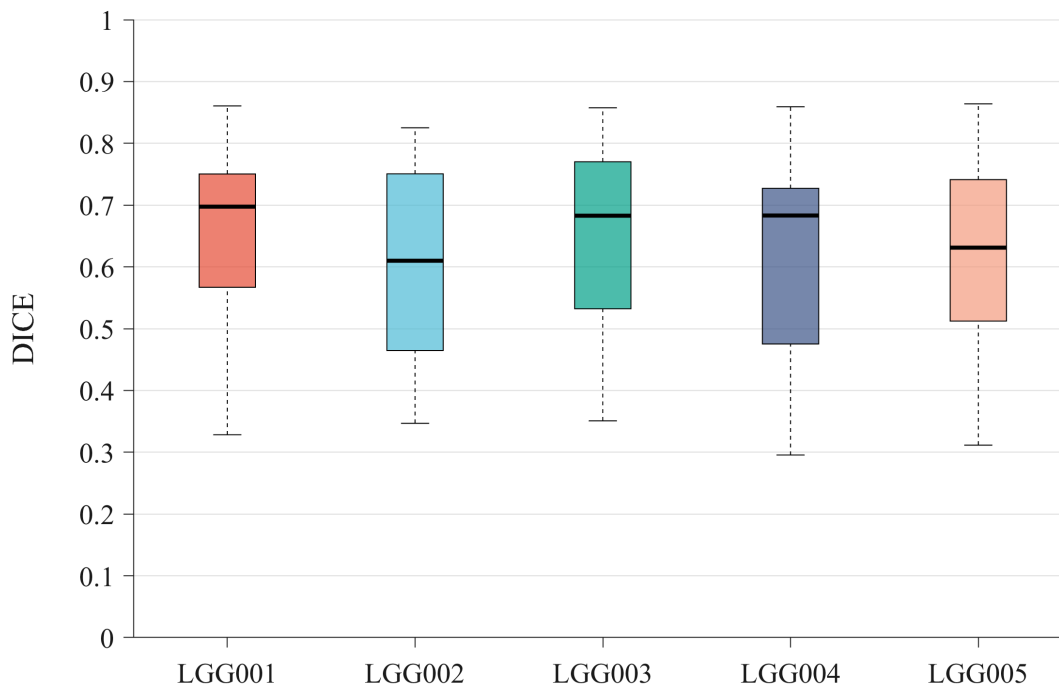


Figura 5.1: Coeficiente DICE obtenido para las distintas redes entrenadas únicamente con casos de LGG en el proceso de validación interna.

los resultados obtenidos utilizando el test Wilcoxon, tal y como se puede observar en la tabla 5.2.

Respecto a la utilización de técnicas de *data augmentation*, tanto para la red con cinco niveles de profundidad (LGG001) como para la de cuatro niveles de profundidad (LGG003), estas no logran mejorar el resultado de las redes. Asimismo, el único cambio apreciable en el uso de este tipo de técnicas es el aumento del rango intercuartil y la disminución del DICE medio, de nuevo de forma no significativa, tal y como se puede observar en los diagramas de LGG002 y LGG004 de la figura 5.1. Todo ello parece indicar que estas transformaciones realizadas, basadas en rotaciones y espejados, no son adecuadas para este conjunto de datos y para la metodología *patch-wise* utilizada, que puede ser de por sí invariante a rotaciones y espejados de los datos de entrada. Es por ello, de aquí en adelante se prescindirá del uso de técnicas de *data augmentation* para estudiar el efecto del resto de características en el desempeño final de cada una de las CNN.

Tabla 5.2: Valores p obtenidos tras aplicar el test Wilcoxon y corregir mediante FDR a los resultados obtenidos para cada una de las redes de segmentación de LGG.

	LGG001	LGG002	LGG003	LGG004	LGG005
LGG001		0.9106	0.9339	0.9286	0.9106
LGG002	0.9106		0.9106	0.9106	0.9286
LGG003	0.9339	0.9106		0.9286	0.9106
LGG004	0.9286	0.9106	0.9286		0.9286
LGG005	0.9106	0.9286	0.9106	0.9286	

Por otro lado, los resultados obtenidos parecen estar afectados por la cantidad de datos disponibles, de modo que las redes no son capaces de “aprender” correctamente el problema que tratan de solucionar, cayendo en posibles casos de sobreajuste. Además, la utilización de distintas arquitecturas no logra suplir este defecto, mostrando un DICE medio similar en la mayoría de los casos que no se utilizaron técnicas de *data augmentation* (LGG001, LGG003 y LGG005).

5.2 Segmentación conjunta de los casos de LGG y HGG

En la presente sección se abordarán todas aquellas redes que fueron entrenadas tanto con casos de HGG como de LGG. Para ello se utilizaron dos metodologías distintas, aquella en la que la red diseñada se encargaría de segmentar cada una de las componentes tumorales (NSR, EP y NR) y aquella encargada de segmentar cada una de las componentes principales mediante redes distintas o operaciones con las máscaras obtenidas por cada una de estas redes mediante la agregación de las diferentes CNN entrenadas.

5.2.1 Segmentación mediante CNN únicas

Tras las primeras evaluaciones realizadas con aquellas redes diseñadas para segmentar únicamente casos de LGG, a continuación se decidió entrenar otras para el conjunto total de los datos, incluyendo casos de todos los distintos grados tumorales. La hipótesis a contrastar era si la inclusión de casos de HGG favorecería al desempeño general de la red, viéndose beneficiada de la información proporcionada por los dos grupos así como por un aumento del número de casos disponibles.

De este modo, tras entrenar las distintas arquitecturas definidas en la sección 4.1, se realizó la validación interna con el 20% de los datos del conjunto que se manejaba, tal y como fue estipulado con anterioridad. Los resultados de esta evaluación interna, consistentes en el cálculo del coeficiente DICE, se encuentran resumidos en la tabla 5.3 y representados en el diagrama de la figura 5.2. Así, se puede observar que en estos casos las redes obtienen mejores resultados que únicamente para los casos de LGG, y superan en todos los casos el 0.7 de DICE medio. Asimismo, el desempeño de todas las redes entrenadas es semejante y de nuevo no existe ninguna diferencia relevante entre los métodos según el test de contraste de hipótesis realizado, tal y como se puede observar en la tabla 5.4. No obstante, de forma general, los experimentos HGGLGG003 y HGGLGG006 presentan un comportamiento ligeramente diferente, como se puede apreciar en el diagrama de caja y bigotes. Estos son los modelos entrenados para los cuales los casos de entrada fueron preprocesados eliminando el campo de inhomogeneidades producido por la

Tabla 5.3: Coeficiente DICE obtenido en el proceso de validación interna para el conjunto de datos con casos de LGG y casos de HGG.

Experimento	Media	Desviación estándar	Mediana	Máximo	Mínimo
HGGLGG001	0.7528	0.1358	0.7677	0.9437	0.4181
HGGLGG002	0.7528	0.1362	0.7714	0.9504	0.3874
HGGLGG003	0.7037	0.1532	0.7255	0.9229	0.3603
HGGLGG004	0.7581	0.1260	0.7748	0.9432	0.4039
HGGLGG005	0.7446	0.1388	0.7633	0.9506	0.3673
HGGLGG006	0.7202	0.1499	0.7450	0.9274	0.3899

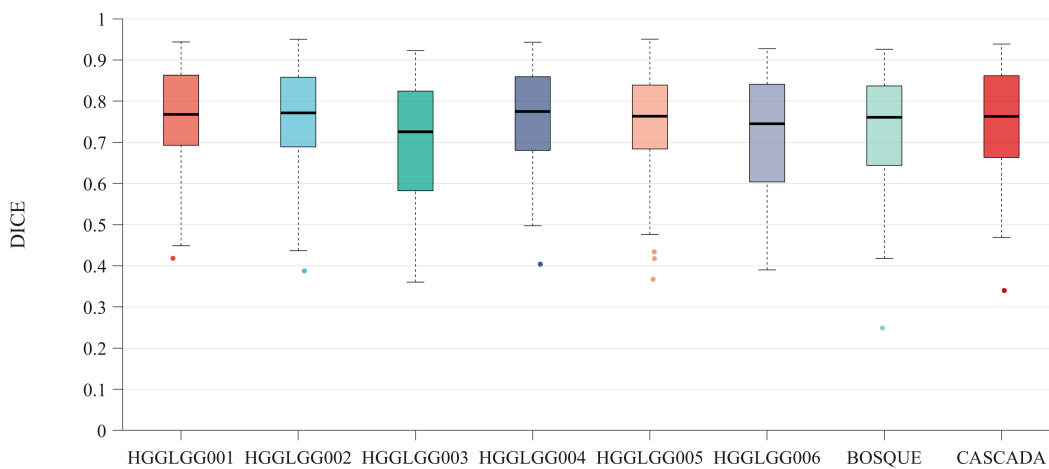


Figura 5.2: Coeficiente DICE obtenido para las distintas redes entrenadas con casos de LGG y HGG en el proceso de validación interna.

fluctuaciones del campo magnético en el interior de la máquina de RM. En estos dos experimentos se puede observar un aumento del rango intercuartil y una disminución del DICE medio, lo que parece indicar que el método utilizado para el filtrado del campo de inhomogeneidades no es el más adecuado y que la red es capaz de “aprender” determinados filtros para combatir este tipo de artefactos. Por otro lado, la disminución de los niveles de profundidad no afecta en exceso al desempeño de las redes, para las cuales el aumento de los niveles de profundidad aumenta ligeramente el DICE obtenido.

