

# Clasificación de tumores cerebrales mediante redes neuronales convolucionales con transferencia de aprendizaje.



Universitat  
Oberta  
de Catalunya

---

Nombre Estudiante

Álvaro Linacero Gracia

MU Bioinf. y Bioest.

Área de trabajo final

Diseño de fármacos y biología estructural

Nombre Tutor/a de TF

Karmele Lopez-de-Ipiña

Profesor/a responsable de la asignatura

Fecha Entrega 14/01/2025



Esta obra está sujeta a una licencia de Reconocimiento [3.0 España de Creative Commons](https://creativecommons.org/licenses/by/3.0/es/)

## FICHA DEL TRABAJO FINAL

Título del trabajo:	Implementación de redes neuronales para el análisis de MRI en el estudio de tumores cerebrales.
Nombre del autor:	Álvaro Linacero Gracia
Nombre del director/a:	Karmele Lopez de Ipina
Fecha de entrega (mm/aaaa):	14-01-2025
Titulación o programa:	Máster en Bioinformática y Bioestadística
Área del Trabajo Final:	Análisis de imágenes médicas mediante IA
Idioma del trabajo:	Castellano
Palabras clave	Convolutional neural networks, Nuclear Magnetic Resonance Imaging, Brain cancer
Resumen del Trabajo	
<p>Los tumores del sistema nervioso central tienen una prevalencia global del 2% entre a nivel mundial. Su localización hace de estos un problema difícil de tratar, ya que el área del cerebro es delicada y de difícil acceso. Con la finalidad de mejorar algunos modelos ya existentes, que permiten el diagnóstico precoz y precisión en la clasificación, se emplearán modelos de redes neuronales convolucionales (CNN) pre-entrenados (EfficientNetB1, Xception, VGG19, etc) en grandes conjuntos de imágenes “imagenet” para resolver problemas de clasificación de distintos subtipos tumorales (Meningioma, Pituitario, Glioma, Glioblastoma) o grado tumoral, Glioma de grado (II, III, IV) y clasificación en función de Supervivencia Global (SG) alta y baja. Adaptándose al problema de clasificación, empleando transferencia de aprendizaje de modelos de CNN para clasificar imágenes que provienen de distintas bases de datos. Los resultados mostrarán el modelo con mejor rendimiento de entrenamiento y validación, así como la sensibilidad, precisión, exactitud, f1-score. Como conclusiones hemos podido comprobar que emplear la estrategia de transferencia de aprendizaje puede ser muy útil para ahorrar tiempo de entrenamiento y se obtienen resultados interesantes, aunque inferiores a los vistos en la bibliografía. El número de imágenes y la variabilidad de estas es un factor crucial en el rendimiento del modelo de clasificación.</p>	
Abstract	
<p>Tumours of the central nervous system have an overall prevalence of 2% worldwide. Their location makes them a difficult problem to treat, as the area of the brain is delicate and difficult to access. In order to improve some existing models, which allow early diagnosis and classification accuracy, pre-trained convolutional neural network (CNN) models (EfficientNetB1, Xception, VGG19, etc) on large image sets ‘imagenet’ to solve classification problems for different tumour subtypes (Meningioma, Pituitary, Glioma, Glioblastoma) or tumour grade, Glioma grade (II, III, IV) and classification according to high and low Overall Survival (OS). Adapting to the classification problem, using CNN model learning transfer to classify images coming from different databases. The results will show the model with the best training and validation performance, as well as sensitivity, precision, accuracy, f1-score. As conclusions we have been able to prove that using the transfer learning strategy can be very useful to save training time and interesting results are obtained, although lower than those seen in the literature. The number of images and their variability is a crucial factor in the performance of the classification model.</p>	

# Índice

1.	Introducción.....	1
1.1.	Contexto y justificación del Trabajo.....	1
1.2.	Objetivos del Trabajo.....	3
1.3.	Impacto en sostenibilidad, ético-social y de diversidad.....	4
1.4.	Enfoque y método seguido.....	5
1.5.	Planificación del Trabajo .....	6
1.6.	Sumario de productos obtenidos: .....	7
1.7.	Descripción de otros capítulos de la memoria: .....	7
2.	Estado del arte .....	8
2.1.	Cáncer cerebral. ....	8
2.1.1.	Meningiomas.....	9
2.1.2.	Tumores hipofisarios .....	9
	Gliomas.....	9
	Glioblastoma (GBM).....	9
2.2.	Pruebas de imagen. ....	10
2.2.1.	Tipos de datos RM .....	11
2.3.	Aplicaciones de la IA al estudio radiológico del cáncer de cerebro. ....	12
2.3.1.	Clasificación de tumores cerebrales.....	13
2.4.	Machine Learning y Deep Learning .....	15
2.4.1.	Machine Learning Clásico: .....	15
2.4.2.	Deep Learning:.....	15
2.4.3.	Transferencia de aprendizaje .....	16
2.4.4.	CNN .....	17
3.	Metodología.....	20
3.1.	Bases de datos MR .....	20
3.2.	Preprocesado .....	21
3.3.	Selección del modelo.....	23
3.4.	Métricas de evaluación: .....	25
3.6.	Optimización de hiper parámetros .....	26
4.	Resultados.....	28

4.2.	Preprocesado de las imágenes .....	29
4.3.	Clasificación de subtipos: Meningioma, Glioma, Tumor hipofisario y no tumor 30	
4.4.	Clasificación Glioma grado II, III, IV.....	34
4.5.	Predicción de supervivencia en GBM. ....	40
5.	Conclusiones y trabajos futuros.....	43
6.	Bibliografía.....	47
	Anexos.....	52

## Índice de Figuras

Figura 1. Función de activación ReLU.....	19
Figura 2.Estructura de modelo VGG19[48]. ....	24
Figura 3. Esquema estructura de bloques EfficienNetB1[51]. ....	25
Figura 4. Muestra de 9 imágenes aleatorias aumentadas aplicando transformaciones. .	30
Figura 5. Imágenes de RM de pacientes sanos y 3 subtipos.....	31
Figura 6. Matrices de confusión. Clasificación subtipos a) VGG19 y b) EfficientNetB1 (30 etapas).....	33
Figura 7. Curvas ROC. a) VGG19 y b) EfficientNetB. i) Clase 1 = Glioma, ii) Clase 2 = Meningioma, iii) Clase 3 = No Tumor, iv) Clase 4 = Tumor Hipofisario. ....	33
Figura 8. Gliomas de grados (II, III, IV). Base de datos REMBRANDT. ....	35
Figura 9. Matriz de confusión y curva ROC para modelo sin normalizar y sin cropping. ....	36
Figura 10. Matriz de confusión y curva ROC de modelo con normalización y cropping. ....	36
Figura 11. Matriz de confusión y curvas ROC. Modelo EfficientNet (*) 15 etapas. ....	38
Figura 12. Gráficos modelo EfficientNetB1(*) optimizado 60 etapas.....	38
Figura 13. Exactitud y pérdida en entrenamiento y validación (60 etapas). ....	39
Figura 14. Matriz de confusión LGG y HGG.....	40
Figura 15. Imágenes de GBM extraídas de la base de datos UPENN de TCIA.....	40
Figura 16. Exactitud y pérdida en entrenamiento y validación con VGG19 (SG).....	42
Figura 17. Matriz de confusión y curva ROC. Predicción SG para GBM. ....	42

# 1. Introducción

## 1.1. Contexto y justificación del Trabajo

El desarrollo de nuevas tecnologías basadas en inteligencia artificial ha revolucionado el estudio de muchas enfermedades, siendo los diferentes tipos de cáncer de las más estudiadas, estos avances permiten ayudar a un diagnóstico precoz, búsqueda de mutaciones específicas, métodos de clasificación basados en características de imagen, clínicas, moleculares, etc.

El cáncer ha sido y sigue siendo una de las principales causas de muerte a nivel mundial. En 2020 la organización mundial de la salud (WHO) indicó que fue la segunda causa de muerte a nivel mundial. Según la agencia internacional de investigación del cáncer la incidencia del cáncer de cerebro a nivel mundial, actualizada al año 2022 fue de 321.731 casos a nivel mundial, 67.559 en Europa y 4.317 en España. Aunque representan un porcentaje menor, aproximadamente un 2% de todos los cánceres a nivel mundial, su estudio es de gran interés, debido a la complejidad que supone el acceso al cerebro y la dificultad de tratamiento[1].

Existen aproximadamente 150 subtipos de tumores en el Sistema Nervioso Central (SNC)[2]. Los principales subtipos de cáncer cerebral son los Glioma que se origina a partir de las células gliales del cerebro, Meningioma que se forma en la membrana que recubre al cerebro y en los alrededores de la médula espinal y tumores generados en la glándula pituitaria (Tumores hipofisarios), que generalmente son tumores benignos.

Los síntomas pueden variar en función del subtipo, ubicación y tamaño del tumor. Siendo los más comunes dolor de cabeza persistente, problemas de equilibrio, concentración, convulsiones, problemas de visión y cambios de comportamiento y de personalidad.

El seguimiento de gran parte de los casos de tumores recibe seguimiento mediante pruebas de imagen radiológica. Las técnicas más empleadas para la detección, diagnóstico y seguimiento de tumores cerebrales son la tomografía computarizada (TC), la electroencefalografía, la ultrasonografía, la tomografía por emisión de positrones, la resonancia magnética (RM) y la tomografía computarizada por emisión monofotónica.

Algunas de estas modalidades de imagen han surgido recientemente y entre las técnicas anteriormente mencionadas, la más empleada es RM por su buena capacidad para aplicar contraste con una leve carga ionizante[2], además de permitir obtener una imagen que informa del tamaño, forma y localización de tumor.

Los métodos de aprendizaje automático (ML) y aprendizaje profundo (DL) han permitido la creación de herramientas que ayudan a mejorar tareas antes realizadas exclusivamente por profesionales sanitarios, como el análisis de resonancias magnéticas para la detección, diagnóstico y monitorización de tumores cerebrales, así como su capacidad para diferenciar entre subtipos tumorales y ser capaz de encontrar anomalías o señales que podrían pasar desapercibidas al ojo humano.

Gracias a la constante investigación en esta área se han obtenido rendimientos muy elevados para los distintos problemas de clasificación: Detección de la presencia de tumor, clasificación tumores en los diferentes subtipos o distinguir el grado del tumor. Hay que aclarar que estos modelos alcanzan buenos rendimientos en tareas de clasificación concretas y no es posible diseñar un modelo de CNN eficiente para todos los problemas de clasificación[3].

Con los resultados tan prometedores que se han obtenido con algunos modelos puede resultar sorprendente que no estén aun implantados en la práctica clínica habitual. Sin embargo, no están exentos de complicaciones y dificultades. Debido a la dificultad a la hora de obtener imágenes con parámetros de adquisición similares[4], los escáneres realizados en cada centro pueden tener diferencias en el tamaño de salida y la calidad de las imágenes debido a la falta de estandarización. Por ello es necesario realizar un trabajo de preprocesado y normalización de las imágenes para reducir la variabilidad de estas, consiguiendo muy buenos resultados. Pese a transformar las imágenes, los modelos siguen perdiendo calidad al emplear un conjunto de prueba proveniente de un centro diferente al conjunto de entrenamiento.

Actualmente hay empresas que comercializan software diseñado para el análisis de imágenes para implantarlo en los centros sanitarios como herramienta para ayudar al trabajo de los radiólogos en la toma de decisiones con procesos relacionados con el



análisis y procesamiento de imágenes RM, aunque la decisión final siempre es llevada a cabo por un humano.

Entre la gran variedad de métodos existentes en análisis de imágenes asistido por ordenador, cada uno tiene sus limitaciones y sus ventajas[5]. Las redes neuronales convolucionales (CNN) han sido durante los últimos años el estándar del análisis de imágenes médicas, realizando tareas como segmentación de áreas tumorales y clasificación de subtipos tumorales. Esta rama de la inteligencia artificial ha llegado a desarrollar modelos de arquitectura neuronal capaces de analizar volúmenes en 3D que permiten alcanzar resultados con precisiones cercanas al 100%. La parte negativa es que necesitan de grandes volúmenes de datos, ordenadores y GPUs muy potentes y mayores tiempos de computación, lo que dificulta su implementación como herramienta rutinaria para análisis de imágenes radiológicas en centros hospitalarios.

Este trabajo está motivado por la necesidad de mejorar la eficacia del diagnóstico de tumores cerebrales, empleando técnicas de aprendizaje profundo que no precisen de tiempos de entrenamiento elevados ni de un gran volumen de imágenes. Sirviéndonos de modelos previamente entrenados en grandes conjuntos de datos y adaptándolos al problema de clasificación.

## 1.2. Objetivos del Trabajo

Objetivo principal: Analizar mediante Técnicas informáticas imágenes RM y entrenar un modelo para que pueda servir de herramienta diagnóstica y pronóstica.

Objetivos secundarios:

- Emplear IA para predecir la supervivencia de los pacientes con cáncer cerebral basándose en características extraídas de RM mediante redes neuronales empleando características extraídas de las imágenes y datos clínicos.
- Comparar el modelo con sistemas de clasificación y de predicción de supervivencia.

### 1.3. Impacto en sostenibilidad, ético-social y de diversidad

- I. Sostenibilidad: Teniendo en cuenta que para el tratamiento de imágenes solamente se necesitan ordenadores y las propias imágenes de RM considero que el único impacto en aspecto de sostenibilidad es el gasto energético, ya que entrenar modelos pesados es costoso energéticamente, aun así, los modelos que se plantean en este trabajo no requieren de demasiados recursos al emplear modelos que ya han sido entrenados.

ODS 9- Industry, innovation and infrastructure.

- II. Comportamiento ético y responsabilidad social (RS): En este trabajo se utilizan datos personales, concretamente Imágenes de RM de personas con alta resolución, estas permiten visualizar el cráneo y algunas estructuras como los ojos y con las que se podría replicar algunas facciones de la persona, por ello al solicitar acceso a las imágenes es necesario firmar un documento con el compromiso de que se van a emplear para fines de investigación.

También aparece el dilema ético de poner por delante la decisión de una máquina a la de un profesional de la salud como es el/la radiólogo/a. De momento los softwares de análisis de imagen no se emplean más que como una herramienta, pero el debate está presente y con los futuros avances será un tema para tratar.

Como ayudaría a mejorar la precisión en el trabajo, considero que encajaría con ODS. 8- Decent work and economic growth.

- III. Diversidad (género entre otros) y derechos humanos: Este trabajo no tiene impacto ni positivo ni negativo en aspectos de género, en cuanto a la diversidad no contamos con las nacionalidades de los pacientes que forman parte del estudio. En cuanto a los derechos humanos: Este tipo de trabajo podría hacer más accesible la medicina a todo el mundo. Una contraparte es que el acceso a los datos sobre una persona puede ir en contra de los derechos de privacidad de cada uno, aunque ya está previsto por los comités de ética en el caso de las imágenes publicadas en bases de datos públicas. ODS 10- Reduced inequalities.

## 1.4. Enfoque y método seguido

En base a la literatura y a los estudios previos en patologías similares se buscaron bases de datos que contuviesen imágenes de resonancia magnética nuclear (RMN) de pacientes con cáncer cerebral con el fin de identificar y clasificar tumores en función de variables, como tiempo de supervivencia, subtipo tumoral, además de características del tumor como la clasificación TNM y datos moleculares, genómicos, proteómicos.

