



# UNIVERSIDAD DE GRANADA

## TRABAJO FIN DE GRADO INGENIERÍA INFORMÁTICA

### Diagnóstico de tumores cerebrales a partir de IRM mediante aprendizaje profundo

---

Diseño e implementación de arquitecturas neuronales para la  
clasificación y la segmentación

**Autor**  
Jaime Castillo Uclés

**Directora**  
Rosa María Rodríguez Sánchez



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE  
TELECOMUNICACIÓN

—  
Granada, Junio de 2024







# **Diagnóstico de tumores cerebrales a partir de IRM mediante aprendizaje profundo : Diseño e implementación de arquitecturas neuronales para la clasificación y la segmentación**

Jaime Castillo Uclés

**Palabras clave:** clasificación, segmentación, red convolucional, U-net, conexiones residuales, resonancia, slice, tumor

## **Resumen**

Poner aquí el resumen.



**Brain tumor diagnosis from MRI images using Deep  
Learning : Design and neuronal architecture implementation  
for classification and segmentation**

Jaime Castillo Uclés

**Keywords:** Keyword1, Keyword2, Keyword3, ....

**Abstract**

Write here the abstract in English.



---

Yo, **Jaime Castillo Uclés**, alumno de la titulación Grado en Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 45924736S, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Jaime Castillo Uclés

Granada a 24 de Junio de 2024 .



---

Dra. **Rosa María Rodríguez Sánchez**, Profesora del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

**Informan:**

Que el presente trabajo, titulado *Diagnóstico de tumores cerebrales a partir de IRM mediante aprendizaje profundo , Diseño e implementación de arquitecturas neuronales para la clasificación y la segmentación* , ha sido realizado bajo su supervisión por **Jaime Castillo Uclés**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 25 de Junio de 2024 .

**La directora:**



# Agradecimientos

Desde el inicio de mi etapa universitaria y en este tan corto período de tiempo, puedo decir que los cambios en mí, en magnitud, han sido positivamente radicales. En este trabajo que representa el final de esta primera etapa, sólo puedo estar profundamente agradecido por todas las personas que se han cruzado de forma positiva en algún punto conmigo. A mi familia, que siempre ha creído en mí, dándome la oportunidad de poder formarme incluso cuando más me ha costado afrontar este proceso de cambio; a los buenos amigos que he hecho en el camino y a todos los buenos profesores de los que he recibido clase o apoyo, que han sido inspiradores para mí.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	1
1.2. Metodología . . . . .	4
1.2.1. Conjunto de datos . . . . .	4
1.2.2. Línea de investigación . . . . .	12
1.2.3. Evaluación y métricas . . . . .	12
<b>2. Estado del arte</b>	<b>17</b>
2.1. Revisión histórica de clasificación binaria de tumores . . . . .	18
2.2. Revisión histórica de segmentación . . . . .	18
2.2.1. Métodos que se enfocan en la arquitectura . . . . .	20
2.2.2. Métodos que tratan el desbalanceo . . . . .	24
2.2.3. Métodos que tratan la información multi-modal . . . . .	27
2.3. Enfoques actuales para la segmentación . . . . .	29
2.3.1. Basados en Transformers . . . . .	29
<b>3. Metodología</b>	<b>31</b>
3.1. Análisis de los recursos disponibles . . . . .	31
3.2. Preprocesado de Datos . . . . .	32
3.2.1. Elección de dimensionalidad de las entradas . . . . .	33
3.2.2. Normalizado de las imágenes . . . . .	33
3.2.3. Recortado de imagen . . . . .	34
3.2.4. Undersampling . . . . .	34
3.3. Elección de modelos . . . . .	35
3.3.1. Codificador y representación latente . . . . .	35
3.3.2. Modelo de clasificación . . . . .	37
3.3.3. Modelo de segmentación . . . . .	37
3.4. Diseño de las arquitecturas . . . . .	37
3.4.1. Funciones de activación . . . . .	37
3.4.2. Construcción del codificador y representación latente .	38
3.4.3. Arquitectura para clasificación . . . . .	38
3.4.4. Arquitectura para segmentación . . . . .	38
3.5. Optimización de las arquitecturas . . . . .	38

---

3.5.1. Optimizador . . . . .	38
3.5.2. Funciones de pérdida . . . . .	38
3.5.3. One-Cycle Policy . . . . .	39
3.6. Validación y garantías . . . . .	39
3.7. Solución teórica a la predicción de la evolución . . . . .	39
<b>4. Experimentación</b>	<b>41</b>
<b>5. Conclusiones y Trabajos Futuros</b>	<b>43</b>
<b>Bibliografía</b>	<b>43</b>
<b>Glosario</b>	<b>49</b>

# Índice de figuras

1.1.	Porcentaje de Glioblastomas y Meningiomas en el conjunto de datos . . . . .	5
1.2.	Cantidad de instancias temporales en el conjunto de datos . .	6
1.3.	Visualización de imágenes MRI de tumores en adultos de origen europeo y americano . . . . .	7
1.4.	Visualización de imágenes MRI de tumores en niños y adultos de origen africano . . . . .	8
1.5.	Visualización de imágenes MRI de tumores en adultos de origen europeo y americano con su segmentación. . . . .	9
1.6.	Visualización de imágenes MRI de tumores en niños y adultos de origen africano con su segmentación. . . . .	9
1.7.	Distribución del tejido tumoral en Gliomas. . . . .	10
1.8.	Distribución del tejido tumoral en Meningiomas. . . . .	11
1.9.	Interpretación del coeficiente de Similaridad Dice . . . . .	14
1.10.	Distancias de Hausdorff entre dos conjuntos . . . . .	15
2.1.	Evolución histórica del estado del arte hasta 2021. . . . .	19
2.2.	Comparación entre arquitecturas de una y múltiples trayectorias. Imagen de [Liu et al., 2023] . . . . .	21
2.3.	Arquitectura de dos vías de [Havaei et al., 2017] . . . . .	22
2.4.	Comparación de distintas arquitecturas encoder-decoder . . .	23
2.5.	Estructura de método en cascada de [Wang et al., 2018] . . .	24
2.6.	Arquitectura del autoencoder regularizador de [Myronenko, 2019]	26
2.7.	Arquitectura de [Zhou et al., 2021] . . . . .	27
2.8.	Modelo especializado en la correlación de las modalidades . .	28
2.9.	Red [Zhou et al., 2021] de fusión de representaciones latentes	28
3.1.	Comparativa de rendimiento de las GPU disponibles . . . .	32
3.2.	Esquema de la arquitectura empleada para inicializar el codificador y representación latente . . . . .	36



# Índice de cuadros

3.1. Porcentaje de imágenes conservadas tras undersampling. . . . 35



# Capítulo 1

## Introducción

Los tumores cerebrales son una de las formas más letales de cáncer. Específicamente, los glioblastomas y sus variantes difusas son los más comunes y agresivos tipos de tumor del sistema nervioso central en adultos. Su alta heterogeneidad en apariencia, forma e histología los convierte en una de las patologías más difíciles de diagnosticar, de tratar y un reto para el campo de la imagen médica.

Desde el punto de vista de la ingeniería y la informática, vemos como sin duda la aplicación de técnicas de Visión por Computador es una de las máximas para la investigación en imagen médica en la actualidad. Sólo considerando su aplicación en el diagnóstico de enfermedades, desde 2008 el número de publicaciones promedio realizadas por año se ha incrementado notablemente tanto que actualmente es diez veces mayor que en sus inicios.

Resultados notables como la inclusión de robots especializados para la cirugía [Cheng et al., 2022] o buenos resultados en competiciones de ciencia de datos que replican la precisión médica en el diagnóstico mediante imagen [Bulten et al., 2022] evidencian esta tendencia. El trabajo conjunto de personal médico e ingenieros promete seguir dando resultados que de forma separada eran inaccesibles.

### 1.1. Objetivos

Con este trabajo se persigue la creación de una arquitectura basada en aprendizaje profundo para equipar a un programa de uso médico, estudiarla y compararla junto a trabajos previos y estado del arte. Este programa tiene el objetivo de la ayuda en la evaluación del diagnóstico y pronóstico de un posible paciente de tumor cerebral y en caso afirmativo, la ayuda en la aplicación de la terapia por radiación.

Se seguirá un planteamiento similar al seguido en la competición **BraTS: Brain Tumor Segmentation 2023** [Baid et al., 2021] históricamente reconocida por ser un benchmark recurrente de las capacidades de las arquitecturas profundas en el campo de la imagen médica.

BraTS es una competición que se define como un conglomerado de diferentes tareas relevantes en el diagnóstico de los tumores cerebrales « Cluster of Challenges ». En 2023 dando especialmente importancia a la generalidad de un modelo que mantenga los resultados anteriores para pacientes más diversos (de diferente origen étnico y de diferentes edades) y con diferentes tipos de tumores. En concreto para 2023, se contemplaron las siguientes tareas por separado: segmentación de glioblastomas en adultos, segmentación de meningiomas en adultos, segmentación de tumores pediátricos, segmentación de tumores de pacientes de origen africano, generación de pruebas faltantes y generación de partes de la imagen.

De forma análoga a esta competición, se plantea conseguir dicho objetivo a partir de la resolución de las siguientes tareas.

1. **Segmentación de los tumores.** La segmentación de la lesión tumoral de los glioblastomas y los meningiomas.
2. **Clasificación entre tipos de tumores.** Clasificación binaria entre glioblastomas y meningiomas. El programa indicará de qué tipo de tumor se trata una prueba dada.
3. **Predicción de la evolución.** La segmentación a corto plazo de la más probable instancia futura a partir de la resonancia actual. Ya no solo se pretende dar una segmentación y clasificación que pudieran aportar valor en las decisiones médicas, sino predecir una nueva segmentación a partir de la resonancia apoyándose en casos similares y en la segmentación de la metástasis de la resonancia actual.

A continuación, detallaremos de una forma más profunda la naturaleza de este planteamiento.

Sólo en los EEUU para 2024 se esperan 25400 nuevos casos de tumor cerebral. La supervivencia de estos a los cinco años es del 33.8% de los pacientes. [cancer.org, 2024]

El cerebro no tiene terminaciones nerviosas. Los pacientes no sienten dolor a causa de un tumor cerebral por sí mismo, lo cual hace que no exista una alerta sobre el paciente que lo motive a buscar ayuda médica en las primeras fases de la patología. Generalmente, acaban buscando ayuda médica por la aparición de otros indicios relacionados difíciles de distinguir de otras patologías agudas y de menor transcendencia como visión borrosa, pérdida

del control, etc. Además, los glioblastomas son tumores de muy rápido crecimiento pueden llegar a estar en una fase avanzada desde su inicio en tan solo 2-3 meses.

Por estos motivos, es común llegar tarde. Tomando mucha importancia el diagnóstico temprano para su superación. Es en este punto donde se tiene el objetivo de evaluar las capacidades del aprendizaje profundo para la segmentación de tumores especialmente los que en sus inicios podrían ser pasados desapercibidos por incluso el ojo médico y eventualmente la segmentación de las zonas potenciales en la aparición de nuevos tumores asociado a una probabilidad.

En general, los tumores cerebrales son difíciles de tratar y son resistentes a terapias convencionales usadas en otros tipos de cánceres como la quimioterapia debido a los desafíos que presenta el cerebro para tolerar ciertos químicos, transportar medicamentos dentro de él y la alta importancia que tiene en este órgano la optimización del uso de tratamientos que puedan ser invasivos. En otras palabras, el uso de tratamientos basados en la extirpación o en la medicación pueden ser arriesgados. Por tanto, el tratamiento más común de estos está basado en la radioterapia.

A la hora de aplicar un tratamiento de radioterapia siempre se tiene el objetivo de ser lo menos invasivo posible. Para ello, el médico debe ser lo más preciso posible en introducir la segmentación correcta en la que se aplicarán los rayos.