Tras el proceso de evaluación interna se decidió evaluar las dos redes de los experimentos HGGLGG001 y HGGLGG004 a través de la plataforma del BRATS, ya que estas habían demostrado un comportamiento ligeramente mejor en la tarea de segmentación que el resto. Para ello, previamente era necesario reentrenar la red con el 100 % de los casos disponibles para aumentar el poder de generalización de la red, entre otros efectos. Una vez reentrenadas las redes, los archivos que contenían las segmentaciones resultantes fueron adaptadas para cumplir las indicaciones de la plataforma y cargadas en esta. Tras un trascurso de tiempo variable entre 1-24 horas, los resultados fueron recibidos. Estos se presentaban en un archivo *csv* con la información del coeficiente DICE obtenido para cada uno de los casos segmentados y, adicionalmente, se proporcionaban otras métricas como la sensibilidad y la especificidad. Todas estas métricas venían dadas para la

Tabla 5.4: Valores p obtenidos tras aplicar el test Wilcoxon y corregir mediante FDR a los resultados obtenidos para cada una de las redes de segmentación de HGG y LGG.

	HGGLGG 001	HGGLGG 002	HGGLGG 003	HGGLGG 004	HGGLGG 005	HGGLGG 006	BOSQUE	CASCADA
HGGLGG 001		0.9774	0.6367	0.9774	0.9774	0.6921	0.6367	0.9774
HGGLGG 002	0.9774		0.6367	0.9774	0.9774	0.6921	0.6367	0.9774
HGGLGG 003	0.6367	0.6367		0.6367	0.6367	0.9089	0.8501	0.6367
HGGLGG 004	0.9774	0.9774	0.6367		0.9675	0.6367	0.6367	0.9774
HGGLGG 005	0.9774	0.9774	0.6367	0.9675		0.8501	0.8501	0.9774
HGGLGG 006	0.6921	0.6921	0.9089	0.6367	0.8501		0.9774	0.6921
BOSQUE	0.6367	0.6367	0.8501	0.6367	0.8501	0.9774		0.6367
CASCADA	0.9774	0.9774	0.6367	0.9774	0.9774	0.6921	0.6367	

componente NR, para el tumor total y para el núcleo tumoral, ya que estas eran las estructuras con las que evaluaban en el BRATS. Cabe destacar que, para cada uno de los experimentos evaluados a través de la plataforma, estos fueron cargados de nuevo tras postprocesar las segmentaciones obtenidas, tal y como se definía en la sección 4.4.

Respecto al experimento HGGLGG001, los resultados obtenidos tras la evaluación externa para las segmentaciones y las segmentaciones postprocesadas se encuentra resumidos en las tablas 5.5-5.6, respectivamente. Asimismo, estos se encuentran representados en el diagrama de la figura 5.3, donde los resultados obtenidos tras el postproceso de las máscaras se encuentra reproducido con mayor transparencia justo al lado del caso sin postprocesar para poder observar el efecto que este filtrado presenta. En esta figura se puede observar como el filtrado posterior de las máscaras obtenidas parece mejora ligeramente las segmentaciones para la totalidad del tumor y el núcleo tumoral, si bien es cierto que la segmentación del NR se ve afectada negativamente. Este hecho puede ser debido a que la componente del NR posee una menor cantidad de píxeles, por lo tanto, de acuerdo con la Ecuación 4.2, la eliminación de algunos píxeles puede causar un efecto notable en el DICE obtenido. No obstante, para ninguna de las componentes evaluadas en esta validación existe una relevancia estadística que afirme que el postproceso de las máscaras afecte positiva o negativamente a los resultados finales, tal y como se puede observar en la tabla 5.9.

Por otro lado, los resultados de la validación externa del experimento HGGLGG004 se encuentran detallados en las tablas 5.7-5.8. Del mismo modo que anteriormente, estos resultados fueron representados en la figura 5.4 y los correspondientes a las segmentaciones postprocesadas se muestran con mayor transparencia. En este caso, los resultados obtenidos son semejantes a los del experimento HGGLGG001, si bien es cierto que el postproceso de las máscaras obtenidas empeoraba levemente el rendimiento en la segmentación del NR, así como presentaba unos resultados sutilmente superiores respecto a la segmentación del núcleo tumoral. De nuevo, estas diferencias resultaron no ser significativas.

Tabla 5.5: Métricas obtenidas para el experimento HGGLGG001 en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.

Métricas	Media	Desviación estándar	Mediana	Cuantil 25	Cuantil 75
DICE NR	0.7331	0.2636	0.8378	0.6819	0.8924
DICE tumor total	0.8624	0.0919	0.8910	0.8449	0.9240
DICE núcleo tumoral	0.7267	0.2451	0.8200	0.6875	0.8915
Sensibilidad NR	0.7609	0.2678	0.8419	0.7245	0.9385
Sensibilidad tumor total	0.8567	0.1395	0.9093	0.8236	0.9447
Sensibilidad núcleo tumoral	0.7233	0.2796	0.8263	0.6448	0.9159
Especificidad NR	0.9982	0.0030	0.9990	0.9982	0.9995
Especificidad tumor total	0.9940	0.0054	0.9950	0.9923	0.9975
Especificidad núcleo tumoral	0.9969	0.0037	0.9981	0.9962	0.9993

Tabla 5.6: Métricas obtenidas para el experimento HGGLGG001 con postproceso de las máscaras en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.

Métricas	Media	Desviación estándar	Mediana	Cuantil 25	Cuantil 75
DICE NR	0.7178	0.2829	0.8352	0.6423	0.8946
DICE tumor total	0.8732	0.0906	0.9105	0.8646	0.9348
DICE núcleo tumoral	0.7235	0.2525	0.8266	0.6796	0.9135
Sensibilidad NR	0.7394	0.2888	0.8363	0.7097	0.9282
Sensibilidad tumor total	0.8409	0.1448	0.9017	0.7924	0.9406
Sensibilidad núcleo tumoral	0.7117	0.2900	0.8263	0.6261	0.9159
Especificidad NR	0.9983	0.0030	0.9992	0.9983	0.9996
Especificidad tumor total	0.9963	0.0033	0.9972	0.9947	0.9989
Especificidad núcleo tumoral	0.9974	0.0035	0.9985	0.9966	0.9994

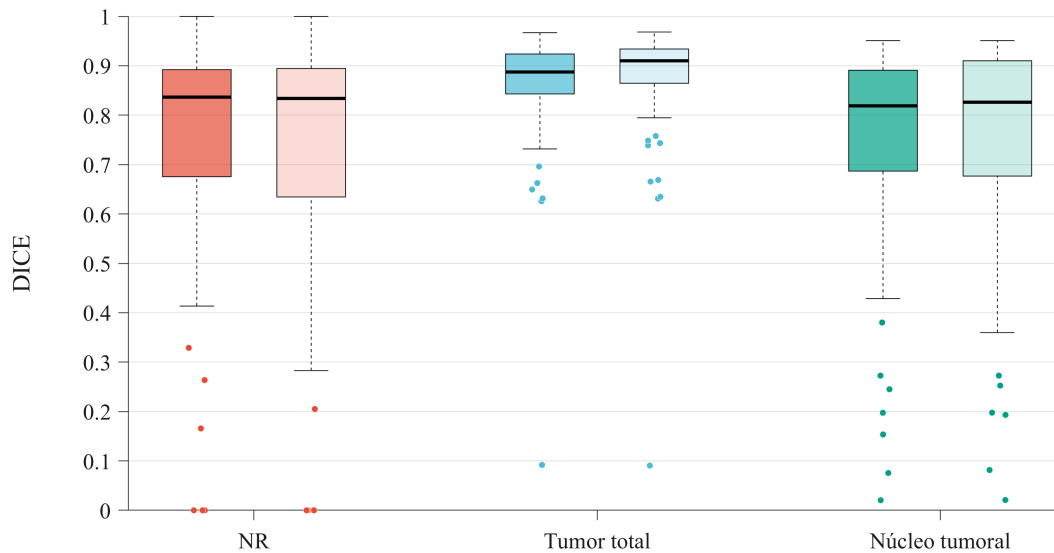


Figura 5.3: Coeficiente DICE obtenido para el experimento HGGLGG001 para el NR, tumor total y núcleo tumoral, durante la evaluación externa.

Tabla 5.7: Métricas obtenidas para el experimento HGGLGG004 en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.

Métricas	Media	Desviación estándar	Mediana	Cuantil 25	Cuantil 75
DICE NR	0.6577	0.3174	0.8138	0.5431	0.8799
DICE tumor total	0.8594	0.0989	0.8911	0.8500	0.9128
DICE núcleo tumoral	0.7476	0.2090	0.8072	0.6842	0.9075
Sensibilidad NR	0.7810	0.2581	0.8765	0.7770	0.9230
Sensibilidad tumor total	0.8805	0.1395	0.9277	0.8784	0.9578
Sensibilidad núcleo tumoral	0.7735	0.2456	0.8763	0.6808	0.9352
Especificidad NR	0.9973	0.0043	0.9986	0.9971	0.9993
Especificidad tumor total	0.9920	0.0064	0.9941	0.9898	0.9966
Especificidad núcleo tumoral	0.9974	0.0035	0.9985	0.9966	0.9994

Tabla 5.8: Métricas obtenidas para el experimento HGGLGG004 con postproceso de las máscaras en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.