El objetivo es combinar técnicas de aprendizaje profundo, específicamente redes neuronales convolucionales (CNN), con análisis estadísticos para evaluar la eficacia del modelo en la clasificación de tumores. Nos basaremos en la literatura que emplea análisis de imágenes con “Deep Learning” encontrados en el estado del arte, por la capacidad que tienen de extraer características de las imágenes, necesarias para resolver problemas de clasificación.

Como metodología para llevar a cabo clasificación de subtipos tumorales empleamos una serie de modelos de Deep Learning pre entrenados con grandes conjuntos de datos (ImageNet), empleando la técnica de transferencia de aprendizaje “Transfer Learning” y adaptando la estructura de la red convolucional para nuestro problema de clasificación.

El trabajo consistirá en aplicar Transferencia de aprendizaje con modelos VGG19, EfficientNetB1, Xception y MobileNetV2 con imágenes en 2D extraídas de bases de datos públicas para entrenar modelos, tratando de encontrar los hiper parámetros que mejoren el rendimiento del modelo y lo adapten al problema de clasificación.

Algunas de las imágenes han sido extraídas de una base de datos de GBM en 3D mediante un script de Python que realizó los cortes, transformándolas a imágenes 2D. También clasificamos las imágenes por contraste y corte para poder centrarnos en un conjunto de ellas.

Tras seleccionar heurísticamente imágenes donde se apreciaba de manera clara la presencia de la masa tumoral se creó un subconjunto de datos de GBM que se incluyó en el problema de clasificación de supervivencia. Un proceso similar fue llevado a cabo con imágenes de la base de datos de Gliomas.

Pese a los últimos avances en análisis de volúmenes en 3D hemos decidido trabajar con imágenes en 2D ya que la capacidad del equipo y el tiempo disponible no permitían trabajar con volúmenes de datos tan grandes.

El tratamiento de los datos clínicos de los pacientes se llevó a cabo con el programa R.

## 1.5. Planificación del Trabajo

### **PEC 2 (11/10/2024 – 07/11/2024)**

Hito 1. Revisar artículos recientes sobre IA en resonancia magnética y detección de tumores y conocer los enfoques más relevantes y tendencias en el uso actuales. (11/10/2024-21/10/2024)

Hito 2. Redactar el estado del arte y revisión de este con la tutora. (22/11/2024 – 02/11/2024)

Hito 3. Completar el estado del arte con las correcciones/aportaciones tras la revisión (03/11/2024 – 07/11/2024)

### **PEC 3 (08/11/2024 – 11/12/2024)**

Hito 4. Definir el conjunto de datos y trabajarlo previo a entrenar al modelo. (09/11/2024 – 15/11/2024)

Hito 5. Iniciar la fase de entrenamiento con el modelo (CNN). (16/11/2024 – 23/11/2024)

Hito 6. Optimización e iteración del modelo, evaluar su rendimiento. (24/11/2024 – 04/12/2024)

Hito 7. Realizar el análisis estadístico de los resultados y realización de gráficos/figuras. (05/12/2024 – 11/12/2024)

### **PEC 4 (12/11/2024 – 14/01/2025)**

Hito 8. Redacción de la memoria (13/12/2024 – 30/12/2024)

Hito 9. Realización de la presentación (31/12/2024- 14/01/2025)

### **PEC 5 (23/01/2025 – 31/01/2025)**

Hito 10. Practicar la presentación (23/01/2025 – 31/01/2025)

Hito 11. Hacer los ajustes finales a las diapositivas si es necesario. (23/01/2025 – 31/01/2025).

## 1.6. Sumario de productos obtenidos:

- 1- Conjunto de imágenes de Glioma y GBM en 2 D extraída de los volúmenes en formato DICOM y NIfTi de las bases de datos REMBRANDT y UPENN-GBM.
- 2- Código para realizar las tareas de obtención de imágenes, clasificación, preprocesado y de entrenamiento del modelo.
- 3- Medidas de rendimiento de los distintos modelos
- 4- Modelo con mejor rendimiento

## 1.7. Descripción de otros capítulos de la memoria:

Introducción: Introducción al cáncer cerebral, contexto, importancia y motivación a la hora de realizar el trabajo.

Estado del arte: Descripción de los principales tipos de tumores cerebrales, de las imágenes de resonancia magnética y de las estrategias basadas en inteligencia artificial más empleadas para su manejo, centrándonos en las redes neuronales convolucionales y la transferencia de aprendizaje.

Metodología: Explicación de la estructura principal de las redes neuronales convolucionales, de sus partes principales, funciones más importantes y medidas de rendimiento.

Resultados: Imágenes extraídas de los volúmenes en tres dimensiones, tablas de métricas para los distintos modelos, figuras que muestran el rendimiento para cada problema de clasificación y el código empleado.

Conclusiones: Los resultados muestran la capacidad de los modelos pre entrenados para resolver problemas de clasificación, pese a ello no alcanzan los resultados esperados,

debido a la variabilidad de las imágenes extraídas manualmente, la complejidad del problema de clasificación o un preprocesamiento de las imágenes mejorable.

Glosario: Descripción de conceptos importantes de la memoria

Bibliografía: Conjunto de citas empleadas para la realización del trabajo.

## 2. Estado del arte

### 2.1. Cáncer cerebral.

Según el Observatorio Mundial del cáncer[1], en el año 2022 confirmó que el cáncer cerebral o asociado al sistema nervioso central afectó a 321.731 personas en todo el mundo y causó 248.500 muertes registradas, ocupando el décimo sexto lugar en mortalidad.

Los diferentes subtipos de tumor tienen diferentes pronósticos, tratamientos y es de gran importancia conocer de cual se trata para dar un mejor tratamiento[6]. Al tratarse el cerebro de una zona de difícil acceso no es habitual la realización de biopsia hasta el momento de la cirugía. Para poder diferenciar entre subtipos se depende del análisis radiológico, datos clínicos y pruebas moleculares.

Las imágenes de RM[7] ayudan a los médicos a evaluar los tumores para planificar el tratamiento posterior, basándose en la forma, el tamaño, el tipo, el grado y la localización del cáncer. Dependiendo del estado del paciente, estos factores pueden variar enormemente. Por lo tanto, el reconocimiento y la clasificación precisos de los tumores cerebrales son fundamentales para un tratamiento adecuado.

Desde hace años el análisis de imágenes con ayuda de programas informáticos es un tema en desarrollo y han sido numerosos los intentos de desarrollar modelos de clasificación para llevar a cabo esta tarea y con el crecimiento de la IA cada vez surgen nuevos avances.

Algunos de los tumores cerebrales primarios más estudiados son los Gliomas, los Meningiomas y los Tumores Hipofisarios.

### 2.1.1. Meningiomas

Los Meningiomas[8] son tumores primarios del SNC. Surgen de las membranas que recubren el cerebro y rodean el sistema nervioso central. Pueden clasificarse en varios grados, siendo el grado I el más común, tumor benigno de crecimiento lento. Meningioma de grado II es de grado intermedio, tiene mayores probabilidades de recurrencia después de ser extirpado y Meningioma de grado III es un tumor maligno y con crecimiento rápido.

### 2.1.2. Tumores hipofisarios

Los tumores de hipófisis, aunque pueden ser benignos y de crecimiento lento como los Adenomas hipofisarios benignos, pueden causar otro tipo de daños. Las enfermedades y síndromes que afectan a la hipófisis pueden causar deficiencias hormonales y puede afectar al funcionamiento de la glándula u órgano controlado por la hormona en cuestión. Puede ocasionar pérdida de visión, dolor de cabeza, problemas de infertilidad y demás enfermedades derivadas de un desajuste hormonal. Esta afección plantea la importancia de diagnóstico precoz que permite seleccionar la medicación y minimizar el sufrimiento de los pacientes.[9,10]

## Gliomas

Gliomas son un tipo de tumor cerebral que se origina en las células gliales, representan un 30% de los tumores cerebrales y gran parte de los malignos. Se pueden clasificar en función del tipo de células gliales que los originan. Astrocitomas (astrocitos), Oligodendrogliomas (oligodendrocitos) y Ependimomas (células ependimarias).[11]

También se pueden dividir en función de su agresividad, estadiándolo en grados, siendo (I-II) grado bajo y (III-IV) grado alto.

### Glioblastoma (GBM)

Es el tumor primario más común en adultos, más agresivo y complejo del SNC. Pese a los avances en la medicina del cáncer, la supervivencia global para esta patología no ha mejorado, siendo la supervivencia media de 16-20 meses y una ratio de supervivencia a los 5 años del 10%. [12]

GBM es la forma más agresiva de glioma, se desarrolla en los astrocitos. Su forma, extensión y localización heterogéneas son algunos de los factores que hacen que estos tumores sean difíciles de extirpar y en algunos casos, inoperables.

Perfil genómico: GBM primario es histológicamente similar a su variante secundaria, difieren en el perfil genético, GBM primario suele presentar características moleculares específicas como mutaciones en EGFR, PTEN, CDKN2A y MDM2, aparece directamente como tumor de grado IV, afectando principalmente a pacientes de edad avanzada y presenta el peor pronóstico. Mientras que GBM secundario suele derivar de otros tumores como astrocitomas de grado II o III y tiene una evolución más lenta y puede presentar mutaciones en pT53 y IDH.[13]

Otra mutación significativa para la patogénesis de los gliomas y el glioblastoma es la isocitrato deshidrogenasa-1 (IDH-1) en el cromosoma 2. El glioblastoma de tipo salvaje es mucho más frecuente que el que presenta mutación en el gen IDH-1 y se considera la forma más agresiva, responde mal a las terapias actuales y tiene un pronóstico peor a cinco años. Los glioblastomas con mutaciones en IDH-1 tienen un pronóstico ligeramente mejor debido a su mejor respuesta a la radioterapia y a la temozolomida.

El aspecto heterogéneo del GBM puede ser a veces difícil de distinguir de otras neoplasias o afecciones no neoplásicas, como los gliomas de grado inferior, incluso utilizando técnicas avanzadas de RM. Por ello se necesitan métodos más precisos para clasificar estos tumores.

Algunos de los síntomas derivados por el crecimiento heterogéneo de la masa tumoral son: que las áreas comprimidas y desplazadas del cerebro pueden ver alteradas sus conexiones y su flujo sanguíneo, provocando síntomas como dolor de cabeza, convulsiones, pérdida de función cognitiva, o alteraciones motoras y sensoriales. Al ser heterogénea puede darse en diferentes áreas del cerebro, lo que dificulta delimitar el tumor, el tratamiento quirúrgico y la radioterapia.

## 2.2. Pruebas de imagen.

La RM tiene un excelente contraste tisular y puede proporcionar información diversa sobre la difusividad del agua, el flujo sanguíneo, el estado de la barrera hematoencefálica



y la celularidad. El análisis de estas pruebas está sujeto al factor subjetivo del análisis humano, precisa de tiempo, posibilidad de errores debidos al cansancio o fatiga visual. Problemas con la imagen, debido a que pueden presentar formas y tamaños diferentes incluso dentro del mismo tipo de tumor, la imagen puede tener ruido, bordes del tumor difusos, lo que dificultará la interpretación de la imagen.[14]

Para reducir estas posibles complicaciones se han desarrollado diferentes modelos de segmentación y clasificación asistidos por ordenador para mejorar la precisión y ayudar al radiólogo a tomar mejores decisiones y reducir el error. En los últimos años se han desarrollado diferentes modelos y estrategias. Los más empleados son los algoritmos de Machine Learning (ML) y Deep Learning (DL), donde la estrategia más empleada es el uso de redes neuronales convolucionales.

### 2.2.1. Tipos de datos RM

Las imágenes de RM en su formato de salida están en formato DICOM, son una serie de archivos que contienen datos de cortes en dos dimensiones que permiten reconstruir el volumen en 3D. Es el método estándar más empleado para trabajar en radiología, son compatibles con softwares de sistemas hospitalarios y contiene metadatos con información del aparato y sus parámetros de adquisición. El otro formato más extendido es NIfTI, fue diseñado para ser empleado en procesamiento e investigación en neuroimagen. Contiene la imagen volumétrica del cerebro en un único archivo en 3D, aunque no puede visualizarse directamente, existen programas que permiten visualizar el volumen y realizar cortes y segmentaciones de manera muy sencilla (3D Slicer).

Imágenes de resonancia magnética multiparamétrica (mpMRI) obtienen una serie de imágenes de salida aplicando contrastes. Las más empleadas para el estudio de tumores cerebrales gracias a su versatilidad son:

**T1** ponderada sin contraste, los tumores suelen aparecer oscuros en comparación con el tejido sano.

**T1-Gd** Contraste con Gadolinio. El tumor y sus bordes se vuelven más visibles

**T2** ponderada proporciona una visión clara de las áreas con alto contenido de agua, como el edema y las zonas tumorales. En este tipo de imágenes los tumores suelen aparecer brillantes, siendo útil para identificar el edema circundante, así como áreas heterogéneas que puedan representar cambios tumorales.

**FLAIR** Similar a T2, pero suprime el líquido cefalorraquídeo (LCR), permitiendo destacar el edema y la infiltración tumoral sin la interferencia del LCR brillante. Es muy útil para identificar la extensión de la infiltración tumoral en el cerebro, especialmente en el caso de tumores cerebrales difusos.

#### **Técnicas adicionales de mpMRI:**

Imagen ponderada por difusión (Diffusion-weighted imaging, DWI).

Imagen ponderada por perfusión (Perfusion-weighted imaging, PWI).

Imagen ponderada por susceptibilidad (Susceptibility-Weighted Imaging, SWI).

Espectroscopía por RM (Magnetic Resonance Spectroscopy, MRS).

### **2.3. Aplicaciones de la IA al estudio radiológico del cáncer de cerebro.**

Un avance consiste en análisis de imágenes médicas en tiempo real durante la cirugía, permiten visualizar en tiempo real que zonas del cerebro están asociadas con funciones motoras o cognitivas, facilitando la extracción del tumor tratando de no afectar a esas zonas.

Los modelos basados en inteligencia artificial tienen un gran potencial en medicina y ofrecen una serie de herramientas que mejoran el tratamiento de algunas enfermedades. Permite un diagnóstico más preciso y una detección precoz, mejor manejo de la enfermedad y estimación del pronóstico y supervivencia.

Hay una gran cantidad de algoritmos que emplean modelos clásicos de Machine Learning y Deep Learning para ayudar a la toma de decisiones y a resolver problemas de clasificación. Es posible integrar una serie de biomarcadores multiómicos como: Proteómica, genómica, metabolómica, radiómica, con características clínicas para crear

modelos que ayuden a estimar la supervivencia, la probabilidad de recurrencia y la respuesta al tratamiento.

### 2.3.1. Clasificación de tumores cerebrales.

Los métodos de clasificación y toma de decisiones basados en IA son de gran utilidad para mejorar la precisión del tratamiento, en el caso del cáncer de cerebro la mayoría de las técnicas se basan principalmente en el análisis de visión computacional de RM, que es la prueba más empleada rutinariamente por los radiólogos en el estudio de afecciones en el cerebro.

Las redes neuronales convolucionales (CNN) son un tipo de redes neuronales cuyo uso principal es el análisis de imágenes. Por su capacidad para extraer características de las imágenes (detectar bordes, detalles sutiles y las subregiones dentro del tumor, como área de edema, núcleo necrótico) son la técnica más empleada para extraer información de las imágenes de RM.