Uno de los objetivos específicos para la ayuda en el tratamiento se basaría en el uso del modelo para automatizar esta tarea ya que podría suponer ahorrar costes en errores humanos y en tiempo a veces escaso para el personal médico cuya tarea se reduciría a corregir dicha segmentación. En nuestro caso, integrándolo en un programa de uso médico.

Por otro lado, de forma general podemos interpretar este trabajo como un **sistema de apoyo a la decisión médica** ya que en ningún caso respecto al desarrollo actual de este problema se pretende sustituir la decisión final del personal médico. Aunque sí aprovechar las capacidades que puede ofrecer el aprendizaje profundo en un mejor diagnóstico a través de caracterizar mejor el tejido afectado: segmentándolo y clasificándolo.

Este sistema de apoyo a la decisión médica se plasmaría en una aplicación de escritorio que el personal médico pueda usar como paso intermedio entre la recogida de las imágenes del escáner y un posible tratamiento de radioterapia.

## 1.2. Metodología

### 1.2.1. Conjunto de datos

Una de las **limitaciones frecuentes en el campo de la imagen médica** es la poca disponibilidad de datos. En general y también para nuestro problema, las dificultades que se presentan a la hora de construir un conjunto de datos médico grande son:

1. **Poca densidad de pruebas médicas.** A pesar de que la densidad de casos es alta, es frecuente que la cantidad de pruebas que se realizan sea mucho más baja. Especialmente, para datos médicos es frecuente encontrarse con pocos datos en magnitud con la necesidad de variabilidad en la muestra que tienen las técnicas típicas de optimización.
2. **La desvinculación de los pacientes de sus datos.** El tratamiento de un dato médico siempre supone la eliminación de cualquier identificativo que ponga en riesgo la privacidad de este. Suponiendo un trabajo adicional para el personal médico que no siempre se puede asumir.
3. **No existe una fuerte centralización de datos.** Al igual que los avances en imagen médica, el interés por la construcción de una base de datos única que recoja los máximo datos posibles es también reciente haciendo que en la actualidad la mayoría de los datos estén distribuidos en muchos centros médicos diferentes con formatos diferentes.

Partimos del conjunto de datos **BraTS** y serán los que únicamente utilizaremos para todo el trabajo ya que son los únicos que se pueden encontrar en la red pidiendo un acceso a ellos de una forma simple e incluso podríamos decir legal.

Hasta nuestro conocimiento, BraTS es el mayor conglomerado de resonancias magnéticas en la actualidad para los desafíos que planteamos en este trabajo. Otros datasets usados por la comunidad para este problema como el incluido en **Medical Segmentation Decathlon** descubrimos que es un subconjunto de **BraTS**.

**BraTS** está patrocinado y organizado por la ASNR (American Society of Neurology), MICCAI (Medical Image Computing and Computer Assisted Intervention Society), National Cancer Institute, la Universidad de Pensilvania e Intel entre otros.

**Los datos de BraTS son heterogéneos** ya no sólo externamente con diferentes tipos tumores (glioblastomas y meningiomas) y de pacientes de orígenes distintos y de diferentes edades sino también internamente con

lesiones más y menos avanzadas en el tiempo (low- and high-grade) y con datos de multitud de centros que han sido escaneados con distintos escáneres.

Definiremos como nuestro conjunto de datos  $X$  a un conjunto de resonancias magnéticas cerebrales completas. Este lo tenemos en formato **nii** : **Neuroimaging Informatics Technology Initiative (NIIfTI)** el cual podremos hacer operar con el directamente con su versión comprimida en **.gz** : **GNU ZIP** gracias al uso de la librería **nibabel** la cual nos permitirá la lectura y conversión de cada resonancia a un 3D-array de NumPy de dimensiones  $240 \times 240 \times 155$  que representa un array de 155 imágenes de resolución  $240 \times 240$  del cerebro del paciente dividida en partes equidistantes.

Por otro lado, definimos el conjunto de etiquetas  $Y$  como otro 3D-array de las mismas dimensiones que en el conjunto de datos donde se muestra la segmentación (Ground Truth) de los tumores.

En términos numéricos, contando solo los datos que tenemos sobre adultos tenemos 1251 resonancias de glioblastomas de 1133 pacientes diferentes y 1000 resonancias de meningiomas de 944 pacientes diferentes. Adicionalmente, tenemos resonancias de glioblastomas pediátricos y de adultos de origen africano que brindan de una distribución mucho más rica al dataset, aunque estos son una minoría siendo 99 y 60 respectivamente.

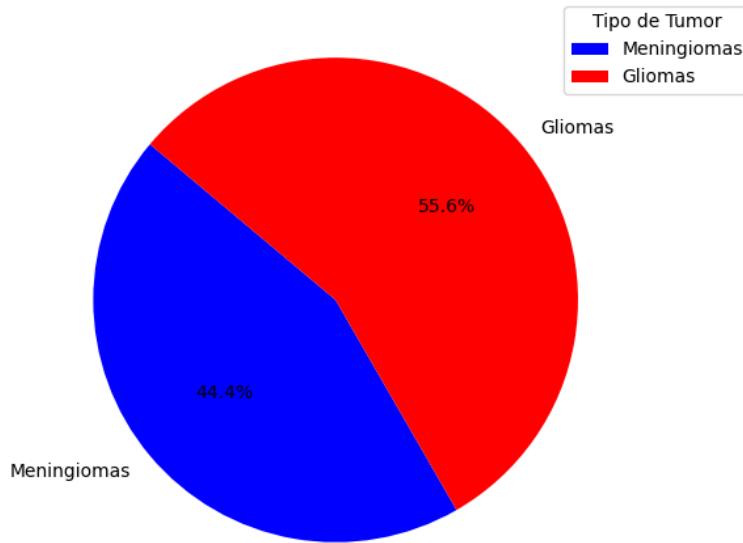


Figura 1.1: Porcentaje de Glioblastomas y Meningiomas en el conjunto de datos

Por otro lado, observamos los pacientes que tienen más de una resonancia versus los que tienen una única resonancia para caracterizar la cantidad

de instancias temporales que tenemos en el dataset.

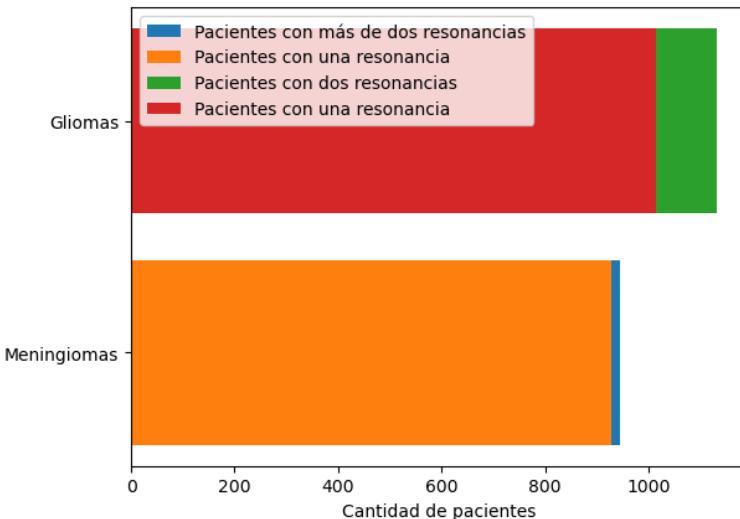


Figura 1.2: Cantidad de instancias temporales en el conjunto de datos

En comparativa, observamos como la mayoría de pacientes no tienen más de una resonancia magnética, dejando sólo una cantidad de 118 pacientes de glioblastoma y de 16 pacientes de meningiomas con varias resonancias en el tiempo.

### Visualizado de datos

A continuación, detallaremos más profundamente los datos haciendo algunas visualizaciones.

Por resonancia magnética para todos los tumores del conjunto de datos tenemos distintos tipos de muestras según las características de la frecuencia empleada en la toma de la resonancia, **T1-weighted** o **T2-weighted**. [Dominic LaBella, 2023]

El tiempo de repetición (TR) es la cantidad de tiempo entre secuencias de pulsos sucesivas aplicadas al mismo segmento. El tiempo hasta el eco (TE) es el tiempo entre la entrega del pulso sobre el tejido y la recepción de la señal de eco.

Las imágenes ponderadas en T1 se producen utilizando tiempos TE y

TR cortos. Por el contrario, las imágenes ponderadas en T2 se producen utilizando tiempos TE y TR más largos.

Las imágenes ponderadas en T1 muestran con más detalle la anatomía normal del tejido blando y la grasa. Las imágenes ponderadas en T2 muestran con más detalle el líquido y alteraciones (p. ej., tumores, inflamación, traumatismo).

En resumen, tenemos cuatro 3D-arrays por resonancia magnética según frecuencia de señal y aplicando o no un agente de contraste:

1. **T1N Pre-contrast T1-weighted** : Resonancia en frecuencia T1 sin suministrarle ningún agente de contraste al paciente.
2. **T1C Post-contrast T1-weighted** : Resonancia en frecuencia T1 suministrándole un agente de contraste al paciente.
3. **T2W T2-weighted** : Resonancia en frecuencia T2 convencional.
4. **T2F T2-weighted Fluid Attenuated Inversion Recovery** : Resonancia en frecuencia T2 en la que se anula la señal proveniente del líquido cefalorraquídeo.

A continuación, observamos las imágenes producidas por una resonancia magnética en las diferentes pruebas para los dos tipos de tumores que tenemos.

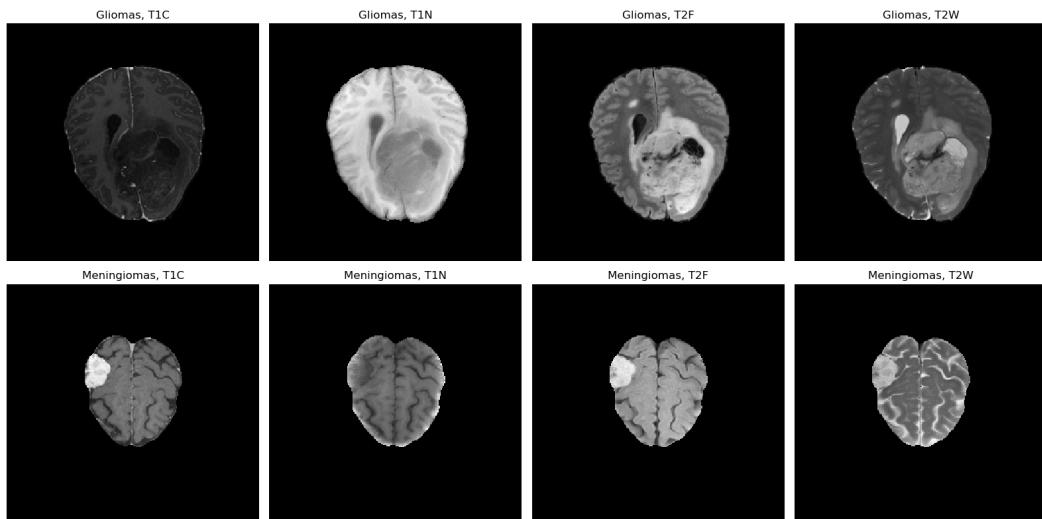


Figura 1.3: Visualización de imágenes MRI de tumores en adultos de origen europeo y americano

Por otro lado, podemos observar algunas imágenes de las resonancias más específicas que tenemos, pediátrica y de adultos de origen africano.