Métricas	Media	Desviación estándar	Mediana	Cuantil 25	Cuantil 75
DICE NR	0.6669	0.3196	0.8265	0.5431	0.8909
DICE tumor total	0.8767	0.0940	0.9089	0.8696	0.9305
DICE núcleo tumoral	0.7475	0.2193	0.8180	0.6653	0.9142
Sensibilidad NR	0.7637	0.2821	0.8740	0.7654	0.9224
Sensibilidad tumor total	0.8647	0.1464	0.9251	0.8387	0.9565
Sensibilidad núcleo tumoral	0.7590	0.2616	0.8763	0.6640	0.9352
Especificidad NR	0.9975	0.0043	0.9987	0.9975	0.9994
Especificidad tumor total	0.9949	0.0043	0.9962	0.9919	0.9986
Especificidad núcleo tumoral	0.9955	0.0057	0.9979	0.9943	0.9988

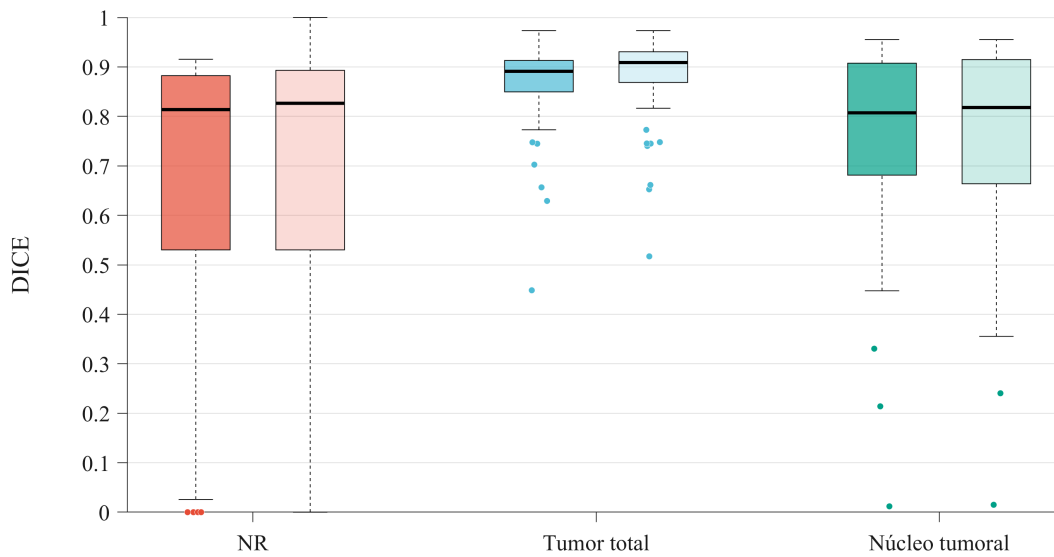


Figura 5.4: Coeficiente DICE obtenido para el experimento HGGLGG004 para el NR, tumor total y núcleo tumoral, durante la evaluación externa.

Tabla 5.9: Valores p obtenidos tras el estudio del efecto del postprocesado de las máscaras de aquellos casos evaluados a través de la plataforma del BRATS.

	NR	Tumor total	Núcleo tumoral
HGGLGG001	0.9786	0.2996	0.7506
HGGLGG004	0.6476	0.1060	0.7370
CASCADA	0.8481	0.6704	0.9595

Tabla 5.10: Coeficiente DICE obtenido en el proceso de validación interna para el conjunto de datos con casos de LGG y casos de HGG para los experimentos realizados mediante la agregación de modelos.

Experimento	Media	Desviación estándar	Mediana	Máximo	Mínimo
BOSQUE	0.7498	0.1280	0.7845	0.9454	0.4090
CASCADA	0.7525	0.1322	0.7629	0.9387	0.3398

Tabla 5.11: Métricas obtenidas para el la cascada de modelos en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.

Métricas	Media	Desviación estándar	Mediana	Cuantil 25	Cuantil 75
DICE NR	0.6970	0.2978	0.8224	0.6035	0.8822
DICE tumor total	0.8705	0.1010	0.9024	0.8682	0.9286
DICE núcleo tumoral	0.7269	0.2338	0.7848	0.6927	0.8805
Sensibilidad NR	0.7571	0.2771	0.8396	0.7341	0.9320
Sensibilidad tumor total	0.8594	0.1559	0.9148	0.8395	0.9583
Sensibilidad núcleo tumoral	0.7673	0.2615	0.8708	0.6948	0.9541
Especificidad NR	0.9978	0.0039	0.9989	0.9980	0.9995
Especificidad tumor total	0.9944	0.0046	0.9957	0.9926	0.9979
Especificidad núcleo tumoral	0.9952	0.0058	0.9968	0.9941	0.9985

5.2.2 Segmentación mediante agregación de modelos

El siguiente aspecto a evaluar era si la división del problema inicial en pequeñas tareas podría beneficiar al modelo y lograr un mejor resultado que con las redes anteriormente expuestas. Para ello, se utilizaron las dos metodologías definidas en la sección 4.2, las cuales consistían en el entrenamiento de múltiples CNN y su posterior agregación, formando lo que fue denominado como “bosque de redes” y “cascada de redes”.

Del mismo modo que anteriormente, en primer lugar se realizó una evaluación interna con el 20% de los datos entrenamiento proporcionados. Los resultados obtenidos de esta validación se encuentran resumidos en la tabla 5.10 y representados en la figura 5.2. En esta figura se puede apreciar como la metodología en cascada proporciona unos resultados ligeramente superiores a la metodología del bosque de redes, y estos resultados se encuentran en sintonía con los obtenidos para las redes únicas. Así, se decidió elegir la metodología en cascada para ser evaluada en la plataforma del BRATS para poder ser comparada con las redes únicas expuestas en la subsección 5.2.1.

Respecto a la metodología en cascada, los resultados obtenidos tras la evaluación externa están presentados en las tablas 5.11-5.12, para las segmentaciones sin postprocesar y postprocesadas, respectivamente. Asimismo, los resultados fueron representados en un diagrama de caja y bigotes, presentado en la figura 5.5, del mismo modo que anteriormente. En este caso, el postproceso de las máscaras no logra mejorar significativamente los resultados obtenidos, afectando ligeramente a la segmentación del NR.

Tabla 5.12: Métricas obtenidas para la cascada de modelos con postproceso de las máscaras en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.

Métricas	Media	Desviación estándar	Mediana	Cuantil 25	Cuantil 75
DICE NR	0.6784	0.3140	0.8195	0.5436	0.8822
DICE tumor total	0.8685	0.1058	0.9084	0.8635	0.9301
DICE núcleo tumoral	0.7176	0.2424	0.7848	0.6893	0.8942
Sensibilidad NR	0.7354	0.2978	0.8293	0.7096	0.9218
Sensibilidad tumor total	0.8445	0.1612	0.9106	0.8097	0.9575
Sensibilidad núcleo tumoral	0.7567	0.2751	0.8708	0.6839	0.9541
Especificidad NR	0.9978	0.0039	0.9990	0.9982	0.9995
Especificidad tumor total	0.9953	0.0042	0.9961	0.9936	0.9987
Especificidad núcleo tumoral	0.9953	0.0057	0.9969	0.9941	0.9985

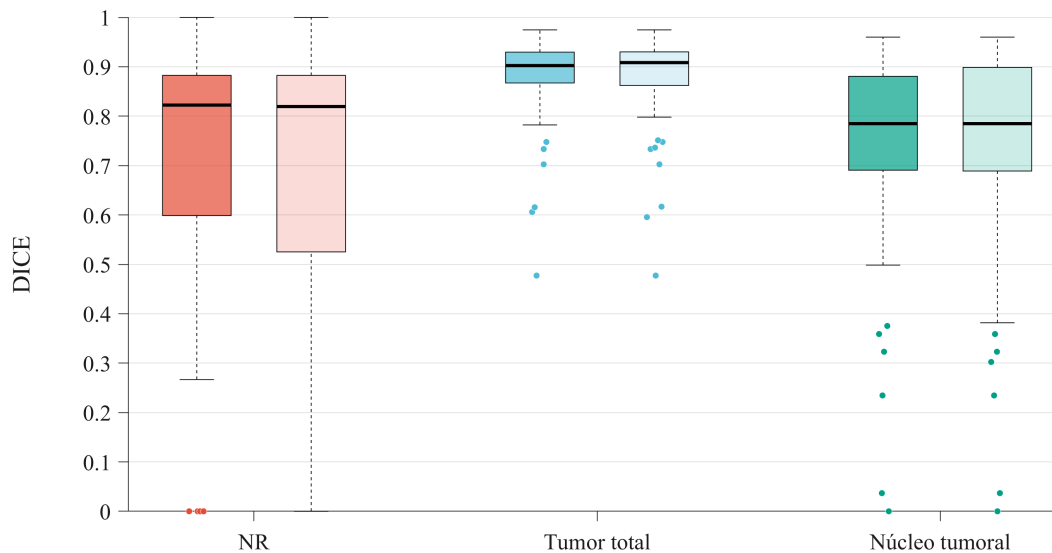


Figura 5.5: Coeficiente DICE obtenido para la metodología en cascada para el NR, tumor total y núcleo tumoral, durante la evaluación externa.

5.3 Discusiones

En el presente proyecto se ha evaluado el desempeño de distintas arquitecturas y metodologías basadas en el uso de CNN para la segmentación de gliomas. Asimismo, se ha estudiado el efecto que tiene el tratar de resolver la tarea anterior diferenciando en función de la gradación del tumor.