Las principales aplicaciones en el análisis de imágenes de RM son: Segmentación de la región específica donde se encuentra el tumor, clasificación de algunos subtipos tumorales a partir de las imágenes en 2D o incluso de volúmenes en 3D[15]. También hay estudios donde clasifica los tumores en alto o bajo grado.

**Clasificación de subtipos:** Hay un elevado número de artículos publicados con arquitecturas diferentes para resolver el problema de clasificación de Glioma, Meningioma, Tumor hipofisario y ausencia de tumor.

DNN con preentrenamiento de red generativa adversarial [16]. Modelo que mejora el contraste de las imágenes, extrae la localización del tumor y analiza las imágenes mediante EfficientNet consiguiendo una exactitud general de clasificación de 98.04%. [17]. Un modelo de CNN personalizada para resolver los dos problemas de clasificación. [18]. Otro modelo de CNN personalizada que ha optimizado todos sus parámetros para resolver los problemas de clasificación (Tumor, no tumor), clasificación (No tumor, Meningioma, Glioma, Tumor hipofisario, Metastasis) con un 92.66% de exactitud y clasificación entre los grados de Glioma (II, III, IV) con un 98.14% de exactitud[3]. Clasificación de tumores (Glioblastoma, Sarcoma y carcinoma broncogénico) mediante

DNN[19]. Uso de red neuronal recurrente (RNN) o CNN para clasificar subtipos tumorales (Tumor metastático, Glioma, Meningioma) con un codificador automático basado en DenseNet para la extracción de características[20]. Empleo de validación cruzada (5-fold cross validation) se obtuvo precisión de (92.3 %) GoogleNet, (97.8 %) SVM y (98.0 %) KNN [21] aunque tuvo dificultades para clasificar los meningiomas. Un modelo se entrenó mediante fine tuning para clasificar los tumores en tres clases. La exactitud de la clasificación fue del 96,9% en un conjunto de datos no visto en el entrenamiento[22].

Mecanismo que mezcla características extraídas con el modelo EfficientNetB7 y MobileNetV3 mediante un mecanismo de co-atención, alcanzando un 98.94% de exactitud. [2]

**Clasificación del grado tumoral en Gliomas:** Los Gliomas se pueden clasificar en función de su grado tumoral (I, II, III, IV) y también agrupando los de menor grado (LGG y HGG).

Empleando un modelo basado en Alexnet modificado se consigue un valor de exactitud global de (0.91) en la clasificación de tres clases (Sano, LGG y HGG) [23]. Un modelo de transfer learning y 5-fold CV con GoogleNet se obtuvo un valor de exactitud en el test de (0.95) y AUC (0.97) con imágenes T1 Gd.[24]

Modelo 3D multi céntrico para la predicción del estado de mutación IDH1 (0.90) AUC, LGG vs HGG (0.91) AUC y clasificación de los grados de Glioma II, III, IV de (.81) AUC[25]. Otro estudio similar emplea 3D-Unet entrenado en la base de datos BRATS 2019, extrae la región tumoral de interés y la introduce a una CNN para predecir el grado, estado de mutación IDH1 y 1p19q. [26]. Clasificación de LGG de HGG con RM, T1-Gd y AUC (0.97). Meta análisis [27]. Modelo compuesto por CNN para la extracción de características y SVM para la clasificación del glioma en LGG y HGG obteniendo una exactitud de 95.83% [28]. Modelo DNN a partir de imágenes de tensores de difusión de 44 LGG y 48 HGG (0,98 AUC)[29]. El mecanismo que mezcla características extraídas con el modelo EfficientNetB7 y MobileNetV3 mediante un mecanismo de co-atención, alcanzando un 98.48% de exactitud. [2]

## 2.4. Machine Learning y Deep Learning

### 2.4.1. Machine Learning Clásico:

Son métodos supervisados, necesitan la selección de las mejores características para llevar a cabo la tarea de clasificación. Primero se lleva a cabo un análisis de los datos, realizándose la limpieza de estos si es necesario, transformaciones, preprocesado y a continuación, se procede a la extracción de características y la selección de estas de manera heurística o empleando alguna estrategia, como análisis de componentes principales. Finalmente se construye el modelo con las muestras etiquetadas.

Una de las debilidades del aprendizaje automático es que no es capaz de extraer características de las imágenes que procesa, sin embargo, estos algoritmos pueden utilizarse solos o en combinación con otros algoritmos de ML o técnicas de selección de características. Mediante aprendizaje profundo se pueden extraer características profundas de los modelos CNN y se pueden utilizar como entrada en distintos clasificadores ML, Gaussian Naïve Bayes (Gaussian NB), Adaptive Boosting (AdaBoost), K-Nearest Neighbors (k-NN), Random forest (RF), Support Vector Machine (SVM), etc.[5,30]

### 2.4.2. Deep Learning:

Son redes con una arquitectura de capas, principalmente capas profundas, que le dan su nombre. Son capaces de extraer las mejores características tras un proceso de entrenamiento sin necesidad de seleccionarlas heurísticamente. [5]

Visión Transformers: La arquitectura “transformer” ha sido una revolución en el procesamiento de lenguaje natural incorpora un mecanismo de autoatención, que evalúa la relación entre diferentes posiciones de una secuencia. Vision Transformer divide las imágenes en parches y los analiza uno a continuación de otro, permitiendo capturar relaciones espaciales. Se han desarrollado modelos para segmentar y clasificar imágenes 2D y 3D.

Modelos híbridos: Empleando modelos híbridos, mezcla de algoritmos de CNN y ML clásico (RF, SVM, NB, K-NN, C-means), los clasificadores basados en conjuntos generan varios clasificadores y los combinan sus predicciones para ayudan a mejorar precisión y robustez y reducir el sobreajuste.

Mecanismos de atención: Es un mecanismo que permite centrar la atención en ciertas partes de la entrada, lo que es especialmente relevante cuando hay entradas más importantes que otras para el modelo. Tipos de atención (Soft attention, Hard attention, Self-attention y Multi-head attention).

Redes Neuronales convolucionales: Se trata de un tipo de red neuronal especializada en extracción de características y clasificación de imagen. Su estructura está inspirada en las conexiones entre neuronas del cerebro humano, siendo las neuronas las unidades básicas de la red, estas neuronas están agrupadas en capas, donde la salida de una capa será la entrada a la siguiente[10].

En el caso de las imágenes, los datos de entrada son valores de intensidad de los píxeles que la componen. La correlación entre los valores de los píxeles próximos entre si aporta información de esa subregión de la imagen, las CNN extraen esta información en forma de patrones aplicando una serie de filtros en las capas convolucionales.

Al ir aumentando el número de capas convolucionales, estas características locales se acumulan para formar características de orden superior[31].

### 2.4.3. Transferencia de aprendizaje

Transferencia de aprendizaje o “Transfer Learning” es una opción para generar un modelo aprovechando una arquitectura de red pre-entrenada (U-net, ResNet, MobileNet, VGG-19, EfficientNet u otra CNN), transfiriendo los pesos de entrenamiento. Se congelan sus capas iniciales y se modifica la capa de salida, a continuación, se configura un modelo de optimización, una función de pérdida y métricas adecuadas y se entrena al modelo en el nuevo conjunto de datos, ajustando solo los pesos de las capas superiores, de manera que las últimas capas y la capa de salida estarán adaptadas a una nueva función de clasificación.

Estos modelos han sido entrenados en conjuntos con un número muy elevado de imágenes de objetos muy diferentes, de las que obtienen la capacidad de extraer características. [7]

Fine-tuning: Permite entrenar no solo las ultimas capas, sino también las intermedias e incluso todas, puede ser útil cuando se trata de mejorar un modelo modificado con transferencia de aprendizaje.

#### 2.4.4. CNN

Neuronas: Son las unidades básicas del modelo, están inspiradas en las neuronas del cerebro humano y en la manera en que se conectan entre ellas para llevar a cabo tareas. Las neuronas de la capa de entrada simplemente pasan la información al siguiente nivel, las que forman las capas densas, reciben información, la procesan y la transmiten a la capa siguiente y las neuronas de la capa de clasificación representan una de las clases del problema.

Las neuronas de las capas densas reciben una serie de entradas, que suman y multiplica por los pesos definidos por el modelo, a continuación, hace pasar la suma ponderada por una función de activación  $\sigma$  (Relu, Selu, SoftMax, etc) dependiendo de que capa se trate y la función que desempeñe. La salida de la neurona se conecta con alguna neurona de la capa siguiente y en caso de la capa completamente conectada se conectan todas las neuronas de la capa de salida con la siguiente.

$$z = \sum_{i=1}^n (x_i * w_i) + b$$

Salida =  $\sigma(z)$

$z$  = suma ponderada de características

$x_i$  = Características de entrada

$w_i$  = Pesos de cada entrada

$b$  = Sesgo

$\sigma$  = Función de activación

Al comienzo del entrenamiento los pesos iniciales son aleatorios. Cada etapa del entrenamiento consiste en hacer pasar todo el conjunto de imágenes de entrenamiento por las diferentes capas, extrayendo las características mediante filtros, hasta llegar a la capa

de clasificación, pasando por las capas de agrupamiento, normalización, drop-out, completamente conectadas y finalmente la capa de salida o de clasificación. Se realizan predicciones para los datos de entrenamiento, midiendo la función de pérdida en cada etapa y se calcula el gradiente de la función de pérdida con respecto a cada peso mediante el algoritmo “back propagation”. Los algoritmos de optimización ajustan los pesos del modelo en función de los gradientes calculados.

**Función de pérdida:** La pérdida[32] se define como el coste por cometer una imprecisión en la clasificación. Cuanto menor sea la pérdida más se ajustarán las predicciones a las etiquetas verdaderas.

- Categorical cross entropy: Función que calcula la pérdida, computando la diferencia entre las etiquetas y las predicciones.

### **Optimizadores:**

- Adam (Adaptive Momentum Estimation): Adam es un potente algoritmo de optimización que combina los puntos fuertes de momentum y RMSprop al tiempo que añade algunas mejoras adicionales. Se ha convertido en una opción popular para aplicaciones de aprendizaje profundo.[33] Ajusta automáticamente las tasas de aprendizaje para cada parámetro.
- SGD (Stochastic Gradient Descent): El algoritmo de descenso de gradiente actualiza los parámetros del modelo de forma iterativa modificándolos en la dirección opuesta al gradiente de la función de pérdida, busca reducir el valor de la función de pérdida en procesos de ML[34]. Actualiza los pesos empleando gradientes calculados con subconjuntos de datos.

### **Estructura de una CNN [3,10,23,32]:**

- Capa de entrada: Los datos tienen que estar en el formato específico para que puedan ser interpretados por las capas convolucionales. Al tratarse de imágenes, estas deberán tener unas dimensiones específicas en número de píxeles (ej. 240 x 240).

**Extracción de características:** Se lleva a cabo en las capas convolucionales.

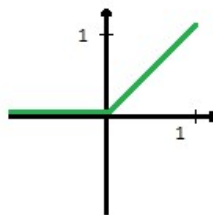
- Capas convolucionales: Son estas las capas donde se extrae la información de las imágenes aplicando una serie de filtros “kernels” que detectan patrones en



las imágenes, recorren la imagen original desplazándose horizontal y verticalmente en función de un parámetro “stride” que determina la cantidad de píxeles que se desplaza el filtro en cada paso. Para cada desplazamiento del filtro se calcula un valor que representa la respuesta del filtro para cada posición, generando un mapa de características (líneas, bordes, texturas, formas obteniendo información de patrones en las imágenes). Un stride de 1 hará que la imagen de salida tenga la misma dimensión, sin embargo, un stride mayor de 1 disminuirá la dimensión de salida. La salida de esta capa se conecta con la función de activación (Relu, Selu, etc).

Rectified Linear Unit (Relu): Mantiene los valores positivos inalterados y los valores negativos los convierte a 0, obteniéndose únicamente las partes con información.

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



**Figura 1. Función de activación ReLU**

- Capa de normalización: Escala los valores obtenidos en la capa anterior, permitiendo acelerar el entrenamiento y estabilizar el proceso de aprendizaje.
- Capa de agrupamiento “pooling layer”: Reduce la dimensionalidad de los mapas de características. Para ello lo divide en parches y selecciona el valor máximo (Max-pooling) o el valor promedio (Average-pooling), devolviendo un mapa con menos datos que mantienen las características más importantes.
- Capa de Dropout: Esta capa desactiva aleatoriamente algunas conexiones de la red, obligando al modelo a no depender de patrones específicos en los datos de entrenamiento.

### **Clasificación:**

- Capa completamente conectada: Conecta la salida de todas las neuronas de la capa anterior con las de la capa siguiente.
- Capa SoftMax: Convierte las salidas de la red neuronal en probabilidades para cada clase en el rango  $[0, 1]$  sumando entre todas las clases la unidad.
- Capa de clasificación: Última capa, tendrá el mismo número de neuronas que las clases del problema de clasificación. La predicción final será la que corresponda a la clase con mayor probabilidad.

### 3. Metodología

#### 3.1. Bases de datos MR

- 1 Brain MRI Images for Tumor Brain Tumor Detection. [35]

Base de datos con imágenes con tumor e imágenes de cerebros sanos.

- 2 Brain Tumor Classification (MRI Scans). [36]

Base de datos que contiene imágenes de cerebros sanos, Meningiomas, tumores de la glándula pituitaria y Gliomas, con más de 65.700 descargas.

- 3 UPENN-GBM[37,38]: Contiene imágenes de muestras de tejido, resonancias magnéticas multiparamétricas de pacientes con Glioblastoma Multiple (GBM), datos radiómicos extraídos de las mismas. Estos datos provienen del sistema de salud de la universidad de Pensilvania, datos de Mutación en IDH1 y de la presencia de O6-Metilguanina-DNA Metiltransferasa.

La base de datos contiene escáneres de RM (Formato Nifti), pre-procesadas y máscaras de segmentación. Los escáneres incluidos en esta base de datos tienen origen y protocolos de adquisición heterogéneos. Se realizó el mismo pre-procesado a todos los volúmenes para que las dimensiones y los tamaños de vóxeles fuesen uniformes a lo largo del estudio. Protocolo (BRATS).

Dentro de las imágenes de RM hay imágenes de segmentación automática, otro conjunto a partir del cual se ha realizado la segmentación automática, imágenes DSC y DTI, imágenes estructurales con los 4 tipos de contrastes con cráneo y las imágenes con los 4 tipos de contrastes, pero sin cráneo.

En este estudio se va a emplear el conjunto de imágenes del conjunto sin cráneo.

- 4 CPTAC-GBM[39,40]: Colección Glioblastoma Multiforme del Consorcio de Análisis Clínico Proteómico de Tumores del instituto nacional del cáncer de EEUU: Contiene imágenes de MR (Formato DICOM), escáneres de tomografía computarizada (CT) y tres bases de datos asociadas de datos proteómicos, genómicos y metabolómicos.
- 5 Rembrandt[41]: Base de datos de Gliomas de alto y bajo grado en forma de espectros de RM en formato DICOM. Además, contiene datos clínicos, clasificación del grado tumoral (II, III, IV) y seguimiento.