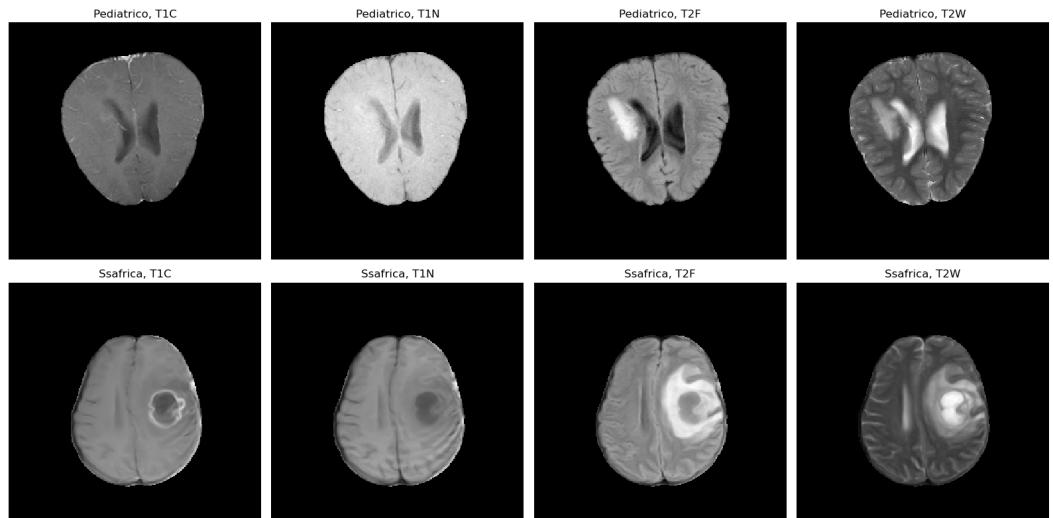


Figura 1.4: Visualización de imágenes MRI de tumores en niños y adultos de origen africano

A continuación, exploraremos el **etiquetado de las resonancias**, es decir, su segmentación realizada por el personal médico colaborador en BraTS.

La etiquetas de la segmentación pueden tomar cuatro valores, los del intervalo  $[0, 3]$  que están relacionados con el tipo de tejido que segmentan. Tenemos tres tipos de tejidos que se relacionan con el valor de etiquetas en el array.

1. **Etiqueta 1. NCR:** Tejido Necrótico. Núcleo del tumor tejido sin vida y usualmente reseco.
2. **Etiqueta 2. ED:** Edema peritumoral. Tejido afectado resultado de la expansión del tumor generalmente acumulación de líquido y tejido sano desplazado.
3. **Etiqueta 3. ET:** Enhancing tumor. Tejido donde se encuentra la principal actividad de expansión del tumor. Área de actividad tumoral más agresiva o prolífica.
4. **Etiqueta 0. Sano:** Tejido sano o la no existencia de tejido es etiquetado con 0.

Podemos ver la segmentación de las resonancias antes visualizadas.

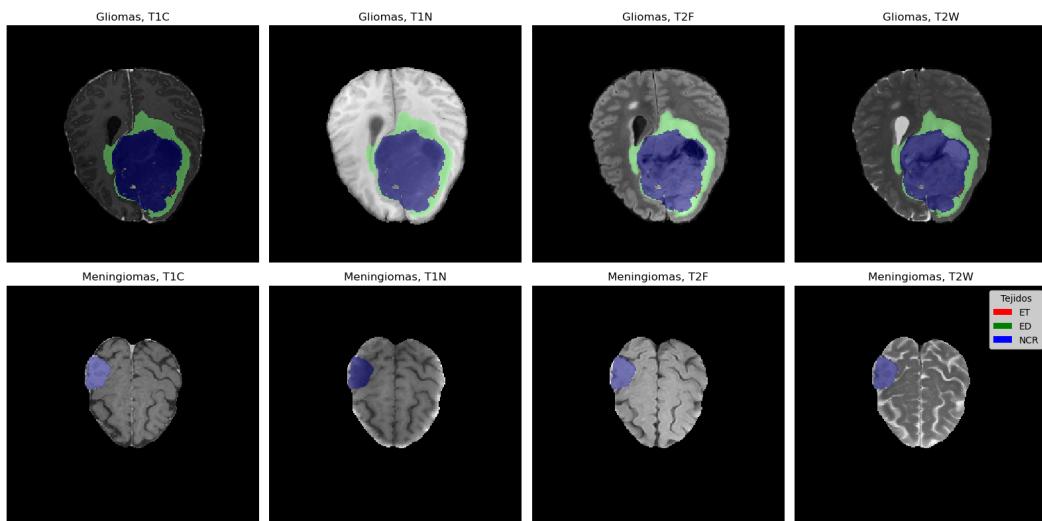


Figura 1.5: Visualización de imágenes MRI de tumores en adultos de origen europeo y americano con su segmentación.

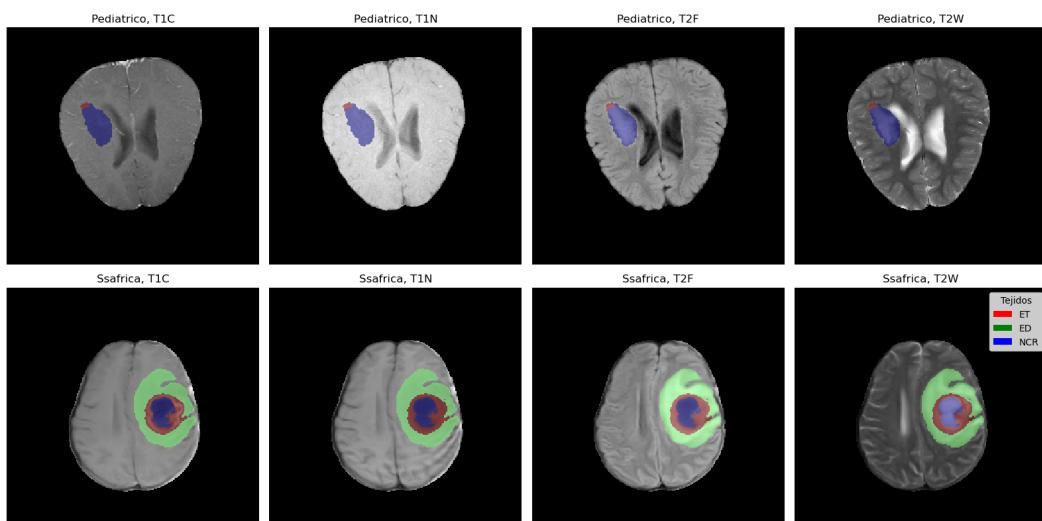


Figura 1.6: Visualización de imágenes MRI de tumores en niños y adultos de origen africano con su segmentación.

A continuación, exploraremos la **localización de los tumores** en todo el conjunto de datos. Buscando responder a la siguiente pregunta significativa para el tránscurso del trabajo: ¿Existe alguna zona del cerebro especialmente afectada? Para responderlo podemos intentar visualizarlo. Crearemos un mapa de calor para los 150 primeros slices marcando la presencia de la lesión, ponderaremos de forma lineal a los tejidos según su importancia.

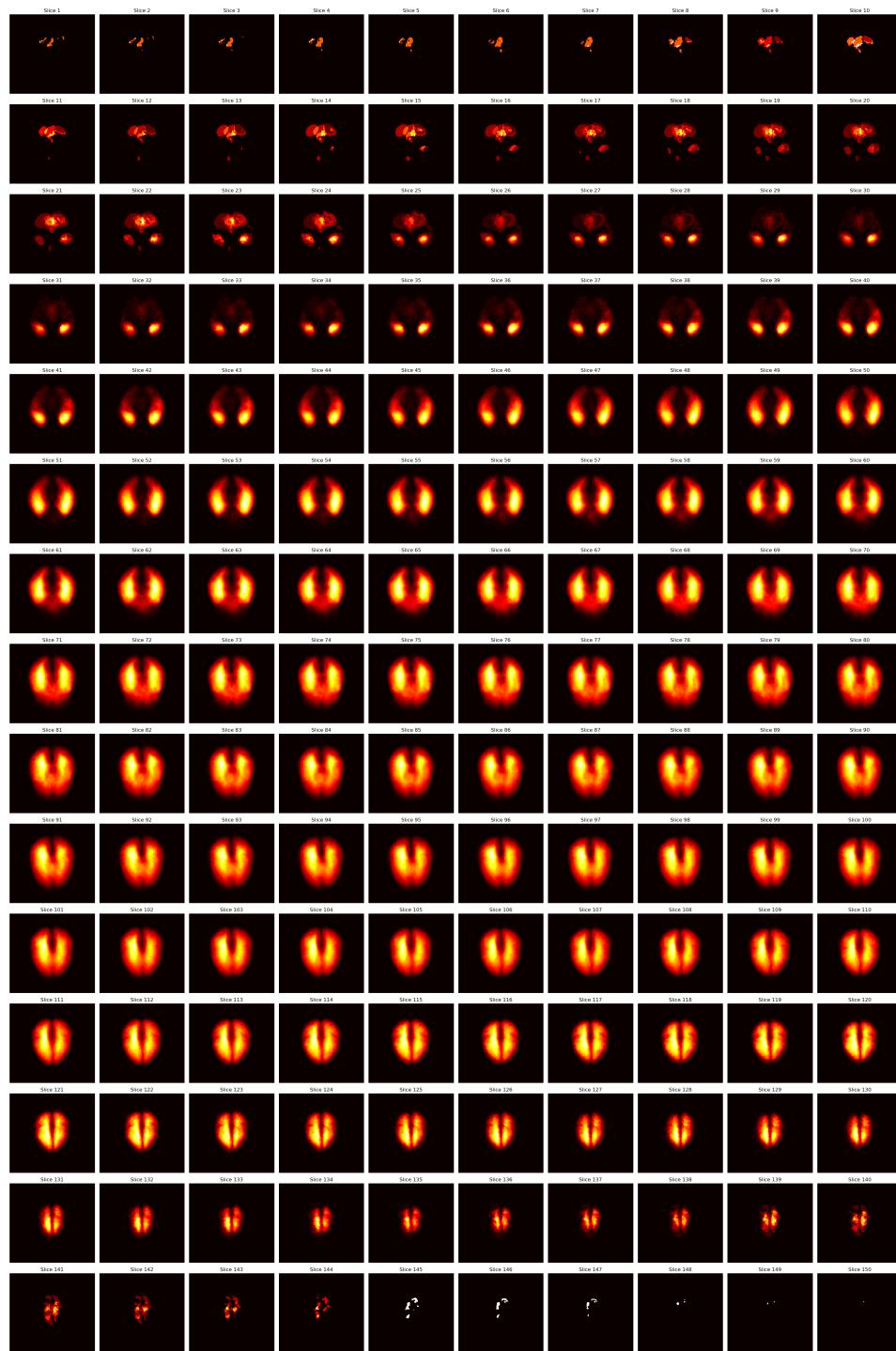


Figura 1.7: Distribución del tejido tumoral en Gliomas.