A la vista de los resultados obtenidos, se puede observar el potencial de las CNN en las tareas de segmentación de tumores cerebrales, donde la propia red ha podido lidiar adecuadamente con los posibles artefactos presentes en las imágenes y alcanzar resultados que se sitúan en la media del estado del arte. El buen comportamiento que han presentado las redes anteriores frente a posibles imágenes ruidosas, superando aquellas redes entrenadas con casos preprocesados, indica que el resultado final depende en gran medida de la técnica de preprocesado utilizada, convirtiéndose esta etapa en una de las más importante de todo proceso de segmentación. Este hecho supone una ventaja ya que, al no presentar una gran necesidad de preprocesar los datos de entrada o el uso de complejos filtros de eliminación de ruido, supone la simplificación de una tarea que de por sí presenta una gran complejidad.

Por otro lado, las estrategias de *data augmentation* son ampliamente utilizadas en la bibliografía en tarea de segmentación. No obstante, en ocasiones estas técnicas no aportan ningún beneficio a las tareas acometidas. En las redes presentadas en la sección 5.1, la adición de una etapa de *data augmentation* basada en la realización de rotaciones y espejados de forma aleatoria a los *patches* extraídos de las imágenes de entrada, empeora los resultados obtenidos por las mismas redes entrenadas sin el uso de esta. La fase de *data augmentation* puede introducir relaciones artificiales en el conjunto de datos en aquellos casos en los que la transformación realizada incorpore nuevos valores a los píxeles mediante, por ejemplo, la adición de ruido. De acuerdo con el ejemplo anterior, si este ruido no es aquel los datos de entrada suelen presentar (v.g. introducción de ruido de sal y pimienta a imágenes de RM), esto puede significar que la red necesita dedicar algunos de sus filtros a la tarea de eliminación de este ruido, impidiendo que estos filtros se dedican a otras tareas como la extracción de características. Es por ello que esta etapa requiere de especial atención a la hora de ser introducida en cualquier metodología de segmentación, al igual que ocurría con las técnicas de preprocesado. Cualquier incoherencia en este tipo de operaciones puede ser fatal para el funcionamiento de la red. Dado que el presente trabajo no tenía como objetivo un estudio meticuloso de los distintos tipos de técnicas que pueden ser utilizadas para este cometido, se decidió realizar únicamente una pequeña prueba basada en rotaciones y espejados y, visto que no lograba mejorar los resultados, se decidió prescindir de esta etapa.

Respecto a las diferentes arquitecturas y metodologías de segmentación utilizadas, ninguna de ellas parecía presentar una gran diferencia frente al resto en el proceso de validación interna para los casos de HGG y LGG, conjuntamente. No obstante, tras la validación externa realizada, la cual estudiaba el desempeño de los modelos en la segmentación del NR, del tumor total y del núcleo tumoral, la arquitectura basada en redes U-NET presenta mejores resultados y algunos beneficios frente al resto. Al presentar una menor complejidad que las W-NET utilizadas en este proyecto, las redes U-NET presentaban una menor predisposición a sufrir sobreajuste. Por otro lado, si comparamos estas redes con las metodologías de agregación de modelos, las redes U-NET presentan un claro beneficio relacionado con el tiempo de computación. Esta característica parece

Tabla 5.13: DICE obtenido por los principales participantes del BRATS 2017.

Referencia	DICE	DICE	DICE
	NR	Tumor total	Núcleo tumoral
[28]	0.7376	0.9009	0.7972
[32]	0.7859	0.9050	0.8378
[33]	0.7756	0.9027	0.8194

no ser importante, pero la realidad es muy distinta. El tiempo necesario para la segmentación de un único caso mediante una red única mediante una GPU como la utilizada en este trabajo, como la U-NET o la W-NET, es tan solo de unos pocos segundos, mientras que si se realiza mediante alguna de las metodologías de agregación anteriormente descritas, el tiempo de cómputo se incrementa hasta varios minutos. No obstante, estos modelos tienen como principal objetivo el ser introducidos en la práctica clínica diaria y servir de ayuda al profesional sanitario. Por desgracia, no todos los servicios de radiología pueden permitirse un ordenador con una GPU con unas prestaciones semejantes a las utilizadas en este proyecto, por lo que el proceso de segmentación debe ser realizado mediante Unidad de Central de Procesamiento (CPU), lo que incrementaría de forma notable e impermisible el tiempo de cómputo.

En las figuras 5.6-5.7 se puede observar ejemplos de segmentaciones que obtuvieron un DICE elevado y bajo en el proceso de validación externa, respectivamente. En esta figura se presentan las segmentaciones realizadas por los tres modelos evaluados externamente, así como los resultados obtenidos tras el postproceso de las máscaras. En la segmentación realizada por el modelo HGGLGG004 en la figura 5.6 puede observarse el efecto de aplicar el postprocesado indicado, donde una pequeña zona segmentada en el centro del cerebro desaparece tras este filtrado.

Finalmente, cabe destacar que los resultados obtenidos se encuentran dentro de los presentados en el estado del arte de segmentación de tumores cerebrales. De este modo, las métricas obtenidas por las redes de este proyecto son ligeramente inferiores a las logradas por los tres mejores trabajos del *challenge* BRATS 2017 en la tarea de validación, resumidas en la tabla 5.13, y muy superior a la media de todos los trabajos presentados en este.

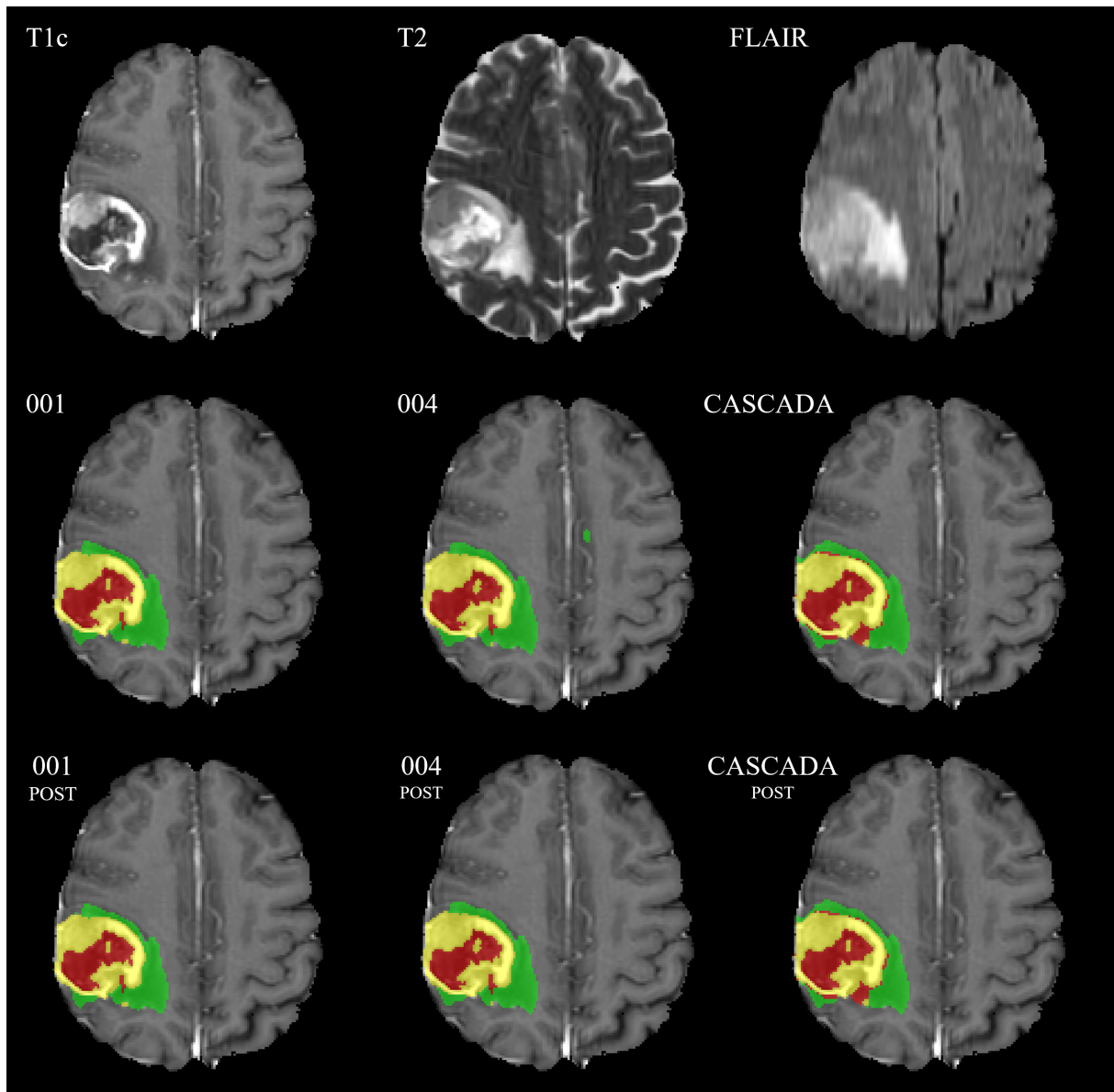


Figura 5.6: Segmentaciones obtenidas para los modelos evaluados externamente de un caso del conjunto de validación que obtuvo un DICE elevado.