### 3.2. Preprocesado

Las imágenes deben estar en el formato adecuado para ser procesadas[42]: Para que las imágenes puedan ser procesadas por modelos se deben introducir con dimensiones y resolución similares, esto variará en función de si analizamos una imagen tridimensional o imágenes bidimensionales.

Si vamos a trabajar con imágenes bidimensionales a partir de una imagen en 3D primero debemos extraer los cortes. Es necesario ajustar las dimensiones de las imágenes al tamaño predefinido por el método de procesamiento de imagen que se vaya a emplear. En 2D el tamaño de las imágenes serán generalmente 224 x 224 o 240 x 240.

El ruido en la imagen puede deberse a una causa humana como el movimiento del paciente durante la realización de la prueba, problemas técnicos de la máquina o interferencias. Para corregirlo se pueden emplear diferentes estrategias, como filtros de mediana, de difusión anisotrópica, filtro de Wiener, de Wavelet, etc. (p.ej. N4ITK).

Desenmascarado: Todas las imágenes fueron obtenidas una vez desenmascaradas. Para llevar a cabo este proceso se empleó la herramienta ‘mri\_deface’ [43]

Extracción del cerebro: Para analizar exclusivamente el cerebro es necesario que eliminar el cráneo de la imagen. La base de datos de UPENN proporciona los volúmenes una vez extraídos el cráneo, para ello se ha empleado un enfoque basado en Deep Learning llamadao BrainMaGe. [44]

Normalización: La intensidad de las imágenes puede variar. Para garantizar que las intensidades tengan un rango de intensidad similar se emplean los métodos de normalización:

- Min-max o escalado a rango fijo (para imágenes de 8 bits por canal), conserva la relación entre los valores originales, pero es sensible a valores atípicos. Muy empleado en CNNs.

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- Z-score (Estandarización de los datos para que tengan media = 0 y desviación estándar = 1), muy útil para variables en diferentes escalas.

$$Z = \frac{x - \mu}{\sigma}$$

Recorte “Cropping”: Técnica que permite seleccionar una región específica, en este caso seleccionaremos únicamente la parte del cráneo, eliminando los bordes negros de la imagen, que no aportan información.

Aumento del Dataset: Las bases de datos públicas de RM no cuentan con muchos pacientes, lo que limita el entrenamiento de los modelos, las redes neuronales necesitan de muchos datos de entrada para tener un buen rendimiento y que el modelo generalice. Para solucionar esta limitación se puede incrementar el número de imágenes aplicando distintas transformaciones (Rotaciones, translaciones, escalado, reorientación vertical/horizontal, ajuste de brillo y contraste) [45].

Transformación a tensores: Estos son estructuras de datos que permiten realizar operaciones matemáticas en múltiples dimensiones. Los tensores organizan la información en dimensiones específicas y permite a las GPUs y TPUs trabajar con conjuntos de imágenes (batches), aplicar técnicas como Dropout o batch normalization, necesarias para mejorar el rendimiento.

Los tensores pueden ser desde 1 escalar, un vector (1D), una matriz (2D), imágenes RGB (3D) o lotes (batches) de imágenes RGB en (4D).

**One hot encoding:** Es un tipo de codificación que transforma los datos categóricos a binarios, donde cada categoría se convierte a un vector donde todos los valores son 0, excepto el que contiene un 1 que equivaldrá a una categoría diferente para cada posición.

**Batch Normalization:** Estabiliza el proceso de aprendizaje normalizando los datos de salida de una capa antes de introducirse en la siguiente[46]. Sirve para reducir el sobreentrenamiento al estandarizar los vectores antes de ser la entrada a la siguiente capa densa eliminando el ruido y permite emplear menores tasas de salida “Dropout”.

### 3.3. Selección del modelo

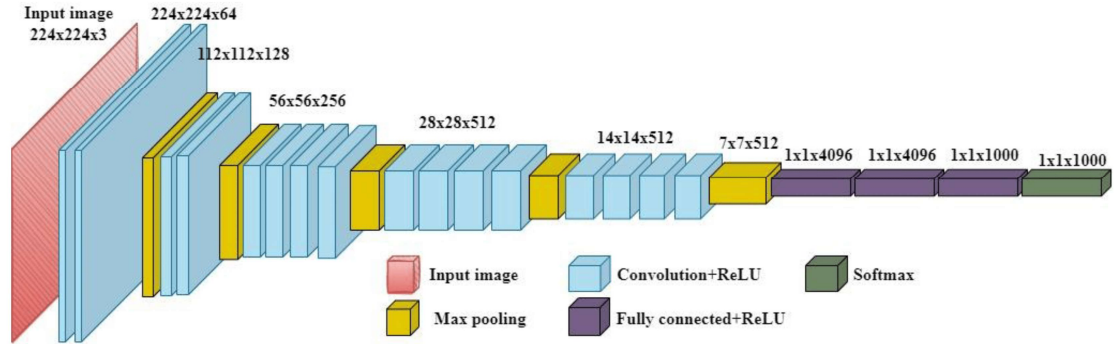
**Imagenet:** Conjunto de imágenes de entrenamiento que contiene 14 millones de imágenes a color de 1000 categorías genéricas con el que se entrenan modelos como (VGG16, VGG19, Xception, MobileNetV2).

**Xception:** Introduce convoluciones separables en profundidad. Basado en la arquitectura Inception, reduce significativamente el número de parámetros y el tiempo de computación. 2017.

**VGG19:** Es una red neuronal convolucional profunda diseñada para tareas de clasificación y reconocimiento de imágenes[47]. La estructura de VGG19 contiene 19 capas profundas, de las cuales 16 son convolucionales con tamaño de kernel 3 x 3, estas componen 5 bloques separados por capas de Max-Pooling con tamaño de kernel 2 x 2 y un stride de 2. La última capa de pooling tiene dimensión de 7 x 7 y 512 filtros. Las capas completamente conectadas constan de dos capas con 4096 neuronas y función de activación Relu y una capa final de salida SoftMax con las neuronas correspondientes a las clases del problema.

Nº bloque	Nº capas convolucionales	Filtros	Dimensión de salida
1	2 + Max-pooling	64	224 x 224
2	2 + Max-pooling	128	112 x 112
3	4 + Max-pooling	256	56 x 56
4	4 + Max-pooling	512	28 x 28
5	4	512	14 x 14
6	Max-pooling	512	7x7

**Tabla 1. Arquitectura de bloques convolucionales VGG19.**



**Figura 2. Estructura de modelo VGG19[48].**

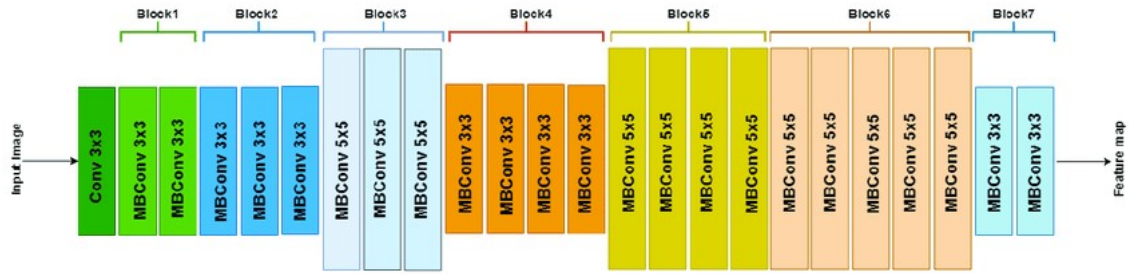
**EfficientNet:** Es una familia de arquitecturas de redes neuronales, existen diferentes versiones, desde la estructura básica (B0) hasta la más potente (B7). La innovación que presenta EfficientNet es que mediante escalado compuesto ajusta uniformemente la anchura, profundidad y resolución para maximizar el rendimiento, permitiendo muy buena relación entre precisión y eficiencia.

La estructura de EfficientNet se caracteriza por estar formado por un tipo de bloques convolucionales “Mobile Inverted Bottleneck Convolution” (MBConv), una capa de agrupamiento global y una capa completamente conectada. Cada bloque tiene una configuración diferente. Tabla 1.

Tiene la mejor relación entre exactitud y número de parámetros comparándolo con el resto de modelos pre entrenados en “ImageNet”. [49] [50].

Bloque	Nº Capas	Tamaño de Kernel	Filtros
1	1	3x3	16
2	2	3x3	24
3	3	5x5	40
4	4	3x3	80
5	4	5x5	112
6	5	5x5	192
7	2	3x3	320

**Tabla 2. Arquitectura de bloques MBConv.**



**Figura 3. Esquema estructura de bloques EfficienNetB1[51].**

**MobileNet:** Diseñado para dispositivos de baja potencia. MobilenetV1, MobilenetV2 y MobilenetV3.

### 3.4. Métricas de evaluación:

El rendimiento del modelo se mide mediante una serie de métricas que sirven para valorar su desempeño. Precisión, sensibilidad, especificidad, exactitud, pérdida, “f1-score”. que se obtienen de la matriz de confusión a partir de la relación entre (VP, VN, FP, FN). Otra medida que se emplea habitualmente es el Area bajo la curva de la representación de la curva “Receiving operating curve” (ROC).

VP = Verdadero Positivo

VN= Verdadero Negativo

FP = Falso Positivo

FN = Falso Negativo

- Exactitud “accuracy”: Mide la relación de aciertos entre casos totales.

$$A = \frac{VP + VN}{VP + VN + FP + FN}$$

- Precisión: Tasa de predicción de positivos. Mide la relación de verdaderos positivos entre todos los que se han clasificados como positivos, correcta o incorrectamente.

$$P = \frac{VP}{VP + FP}$$

- Sensibilidad (recall): Mide la tasa de verdaderos positivos entre los que son realmente positivos. Ratio de verdaderos positivos (TPR).

$$S = \frac{VP}{VP + FN}$$

- Especificidad: Mide la tasa de verdaderos negativos entre los casos que son realmente negativos. Ratio de falsos negativos (FPR).

$$FPR = \frac{VN}{VN + FP}$$

- F1-score: Es una medida que permite conocer el rendimiento del modelo en función del “recall” y la precisión.

$$F1 = 2 * \left( \frac{P * S}{P + S} \right)$$

Hay muchos factores que pueden hacer que un modelo funcione de manera óptima, las redes convolucionales son complicadas y tienen el problema de “caja negra” que no permite conocer exactamente que está valorando el modelo de manera intuitiva.

### 3.6. Optimización de hiper parámetros

Las CNNs tienen estructuras complejas y es difícil predecir la mejor combinación de hiper parámetros para resolver el problema de clasificación, la obtención de resultados exitosos depende en gran medida de la elección de los hiper parámetros.

El resumen del modelo muestra cuantos parámetros tiene la red y cuantos de ellos son entrenables. Estos parámetros se pueden clasificar en:



- **Parámetros arquitectónicos:** Los que pertenecen a las capas de agrupamiento, capas convolucionales, el número de filtros, tamaño de los filtros, la función de activación.
- **Hiper parámetros:** Tamaño de lote, tasa de aprendizaje, número de etapas, tamaño de k en validación cruzada, valor de dropout, elección del optimizador de la función de pérdida.

**Optimización Bayesiana:** Método para ajustar los hiper parámetros (número de capas, la tasa de aprendizaje, el tamaño del lote, épocas, el tipo de optimizador del modelo), selecciona las combinaciones de hiper parámetros en función de un modelo probabilístico que actualiza con cada evaluación, en lugar de probar todas las combinaciones posibles.

**Algoritmo de búsqueda de cuadrícula (Grid Search):** Prueba todas las combinaciones de hiperparámetros que se hayan definido previamente y selecciona la combinación que maximice el rendimiento del modelo.

**Sobre-entrenamiento (Overfitting):** Ocurre cuando un modelo de aprendizaje automático se ajusta demasiado bien a reconocer las características del grupo de entrenamiento, llegando a memorizar los datos de entrenamiento y perdiendo la capacidad de generalizar. Errores de entrenamiento muy bajos y errores de validación muy altos son característicos de este suceso.

Para tratar de prevenirlo se pueden emplear técnicas como normalización de los datos, aumento de datos, uso de la técnica de validación cruzada, emplear “dropout” en el entrenamiento, simplificar el modelo, detener el modelo cuando el error del conjunto de validación comience a aumentar.

## 4. Resultados

El código empleado para la extracción y clasificación de las imágenes, recorte de áreas con fondo negro ajustando la imagen a los bordes del cráneo, así como la construcción de los diferentes modelos está disponible en un repositorio de GitHub, disponible en el apartado Anexos.

Los modelos de clasificación empleados han sido implementados mediante el software Python, haciendo uso de la librería “Keras” de “TensorFlow”, para ejecutar los modelos empleamos la herramienta Jupiter Notebook de Anaconda.

Todas las etapas del trabajo, desde la descarga de los datos, sus transformaciones, clasificación, entrenamiento de los modelos, evaluación de los mismos y la maquetaación del trabajo se ha llevado a cabo en un ordenador HP, procesador Intel® Core(TM) Ultra7 155H 3.80 GHz, 32 GB de memoria RAM, sin acceso a GPU. La versión de Python empleada fue Python 3.10.13.

La extracción y clasificación de las imágenes se realizó mediante diferentes scripts en lenguaje “Python” según el formato de origen (DICOM o NifTi) de los volúmenes en 3D. (Anexo X).

Una vez extraídas las imágenes, se llevó a cabo el entrenamiento de algunos modelos para resolver varios problemas de clasificación (Clasificación de subtipos, diferenciación de grado tumoral y estimación de supervivencia). Se aplicaron sobre imágenes de resonancia magnética en 2D en tonos grises.

Se emplean modelos pre entrenados en el conjunto de imágenes “imagenet”, las capas de extracción de características se congelan durante el entrenamiento, añadiendo una capa de agrupamiento “max-pooling” y una serie de capas densas y de normalización conectadas a la salida del modelo pre-entrenado y finalizando con una capa de clasificación.

Estructura base del modelo:

- Capa densa de 2560 neuronas y función de activación “relu”.
- Capa de normalización por lotes.
- Capa densa de 1280 neuronas y función de activación “relu”.

- Capa de normalización por lotes.
- Capa densa de 1024 neuronas y función de activación “relu”.
- Capa densa de (n) neuronas y función de activación softmax.

Para el modelo de clasificación múltiple se empleará el optimizador Adam y la función de pérdida “Categorical crossentropy” y para el modelo de clasificación binaria la función de pérdida “Sparse-Categorical crossentropy”.

## 4.2. Preprocesado de las imágenes

Es importante llevar a cabo un tratamiento a las imágenes antes de introducirlas al modelo de CNN para el entrenamiento. Es importante que tengan el mismo formato, misma escala de valores y en caso de las imágenes unos valores de intensidad en la misma escala.

Normalización: Se aplicó el método (Min-Max). Escalado de los valores de intensidad de los píxeles a un rango entre 0 y 1. Este valor se multiplica por 255, el valor máximo de la intensidad para un píxel.