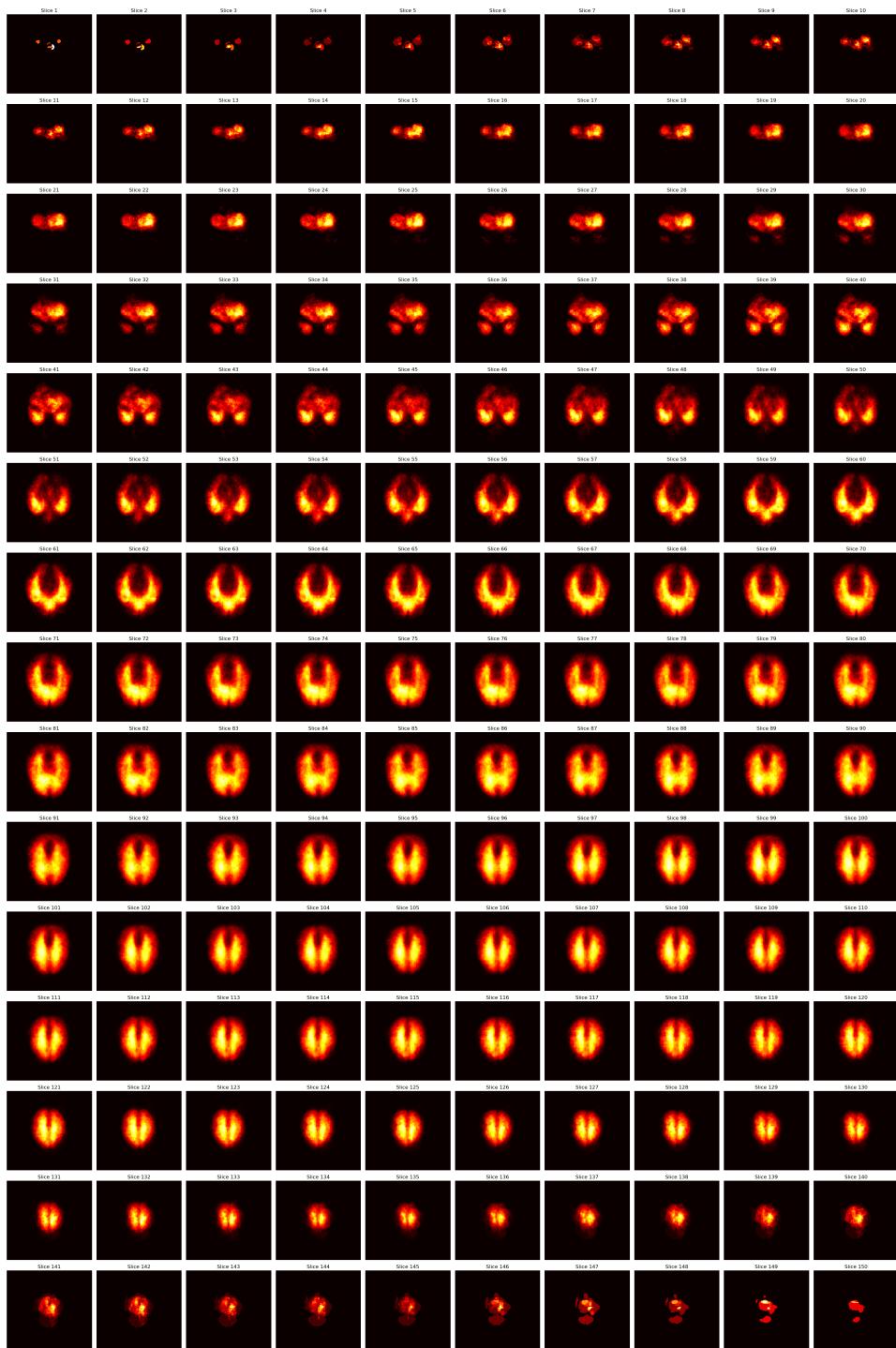


Figura 1.8: Distribución del tejido tumoral en Meningiomas.

Tras la visualización, no se observa una zona especialmente marcada ni en glioblastomas ni en meningiomas. Aunque sí ciertos detalles, como era de esperar en los primeros slices (que son de la base del cerebro) la zona de los meningiomas solo presenta tejido tumoral en la parte posterior esto es debido que a los meningiomas son tumores de desarrollo entre el cráneo y el cerebro, no habiendo ahí tejido posible de esta forma. Esto también puede explicar porque los gliomas tienen una distribución más centrada en los núcleos de los dos hemisferios del cerebro y los meningiomas más distribuido pero con una ligera tendencia en la parte frontal del cerebro donde más líquido cefalorraquídeo se concentra. A parte, de estos detalles íntegros a la naturaleza de estos tumores vemos como siguen una **localización independientemente distribuida**.

### 1.2.2. Línea de investigación

En este trabajo definiremos como línea base la construcción de un arquitectura encoder-decoder totalmente convolucional que se ajuste a los datos de entrenamiento. Para posteriormente utilizar el codificador y representación latente del autoencoder para todas las tareas. Conectándole:

1. **Para clasificación** añadiendo una red densamente conectada.
2. **Para segmentación** un decodificador totalmente convolucional transformando la arquitectura en un autoencoder V-net (U-net con conexiones residuales).
3. **Para predicción** un decodificador basado en la arquitectura Transformer.

La tarea de predicción sólo se planteará a nivel teórico por una baja existencia de datos con relación temporal.

### 1.2.3. Evaluación y métricas

Para la evaluación se distinguirán entre tres subconjuntos de datos: entrenamiento, validación y test. Se utilizará un conjuntos fijos basado en **hold out** común a todas las tareas por motivos de eficiencia no podemos permitirnos usar validación cruzada o distintos conjuntos variables. Estos conjuntos serán separados mediante la utilización de archivos CSV con las rutas absolutas de cada ejemplo para cada uno de ellos.

Seguirán la siguiente distribución: un 49 % para entrenamiento, 21 % para validación y un 30 % para test del conjunto total de datos (glioblastomas y meningiomas juntos) haciendo un reparto aleatorio entre ellos. Se sigue la

regla de partición: 70 % entrenamiento + validación y 30 % test. Además de 70 % entrenamiento y 30 % validación. La finalidad de cada subconjunto es para entrenamiento ajustar los modelos a este subconjunto de datos, validación elegir el mejor modelo de los probados, y test dar el resultado final garantizador de la bondad de los modelos.

### Métricas para clasificación

Para clasificación las métricas usadas serán **accuracy balanceado** y **accuracy** que podremos calcular a través de la construcción de la matriz de confusión.

1. **Accuracy.** Mide cuántas predicciones del modelo son correctas respecto el total del predicciones. Se calcula como:

$$\text{Accuracy} = \frac{\text{Número de predicciones correctas}}{\text{Total de predicciones}} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Accuracy balanceado.** En problemas de clasificación para manejar conjuntos de datos desbalanceados, donde las clases no están igualmente representadas. Se define como el promedio de las tasas de acierto (recall) obtenidas en cada clase, lo que ayuda a proporcionar una evaluación más justa y equilibrada del desempeño del modelo cuando las clases tienen diferentes tamaños.

De forma general para un problema de clasificación de  $N$  clases se calcula como:

$$\text{Balanced Accuracy} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i}$$

De forma específica, para clasificación binaria  $N = 2$  podemos expresarlo como:

$$\text{Balanced Accuracy} = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

### Métricas para segmentación

Siguiendo la definición de evaluación de BraTS 2023, la evaluación de los tumores en su forma original comprende la segmentación de tres zonas diferenciadas según los tipos de tejidos que hay en ella. En BraTS se estudian los siguientes tres conjuntos de tejidos tumorales.

1. **Enhancing Tumor ET**: Sólo incluye al tejido ET.
2. **Tumor Core TC**: Incluye al tejido ET y al tejido NCR.
3. **Whole Tumor WT**: Incluye a todos los tejidos enfermos: ET, NCR y ED. La segmentación de toda la lesión.

Sin embargo, en este trabajo hacemos una reducción del problema por falta de recursos a **la segmentación únicamente del conjunto Whole Tumor**. De esta forma, la misión que tendremos es la de segmentar tejido enfermo o lesionado en todo el volumen de la resonancia. En otras palabras, diferenciar con la segmentación el tejido enfermo del tejido sano o la no existencia de tejido.

Para ello, se utilizará tres métricas. Las dos primeras utilizadas en todas las competiciones de BraTS y la tercera adicionalmente utilizada en los trabajos que han ido conformando el estado del arte como [Zhou et al., 2021].

1. **Similaridad Dice** : Mide la similaridad de dos conjuntos a través de la intersección de ambos respecto el tamaño total de los dos conjuntos.

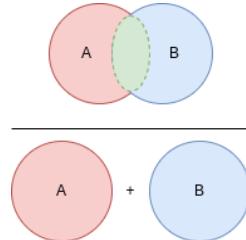


Figura 1.9: Interpretación del coeficiente de Similaridad Dice

Podemos expresarlo como:

$$Dice(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}$$

Donde  $A$  es la segmentación que proporcionará nuestro modelo y  $B$  la verdadera.

De forma más detallada, para una resonancia completa podemos expresarlo como:

$$Dice(A, B) = 2 \cdot \frac{\sum_{i=1}^N \sum_{j=1}^C A_{ij} B_{ij} + \epsilon}{\sum_{i=1}^N \sum_{j=1}^C A_{ij} + B_{ij} + \epsilon}$$

Donde  $N$  es el conjunto de todas las slices de la resonancia,  $C$  el conjunto de clases en nuestro caso  $C = 2$ ,  $A_{ij}$  es el valor de la predicción

en el pixel  $i$  para clase  $j$ .  $B_{ij}$  es el valor real en el pixel  $i$  para la clase  $j$ .  $\epsilon$  es una pequeña constante para evitar dividir entre 0.

2. **Distancia Hausdorff**: Esta métrica tiene el objetivo de medir geométricamente la mayor distancia resultada entre  $A$  la predicción del modelo y la verdadera segmentación  $B$ , siendo  $A$  y  $B$  conjuntos de puntos o píxeles.

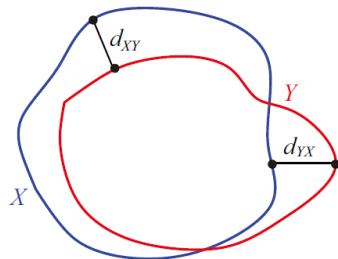


Figura 1.10: Distancias de Hausdorff entre dos conjuntos

Podemos enunciar la distancia máxima de Hausdorff:

$$H(A, B) = \max \left\{ \max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b) \right\}$$

Donde  $a, b$  son puntos concretos en los conjuntos,  $d(a, b)$  la distancia euclidiana entre los dos puntos y  $A, B$  los conjuntos de puntos.

Obtendremos la distancia de Hausdorff media de todos los slices de cada resonancia:

$$\bar{H}(A, B) = \frac{1}{|N|} \sum_{s \in N} H(A, B)$$

De esta forma, a menor distancia de Hausdorff la segmentación salida y real son geométricamente más parecidas.

3. **Sensibilidad o Recall**: Mide la proporción de casos positivos que fueron correctamente identificados por el modelo. En otras palabras y para nuestro problema, mide cuánto la segmentación predicción coincide con la segmentación verdadera, cuantificado en porcentaje.

Podemos expresarlo como:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Donde  $TP$  son los píxeles que son verdaderos positivos y  $FN$  los píxeles que son falsos negativos.



## Capítulo 2

### Estado del arte

En este capítulo estudiaremos y analizaremos los diferentes enfoques dados históricamente para nuestras tareas, **clasificación de tumores cerebrales** y **la segmentación de tumores cerebrales**. Abordando desde el inicio del estudio del problema pasando por la explosión de métodos basados en Aprendizaje profundo con la constitución de **BraTS** hasta nuestros días. Se pondrá especial énfasis a las soluciones actuales comparándolas desde sus diferencias en metodología y perspectiva.

Por un lado, la clasificación entre los dos tipos de tumores no es tan relevante a la hora de diseñar un sistema de ayuda a la toma de decisión ya que clínicamente sí existe una característica diferencial entre ambos, su **localización**. Los meningiomas aparecen entre el cráneo y el cerebro no internamente en el cerebro como los glioblastomas. Esto hace que un médico pueda distinguirlos sin requerir una gran asistencia para la mayoría de los casos. No obstante, la clasificación de por sí ayuda a la toma de decisiones en el tratamiento pero no es el elemento crítico para la supervivencia del paciente que depende de la eliminación del tumor donde su segmentación toma un papel crucial. Por esto, veremos como el estado del arte del problema de segmentación es mucho mayor que el del problema de clasificación.

Por otro lado, la tarea de predicción de la evolución del tumor debido a la baja densidad de instancias temporales de los datos existentes hacen que este problema no sea tratado en trabajos. No hemos encontrado literatura al respecto.

Los grandes esfuerzos se han realizado en entorno a la segmentación, ya que otras tareas que conforman su diagnóstico se verían arrastradas.

## 2.1. Revisión histórica de clasificación binaria de tumores

## 2.2. Revisión histórica de segmentación

Diferentes dificultades han sido las que a pesar de años de desarrollo aún encontrar un algoritmo para la segmentación de tumores cerebrales sea algo mejorable.