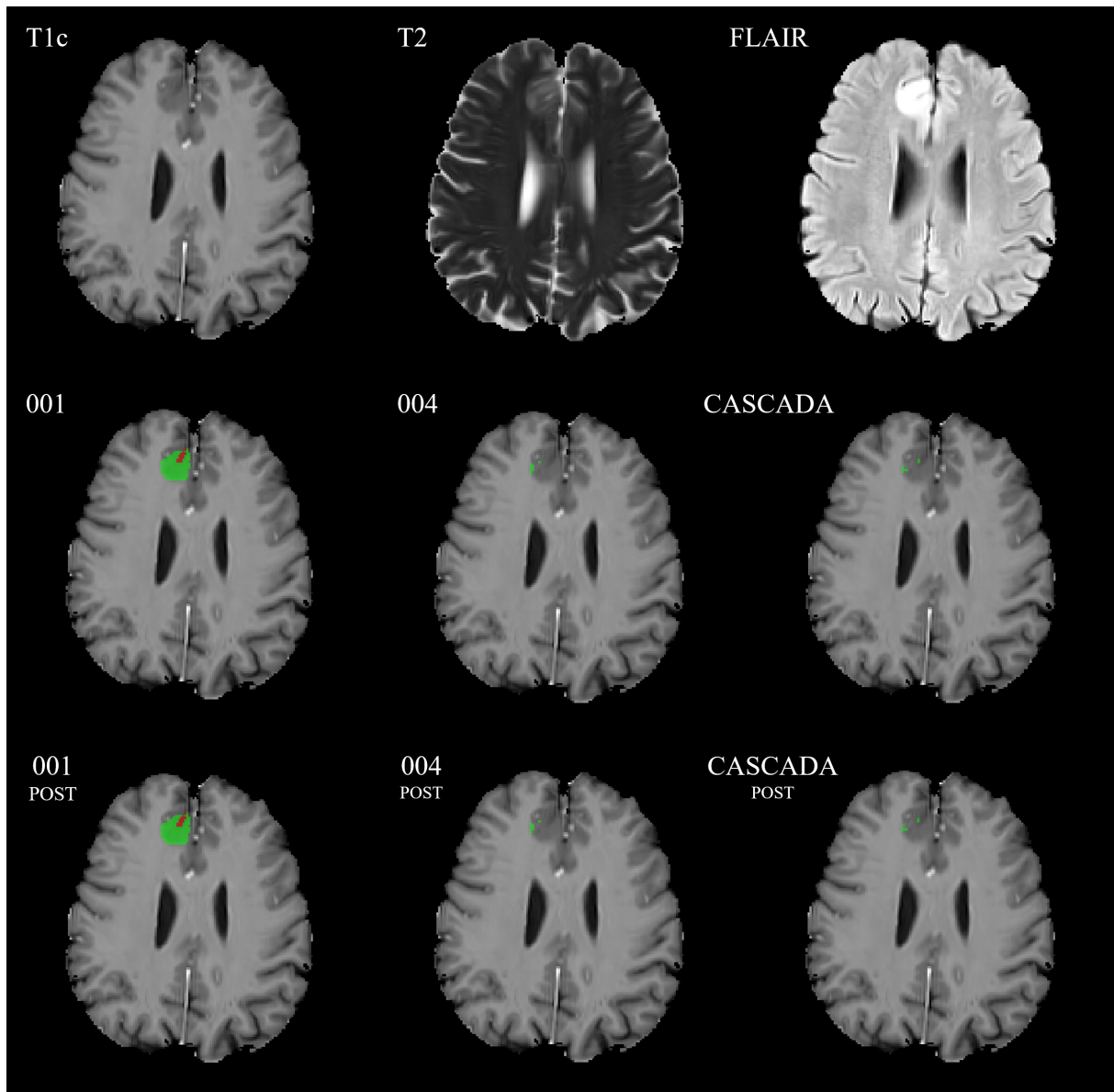


Figura 5.7: Segmentaciones obtenidas para los modelos evaluados externamente de un caso del conjunto de validación que obtuvo un DICE bajo.

Conclusiones y líneas futuras

En este capítulo se expondrán las conclusiones finales obtenidas de la realización del presente proyecto así como futuras líneas de trabajo para continuar y mejorar el mismo.

6.1 Conclusiones

En el presente proyecto se ha abordado la tarea de segmentación de gliomas mediante el uso de CNN. En él se han realizado una serie de experimentos utilizando distintas arquitecturas, metodologías y estrategias para tratar de obtener el mejor resultado posible, no solo en términos de precisión sino también enmarcándolos en el contexto en el que este tipo de redes serían utilizadas, como es el entorno hospitalario. De este modo, a la vista de los resultados obtenidos se pueden obtener las siguientes conclusiones:

- Las CNN entrenadas con casos de LGG y HGG presentan mejores resultados que aquellas dirigidas a la segmentación de un único grado tumoral, ya que se ven beneficiadas de la información proporcionada por los diferentes grados de glioma presentes en la base de datos y del aumento considerable del número de casos utilizados en la etapa de entrenamiento.
- La red que mejores resultados ha obtenido en el presente proyecto se trata de la U-NET identificada con el nombre HGGLGG001. Esta logró un DICE medio de 0.7331, 0.8624 y 0.7267, para el NR, tumor total y núcleo tumoral, respectivamente
- No obstante, la utilización de distintas metodologías y arquitecturas no proporcionaron diferencias en los resultados estadísticamente significativas.
- La etapa de preprocesamiento de las imágenes, basada en el filtrado del campo de inhomogeneidades, requiere de una especial atención y de un estudio meticuloso para garantizar que no se elimine información relevante que pueda perjudicar el desempeño global del modelo creado.
- Las técnicas de *data augmentation* utilizadas en este proyecto no lograron mejorar el desempeño de las redes.

- El postprocesamiento de las máscaras no logra mejorar significativamente los resultados obtenidos, si bien es cierto que requieren de pocos recursos de computación y pueden ser implementadas para otorgar una mayor coherencia a las segmentaciones obtenidas.
- Los resultados obtenidos en el presente proyecto son equiparables a los reportados en el estado del arte y los métodos de segmentación utilizados se presentan como una herramienta útil para la segmentación de tumores cerebrales.
- Las CNN poseen el potencial suficiente para automatizar la tarea de segmentación de gliomas de una forma precisa, equiparable a la realizada por radiólogos expertos.
- La tarea de segmentación de gliomas es realizada de forma eficiente mediante CNN, disminuyendo considerablemente el tiempo dedicado a ella si se hiciera de modo manual.
- La inclusión de este tipo de metodologías en la práctica clínica diaria supondría una mejora en el diagnóstico y pronóstico del paciente, así como una liberación de carga de trabajo del personal sanitario encargado de realizar esta tarea.

6.2 Líneas futuras

En el corto plazo, una de las líneas a investigar con mayor dedicación se trata del uso de técnicas de preprocesamiento de datos y de *data augmentation*, ya que en el presente proyecto estas han presentado un papel secundario. Como ya se ha comentado con anterioridad, estas etapas requieren de un estudio meticuloso de los efectos que producen y las transformaciones que realizan sobre los datos de entrada. En el caso de el preprocesamiento de los datos, en este trabajo tan solo se ha tratado de eliminar el artefacto producido por la inhomogeneidad del campo magnético durante el proceso de adquisición de las imágenes. Para ello se decidió utilizar un filtro ampliamente utilizado en la bibliografía, como es el filtro N4. Puede que este filtro produzca un filtrado excesivo en alguno de los casos, eliminando las posibles relaciones de intensidad presentes entre las distintas componentes tumorales que se tratan de segmentar. Los resultados podrían ser mejorados si se utiliza otro tipos de filtros de inhomogeneidades, así como una etapa previa de filtrado de ruido.

Respecto a las técnicas de *data augmentation*, existen una infinidad de posibles transformaciones de los datos que pueden ser utilizadas. No obstante, es necesario realizar un estudio de como estas afectan en cada una de las tareas a realizar y utilizar únicamente aquellas para las que la red se vea beneficiada. Asimismo, en los últimos años ha aparecido un nuevo tipo de redes, denominadas *Generative Adversarial Net* (GAN) e introducidas por primer vez por Goodfellow et al. [41], las cuales pueden ser utilizadas como técnicas de *data augmentation*. Estas redes se presentan como una gran novedad y con un gran potencial para la realización de esta tarea.

En el largo plazo, la creación de una base de datos propia de casos de gliomas sería de gran ayuda para aumentar la cantidad de casos de los que se dispone hoy en día. Esto, tal y como se ha podido observar en el proyecto, mejoraría considerablemente los resultados obtenidos, otorgando un mayor poder de generalización a las redes entrenadas. Otro de los caminos a explorar se trata del uso de redes preentrenadas con el fin de analizar si estas presentan alguna ventaja frente a las redes diseñadas y entrenadas única y exclusivamente para la segmentación de gliomas.

Bibliografía

- [1] N. M. Borden, S. E. Forseen y C. Stefan, *Imaging Anatomy of the Human Brain: A Comprehensive Atlas Including Adjacent Structures*. Springer Publishing Company, 2015, ISBN: 9781936287741 (vid. págs. 3, 4).
- [2] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, ép. Guyton Physiology Series. Elsevier, 2016, ISBN: 9781455770052 (vid. pág. 3).
- [3] E. C. Holland, “Progenitor cells and glioma formation.”, eng, *Current opinion in neurology*, vol. 14, n.º 6, págs. 683-688, dic. de 2001, ISSN: 1350-7540 (Print) (vid. pág. 5).
- [4] S. Ferguson y M. S. Lesniak, “Percival Bailey and the classification of brain tumors.”, eng, *Neurosurgical focus*, vol. 18, n.º 4, e7, abr. de 2005, ISSN: 1092-0684 (Electronic) (vid. pág. 5).
- [5] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues y D. W. Ellison, “The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.”, eng, *Acta neuropathologica*, vol. 131, n.º 6, págs. 803-820, jun. de 2016, ISSN: 1432-0533 (Electronic). DOI: 10.1007/s00401-016-1545-1 (vid. págs. 5, 6).
- [6] J. W. Kernohan y R. F. Mabon, “A simplified classification of the gliomas.”, eng, *Proceedings of the staff meetings. Mayo Clinic*, vol. 24, n.º 3, págs. 71-75, feb. de 1949, ISSN: 0092-699X (Print) (vid. pág. 5).
- [7] C. Daumas-Duport, B. Scheithauer, J. O’Fallon y P. Kelly, “Grading of astrocytomas. A simple and reproducible method.”, eng, *Cancer*, vol. 62, n.º 10, págs. 2152-2165, nov. de 1988, ISSN: 0008-543X (Print) (vid. pág. 5).
- [8] B. W. S. (P. Kleihues P. C. Burger, *Histological Typing of Tumours of the Central Nervous System*, 2.^a ed., ép. WHO. World Health Organization. International Histological Classification of Tumours. Springer-Verlag Berlin Heidelberg, 1993, pág. 114, ISBN: 978-3-540-56971-8. DOI: 10.1007/978-3-642-84988-6 (vid. pág. 5).