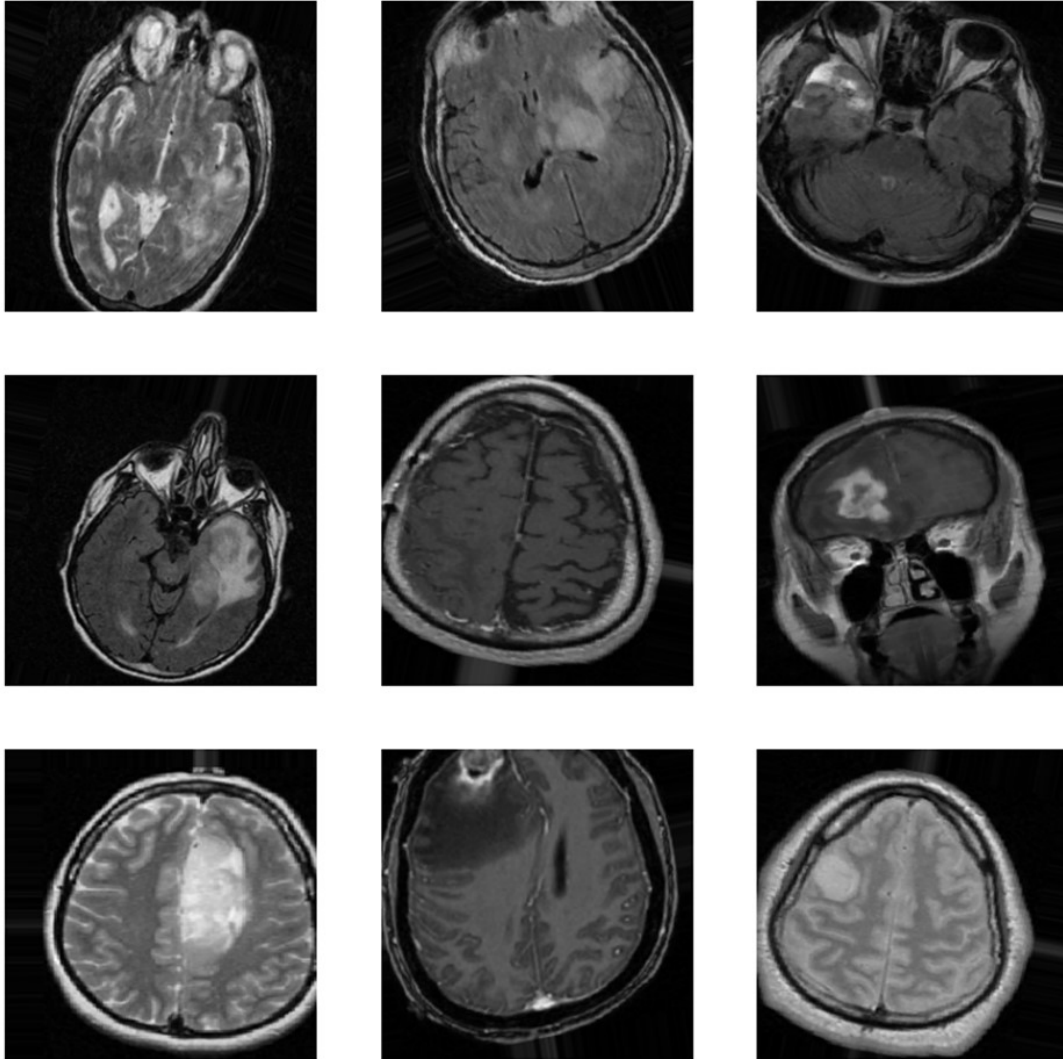
Se pasa el conjunto de imágenes por una máscara que recorta los bordes negros de las resonancias, estos no aportan información. Dentro de la escala de grises (0-255) consideramos un píxel “no negro” a partir del valor 20 de intensidad, consiguiendo un mejor ajuste de las imágenes.

Tratamiento de imágenes: Empleando la librería “ImageDataGenerator” se aplican transformaciones. Esta librería permite realizar ajustes sobre cada lote en tiempo real. Para introducir las imágenes a las redes neuronales pre-entrenadas el formato de entrada de las imágenes debe ser “rgb”.

- Redimensiona la imagen a 240 x 240 píxeles.
- Transforma de escala de grises a formato de 3 canales “rgb”.
- Establece un tamaño de lote.

También permite el aumento del número de imágenes empleadas en el entrenamiento y validación aplicando una serie de transformaciones sin necesidad de guardar las imágenes transformadas en el equipo.

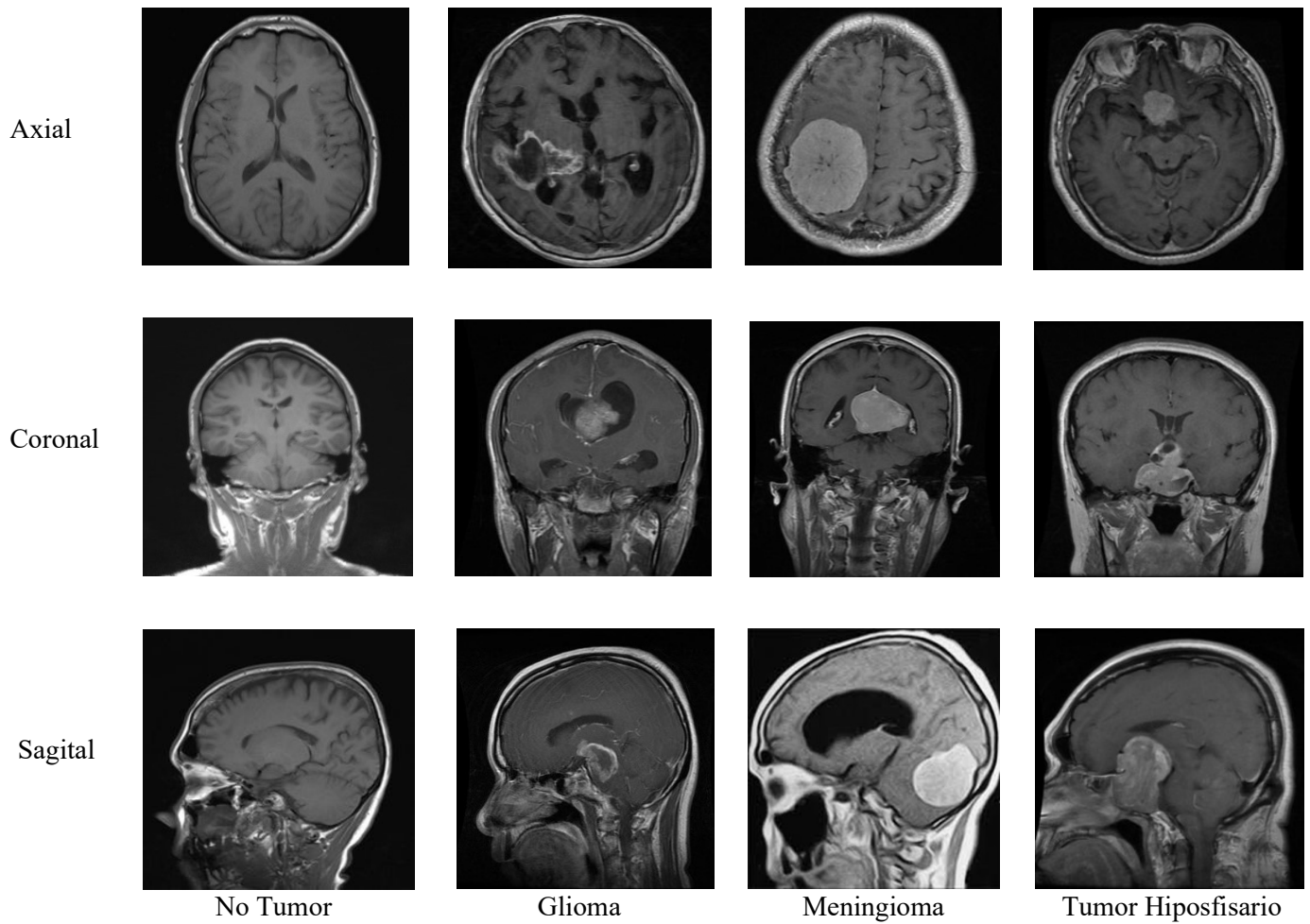
- Rotaciones en un rango de 20°.
- Modificación del ancho y alto de la imagen de 10%.
- Inversiones horizontales.
- Zoom en un rango del 10%.



*Figura 4. Muestra de 9 imágenes aleatorias aumentadas aplicando transformaciones.*

#### 4.3. Clasificación de subtipos: Meningioma, Glioma, Tumor hipofisario y no tumor

Imágenes provenientes de base de datos **Brain Tumor Classification (MRI)**[36] de la plataforma Kaggle que ha sido empleada en multitud de artículos. Contiene imágenes de distintos subtipos de tumor cerebral (Gliomas, Meningiomas y tumores Hipofisarios) y de pacientes sanos. Figura 2.



**Figura 5. Imágenes de RM de pacientes sanos y 3 subtipos.**

Tras realizar pruebas con una serie de modelos y diferentes parámetros, se escogieron los modelos con mejores métricas de rendimiento. Tabla 3. A los modelos VGG19 y EfficientNetB1 se les aplicó el optimizador de hiper parámetros (Búsqueda en cuadrícula “Grid Search”). Se llevó a cabo el entrenamiento con las distintas combinaciones de hiper parámetros previamente seleccionados para encontrar la combinación con mejor rendimiento.

LR = Learning Rate o Tasa de aprendizaje. [0.00001, 0.00005, 0.0001].

Batch Size = Tamaño de lote [16, 32, 64].

Epocs = Etapas de aprendizaje [10, 15].

	<b>Modelo</b>	<b>Etapas</b>	<b>LR</b>	<b>Batch</b>	<b>accuracy</b>	<b>val_accuracy</b>	<b>loss</b>	<b>val_loss</b>
<b>1</b>	VEGG19	8	0.0001	32	0.88	0.83	0.31	0.42
<b>2</b>	VEGG19	8	0.001	32	0.95	0.82	0.12	0.35
<b>3</b>	EfficientNetB1	16	0.0001	32	0.95	0.90	0.12	0.32
<b>4</b>	EfficientNetB1	8	0.001	32	0.89	0.78	0.29	0.62
<b>5</b>	Xception	8	0.0001	32	0.80	0.77	0.49	0.55
<b>6</b>	Xception	16	0.001	16	0.75	0.70	0.64	0.75
<b>7</b>	MobileNetV2	8	0.0001	32	0.93	0.86	0.19	0.35
<b>8</b>	MobileNetV2	16	0.0001	16	0.95	0.86	0.13	0.44

**Tabla 3. Medidas de rendimiento para varios modelos.**

Tras realizar la optimización de hiper parámetros para los modelos VGG19 y EfficientNetB1. Tabla 4. Se observó un ligero aumento del rendimiento para VGG19, en cambio con EfficientNetB1 fue muy similar.

Para tratar de alcanzar el modelo con mejor desempeño se llevó a cabo el entrenamiento para 30 etapas con los hiper parámetros optimizados para ambos modelos, se llevó a cabo el entrenamiento sin encontrar mejora en las métricas de entrenamiento y validación. MobileNetV2 tuvo resultados prometedores en las pruebas iniciales, no obstante, descendieron sus valores de sus métricas considerablemente. Tabla 5.

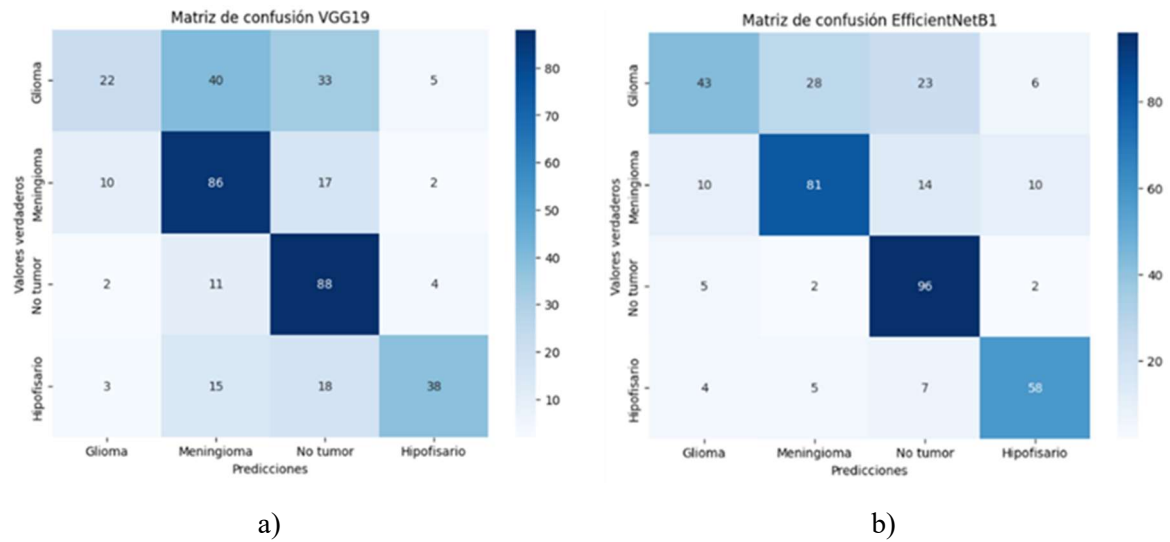
Se aplicó “dropout” en los modelos que aparecen en las primeras pruebas de modelos (Tabla 3), obteniendo rendimientos peores a los mostrados en la tabla. Repetimos la prueba con el entrenamiento en 30 etapas y el resultado no mejoró. Tabla 5.