1. **Incertidumbre en la localización :** Como vimos no existe una zona concreta en general para la aparición de los tumores cerebrales. A excepción, de los meningiomas localizados en zonas superficiales del cerebro y aún siendo una región muy amplia, incluso ya desarrollado un tumor pueden aparecer otros localizados en regiones muy distintas de la original.
2. **Incertidumbre en la morfología :** A diferencia de otras patologías, cada tumor cerebral presenta un tamaño y forma completamente distintas y donde en principio no se puede apreciar un patrón distintivo. Esto hace que sea muy complicado y generalmente aporte malos resultados, la construcción de sistemas basados reglas u otras aproximaciones que no incluyen una componente de aprendizaje.
3. **Bajo contraste :** Una buena resolución y contraste son características muy importantes para entender la información de una imagen. Las imágenes IRM producidas en una resonancia debido a proyecciones de imagen y procesos de tomografía usualmente ofrecen una baja resolución y contraste haciendo más difícil la definición de bordes entre diferentes tejidos de la imagen. Una segmentación precisa es difícil de conseguir.
4. **Sesgo en las etiquetas.** Existen indicios para pensar que las etiquetas proporcionadas pueden presentar ruido. El proceso de segmentado por parte del personal médico depende de su experiencia profesional lo cual puede llevar a cometer errores. Por ejemplo, se han presentado eventualmente discrepancias entre distintos anotadores: algunos tienden a conectar todas las pequeñas regiones de un tejido mientras que otros las segmentan de forma más precisa y separada.
5. **Desbalanceo en el tejido :** Dentro de la segmentación entre los diferentes tipos de tejidos, usualmente existe un el tejido enfermo y que compone la lesión tumoral es usualmente más pequeño que el tejido sano. Esto podría afectar en el proceso de aprendizaje creando rechazo a identificar al tejido enfermo.

**6. Desbalanceo entre pacientes :** En el conjunto de datos tenemos muchos pacientes de norteamérica y de ascendencia blanca, pero pocos de otros orígenes como el africano. Además, de tener un sesgo claro de edad ya que existen pocos casos en niños. Esta falta de datos puede impedir que exista una buena generalización para estos casos más aislados.

A continuación, se presenta una revisión histórica sobre la segmentación de tumores cerebrales hasta 2021 apoyada en [Liu et al., 2023]. Se presenta una línea del tiempo con los principales trabajos de estudio.

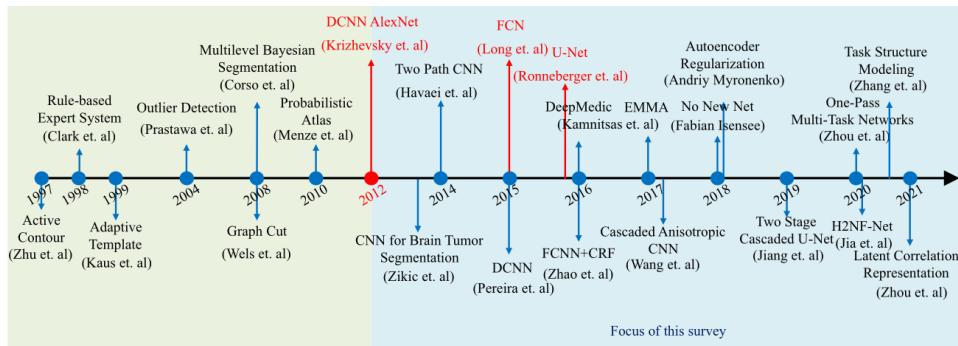


Figura 2.1: Evolución histórica del estado del arte hasta 2021.

En la década de 1990 investigadores como [Zhu and Yan, 1997] fueron pioneros al utilizar una red Hopfield con un modelo de contornos activos para extraer los bordes del tumor. Sin embargo, incluso el entrenamiento de una pequeña red como esta era algo computacionalmente costoso por las limitaciones de la época. Desde 1990 hasta 2012, los métodos que iban surgiendo para la segmentación de tumores cerebrales estaban basados en métodos clásicos de aprendizaje con características extraídas a mano, sistemas expertos que se apoyaban en los histogramas de la imagen, plantillas para la segmentación y modelos gráficos.

A pesar de ser un gran paso inicial, tenían grandes deficiencias. Por ejemplo, la mayoría de ellos sólo se centraba en la segmentación de todo el tumor lo cual lleva a un modelo poco útil. Por otro lado, en los modelos basados en características extraídas se hacía muy tedioso poder usarlos eficazmente ya que este paso de extracción dependía de conocimiento previo experto que en ningún momento se pudo llegar a representar en un modelo. En último lugar, los mismos problemas que compartimos hoy en día sobre el desbalanceo y la incertidumbre del problema eran mucho más agresivos.

Tras 2012 con la revolución del Deep Learning, se introducen nuevas tecnologías (Redes neuronales convolucionales y U-net) que mejorarán los

resultados obtenidos hasta el momento. Se empezarán a construir arquitecturas encoder-decoder convolucionales para conseguir pipelines completos para la segmentación. El aprendizaje profundo toma el problema de lleno proclamándose el enfoque que define el estado del arte.

Podemos clasificar las soluciones basadas en aprendizaje profundo apor- tadas en tres categorías según el principal problema para que el están pesa- das. Sin embargo, como veremos en las soluciones más actuales lo ideal es tratar con los tres problemas.

### 2.2.1. Métodos que se enfocan en la arquitectura

Para poder obtener redes que automáticamente extraen características discriminativas a altas dimensiones es necesario un efectivo diseño de módu- los y arquitecturas. Por un lado, se pretende que la arquitectura sea capaz de aprender las características distintivas de los tejidos y a localizar regiones de interés por medio de añadir profundidad a la red, a través de mecanis- mos de atención o la fusión de características entre las resonancias. Por otro lado, se pretende minimizar la cantidad de parámetros entrenables de la red o conseguir un entrenamiento más rápido.

#### Diseño de bloques especializados

Los primeros trabajos que tenían este objetivo comenzaron por basarse en arquitecturas bien conocidas como AlexNet o VGGNet a través del uso de una única imagen de la resonancia completa como entrada de la red.

Para la mejora de resultados, se optó por introducir todas la secuencia de imágenes de una resonancia como entrada de la red y añadir más capas convolucionales. Con ello, teníamos redes más profundas pero que pronto empezaban a sufrir los problemas de la explosión y desvanecimiento del gradiente durante el proceso de entrenamiento. Para ayudar a lidiar con estos problemas, se introdujo a las redes, **conexiones residuales** [Chang, 2016]. Conectando la entrada de la red con su salida, convergiendo más rápido y con mejores resultados.

Este proceso de aumento de profundidad con conexiones residuales no sería definitivo porque también conlleva el sacrificio de resolución espacial. Se reemplazaría en trabajos siguientes, el uso de la convolución simple por convoluciones dilatadas. El **uso de convoluciones dilatadas** traería el au- mento del espacio receptivo (ya que se aplica una convolución a un espacio mayor de la imagen) sin necesidad de introducir parámetros a la red. La con- volución dilatada se vería especialmente útil por ejemplo en la segmentación de áreas grandes como suele ocupar el tejido ED (edema tumoral).

Respecto conseguir una buena eficiencia en tiempo de entrenamiento es conocido aplicar un reordenamiento en memoria de las imágenes de la resonancia similares (p. ej. el mismo slice en las 4 pruebas) de forma que se reduzcan la comunicación entrada-salida con GPU. Adicionalmente, autores como [Brügger et al., 2019] utilizan **conexiones reversibles** en la red de forma que durante el proceso de backpropagation (backward pass) no se necesite memoria adicional para guardar las activaciones intermedias. Por último, para ahorrar en eficiencia se sustituye la convolución standard por la combinación de **convoluciones separables**.

### Diseño de arquitecturas efectivas

La mayoría de los trabajos de recorrido histórico se encasillan en alguno de los siguientes dos enfoques de arquitectura: **redes neuronales convolucionales** para extraer características de la imagen y clasificar los patches o píxeles de la imagen según las etiquetas de los tejidos posibles o **redes encoder-decoder** en las cuales se puede definir un pipeline completo convolucional sin la necesidad de la agregación de capas totalmente conectadas.

#### 1. Redes neuronales convolucionales de una/múltiples trayectorias

A diferencia de una red convolucional de una única trayectoria, las redes de trayectoria múltiples tienen la capacidad de extraer diversas características a diferentes escalas. Estas características se combinan para su posterior procesamiento, usualmente en capas totalmente conectadas, permitiendo a las redes aprender tanto características globales como locales.

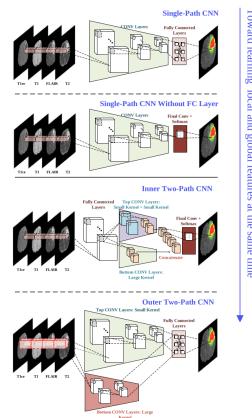


Figura 2.2: Comparación entre arquitecturas de una y múltiples trayectorias. Imagen de [Liu et al., 2023]

Por ejemplo, [Havaei et al., 2017] desarrollaron una estructura de dos vías que integra información tanto local como global del tumor, utilizando núcleos de convolución de diferentes tamaños.

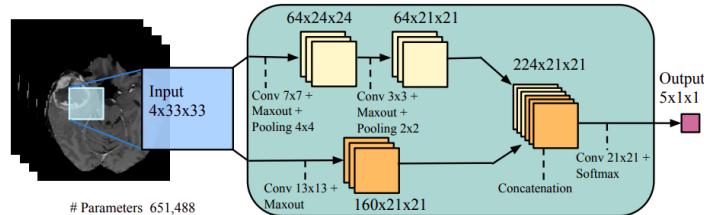


Figura 2.3: Arquitectura de dos vías de [Havaei et al., 2017]

Otros enfoques, como el de [Kamnitsas et al., 2017], optan por aprender información global y local desde la entrada misma, utilizando redes de doble vía, patches de diferentes tamaños y pequeños núcleos de convolución.

Este tipo de arquitecturas fueron una de las primeras aproximaciones que empezaban adaptarse con éxito a las complejidades de la segmentación de tumores cerebrales. Sin embargo, veremos como la dificultad de un buen ajuste en el diseño de estas arquitecturas todavía seguía siendo un problema.

## 2. Arquitecturas Encoder-Decoder

Las redes de una/múltiples trayectorias toman como input un patch de una cierta región de la imagen y dan como output la clasificación del tejido que existe en ese patch. Este enfoque hace que obtener una buena arquitectura que haga la transformación de los patches a información categórica sea complicado por varios motivos:

- a) Existe una gran **dependencia** entre el tamaño y calidad de los patches, y los resultados que ofrecería la arquitectura.
- b) Toda la transformación de características visuales (aunque, reducidas) a información categórica estaría concentrada en las capas totalmente conectadas. Las capas totalmente conectadas de un tamaño razonables para una capacidad de memoria usualmente utilizada **no puede totalmente representar un espacio de características tan grande**.
- c) Si necesitamos tener distintas redes separadas, el proceso de ajuste de cada una de ellas es independiente. Esto lo podemos interpretar como un coste añadido en términos de **eficiencia**.

Para superar estos problemas en los siguientes trabajos se empieza a utilizar **FCN Redes neuronales totalmente convolucionales** y

**U-net** basadas en arquitecturas encoder-decoder, de forma que se establece un pipeline completo desde la imagen a la segmentación.

Una de los tipos más importantes de FCN para este problema es U-net. U-net consiste en la creación de conexiones entre el encoder y el decoder. Permitiendo una vinculación directa en el proceso de reducción y ampliación de dimensionalidad. Estas conexiones reciben el nombre de **Skip Connections** y pueden ayudar a las capas del decoder a recuperar detalles visuales aprendidos en el encoder, llevando a una segmentación más precisa.

[Isensee et al., 2018] utilizan una U-Net dándole aún más énfasis a la tarea de una segmentación utilizando una función de pérdida basada en la similaridad Dice.

Similar a las skip connections antes mencionado, el uso de conexiones residuales y skip connections permiten el paso de características de alto y bajo nivel para una mejor segmentación final.