- [9] J. J. Plascak y J. L. Fisher, “Area-Based Socioeconomic Position and Adult Glioma: A Hierarchical Analysis of Surveillance Epidemiology and End Results Data”, *PLoS ONE*, vol. 8, n.º 4, J. S. Castresana, ed., e60910, abr. de 2013, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0060910 (vid. pág. 6).
- [10] H. Ohgaki y P. Kleihues, “Epidemiology and etiology of gliomas”, *Acta Neuropathologica*, vol. 109, n.º 1, págs. 93-108, ene. de 2005, ISSN: 1432-0533. DOI: 10.1007/s00401-005-0991-y (vid. págs. 6, 11).
- [11] S. Antoni, I. Soerjomataram, B. Møller, F. Bray y J. Ferlay, “An assessment of GLOBOCAN methods for deriving national estimates of cancer incidence”, *Bulletin of the World Health Organization*, vol. 94, n.º 3, págs. 174-184, mar. de 2016, ISSN: 0042-9686. DOI: 10.2471/BLT.15.164384 (vid. pág. 6).
- [12] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes y K. V. Leemput, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”, *IEEE Transactions on Medical Imaging*, vol. 34, n.º 10, págs. 1993-2024, 2015, ISSN: 0278-0062 VO - 34. DOI: 10.1109/TMI.2014.2377694 (vid. págs. 6, 11, 13, 17, 18, 20).
- [13] J. Undabeitia, S. Torres-Bayona, N. Samprón, M. Arrázola, A. Bollar, M. Armendariz, P. Torres, I. Ruiz, M. Caballero, L. Egaña, A. Querejeta, J. Villanua, E. Pardo, I. Etxegoien, G. Liceaga, M. Urtasun, M. Michan, J. Emparanza, P. Aldaz, A. Matheu y E. Úrculo, “Costes indirectos asociados al glioblastoma. Experiencia en un centro”, *Neurología*, vol. 33, n.º 2, págs. 85-91, mar. de 2018, ISSN: 0213-4853. DOI: 10.1016/J.NRL.2016.05.003 (vid. págs. 6, 7).
- [14] R. A. y L. M., “Burden and cost of neurological diseases: a European North–South comparison”, *Acta Neurologica Scandinavica*, vol. 132, n.º 1, págs. 16-22, DOI: 10.1111/ane.12339 (vid. pág. 7).
- [15] A. R. Giovagnoli, R. F. Meneses, A. Silvani, I. Milanesi, L. Fariselli, A. Salmaggi y A. Boiardi, “Quality of life and brain tumors: what beyond the clinical burden?”, *Journal of Neurology*, vol. 261, n.º 5, págs. 894-904, mayo de 2014, ISSN: 1432-1459. DOI: 10.1007/s00415-014-7273-3 (vid. pág. 7).

- [16] E. Alpaydin y F. Bach, *Introduction to Machine Learning*. MIT Press, 2009, ISBN: 9780262303262 (vid. pág. 7).
- [17] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdal, S. Turcotte, C. J. Pal, S. Kadoury y A. Tang, “Deep Learning: A Primer for Radiologists”, *RadioGraphics*, vol. 37, n.º 7, págs. 2113-2131, 2017. DOI: 10.1148/rg.2017170077 (vid. págs. 7, 8).
- [18] V. K. Ho, J. C. Reijneveld, R. H. Enting, H. P. Bienfait, P. Robe, B. G. Baumert y O. Visser, “Changing incidence and improved survival of gliomas”, *European Journal of Cancer*, vol. 50, n.º 13, págs. 2309-2318, sep. de 2014, ISSN: 0959-8049. DOI: 10.1016/J.EJCA.2014.05.019 (vid. pág. 11).
- [19] *MICCAI BraTS 2017: People | Section for Biomedical Image Analysis (SBIA) | Perelman School of Medicine at the University of Pennsylvania* (vid. pág. 11).
- [20] E. Eisenhauer, P. Therasse, J. Bogaerts, L. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe y J. Verweij, “New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)”, *European Journal of Cancer*, vol. 45, n.º 2, págs. 228-247, ene. de 2009, ISSN: 0959-8049. DOI: 10.1016/J.EJCA.2008.10.026 (vid. pág. 11).
- [21] P. Y. Wen, D. R. Macdonald, D. A. Reardon, T. F. Cloughesy, A. G. Sorensen, E. Galanis, J. Degroot, W. Wick, M. R. Gilbert, A. B. Lassman, C. Tsien, T. Mikkelsen, E. T. Wong, M. C. Chamberlain, R. Stupp, K. R. Lamborn, M. A. Vogelbaum, M. J. van den Bent y S. M. Chang, “Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group.”, eng, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 28, n.º 11, págs. 1963-1972, abr. de 2010, ISSN: 1527-7755 (Electronic). DOI: 10.1200/JCO.2009.26.3541 (vid. pág. 11).
- [22] D. Hanahan y R. A. Weinberg, “Hallmarks of cancer: the next generation.”, eng, *Cell*, vol. 144, n.º 5, págs. 646-674, mar. de 2011, ISSN: 1097-4172 (Electronic). DOI: 10.1016/j.cell.2011.02.013 (vid. pág. 14).
- [23] J. Juan-Albarracín, E. Fuster-Garcia, J. V. Manjón, M. Robles, F. Aparici, L. Martí-Bonmatí y J. M. García-Gómez, “Automated Glioblastoma Segmentation Based on a Multiparametric Structured Unsupervised Classification”, *PLOS ONE*, vol. 10, n.º 5, e0125143, mayo de 2015. DOI: 10.1371/journal.pone.0125143 (vid. pág. 14).
- [24] J. Juan-Albarracín, E. Fuster-Garcia, A. Pérez-Girbés, F. Aparici-Robles, Á. Alberich-Bayarri, A. Revert-Ventura, L. Martí-Bonmatí y J. M. García-Gómez, “Glioblastoma: Vascular Habitats Detected at Preoperative Dynamic Susceptibility-weighted Contrast-enhanced Perfusion MR Imaging Predict Survival”, *Radiology*, vol. 287, n.º 3, págs. 944-954, 2018. DOI: 10.1148/radiol.2017170845 (vid. pág. 14).

- [25] L. Jan, L. Teresa, I. A. J., S. A. W., H. Arend, V. Dirk, S. J. A. K. y V. H. Sabine, “Nosologic imaging of the brain: segmentation and classification using MRI and MRSP”, *NMR in Biomedicine*, vol. 22, n.º 4, págs. 374-390, 2008. DOI: 10.1002/nbm.1347 (vid. pág. 14).
- [26] A. Torregrosa Lloret, “Proyecto de generación no supervisada de imágenes nosológicas usando análisis multivariante para el estudio del cáncer de próstata.”, Tesis doct., Universitat Politècnica de València, jul. de 2017 (vid. pág. 14).
- [27] B. H. Menze, K. Van Leemput, D. Lashkari, T. Riklin-Raviv, E. Geremia, E. Alberts, P. Gruber, S. Wegener, M.-A. Weber, G. Szekely, N. Ayache y P. Golland, “A Generative Probabilistic Model and Discriminative Extensions for Brain Lesion Segmentation—With Application to Tumor and Stroke.”, eng, *IEEE transactions on medical imaging*, vol. 35, n.º 4, págs. 933-946, abr. de 2016, ISSN: 1558-254X (Electronic). DOI: 10.1109/TMI.2015.2502596 (vid. pág. 14).
- [28] K. Kamnitsas, W. Bai, E. Ferrante, S. G. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. C. H. Lee, B. Kainz, D. Rueckert y B. Glocker, “Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation”, *CoRR*, vol. abs/1711.0, 2017. arXiv: 1711.01468 (vid. págs. 15, 56).
- [29] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. Nori, A. Criminisi, D. Rueckert y B. Glocker, *DeepMedic for Brain Tumor Segmentation*. abr. de 2016, págs. 138-149, ISBN: 978-3-319-55523-2. DOI: 10.1007/978-3-319-55524-9_14 (vid. pág. 15).
- [30] J. Long, E. Shelhamer y T. Darrell, “Fully Convolutional Networks for Semantic Segmentation”, *CoRR*, vol. abs/1411.4, 2014 (vid. pág. 15).
- [31] O. Ronneberger, P. Fischer y T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, *CoRR*, vol. abs/1505.0, 2015 (vid. pág. 15).
- [32] G. Wang, W. Li, S. Ourselin y T. Vercauteren, “Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks”, *CoRR*, vol. abs/1709.0, 2017. arXiv: 1709.00382 (vid. págs. 15, 56).
- [33] F. Isensee, P. Kickingeder, W. Wick, M. Bendszus y K. H. Maier-Hein, “Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge”, *CoRR*, vol. abs/1802.1, 2018 (vid. págs. 16, 56).
- [34] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani y C. Davatzikos, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features.”, eng, *Scientific data*, vol. 4, pág. 170117, sep. de 2017, ISSN: 2052-4463 (Electronic). DOI: 10.1038/sdata.2017.117 (vid. pág. 17).