	<b>Modelo</b>	<b>Etapas</b>	<b>LR</b>	<b>Batch</b>	<b>accuracy</b>	<b>val_accuracy</b>	<b>loss</b>	<b>val_loss</b>
<b>1</b>	VEGG19	15	0.0001	32	0.97	0.89	0.09	0.33
<b>2</b>	EfficientNetB1	15	0.0001	64	0.93	0.89	0.19	0.29

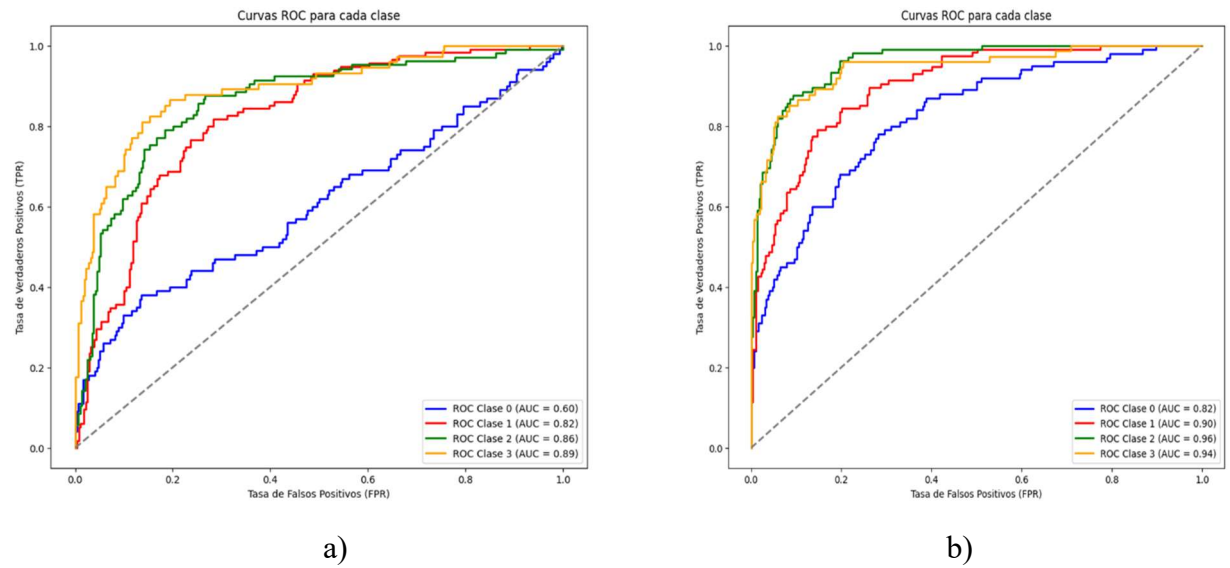
**Tabla 4. Medidas de rendimiento con hiperparámetros optimizados.**

	<b>Modelo</b>	<b>Etapas</b>	<b>LR</b>	<b>Batch</b>	<b>accuracy</b>	<b>val_accuracy</b>	<b>loss</b>	<b>val_loss</b>
<b>1</b>	MobileNetV2	30	0.0001	16	0.79	0.75	0.58	0.77
<b>2</b>	EfficientNetB1	30	0.0001	64	0.96	0.90	0.09	0.33
<b>3</b>	EfficientNetB1*	30	0.0001	64	0.85	0.84	0.39	0.30
<b>4</b>	VGG19	30	0.0001	32	0.98	0.87	0.47	0.49

**Tabla 5. Medidas de rendimiento. (\*) Dropout después de capa densa (0.5, 0.5, 0.3).**



**Figura 6. Matrices de confusión. Clasificación subtipos a) VGG19 y b) EfficientNetB1 (30 etapas).**



**Figura 7. Curvas ROC. a) VGG19 y b) EfficientNetB1. i) Clase 1 = Glioma, ii) Clase 2 = Meningioma, iii) Clase 3 = No Tumor, iv) Clase 4 = Tumor Hipofisario.**

Modelo	Etapas	LR	Batch	accuracy	val_accuracy	loss	val_loss
EfficientNetB1	60	0.0001	64	0.98	0.89	0.06	0.39

**Tabla 6. Métricas modelo EfficientNetB1 con hiperparámetros optimizados (60 etapas).**

Subtipo	Precisión	Recall	F1-score	Precisión	Recall	F1-score
Glioma	0.59	0.22	0.32	0.69	0.53	0.53
Meningioma	0.57	0.75	0.64	0.70	0.70	0.70
No tumor	0.56	0.84	0.67	0.69	0.91	0.78
T. Hipofisario	0.78	0.51	0.62	0.76	0.78	0.77
	VGG19			EfficientNetB1		

**Tabla 7. Precisión, sensibilidad y f1-score de los modelos VGG19 y EfficientNetB1.**

#### 4.4. Clasificación Glioma grado II, III, IV.

Para la clasificación de Gliomas en función del grado tumoral también se empleó la técnica de transferencia de aprendizaje, con los modelos pre entrenados estudiados en el apartado anterior.

La base de datos Rembrandt cuenta con 127 pacientes a los que se les realizó un examen de resonancia magnética previo a la resección del tumor. 41 padecían Glioma de grado II, 25 de grado III y 43 de grado IV (GBM).

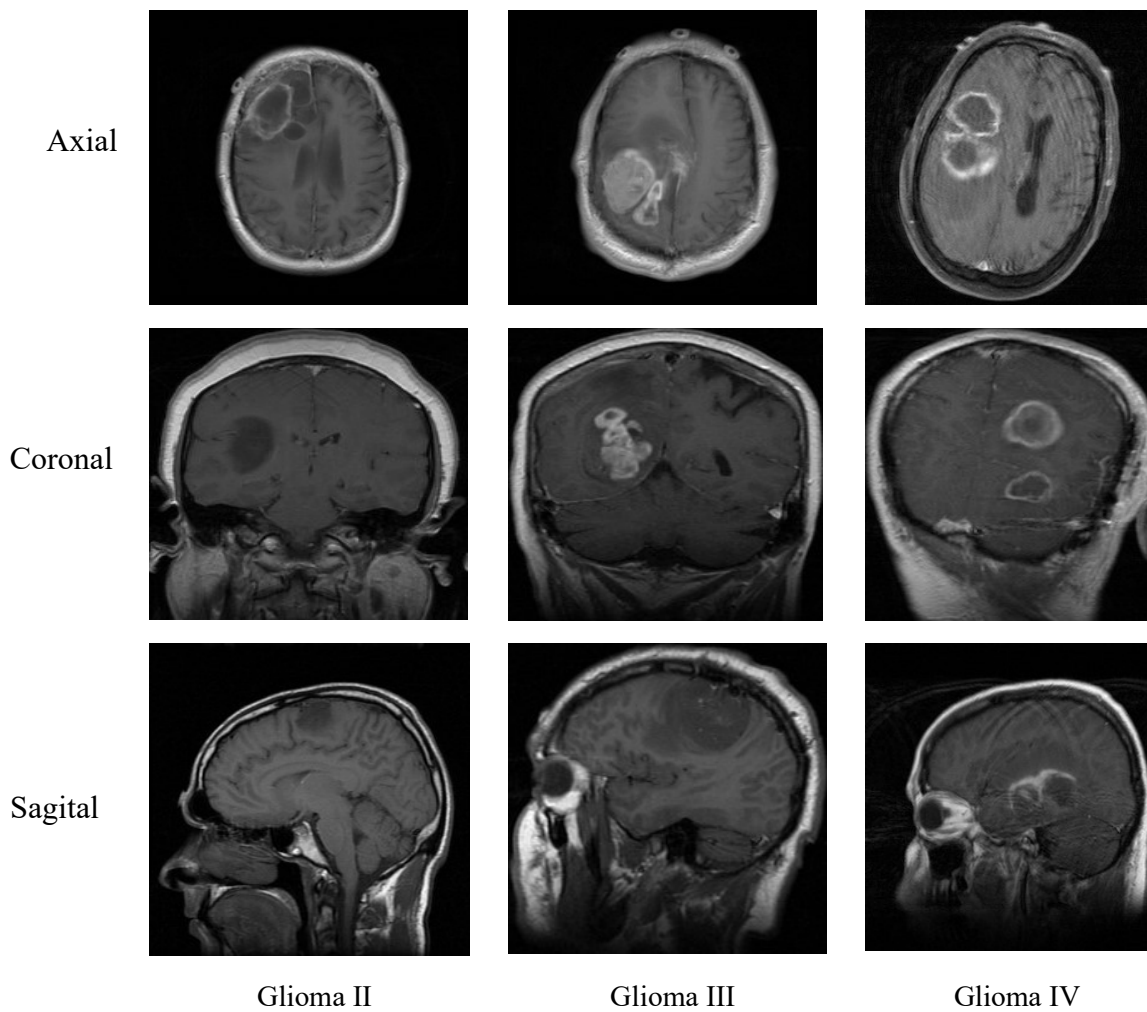
Sexo: Solo 79 pacientes tienen dato del sexo, siendo 36 Mujeres (45.6%) y 43 Hombres (54.4%).

Supervivencia: La supervivencia media para los tres grados de Glioma fueron 37,89 meses. Cuanto mayor fue el grado tumoral se observan menores tiempos de supervivencia.

- Glioma II: 58,7 meses
- Glioma III: 38,5 meses
- Glioma IV: 19,3 meses

En el caso del glioma de IV grado, también conocido como Glioblastoma multiforme observamos que la supervivencia media fue de aproximadamente un año y medio, coincidiendo con la literatura.





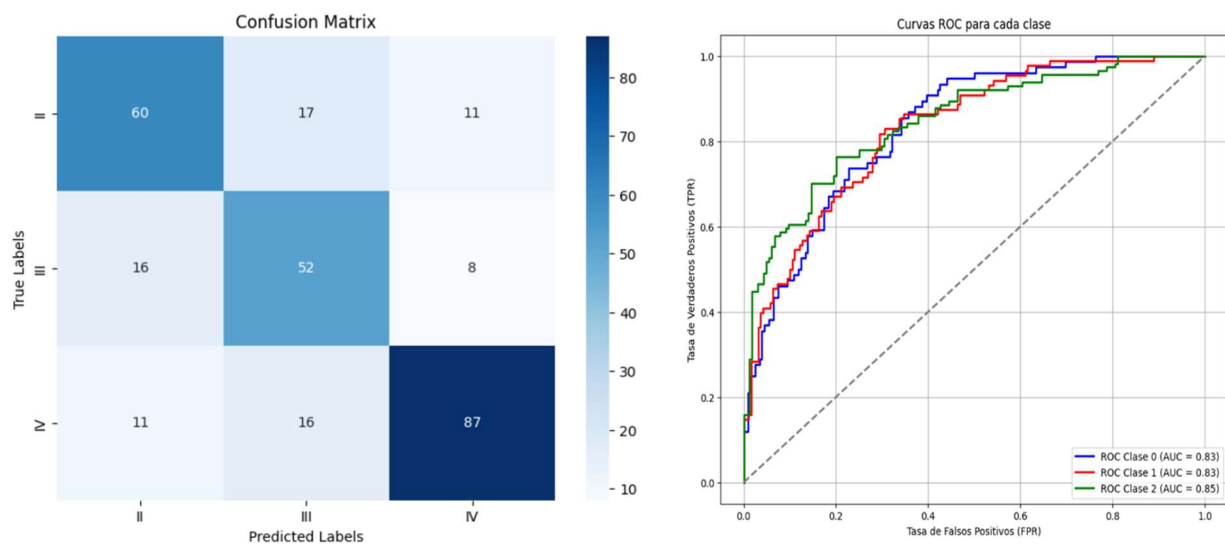
***Figura 8. Gliomas de grados (II, III, IV). Base de datos REMBRANDT.***

La extracción de los espectros de resonancia magnética en formato DICOM se convirtieron a imágenes en formato png mediante un script en lenguaje “Python”.

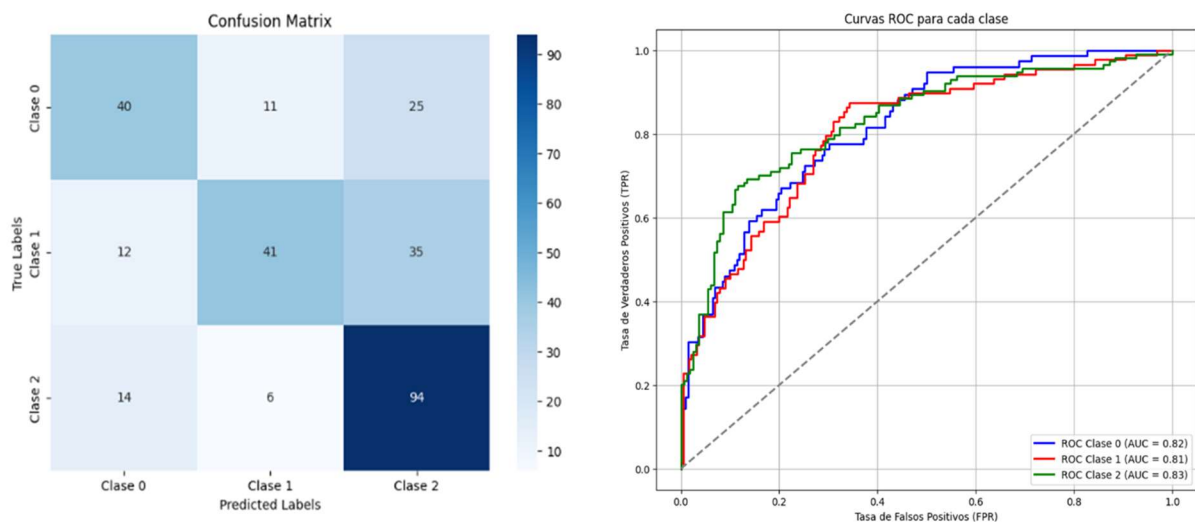
Al emplear imágenes que no han sido descargadas directamente, sino extraídas y seleccionadas heurísticamente de los volúmenes originales, se quería comprobar el rendimiento del modelo con las imágenes directamente extraídas y una vez normalizadas y recortadas con el modelo VGG19. Se observaron mejores resultados de exactitud de validación en los modelos sin normalización.

	Modelo	Etapas	LR	Batch	accuracy	val_accuracy	loss	val_loss
1	VEGG19	16	0.0001	32	0.98	0.67	0.088	0.93
2	VEGG19(+)	16	0.0005	32	0.93	0.65	0.23	0.85
3	VEGG19(*)	16	0.0001	32	0.94	0.71	0.16	0.76
4	VEGG19(+)(*)	16	0.0005	32	0.90	0.63	0.30	0.83

**Tabla 8. Modelos de clasificación del grado de Glioma con VGG19. (+) Recorte, (\*) Normalización.**



**Figura 9. Matriz de confusión y curva ROC para modelo sin normalizar y sin cropping.**



**Figura 10. Matriz de confusión y curva ROC de modelo con normalización y cropping.**

	<b>Modelo</b>	<b>Etapas</b>	<b>LR</b>	<b>Batch</b>	<b>accuracy</b>	<b>val_accuracy</b>	<b>loss</b>	<b>val_loss</b>
<b>1</b>	EfficientNetB1	16	0.0001	32	0.98	0.67	0.088	0.93
<b>2</b>	EfficientNetB1	16	0.0005	32	0.93	0.65	0.23	0.85

**Tabla 9. Clasificación grado Glioma con EfficientNetB1.**

A los modelos VGG19 y EfficientNetB1 se les aplicó el optimizador de hiper parámetros (Búsqueda en cuadrícula “Grid Search”), donde se llevó a cabo el entrenamiento con las distintas combinaciones de hiper parámetros previamente seleccionados para encontrar la combinación con mejor rendimiento y los valores más altos para exactitud de validación fueron 0.68 y 0.73 respectivamente. La optimización (\*) mejoró la exactitud de validación para el modelo EfficientNetB1, sin embargo, los valores de precisión sensibilidad y f1-score fueron inferiores. Tabla 13.

LR = Learning Rate o Tasa de aprendizaje. [0.00005, 0.0001, 0.0005]

Batch Size = Tamaño de lote [16, 32, 64]

Epocs = Etapas de aprendizaje [10, 15]

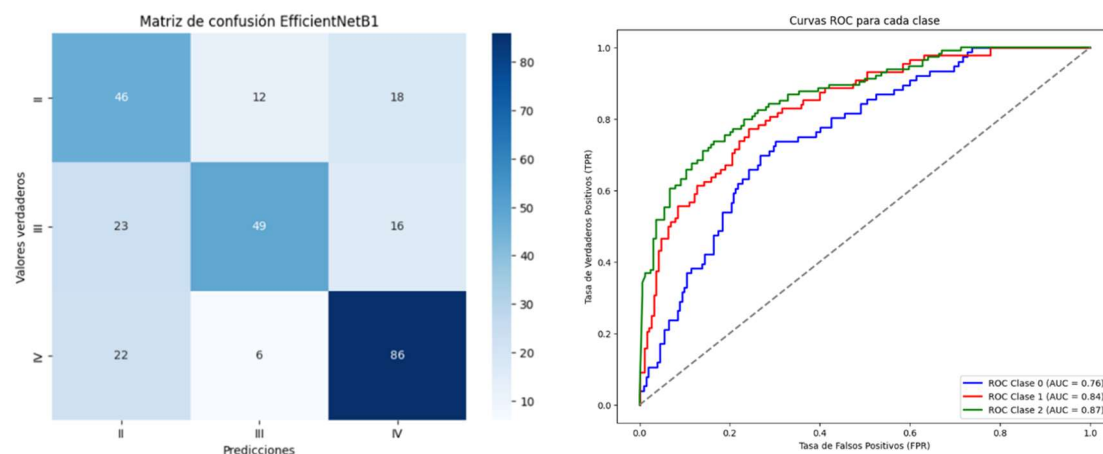
	<b>Modelo</b>	<b>Etapas</b>	<b>LR</b>	<b>Batch</b>	<b>accuracy</b>	<b>val_accuracy</b>	<b>loss</b>	<b>val_loss</b>
<b>1</b>	VEGG19	15	0.0005	16	0.93	0.68	0.26	0.73
<b>2</b>	EfficientNetB1	15	0.0005	64	0.82	0.73	0.45	0.87

**Tabla 10. Modelos con parámetros optimizados.**

Se realizó una prueba entrenando el modelo EfficientNetB1 durante 60 etapas, obteniendo menor exactitud de validación que con el mismo modelo en 15 etapas.