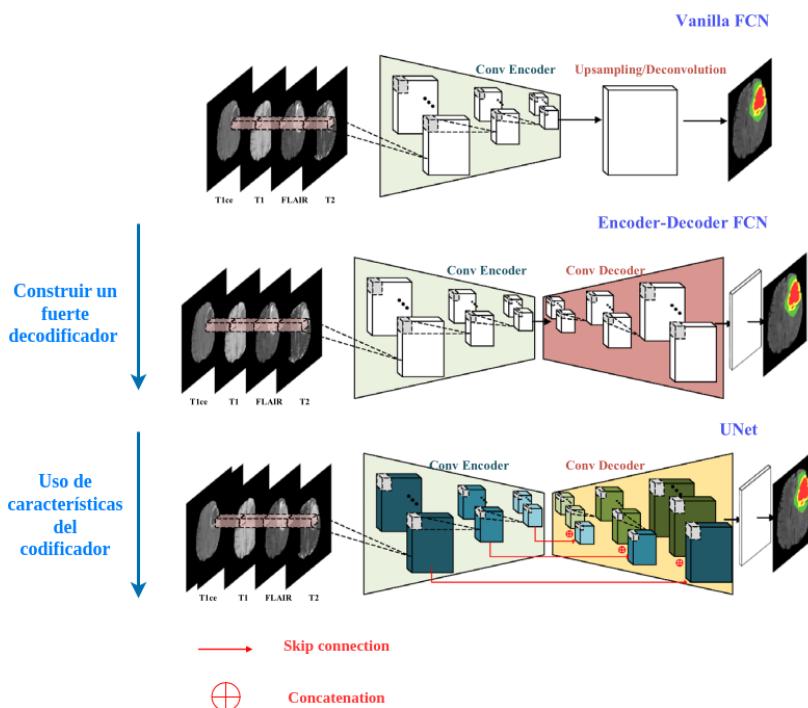


Figura 2.4: Comparación de distintas arquitecturas encoder-decoder

### 2.2.2. Métodos que tratan el desbalanceo

Como anunciábamos anteriormente el alto desbalanceo de los diferentes tejidos presentes en el cerebro de un paciente puede tener un impacto negativo en el proceso de entrenamiento. Motivados por métodos como los sistemas multi-expertos, se empezó a construir métodos específicos para este problema.

Podemos diferenciar en:

1. **Diseños sobre la arquitectura:** Redes en cascada, ensamblado de modelos y arquitecturas multi-tarea.
2. **Mejorar el entrenamiento:** Funciones de pérdida especializadas.

#### Redes en cascada

Una red en cascada es un conjunto de redes más pequeñas ordenadas en las cuales el output de la red anterior sirve como una input a la siguiente, formando una «cascada de redes». De esta forma, podemos tener redes especializadas en distintos niveles.

Las primeras redes de la cascada especializadas a características de más alto nivel y las siguientes de más bajo.

Por ejemplo, en [Wang et al., 2018] se utilizan tres redes especializadas para los tres regiones de tejidos definidas por BraTS. Empezando por la región más grande hasta la más pequeña.

Su primera red WNet segmenta a Whole Tumor, toda la lesión. La siguiente TNet segmenta al núcleo del tumor. Finalmente, Enet a la parte activa del tumor.

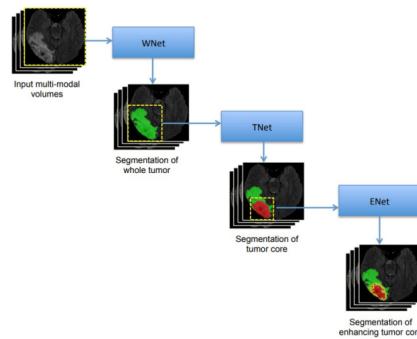


Figura 2.5: Estructura de método en cascada de [Wang et al., 2018]

La ventaja de este modelo es evitar la interferencia de las clases desbalanceadas, ya que cada red trata su clase como un problema de segmentación

binaria.

Sin embargo, hace que la redes dependientes de otras dependan también de sus resultados. Si la primera red obtiene malos resultados, todas las siguientes redes se verán afectadas por ella.

### Ensamblado de modelos

Una de las consecuencias que tiene el uso de una sola red es que está altamente influenciada por la elección de su hiperparámetros. Con el objetivo de obtener un más robusto y general modelo para la segmentación se puede combinar la salida de múltiples redes, ensamblarlas.

El ensamblado de modelos aumentaría el espacio de hipótesis del modelo final evitando, la caída en óptimos locales debido a el desbalanceo de datos.

EMMA de [Kamnitsas et al., 2018] es uno de los primeros modelos para segmentación de tumores que es un ensamblado de varias redes. EMMA utiliza tres modelos: DeepMedic, una red FCN y una U-net para dar el output de los tres con una mayor confianza.

[Jiang et al., 2020] ganadores de BraTS2019 adoptaron una estrategia de ensamblado con 12 modelos obteniendo entorno 0.6 – 1 % mejores resultados que el mejor único modelo.

### Arquitecturas multi-tarea

Todo lo descrito en esta revisión histórica gira entorno a la segmentación de tumores. Sin embargo, la desventaja que puede tener enfocarnos en esta sola tarea es que quizás los modelos específicos para segmentación ignoran información útil en las imágenes para otras tareas, que indirectamente pueda ayudar a obtener una mejor generalización en la segmentación de tumores.

Por un lado, esta idea radica en la suposición de que los modelos que aprenden más tareas están aumentando su aprendizaje en el dominio del problema y esto debería ser beneficioso para todas las tareas. Por otro lado, de una forma más justificada, sabemos que nos enfrentamos a cierto ruido que desconocemos en los datos y etiquetas por tanto si entrenamos para múltiples tareas en conjunto el modelo aprende representaciones más generales reduciendo el riesgo de sobreajuste. Añadir tareas a la arquitectura y aprenderlas en conjunto podría tener **un efecto regularizador**.

Un claro ejemplo de esto es [Myronenko, 2019] que usa como tarea complementaria la reconstrucción de la resonancia de entrada mediante un auto-encoder. Teniendo un efecto regularizador sobre los parámetros compartidos del encoder que a diferencia de regularizaciones L1 o L2 que explícitamente añaden una penalización para evitar el sobreajuste, la tarea nueva añade

una penalización en la dirección en la que ambas tareas son optimizadas reduciendo el espacio de búsqueda de los parámetros entrenables de la red.

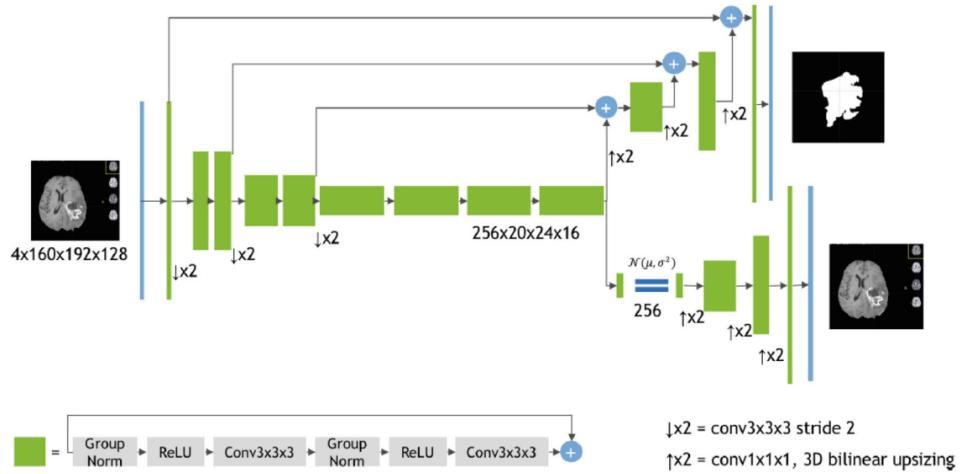


Figura 2.6: Arquitectura del autoencoder regularizador de [Myronenko, 2019]

### Funciones de pérdida especializadas

De forma más detallada, el problema del desbalanceo entre los diferentes tejidos se manifiesta durante el proceso de entrenamiento, en un gradiente excesivamente influenciado por los tejidos mayoritarios. Por ello, atacando directamente al problema multitud de trabajos proponen funciones de pérdida especializadas.

Funciones de pérdida estándar en este problema incluyen categorical cross-entropy, cross-entropy y dice loss  $D_L$ .

Una de las aproximaciones es el uso de utilizar una función de pérdida balanceada. Por ejemplo, añadir una penalización en función de la presencia del tejido segmentado para mitigar su escasa presencia respecto el total.

Otro enfoque se basa en la combinación de diferentes funciones de pérdida en una nueva. Por ejemplo, una nueva función de pérdida de cross-entropy a nivel de píxel y dice loss podría ser su media.

En general, funciones de pérdida que eviten el desbalanceo y mejoren el nivel de atención de las arquitecturas es beneficioso a todo tipo de problemas. Por ello, a diferencia de seguir funciones clásicas como cross-entropy, [Lin et al., 2017] proponen una nueva función llamada **Focal Loss** que será vista en años recientes en combinación con Dice Loss para diversos problemas de segmentación.

### 2.2.3. Métodos que tratan la información multi-modal

Las imágenes asociadas a una resonancia contienen diferentes tipos de imagen según las características de la frecuencia y contraste suministrado al paciente en su toma. Esta forma de proceder en la toma de resonancias es debido a las limitaciones de las imágenes IRM de poder representar y al menos para el ojo humano visualizar todos los tejidos importantes en el diagnóstico. Por ello, surge como idea clave tener métodos que tengan los objetivos de poder fusionar, relacionar y incluso distinguir en importancia las diferentes modalidades de imagen.

Otras arquitecturas basadas en autoencoders como [Myronenko, 2019] únicamente fusionan las cuatro modalidades como los canales de una imagen para un mismo slice concatenando las cuatro pruebas en la misma entrada, obteniendo entradas de dimensiones  $H \times W \times 4$  en caso de 2D y  $H \times W \times D \times 4$  en caso de 3D.

Sin embargo, usar concatenación o adición como método de fusión de los cuatro métodos no permitiría a la red de una forma directa aprender semánticamente la relación entre ellas. Por ello, en trabajos recientes se han adoptado mecanismos de atención aplicados a hacer aprender a la red de forma más robusta las diferentes modalidades e información espacial.

[Zhou et al., 2021] proponen también una arquitectura encoder-decoder con la particularidad de crear un encoder y decoder específico para cada una de las cuatro posibles representaciones, teniendo un espacio latente donde se fusiona la información de salida de los cuatro encoder dando un tratamiento especial a la fusión de las diferentes pruebas.

A continuación, podemos ver la arquitectura específica usada.

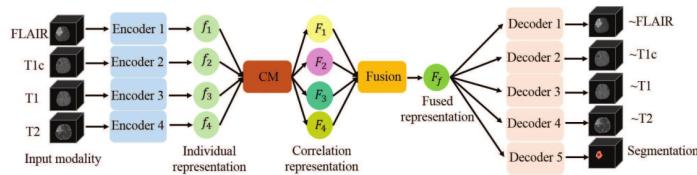


Figura 2.7: Arquitectura de [Zhou et al., 2021]

Por un lado, transforma las representaciones individuales a representaciones correlacionadas. A través de lo que denominan **correlation model**.

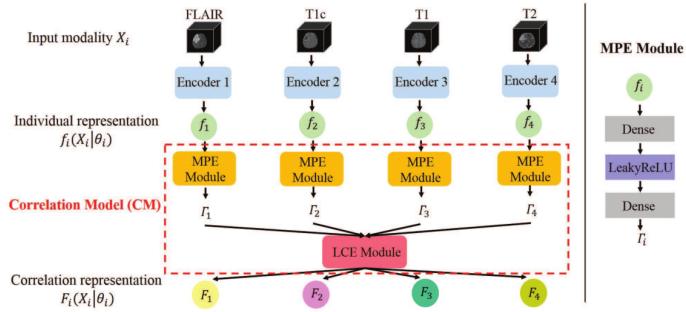


Figura 2.8: Modelo especializado en la correlación de las modalidades

El **correlation model** se compone de dos partes: módulo de estimación de parámetros (MPE) y un módulo de expresión de correlación lineal (LCE).