- [35] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu y X. Zheng, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”, *CoRR*, vol. abs/1603.0, 2016. arXiv: 1603.04467 (vid. pág. 21).
- [36] K. He, X. Zhang, S. Ren y J. Sun, “Deep Residual Learning for Image Recognition”, *CoRR*, vol. abs/1512.0, 2015 (vid. pág. 26).
- [37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *Journal of Machine Learning Research*, vol. 15, págs. 1929-1958, 2014 (vid. pág. 34).
- [38] D. P. Kingma y J. Ba, “Adam: A Method for Stochastic Optimization”, *CoRR*, vol. abs/1412.6, 2014 (vid. pág. 39).
- [39] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich y J. C. Gee, “N4ITK: Improved N3 Bias Correction”, *IEEE transactions on medical imaging*, vol. 29, n.º 6, págs. 1310-1320, jun. de 2010, ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2046908 (vid. pág. 41).
- [40] Y. Benjamini e Y. Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”, *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, n.º 1, págs. 289-300, 1995, ISSN: 00359246 (vid. pág. 43).
- [41] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville e Y. Bengio, “Generative Adversarial Networks”, *ArXiv e-prints*, jun. de 2014. arXiv: 1406.2661 [stat.ML] (vid. pág. 60).

Lista de figuras

1.1.	Esquema de las distintas regiones cerebrales.	4
1.2.	Organización de la sustancia blanca y sustancia gris en el cerebro. En la figura se puede observar una imagen potenciada en T1, T2 y un esquema de la distribución de la sustancia gris y blanca en el cerebro, respectivamente. Adaptada de [1]. . .	4
1.3.	Tasas estimadas de casos incidentes cáncer de cerebro y del sistema nervioso central, para ambos sexos, por edad, en todo el mundo en 2012. Fuente [11], disponible en http://gco.iarc.fr/today/home	6
1.4.	Esquema de un ejemplo de CNN para segmentación de sustancia gris y sustancia blanca a partir de una imagen T1 postcontraste.	8
1.5.	Resumen del proceso de obtención de la segmentación mediante el uso de un modelo lineal.	9
1.6.	Resumen del proceso de optimización de una red neuronal. En rojo se muestra el grafo diseñado formado por un conjunto de operaciones y en azul se muestra el grafo derivado de la utilización de la regla de la cadena, obtenido para el cálculo de los gradientes de los pesos.	10
3.1.	Ejemplo de imágenes correspondientes a un caso de HGG.	19
3.2.	Ejemplo de imágenes correspondientes a un caso de LGG.	19
4.1.	Características de las capas convolucionales.	24
4.2.	Ejemplos de la operación <i>max pooling</i>	25
4.3.	Arquitectura U-NET con cinco niveles de profundidad.	26
4.4.	Arquitectura U-NET con cuatro niveles de profundidad.	27
4.5.	Estructura de un <i>Residual Block</i>	28
4.6.	Estructura de un <i>Simple Block</i>	28
4.7.	Estructura de un <i>Inverted Simple Block</i>	28

4.8. Arquitectura W-NET con tres y cuatro niveles de profundidad, representados en las redes a y b , respectivamente.	29
4.9. Arquitectura de red tubular.	30
4.10. Tareas de segmentación a resolver por el bosque de redes.	31
4.11. Estructura de un <i>Dense Block</i>	32
4.12. Arquitectura de red utilizada para cada una de las CNN diseñadas en el bosque de redes.	33
4.13. Estrategia en cascada para la segmentación de gliomas.	35
4.14. Proceso de preparación de los datos y obtención de los <i>batches</i> de entrenamiento.	36
4.15. Proceso de obtención de los <i>patches</i> de una imagen.	37
4.16. Estrategia de entrenamiento utilizada para entrenar las diferentes CNNs para segmentación de tumor cerebral.	38
5.1. Coeficiente DICE obtenido para las distintas redes entrenadas únicamente con casos de LGG en el proceso de validación interna.	46
5.2. Coeficiente DICE obtenido para las distintas redes entrenadas con casos de LGG y HGG en el proceso de validación interna.	48
5.3. Coeficiente DICE obtenido para el experimento HGGLGG001 para el NR, tumor total y núcleo tumoral, durante la evaluación externa.	51
5.4. Coeficiente DICE obtenido para el experimento HGGLGG004 para el NR, tumor total y núcleo tumoral, durante la evaluación externa.	52
5.5. Coeficiente DICE obtenido para la metodología en cascada para el NR, tumor total y núcleo tumoral, durante la evaluación externa.	54
5.6. Segmentaciones obtenidas para los modelos evaluados externamente de un caso del conjunto de validación que obtuvo un DICE elevado.	57
5.7. Segmentaciones obtenidas para los modelos evaluados externamente de un caso del conjunto de validación que obtuvo un DICE bajo.	58

Lista de tablas

3.1.	Descripción de las imágenes del conjunto de datos de entrenamiento y validación.	18
3.2.	Especificaciones técnicas de la GPU NVIDIA [®] GTX Titan X.	21
4.1.	Tabla de verdad resumiendo la metodología de agregación de las distintas redes del bosque de redes.	31
4.2.	Experimentos realizados con U-NET para la segmentación de LGG.	40
4.3.	Experimentos realizados con U-NET para la segmentación de LGG y HGG. . . .	41
4.4.	Experimentos realizados con W-NET para la segmentación de LGG y HGG. . . .	41
4.5.	Experimento realizado con redes tubulares para la segmentación de LGG.	42
4.6.	Principales parámetros establecidos para la etapa de entrenamiento.	42
5.1.	Coefficiente DICE obtenido en el proceso de validación interna para los casos de LGG.	46
5.2.	Valores p obtenidos tras aplicar el test Wilcoxon y corregir mediante FDR a los resultados obtenidos para cada una de las redes de segmentación de LGG.	47
5.3.	Coefficiente DICE obtenido en el proceso de validación interna para el conjunto de datos con casos de LGG y casos de HGG.	48
5.4.	Valores p obtenidos tras aplicar el test Wilcoxon y corregir mediante FDR a los resultados obtenidos para cada una de las redes de segmentación de HGG y LGG.	49
5.5.	Métricas obtenidas para el experimento HGGLGG001 en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.	50
5.6.	Métricas obtenidas para el experimento HGGLGG001 con postproceso de las máscaras en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.	50
5.7.	Métricas obtenidas para el experimento HGGLGG004 en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.	51

5.8. Métricas obtenidas para el experimento HGGLGG004 con postproceso de las máscaras en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.	52
5.9. Valores p obtenidos tras el estudio del efecto del postprocesado de las máscaras de aquellos casos evaluados a través de la plataforma del BRATS.	52
5.10. Coeficiente DICE obtenido en el proceso de validación interna para el conjunto de datos con casos de LGG y casos de HGG para los experimentos realizados mediante la agregación de modelos.	53
5.11. Métricas obtenidas para el la cascada de modelos en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.	53
5.12. Métricas obtenidas para la cascada de modelos con postproceso de las máscaras en el proceso de validación externa para el conjunto de datos con casos de LGG y casos de HGG.	54
5.13. DICE obtenido por los principales participantes del BRATS 2017.	56
7.1. Tiempo empleado por el ingeniero biomédico en la ejecución de las tareas del proyecto.	76
7.2. Tiempo empleado por el cotutor en la ejecución de las tareas del proyecto.	77
7.3. Tiempo empleado por el tutor en la ejecución de las tareas del proyecto.	77
7.4. Coste total de recursos humanos.	77
7.5. Coste las licencias de <i>software</i> utilizadas en el proyecto.	78
7.6. Coste del equipamiento <i>hardware</i> utilizado en el proyecto.	78
8.1. Coste de ejecución material.	79
8.2. Coste total del proyecto.	80

Lista de acrónimos

BRATS *BRAIn Tumor Segmentation.*

CNN Redes Neuronales Convolucionales.

CPU Unidad de Central de Procesamiento.

DB *Dense Block.*

DCA-SVFMM *Directional Class Adaptive Spatially Varying Finite Mixture Model.*

DSC *Dynamic Suceptibility Contrast.*

EM *Expectation-Maximitation.*

EMMA *Ensemble of Multiple Models and Architectures.*

EP Edema Peritumoral.

FDR *False Discovery Rate.*

FLAIR Atenuación de FLuidos por Inversión Recuperación.

GAN *Generative Adversarial Net.*

GPU Unidad de Procesamiento Gráfico.

HGG Glioma de Alto Grado.

IA Inteligencia Artificial.

IRM Imagen por Resonancia Magnética.

ISB *Inverted Simple Block.*

LGG Glioma de Bajo Grado.

MICCAI *Medical Image Computing and Computer Assisted Intervention.*

NN Núcleo Necrótico.

NR Núcleo tumoral con Realce de contraste.

NSR Núcleo tumoral Sin Realce de contraste.

OMS Organización Mundial de la Salud.

RB *Residual Block.*

rCBF *relative Cerebral Blood Flow.*

rCBV *relative Cerebral Blood Volume.*

RECIST *Response Evaluation Criteria in Solid Tumors.*

ReLU *Rectified Linear Units.*

RM Resonancia Magnética.

SB *Simple Block.*

SNC Sistema Nervioso Central.

SNP Sistema Nervioso Periférico.

TS Tejido Sano.