Glioma	Precisión	Recall	F1-score	Precisión	Recall	F1-score
II	0.69	0.68	0.69	0.61	0.53	0.56
III	0.61	0.68	0.65	0.71	0.47	0.56
IV	0.82	0.76	0.79	0.61	0.82	0.70
	VGG19			VGG19 (+)(*)		

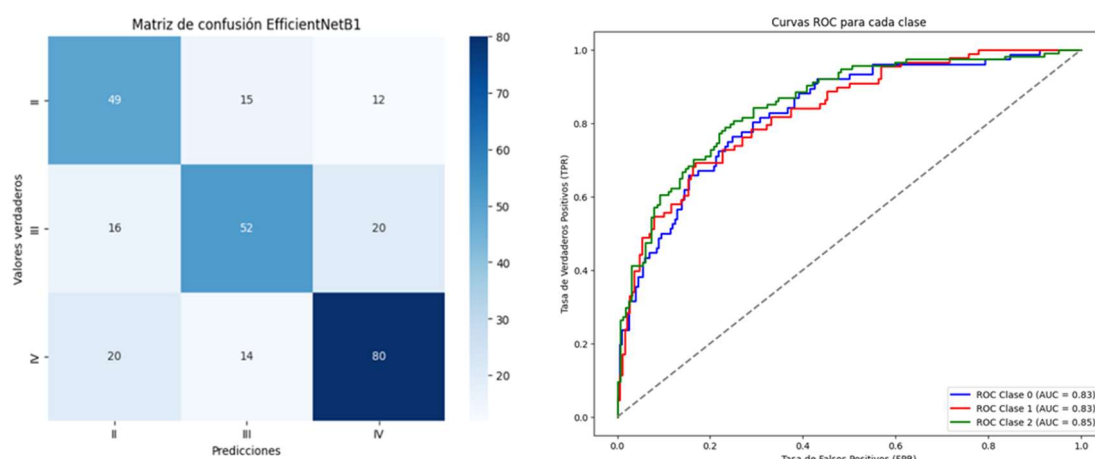
**Tabla 11. Precisión, sensibilidad y f1-score con VGG19. (+) Recorte, (\*) Normalización.**



**Figura 11. Matriz de confusión y curvas ROC. Modelo EfficientNet (\*) 15 etapas.**

	Modelo	Etapas	LR	Batch	accuracy	val_accuracy	loss	val_loss
1	EfficientNetB1(*)	15	0.0005	64	0.92	0.71	0.23	1.65
2	EfficientNetB1(*)	60	0.0005	64	0.95	0.70	0.15	0.70

**Tabla 12. Modelo EfficientNetB1 (\*) 60 etapas.**



**Figura 12. Gráficos modelo EfficientNetB1(\*) optimizado 60 etapas.**

Glioma	Precisión	Recall	F1-score	Precisión	Recall	F1-score
II	0.51	0.61	0.55	0.58	0.64	0.61
III	0.73	0.56	0.63	0.64	0.59	0.62
IV	0.72	0.75	0.74	0.71	0.70	0.71
	EfficientNetB1 (*) 15 etapas			EfficientNetB1 (*) 60 etapas		

**Tabla 13. Precisión, sensibilidad y f1-score. Modelo EfficientNet con parámetros optimizados (\*) 15 y 60 etapas**

Otra clasificación habitual del Glioma es en Glioma de bajo grado (LLG) o Glioma de alto grado (HGG). Según la OMS los Gliomas de grado II y III corresponden a LLG (66) y los de grado IV a HGG (43).

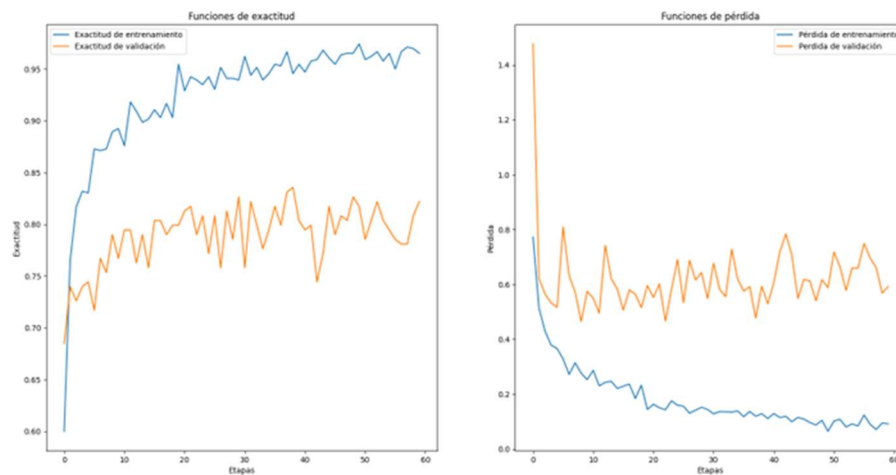
Para esta nueva clasificación se emplearon los mismos modelos empleados anteriormente, ahora se trata de un problema de clasificación binaria, se sustituye la función de pérdida por “sparse categorical crossentropy”.

Se han probado numerosas combinaciones de parámetros y el máximo rendimiento se ha conseguido con el modelo de 60 etapas mostrado a continuación.

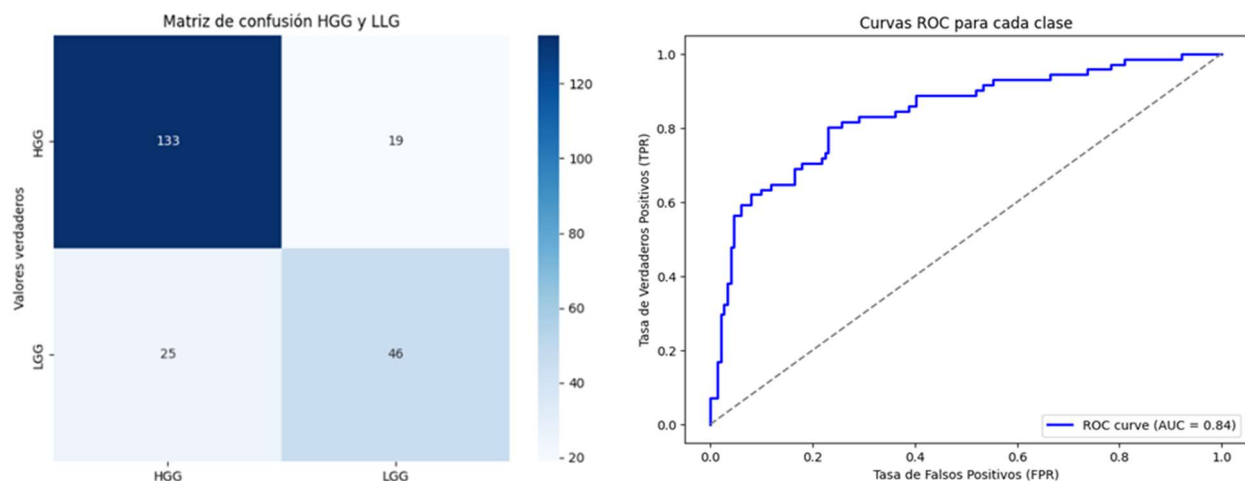
	Precisión	Recall	F1-score
HGG	0.84	0.88	0.86
LGG	0.71	0.65	0.68

**Tabla 14. Precisión, sensibilidad y f1-score en HGG y LGG.**

Tras realizar varias pruebas se obtuvo el mejor rendimiento con el modelo entrenado durante 60 etapas, aunque se puede observar que la gráfica de exactitud aumenta durante el entrenamiento y la de pérdida disminuye durante la validación hasta que empieza a fluctuar arriba y abajo sin estabilizarse. Figura 13.



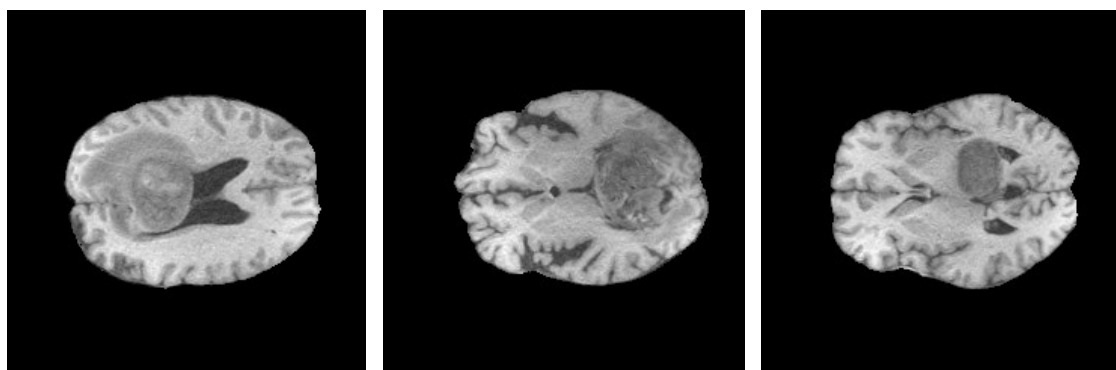
**Figura 13. Exactitud y pérdida en entrenamiento y validación EfficientNetB1 (60 etapas).**



**Figura 14. Matriz de confusión LGG y HGG. EfficientNetB1 60 etapas.**

## 4.5. Predicción de supervivencia en GBM.

Después de convertir los volúmenes en formato NIfTi a imágenes en dos dimensiones, se empleó un script para organizarlas en carpetas en función del tipo de contraste aplicado (T1, T1-GD, T2 y T2-FLAIR) y el corte (Axial, Coronal o Sagital). Debido al inmenso número de imágenes generadas, se seleccionaron las imágenes correspondientes a un contraste (T1) y un corte específico (Axial).



**Figura 15. Imágenes de GBM extraídas de la base de datos UPENN de TCIA.**

Seleccionamos el contraste T1 y el corte Axial, donde hubo un total de 40.424 imágenes. Tras seleccionar de manera heurística varias imágenes de cada paciente donde se pudiese apreciar claramente la masa tumoral, aunque en algunos casos no fue posible identificarla (en estos casos no se seleccionó ninguna imagen).

Tras este proceso se obtuvieron 1713 imágenes de 588 pacientes, divididos en Supervivencia Global (SG) superior a la media (18 meses).

	<b>Imágenes</b>	<b>Pacientes totales</b>	<b>Pacientes con imágenes</b>
<b>SG</b>	1713	644	588
<b>SG alta</b>	537	199	190
<b>SG baja</b>	1116	445	378

**Tabla 15. Datos de supervivencia. Base de datos UPPEN.**

Se trató de obtener un método de clasificación binaria para predecir si la SG de los pacientes fue superior a la media, a partir de imágenes de resonancia magnética de los tumores cerebrales (GBM) extraídas manualmente a partir de los volúmenes en tres dimensiones.

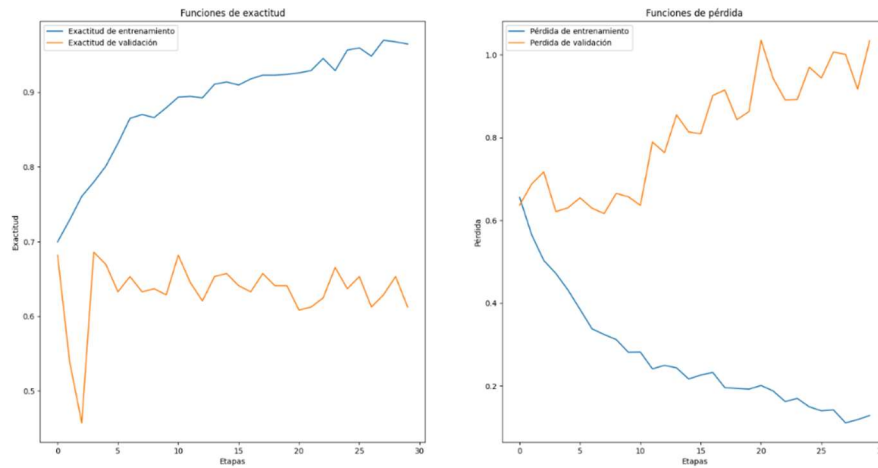
Empleamos los modelos VVG19 y EfficientNetB1 de manera similar a como se realizó con los modelos de apartados anteriores, congelando las primeras capas y entrenando las últimas con las imágenes de GBM para clasificarlas en función del pronóstico de supervivencia.

VGG19 ofreció resultados poco prometedores en sus métricas de entrenamiento, ya que la exactitud de validación no aumentó respecto de la inicial y la función de pérdida de validación aumenta conforme avanza el entrenamiento.

Modelo	Etapas	LR	Batch size	Accuracy	Val accuracy	Loss	Loss val
VGG19	30	0.0005	32	0.91	0.6	0.20	1.02

**Tabla 16. Métricas modelo VVG19 clasificación SG.**

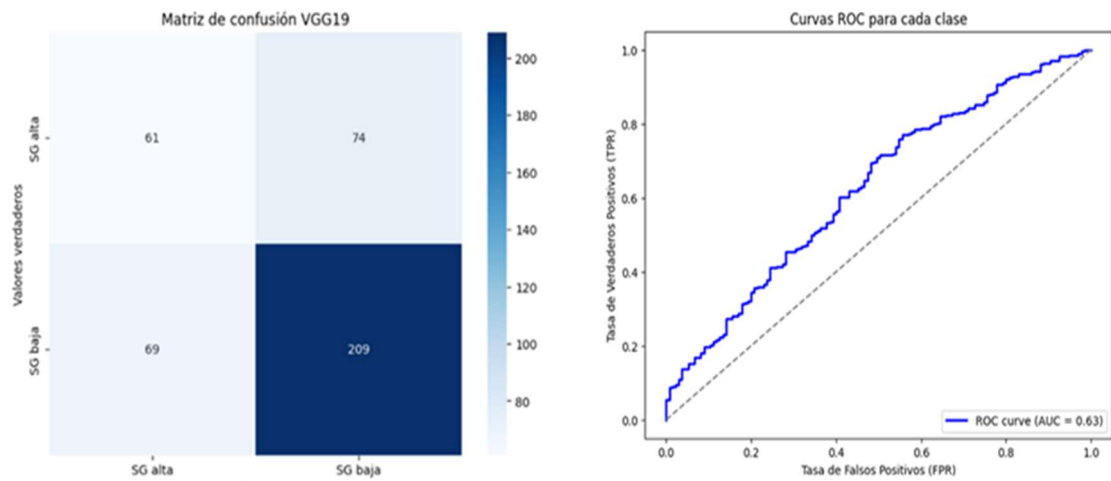
Las medidas de rendimiento de la clasificación binaria muestran que el modelo logra distinguir ligeramente entre los dos grupos, aunque la exactitud de validación no aumentó durante el entrenamiento.