El módulo de estimación de parámetros se compone de una red totalmente conectada que vincula cada representación salida de cada encoder con unos parámetros  $\Gamma_i = \{\alpha_i, \beta_i, \gamma_i, \delta_i\}$

El módulo de expresión de la correlación lineal (LCE) utiliza estos parámetros para obtener una versión correlacionada de cada representación individual aplicando:

$$F_i(X_i|\theta_i) = \alpha_i \odot \gamma_i f_j(X_j|\theta_j) + \beta_i \odot f_k(X_k|\theta_k) + \gamma_i \odot f_m(X_m|\theta_m) + \delta_i, \quad (i \neq j \neq k \neq m)$$

Tras ello, se fusiona las representaciones correlacionadas resultado. Permitiendo al modelo manejar de forma explícita la información multi-modal y dándole robustez ante pruebas faltantes.

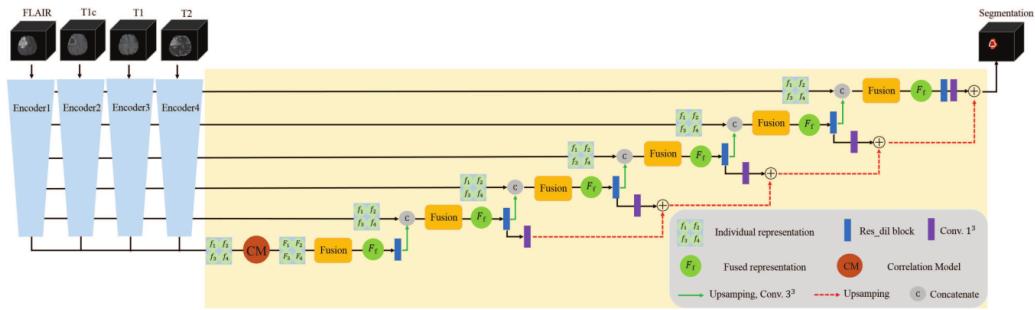


Figura 2.9: Red [Zhou et al., 2021] de fusión de representaciones latentes

Si bien esta arquitectura da ligeramente peores resultados que [Myronenko, 2019] define un paso más en el estado del arte al usar menos recursos computacionales.

## 2.3. Enfoques actuales para la segmentación

Las soluciones más relevantes presentadas en la revisión histórica que se ha hecho anteriormente se basan en la aplicación de la convolución sobre las imágenes de resonancia magnética. En el diagnóstico de tumores cerebrales ha tenido largo recorrido el uso de redes neuronales convolucionales.

Con la inclusión de las arquitecturas transformadoras se planteó un nuevo modelo que podía traer ventajas significativas. No siendo la imagen médica y en concreto este problema una excepción.

Con la adaptación de los transformers al campo de la visión, los Vision Transformers podría ser un modelo más unificador, paralelizable y que ofreciera mejores resultados que las redes convolucionales al romper con la localidad que supone el uso de convoluciones.

En las soluciones más recientes de la segmentación de tumores cerebrales se introduce el uso de Vision Transformers con estas expectativas.

### 2.3.1. Basados en Transformers

A continuación, se presentan las soluciones principales que hacen uso de una arquitectura basada en Transformers para la segmentación de tumores cerebrales.



# Capítulo 3

## Metodología

En este capítulo describimos en profundidad todos los pasos seguidos en los métodos empleados en el trabajo y su justificación. Posteriormente, se aplicarán en la experimentación.

### 3.1. Análisis de los recursos disponibles

Para la realización de este trabajo debemos considerar los recursos hardware disponibles para la inferencia de los modelos pero sobre todo para el entrenamiento de los modelos.

1. **Hardware en entrenamiento.** Para el desarrollo de toda la experimentación, entrenamiento de los modelos y validación nos valdremos de los recursos que gratuitamente ofrece Kaggle, una plataforma de ciencia de datos propiedad de Google.

El recurso más importante que ofrece Kaggle y razón de su uso es que nos permite el uso de su gráfica NVIDIA Tesla P100 por 30 horas semanales. Con ella, podemos entrenar los modelos y hacer una inferencia rápida para validación en tiempo razonable.

Por nosotros mismos sólo disponíamos de un ordenador personal que aunque con mejor disponibilidad de memoria en disco  $\approx 2\ TB$  que la ofrecida por Kaggle  $\approx 100\ GB$ , nuestra gráfica NVIDIA GeForce RTX 2060 tiene inmensamente menores prestaciones que la ofrecida en Kaggle. A continuación, mostramos una gráfica de rendimiento sobre las características de ambos dispositivos para cuantificar este hecho.

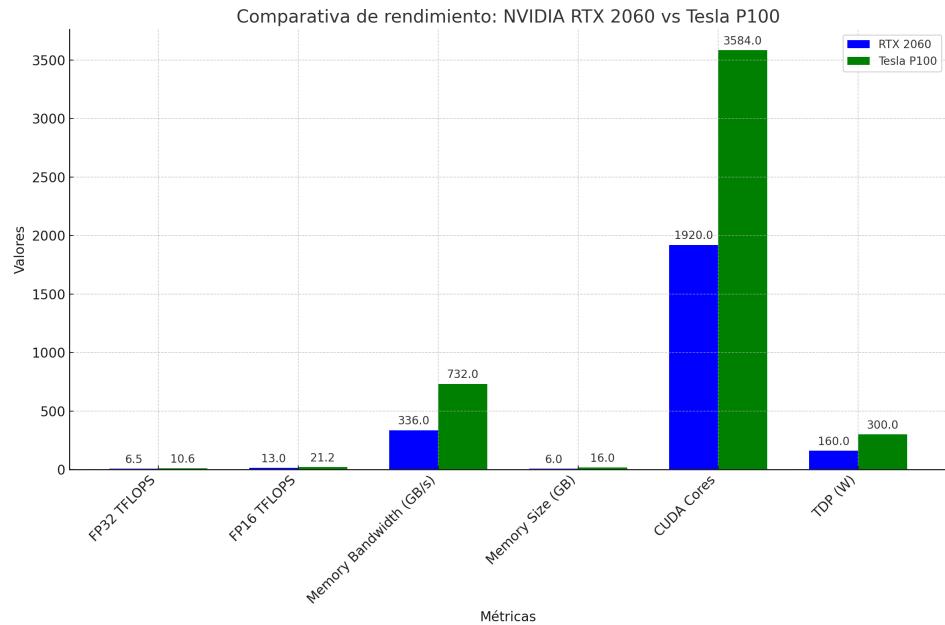


Figura 3.1: Comparativa de rendimiento de las GPU disponibles

En todas las características la gráfica ofrecida por Kaggle supera a la nuestra. Por lo que optaremos por usarla en todo el trabajo.

2. **Hardware para inferir con los modelos.** Para la construcción de la interfaz y su uso si hemos usado nuestro dispositivo personal.

De forma general, el hardware necesario para inferir con los modelos es cualquier PC de uso personal que disponga de una GPU de rendimiento similar o mejor que el nuestro y cumpla las siguientes dependencias (link a apéndice de uso del programa).

### 3.2. Preprocesado de Datos

En este apartado se explicará el preprocesamiento que se ha aplicado a las resonancias magnéticas para convertirlas a entradas de los modelos.

Partiendo de nuestro conjunto de datos que presentamos en la introducción obtenido de la competición BraTS en Synapse. Ya vemos como las resonancias presentan características favorables para ser una entrada a la red.

1. **Dimensiones estandarizadas :** Todas las resonancias (adultos, niños, diferente tipo de tumor) presentan las mismas dimensiones.

2. **Imágenes estandarizadas** : Todas las resonancias se han hecho con el mismo estándar de escáner, todas presentan el mismo rango para su visualización.
3. **No existen valores faltantes** : Observamos como el conjunto de datos es completo en su definición, todas las resonancias de cada paciente tienen las mismas cuatro pruebas.

### 3.2.1. Elección de dimensionalidad de las entradas

Una de las elecciones cruciales al inicio del trabajo es la dimensionalidad de las entradas de la red (determina la forma en la que preprocesaremos los datos). ¿Es mejor trabajar en 2D con una única imagen como entrada o en 3D con todo el conjunto de imágenes de una resonancia como entrada?

En un primer momento debido a que la mayoría de literatura actual trabaja en 3D, intentamos trabajar en 3D. Para ello, construimos un primer modelo inicial de arquitectura para 2D y una forma de transformarlo a 3D es simplemente duplicar cada capa por la cantidad de imágenes en cada resonancia. Por lo que, el tamaño de nuestro modelo inicial 2D  $SIZE_{2D}$  sólo debíamos multiplicarlo por la cantidad de imágenes en una resonancia  $SIZE_{3D} = SIZE_{2D} \times 155$ . Debido que el máximo de memoria RAM que disponíamos en la Tesla P100 de Kaggle es 16 GB siguiendo esa regla, el máximo tamaño que podría ocupar un modelo inicial 2D sería:

$$MAXSIZE_{2D} = \frac{16 \text{ GB}}{155} = 0.10323 \text{ GB} = 105.7 \text{ MB}$$

Esta cantidad de memoria para una arquitectura que obtenga resultados competentes en la actualidad de técnicas que existen no es viable. Viéndonos obligados ante la escasez de memoria en GPU a enfocar nuestros esfuerzos en una dimensionalidad 2D. Tendremos como entrada a las arquitecturas **una única imagen la correspondiente a la vista axial de las resonancias**.

### 3.2.2. Normalizado de las imágenes

Las imágenes que componen las resonancias son mapas en escala de gris donde un píxel de la imagen puede tomar un valor de gris en el intervalo [0, 256]. Entre las imágenes de distintas resonancias se encuentra una misma distribución de valores de píxeles para representar la misma información. Sin embargo, el proceso de entrenamiento no deja de ser un proceso de optimización y puede que este rango sea aún demasiado grande.

Adicionalmente, para evitar posibles píxeles erróneos en la toma de las imágenes que podamos interpretar como outliers que tengan un impacto

negativo en el entrenamiento y para hacer las imágenes más interpretables se aplica a las imágenes normalización Z-score o estandarización.

$$X_{std}^i = \frac{x^i - mean}{std}$$

### 3.2.3. Recortado de imagen

Podría ser razonable reducir las dimensiones de las imágenes para hacer a nuestros datos menos pesados. Sin embargo, se opta por no hacerlo por seguridad y escalabilidad. BraTS fija esas dimensiones en base del estándar en una resonancia magnética, así para cualquier paciente se garantiza que la imagen de su cerebro se puede representar en una resonancia en unas condiciones de resolución iguales al resto de pacientes.

Si recortamos las imágenes de forma cuadrada al cerebro más grande de todas las resonancias, podríamos encontrarnos en inferencia con un cerebro mayor que no se podría representar en una imagen. Es necesario dejar cierto margen, optando por respetar el margen inicial que marcan los organizadores médicos de BraTS.

### 3.2.4. Undersampling

En el estado del arte ya mencionamos que existía un desbalanceo entre tejido sano y tejido enfermo. En la mayoría de resonancias existe una mayor proporción de tejido sano que de tejido enfermo. Esto no sólo podía introducir un sesgo en los algoritmos de segmentación sino que aumenta mucho el coste computacional de entrenar a los modelos por el exceso de imágenes que no contienen la información de una lesión tumoral.

El tratamiento del desbalanceo mediante undersampling siempre es una medida agresiva ya que podría eliminar información que a priori no consideramos relevante y si lo es.

Sin embargo, en nuestro problema aplicaremos undersampling con el principal objetivo de reducir los tiempos de entrenamiento y poder tener una arquitectura más profunda manteniendo tiempos razonables. Intentando aliviar de paso el problema del desbalanceo. A continuación explicamos en detalle como se ha llevado a cabo.

Para todo nuestro conjunto de datos  $X$  creamos archivos CSV para cada partición (entrenamiento, validación y test) donde cada fila de cada archivo CSV representa a una imagen o entrada a la red.