Parte II

Presupuesto

Presupuesto desglosado

En el este capítulo se desglosará el presupuesto necesario para la elaboración del presente proyecto, incluyendo costes de recursos humanos y costes de material.

7.1 Objetivo y estructura

El objetivo principal del presente presupuesto desglosado se trata de establecer los costes que este proyecto ha supuesto de una forma estructurada. Asimismo, se pretende que este sirva de guía en la estimación presupuestario de otros proyectos similares.

Para la elaboración del mismo se decidió dividir los costes en función de costes relacionados con los recursos humanos necesarios para la consecución del proyecto, así como los costes en equipamiento y material que supuso la realización de este. Dichos costes en recursos humanos serán justificados por las actividades realizadas y el tiempo dedicado a ellas por cada una de las partes implicadas.

Respecto a los costes en equipamiento y materiales, estos fueron presupuestados teniendo en cuenta el coste imputable de los mismos según el tiempo de uso utilizando la Ecuación 7.1. En esta donde C hace referencia al coste imputable, c al precio unitario por coste del equipo, t al tiempo de uso de cada equipo y T el periodo de vida útil del equipo o material, el cual se decidió establecer en cinco años, ya que este se considera un periodo de tiempo en el cual los equipos utilizados quedarían obsoletos dado el rápido crecimiento del campo objeto de la investigación.

$$C = \frac{c \cdot t}{T} \quad (7.1)$$

Una vez obtenidos los costes de personal y material, el presupuesto de ejecución material fue obtenido como la suma de los mismos. Posteriormente, a este presupuesto se le añadió un 16 % de gastos, así como un 6 % de beneficio industrial, obteniendo así el llamado presupuesto de ejecución por contrata. Finalmente, el presupuesto total se obtuvo añadiendo el 21 % de IVA al presupuesto de ejecución por contrata.

7.2 Coste de recursos humanos

El coste relacionado con los recursos humanos precisados constituye la mayor parte del presupuesto requerido para la elaboración del presente proyecto. Así, para ello fue necesario el trabajo de un ingeniero biomédico bajo la supervisión de un tutor con el título de Doctor en Filosofía (PhD) y un cotutor con el título de Maestría en Ciencias (MSc).

El ingeniero biomédico fue el encargado en todo momento del diseño, implementación y evaluación de todos los modelos utilizados en el proyecto para la segmentación automática de tumores cerebrales. Por otro lado, los tutores desempeñaron una importante labor de supervisión y guía durante todo el trascurso del trabajo, así como una labor docente relacionada con la naturaleza de los tumores cerebrales, el aprendizaje profundo y las CNN.

A continuación, se decidió justificar el número de horas dedicadas por cada individuo implicado en el desarrollo de este proyecto, con el fin de fundamentar los costes asociados a los recursos humanos. De este modo, las tareas realizadas por el ingeniero biomédico se encuentran justificadas en la tabla 7.1. En ella se puede observar como el diseño e implementación de metodologías engloba la mayor parte del tiempo dedicado en el proyecto. Esto es debido a que esta tarea conlleva el entrenamiento de las distintas redes creadas en el trabajo, y este requiere de una gran cantidad de horas para cada una de ellas. Por otro lado, la tarea relacionada con el aprendizaje también supuso una cantidad considerable de horas de dedicación ya que en esta se incluyen la realización de un curso gratuito sobre aprendizaje profundo impartido por Google a través su plataforma Udacity¹, así como la docencia impartida por parte del cotutor y el aprendizaje propio. De este modo, el tiempo empleado por el ingeniero biomédico ascendió a un total de 500 horas. Por otro lado, el tiempo dedicado por el cotutor y el tutor para la realización de las tareas necesarias para la consecución del presente proyecto se encuentran resumidas en las tablas 7.2-7.3. Finalmente, **el coste total dedicado a los recursos humanos fue de 19 800.00€** y se encuentra detallado en la tabla 7.4.

Tabla 7.1: Tiempo empleado por el ingeniero biomédico en la ejecución de las tareas del proyecto.

CONCEPTO	UNIDAD	Nº DE UNIDADES
Aprendizaje e introducción a las CNN	Horas	100
Planificación del proyecto	Horas	10
Diseño e implementación de metodologías de segmentación	Horas	250
Evaluación de las metodologías	Horas	50
Análisis de los resultados obtenidos	Horas	20
Redacción del proyecto	Horas	60
Revisión final del documento	Horas	10
TOTAL	Horas	500

¹Curso disponible en <https://eu.udacity.com/course/deep-learning--ud730>

Tabla 7.2: Tiempo empleado por el cotutor en la ejecución de las tareas del proyecto.

CONCEPTO	UNIDAD	Nº DE UNIDADES
Docencia	Horas	20
Supervisión y guía	Horas	30
Revisión del documento	Horas	10
TOTAL	Horas	60

Tabla 7.3: Tiempo empleado por el tutor en la ejecución de las tareas del proyecto.

CONCEPTO	UNIDAD	Nº DE UNIDADES
Supervisión y guía	Horas	30
Revisión del documento	Horas	10
TOTAL	Horas	40

Tabla 7.4: Coste total de recursos humanos.

CONCEPTO	UNIDAD	Nº DE UNIDADES	PRECIO UNITARIO (€/hora)	TOTAL (€)
Coste ingeniero biomédico	Horas	500	30.00	15 000.00
Coste cotutor	Horas	60	40.00	2 400.00
Coste tutor	Horas	40	60.00	2 400.00
TOTAL				19 800.00

7.3 Coste de equipamiento

Respecto al equipamiento necesario para la realización del proyecto, dado que este fue realizado mediante la librería *open source* TensorFlowTM y desarrollado en Python, ambos de coste totalmente gratuito, tal y como se especifica en la sección 3.3. Por otro lado, la memoria del proyecto fue íntegramente escrita mediante L^AT_EX, lo cual no supuso ningún coste adicional. Respecto al sistema operativo utilizado para la elaboración del proyecto, se hizo uso de una licencia Microsoft[®] Windows 10 Pro. El coste total imputable de las licencias utilizadas en el presente proyecto se encuentran detalladas en la tabla 7.5.

Tabla 7.5: Coste las licencias de *software* utilizadas en el proyecto.

CONCEPTO	UNIDAD	Nº DE UNIDADES	VIDA ÚTIL DE LA LICENCIA (años)	PERIODO DE USO (años)	PRECIO UNITARIO sin IVA (€)	COSTE IMPUTABLE sin IVA (€)
Microsoft [®] Windows 10 Pro	u	1	5	0.5	230.58	23.06
TensorFlow TM	u	1	5	0.5	0.00	0.00
Python 3.6	u	1	5	0.5	0.00	0.00
L ^A T _E X	u	1	5	0.5	0.00	0.00
TOTAL						23.06

De este modo, el principal coste de esta sección fue abarcado por el equipamiento *hardware*. Este comprendía la utilización de una de una GPU NVIDIA[®] GTX Titan X y un ordenador MSI GP62 6QF *Leopard Pro* (Core i7, 16Gb de RAM), como se puede observar en la sección 3.2. El coste total del equipamiento *hardware* se encuentra resumido en la tabla 7.6.

Tabla 7.6: Coste del equipamiento *hardware* utilizado en el proyecto.

CONCEPTO	UNIDAD	Nº DE UNIDADES	VIDA ÚTIL DE LA LICENCIA (años)	PERIODO DE USO (años)	PRECIO UNITARIO sin IVA (€)	COSTE IMPUTABLE sin IVA (€)
MSI GP62 6QF <i>Leopard Pro</i> (Core i7, 16Gb de RAM)	u	1	5	0.5	805.79	80.58
GPU NVIDIA [®] GTX Titan X	u	1	5	0.5	884.96	88.50
TOTAL						169.08

Cabe destacar que todos los costes imputables, tanto del equipamiento *hardware* como *software*, fueron calculado mediante la Ecuación 7.1 considerando como periodo de vida útil cinco años y periodo de uso seis meses. Así, **el coste total de equipamiento fue de 192.14€.**

Resumen del presupuesto

En este capítulo se se realizará un resumen de todos los costes implicados en la ejecución del proyecto, previamente detallados en el presupuesto desglosado.

8.1 Presupuesto de ejecución material

El presupuesto de ejecución material de este proyecto fue calculado como la suma de los costes de recursos humanos y los costes de material. De este modo, **el coste de ejecución material ascendió a la cifra de 19 992.14€**, el cual se encuentra detallado en la tabla 8.1.

Tabla 8.1: Coste de ejecución material.

CONCEPTO	TOTAL (€)
Coste de recursos humanos	19 800.00
Coste <i>hardware</i>	169.08
Coste <i>software</i>	23.06
TOTAL	19 992.14

8.2 Presupuesto de ejecución por contrata

El presupuesto de ejecución por contrata de este proyecto se encuentra calculado como la suma de los costes de ejecución material, los gastos generales (16 %) y el beneficio industrial(6 %). Así, el coste de ejecución material se encuentra detallado en la tabla 8.2 y fue de **24 390.41€**.

8.3 Presupuesto final del proyecto

El presupuesto final de este proyecto se encuentra calculado como la suma de los costes de ejecución por contrata y el IVA (21%) correspondiente. De este modo, **el coste total del proyecto fue de 29 512.38€**, el cual se encuentra resumido en la tabla 8.2.

Tabla 8.2: Coste total del proyecto.

CONCEPTO	TOTAL (€)
Presupuesto de ejecución material	19 992.14
Gastos generales	3198.74
Beneficio industrial	1199.53
Presupuesto de ejecución por contrata	24 390.41
IVA	5121.97
TOTAL	29 512.38