Estos archivos en formato CSV contendrán únicamente el conjunto de imágenes que contienen lesión tumoral de todas las resonancias  $N$  más una

parte seleccionada aleatoriamente de imágenes sin lesión de tamaño  $\frac{|N|}{2}$ . De esta forma, nos quedamos con todas las imágenes con información de lesión y con una parte representativa y balanceada sin lesión para no sesgar al modelo a segmentar en todas las imágenes.

En estos archivos CSV existen las siguientes columnas:

1. **Rutas absolutas** : En 4 columnas están las rutas absolutas a los archivos .nii de cada prueba de cada resonancia.
2. **Número de slice** : Se guarda el número de slice en la que se localiza esa imagen dentro de la resonancia. Este campo es necesario para poder extraer la imagen.
3. **Etiqueta** : Para el problema de clasificación es necesario guardar la etiqueta que identifica a cada imagen, 0 para Glioblastoma, 1 para Meningioma y 2 para No Tumor.

Tras aplicar este undersampling nos quedamos con 74487 imágenes en entrenamiento, 31899 en validación y 45354 en test. En la siguiente tabla podemos ver recogida esta información también términos de porcentaje respecto las imágenes totales.

Partición	Pacientes	Imágenes	Porcentaje %
Entrenamiento	1033	74487	46.52
Validación	442	31899	46.56
Test	632	45354	46.3

Cuadro 3.1: Porcentaje de imágenes conservadas tras undersampling.

### 3.3. Elección de modelos

A continuación pasamos a discutir los modelos y técnicas empleadas para la creación de las arquitecturas. En este trabajo como al igual que en parte del estado del arte combinaremos técnicas de aprendizaje no supervisado y supervisado.

#### 3.3.1. Codificador y representación latente

Ponemos ahora el foco en un modelo en principio pensado para el aprendizaje sin etiquetas, los **autoencoders**.

Los autoencoders son arquitecturas encoder-decoder con la finalidad de aprender las características de un conjunto de datos o distribución. Por ejem-

pleo, siendo esto útil para obtener modelos generativos como los autoencoders variacionales.

Los autoencoders se formularon inicialmente como una generalización no lineal del análisis de componentes principales (PCA) por su poder para reducir la dimensionalidad. En este trabajo lo incluiremos como modelo base para aplicar aprendizaje no supervisado y que teóricamente presentarían notables ventajas de cara obtener una mayor convergencia y generalización en el proceso de entrenamiento.

A continuación, mostramos un esquema explicativo de las partes implicadas en la arquitectura para construir el codificador y representación latente.

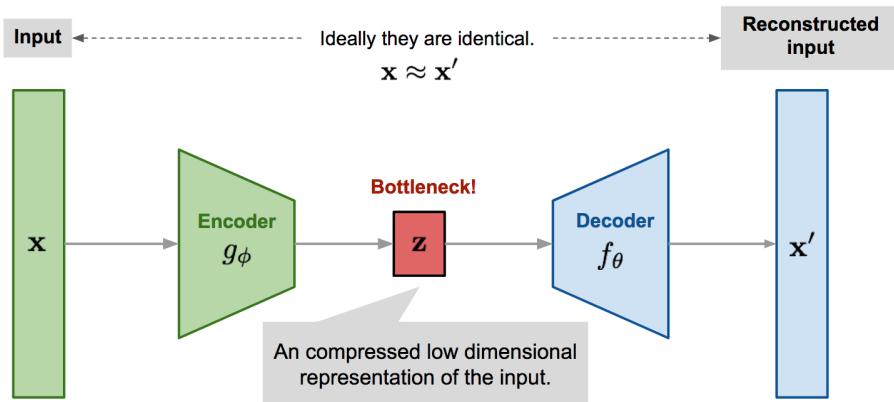


Figura 3.2: Esquema de la arquitectura empleada para inicializar el codificador y representación latente

El objetivo tras aplicar este esquema es construir un fuerte codificador que reduzca la dimensionalidad y una representación latente (bottleneck) que comprima las características principales de las imágenes del conjunto de entrenamiento.

En términos del proceso de optimización que es el entrenamiento, de forma no supervisada se están inicializando los pesos de la red para parecerse al conjunto de imágenes. [Zeiler and Fergus, 2014] estudia como los diferentes filtros de las redes neuronales convolucionales al entrenarse ante una tarea de clasificación acaban replicando las características generales y específicas de las imágenes de las que son entrenados. Por ello, se llega a la conclusión de que una heurística importante dentro de las redes neuronales convolucionales es que **los filtros se parecen a las imágenes**.

Esta razón explica de forma teórica como un proceso de ajuste previo al conjunto de entrenamiento elimina el coste computacional de búsqueda de los pesos via descenso del gradiente y backpropagation que requiere el ajuste

de la red por sí misma a las imágenes sólo a partir de las etiquetas.

### 3.3.2. Modelo de clasificación

### 3.3.3. Modelo de segmentación

## 3.4. Diseño de las arquitecturas

En el siguiente apartado detallaremos los módulos y capas que componen a las arquitecturas así como las componentes importantes.

### 3.4.1. Funciones de activación

Para todas las arquitecturas (segmentación, clasificación y el autoencoder para construir el codificador), se optará por el uso de la función de activación ReLU (Rectified Linear Unit), definida como:

$$\text{ReLU}(x) = \max(0, x)$$

Donde  $x$  es la entrada a la función ReLU.

A continuación, enunciamos algunas de las razones de su elección y su reconocida robustez para una gran amplitud de problemas en aprendizaje profundo.

1. **No Linealidad:** ReLU introduce no linealidad en las redes neuronales, lo cual es crucial para que las redes puedan aprender y modelar relaciones y características complejas en los datos. Esta no linealidad es esencial para tareas como la clasificación y la segmentación, donde las relaciones entre los datos son inherentemente no lineales.
2. **Gradiente Constante:** Para valores positivos de  $x$ , la derivada de ReLU es constante e igual a 1. Esto evita el problema del desvanecimiento del gradiente en redes profundas, donde el gradiente puede volverse extremadamente pequeño en funciones de activación saturadas como la sigmoide y la tangente hiperbólica.

$$\frac{\partial \text{ReLU}(x)}{\partial x} = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Esto facilita el entrenamiento de redes más profundas y la convergencia más rápida durante el proceso de optimización.

3. **Eficiencia Computacional:** La función ReLU es eficiente en términos computacionales. Su implementación es simple (una comparación y una operación de máximo) y no involucra cálculos costosos como funciones exponenciales.

La elección de ReLU como función de activación común a través de estas arquitecturas se basa en sus propiedades matemáticas que promueven la eficiencia, la no linealidad y la estabilidad del gradiente.

### 3.4.2. Construcción del codificador y representación latente

### 3.4.3. Arquitectura para clasificación

### 3.4.4. Arquitectura para segmentación

## 3.5. Optimización de las arquitecturas

En la siguiente sección detallaremos las funciones de pérdida usadas en cada arquitectura y la metodología seguida para llevar a cabo el entrenamiento de la red.

### 3.5.1. Optimizador

### 3.5.2. Funciones de pérdida

#### Función de pérdida para la reconstrucción de imágenes

Usamos el **error absoluto medio** (MAE) de los píxeles de salida reconstruidos y los verdaderos como función de pérdida para realizar la reconstrucción de las imágenes. El MAE mide la magnitud promedio de las diferencias absolutas entre los valores predichos por el modelo y los valores reales. Esta métrica es útil para evaluar la precisión de la reconstrucción de las imágenes, ya que considera todas las diferencias de manera uniforme, sin penalizar más las diferencias grandes que las pequeñas.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Donde  $n$  es el número total de píxeles en la imagen,  $y_i$  es el valor verdadero del píxel  $i$  y  $\hat{y}_i$  es el valor reconstruido o predicho del píxel  $i$ .

En el contexto de la reconstrucción de imágenes, cada  $y_i$  representa la intensidad del píxel en la imagen original, y cada  $\hat{y}_i$  representa la intensidad

del píxel en la imagen reconstruida por el modelo. El MAE proporciona una medida directa de cuán cerca están las intensidades de los píxeles reconstituidos de las intensidades reales.

También usaremos la misma pérdida como métrica de bondad de ajuste del modelo. Esto significa que, además de utilizar el MAE para optimizar el modelo durante el entrenamiento, también lo emplearemos para evaluar el desempeño del modelo en la reconstrucción de imágenes. Utilizar el MAE como métrica de evaluación nos permite tener una interpretación clara y consistente de cómo de bien se está desempeñando el modelo en términos de error promedio de los píxeles reconstituidos.

**Función de pérdida para clasificación**

**Función de pérdida para segmentación**

### **3.5.3. One-Cycle Policy**

## **3.6. Validación y garantías**

## **3.7. Solución teórica a la predicción de la evolución**



## Capítulo 4

# Experimentación



## **Capítulo 5**

# **Conclusiones y Trabajos Futuros**



# Bibliografía

- [Baid et al., 2021] Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F. C., Pati, S., et al. (2021). The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*.
- [Bakas et al., 2017] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozyczki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13.
- [Brügger et al., 2019] Brügger, R., Baumgartner, C. F., and Konukoglu, E. (2019). A partially reversible u-net for memory-efficient volumetric image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, pages 429–437. Springer.
- [Bulten et al., 2022] Bulten, W., Kartasalo, K., Chen, P.-H. C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., Steiner, D. F., van Boven, H., Vink, R., et al. (2022). Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, 28(1):154–163.
- [cancer.org, 2024] cancer.org (2024). American cancer society, cancer statistics center. 17 de marzo de 2024.
- [Cao et al., 2022] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M. (2022). Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer.
- [Chang, 2016] Chang, P. D. (2016). Fully convolutional deep residual neural networks for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and*

*mTOP 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 2*, pages 108–118. Springer.

[Cheng et al., 2022] Cheng, Q., Dong, Y., et al. (2022). Da vinci robot-assisted video image processing under artificial intelligence vision processing technology. *Computational and Mathematical Methods in Medicine*, 2022.

[Dominic LaBella, 2023] Dominic LaBella, e. a. (2023). The asnr-miccai brain tumor segmentation (brats) challenge 2023: Intracranial meningioma.

[Havaei et al., 2017] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., and Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31.

[Isensee et al., 2018] Isensee, F., Kneidingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. H. (2018). Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*, pages 287–297. Springer.

[Jiang et al., 2020] Jiang, Z., Ding, C., Liu, M., and Tao, D. (2020). Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5*, pages 231–241. Springer.

[Kamnitsas et al., 2018] Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al. (2018). Ensembles of multiple models and architectures for robust brain tumour segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*, pages 450–462. Springer.

[Kamnitsas et al., 2017] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B. (2017). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78.

- [Lin et al., 2017] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.
- [Liu et al., 2023] Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., Zhang, X., Jin, Y., and Zhou, H. (2023). Deep learning based brain tumor segmentation: a survey. *Complex & intelligent systems*, 9(1):1001–1026.
- [Menze et al., 2014] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024.
- [Myronenko, 2019] Myronenko, A. (2019). 3d mri brain tumor segmentation using autoencoder regularization. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pages 311–320. Springer.
- [Wang et al., 2018] Wang, G., Li, W., Ourselin, S., and Vercauteren, T. (2018). Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*, pages 178–190. Springer.
- [Zeiler and Fergus, 2014] Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I 13*, pages 818–833. Springer.
- [Zhou et al., 2021] Zhou, T., Canu, S., Vera, P., and Ruan, S. (2021). Latent correlation representation learning for brain tumor segmentation with missing mri modalities. *IEEE Transactions on Image Processing*, 30:4263–4274.
- [Zhu and Yan, 1997] Zhu, Y. and Yan, Z. (1997). Computerized tumor boundary detection using a hopfield neural network. *IEEE transactions on medical imaging*, 16(1):55–67.



# Glosario

newglossaryentryprevalencia name=prevalencia, description=En epidemiología, proporción de personas que sufren una enfermedad con respecto al total de la población en estudio newglossaryentrymaths name=mathematics, description=Mathematics is what mathematicians do