**Figura 16. Exactitud y pérdida en entrenamiento y validación con VGG19 (SG).**

	Precisión	Recall	F1-score
SG alta	0.47	0.45	0.46
SG baja	0.74	0.75	0.75

**Tabla 17. Medidas de evaluación, clasificación SG.**



**Figura 17. Matriz de confusión y curva ROC. Predicción SG para GBM.**



## 5. Conclusiones y trabajo futuro

En base a los resultados obtenidos hemos podido comprobar la capacidad de modelos de CNN pre-entrenados para resolver tareas de clasificación multi clase. Hemos planteado diferentes problemas de clasificación y obtuvimos diferentes resultados para cada uno.

La clasificación en función de subtipos fue la que ofreció mejores resultados, aunque no alcanzaron a ser tan buenos como los encontrados en la bibliografía para el mismo conjunto de datos.

Esto se debe a las distintas metodologías aplicadas[46], distintas transformaciones al aumentar las imágenes, número de neuronas en la red y diferentes combinaciones de hiper parámetros.

En la clasificación del grado tumoral de Gliomas se obtuvieron resultados inferiores a los esperados. Se trata de un problema de clasificación ligeramente más complejo, ya que se trata del mismo tipo de tumor. Para poder clasificar el grado tumoral se basa en características más complejas y para ello necesita de un número grande de imágenes y de un modelo que sea capaz de extraer características complejas para alcanzar el máximo rendimiento.

Al realizarse la extracción de las imágenes a partir de los volúmenes de manera manual se seleccionó un número de imágenes inferior al necesario para obtener un rendimiento óptimo.

Para la clasificación de GBM en función de su supervivencia global mayor o menor a la media encontramos resultados esperables, ya que la SG es una situación demasiado compleja y con demasiada variabilidad como para poderse diferenciar claramente a partir de cortes en 2D de RM, sin embargo, los resultados indicaron una cierta relación, pese a la sencillez del modelo y la complejidad del problema.

El objetivo principal se cumplió, ya que los modelos planteados cumplen el propósito original, aunque los resultados son mejorables.

Los objetivos inicialmente planteados pretendían emplear las metodologías actuales para cubrir el problema de clasificación mediante imagen del cáncer de cerebro. Cuando se

fue profundizando en la metodología se encontraron dificultades para abordar el problema de clasificación. La carga computacional, el tiempo y el volumen de datos óptimo para obtener un rendimiento alto eran demasiado elevados para entrenar modelos profundos desde cero con las herramientas al alcance.

Los objetivos secundarios fueron ligeramente modificados, en vez de predecir la supervivencia global de los pacientes de cáncer de cerebro en general nos centramos en los pacientes de Glioblastoma multiforme. Al tener datos de GBM de otras bases de datos habría sido interesante utilizarlas como grupo de prueba o entrenar un modelo con imágenes de diferentes bases de datos, pero las imágenes de la base de datos UPENN habían sido procesadas sin cráneo, mientras que las de CPTAC y REMBRANDT presentaban cráneo.

Emplear IA para predecir la supervivencia de los pacientes con cáncer cerebral basándose en características extraídas de RM mediante redes neuronales convolucionales empleando características extraídas de las imágenes y datos clínicos. Debido a la escasez de datos clínicos disponibles para las imágenes decidimos centrarnos en los problemas de clasificación principales, aunque quedará como trabajo futuro.

Al realizar únicamente la clasificación de SG alta y SG baja basándonos en imágenes en 2D las comparaciones con otros estudios de SG en GBM son muy diferentes, ya que incluyen datos clínicos, genómicos, radiómicos y proteómicos. Nos habría gustado abordar este enfoque[52–55]. Lamentablemente el tiempo limitó el alcance de este trabajo.

No ha sido posible seguir la planificación al pie de la letra. El acceso a algunas de las bases de datos empleadas requiere de una solicitud para poder acceder a las imágenes, este proceso tarda algunos días.

Una vez se obtuvieron las imágenes y se extrajeron los datos clínicos de la base de datos CPTAC se descartó y se sustituyó por la base de datos REMBRANDT, lo que retrasó algunos días la obtención de los resultados.

La extracción de las imágenes de cerebro y la selección manual de los cortes que presentaban masa tumoral fue laborioso y requiso de un tiempo considerable de trabajo.

Como se comenta anteriormente hubo cambios en el planteamiento del trabajo, cambio en una de las bases de datos, optar por emplear modelos pre entrenados, trabajar con imágenes en 2D.

La variabilidad es un factor que afecta significativamente a la generalización de modelos de inteligencia artificial.

- Los diferentes equipos empleados, parámetros de adquisición o el tipo de contraste aplicado introducen variabilidad en las imágenes, que, junto con el ruido,
- En el estudio de GBM, pese a emplear un único contraste y corte se obtuvo un rendimiento bajo, en parte, debido a la heterogeneidad del GBM en cuanto a tamaño, forma y características internas, lo que lo hace más difícil de ser clasificado.
- Sus características internas (necrosis, edema circundante o la vascularización) son características complejas, difíciles de interpretar al analizar cortes en dos dimensiones, ya que se pierde información espacial al emplear modelos y datos en 2D, que aprenden patrones locales, siendo insuficientes para capturar patrones complejos.
- La estimación de la supervivencia global o la determinación de la presencia de una mutación en un gen son casos complejos, que precisan de mucha información específica, siendo un problema difícil de afrontar. La falta de patrones tridimensionales, un limitado número de imágenes,

Trabajo futuro: De haber contado con más tiempo se habría entrenado los modelos aplicando Fine Tunning, probando más combinaciones de hiper parámetros, seleccionando más imágenes y empleando métodos mas avanzados de extracción de características.

Otro posible enfoque para afrontar este problema sería extraer características de las imágenes, concatenando la salida con datos clínicos (edad, sexo, presencia de mutaciones) u otro tipo de datos (genómicos, radiómicos, proteómicos) o emplear un modelo para cada tipo de dato y combinarlos en una capa final creando un modelo integrado para clasificación.

## 6. Glosario

Slide: Desplazamiento del filtro de convolución en cada movimiento, si el slide es mayor de 1 se reduce la dimensión del mapa de características con respecto a la imagen de entrada.

Kernel: Se le llama así al filtro aplicado en operaciones de convolución.

Dropout: Una técnica que desactiva algunas neuronas durante el aprendizaje, ayudando a disminuir el sobre ajuste del modelo.

Batch Normalization: Normalización de la salida de una capa por lotes antes de pasar por la función de activación.

Rgb: Formato de 3 colores (Red, Green, Blue) para indicar los colores en las imágenes a color.

## 7. Bibliografía

1. Cancer Today [Internet]. [cited 2024 Dec 29]. Available from: <https://gco.iarc.who.int/today/>
2. Celik F, Celik K, Celik A. Enhancing brain tumor classification through ensemble attention mechanism. *Sci Rep*. 2024 Sep 27;14(1):22260.
3. Irmak E. Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework. *Iran J Sci Technol Trans Electr Eng*. 2021 Sep 1;45(3):1015–36.
4. Lee EH, Han M, Wright J, Kuwabara M, Mevorach J, Fu G, et al. An international study presenting a federated learning AI platform for pediatric brain tumors. *Nat Commun*. 2024 Sep 2;15(1):7615.
5. Akinyelu AA, Zaccagna F, Grist JT, Castelli M, Rundo L. Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey. *J IMAGING*. 2022 Aug;8(8):205.
6. Badža MM, Barjaktarović MČ. Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. *Appl Sci*. 2020 Jan;10(6):1999.
7. Khan MA, Ashraf I, Alhaisoni M, Damaševičius R, Scherer R, Rehman A, et al. Multimodal Brain Tumor Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists. *Diagnostics*. 2020 Aug;10(8):565.
8. Meningioma [Internet]. 2019 [cited 2025 Jan 13]. Available from: <https://www.cancer.gov/rare-brain-spine-tumor/espanol/tumores/meningioma>
9. Tratamiento de los tumores de hipófisis [Internet]. 2008 [cited 2025 Jan 13]. Available from: <https://www.cancer.gov/espanol/tipos/hipofisis/paciente/tratamiento-hipofisis-pdq>
10. Kumar S, Kumar D. Human brain tumor classification and segmentation using CNN. *Multimed Tools Appl*. 2023 Feb 1;82(5):7599–620.
11. Definition of glioma - NCI Dictionary of Cancer Terms - NCI [Internet]. 2011 [cited 2025 Jan 13]. Available from: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/glioma>
12. Stupp R, Taillibert S, Kanner A, Read W, Steinberg D, Lhermitte B, et al. Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. *JAMA*. 2017 Dec 19;318(23):2306–16.
13. Ahmed HS, Devaraj T, Singhvi M, Dasan TA, Ranganath P. Radio-anatomical evaluation of clinical and radiomic profile of multi-parametric magnetic resonance imaging of de novo glioblastoma multiforme. *J Egypt Natl Cancer Inst*. 2024 Apr 22;36(1):13.

14. Resonancia magnética - Temas especiales [Internet]. Manual Merck versión para profesionales. [cited 2025 Jan 13]. Available from: <https://www.merckmanuals.com/es-us/professional/temas-especiales/principios-de-estudios-por-la-imagen-radiológicas/resonancia-magnética>
15. Mzoughi H, Njeh I, Wali A, Slima MB, BenHamida A, Mhiri C, et al. Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI Gliomas Brain Tumor Classification. *J Digit Imaging*. 2020 Aug 1;33(4):903–15.
16. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images | Request PDF. ResearchGate [Internet]. 2024 Oct 22 [cited 2025 Jan 10]; Available from: [https://www.researchgate.net/publication/337494284\\_Deep\\_neural\\_network\\_with\\_generative\\_adversarial\\_networks\\_pre-training\\_for\\_brain\\_tumor\\_classification\\_based\\_on\\_MR\\_images](https://www.researchgate.net/publication/337494284_Deep_neural_network_with_generative_adversarial_networks_pre-training_for_brain_tumor_classification_based_on_MR_images)
17. Guan Y, Aamir M, Rahman Z, Ali A, Abro WA, Dayo ZA, et al. A framework for efficient brain tumor classification using MRI images. *Math Biosci Eng*. 2021;18(5):5790–815.
18. Sultan HH, Salem NM, Al-Atabany W. Multi-Classification of Brain Tumor Images Using Deep Neural Network. *IEEE Access*. 2019;7:69215–25.
19. Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM. Classification using deep learning neural networks for brain tumors. *Future Comput Inform J*. 2018 Jun 1;3(1):68–71.
20. Zhou Y, Li Z, Zhu H, Chen C, Gao M, Xu K, et al. Holistic Brain Tumor Screening and Classification Based on DenseNet and Recurrent Neural Network. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T, editors. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing; 2019. p. 208–17.
21. Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. *Comput Biol Med*. 2019 Aug;111:103345.
22. Brain Tumor/Mass Classification Framework Using Magnetic-Resonance-Imaging-Based Isolated and Developed Transfer Deep-Learning Model [Internet]. [cited 2024 Dec 29]. Available from: <https://www.mdpi.com/1424-8220/22/1/372>
23. Khawaldeh S, Pervaiz U, Rafiq A, Alkhawaldeh RS. Noninvasive Grading of Glioma Tumor Using Magnetic Resonance Imaging with Convolutional Neural Networks. *Appl Sci*. 2018 Jan;8(1):27.
24. Yang Y, Yan LF, Zhang X, Han Y, Nan HY, Hu YC, et al. Glioma Grading on Conventional MR Images: A Deep Learning Study With Transfer Learning. *Front Neurosci*. 2018;12:804.
25. van der Voort SR, Incekara F, Wijnenga MMJ, Kapsas G, Gahrman R, Schouten JW, et al. Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning. *Neuro-Oncol*. 2023 Feb 1;25(2):279–89.
26. Decuyper M, Bonte S, Deblaere K, Holen RV. Automated MRI based pipeline for glioma segmentation and prediction of grade, IDH mutation and 1p19q co-deletion

[Internet]. arXiv; 2020 [cited 2025 Jan 10]. Available from: <http://arxiv.org/abs/2005.11965>

27. Sun W, Song C, Tang C, Pan C, Xue P, Fan J, et al. Performance of deep learning algorithms to distinguish high-grade glioma from low-grade glioma: A systematic review and meta-analysis. *iScience*. 2023 Jun 16;26(6):106815.
28. Latif G, Ben Brahim G, Iskandar DNFA, Bashar A, Alghazo J. Glioma Tumors' Classification Using Deep-Neural-Network-Based Features with SVM Classifier. *Diagn Basel Switz*. 2022 Apr 18;12(4):1018.
29. Vidyadharan S, Prabhakar Rao BVVSN, Perumal Y, Chandrasekharan K, Rajagopalan V. Deep Learning Classifies Low- and High-Grade Glioma Patients with High Accuracy, Sensitivity, and Specificity Based on Their Brain White Matter Networks Derived from Diffusion Tensor Imaging. *Diagn Basel Switz*. 2022 Dec 19;12(12):3216.
30. Kang J, Ullah Z, Gwak J. MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. *Sensors*. 2021 Jan;21(6):2222.
31. Paul JS, Plassard AJ, Landman BA, Fabbri D. Deep learning for brain tumor classification. In: Krol A, Gimi B, editors. Orlando, Florida, United States; 2017 [cited 2025 Jan 10]. p. 1013710. Available from: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2254195>
32. Ait Amou M, Xia K, Kamhi S, Mouhafid M. A Novel MRI Diagnosis Method for Brain Tumor Classification Based on CNN and Bayesian Optimization. *Healthc Basel Switz*. 2022 Mar 8;10(3):494.
33. Adaptive moment estimation (Adam) - File Exchange - MATLAB Central [Internet]. [cited 2025 Jan 14]. Available from: [https://es.mathworks.com/matlabcentral/fileexchange/136679-adaptive-moment-estimation-adam?s\\_tid=FX\\_rc2\\_behav](https://es.mathworks.com/matlabcentral/fileexchange/136679-adaptive-moment-estimation-adam?s_tid=FX_rc2_behav)
34. Academy E. ¿Cómo actualiza el algoritmo de descenso de gradiente los parámetros del modelo para minimizar la función objetivo y qué papel juega la tasa de aprendizaje en este proceso? - Academia EITCA [Internet]. EITCA Academy. 2024 [cited 2025 Jan 14]. Available from: <https://es.eitca.org/inteligencia-artificial/eitc-ai-adl-aprendizaje-profundo-avanzado/optimizaci%C3%B3n/optimizaci%C3%B3n-para-aprendizaje-autom%C3%A1tico/optimizaci%C3%B3n-de-la-revisi%C3%B3n-de-ex%C3%A1menes-para-el-aprendizaje-autom%C3%A1tico/%C2%BFC%C3%B3mo-actualiza-el-algoritmo-de-descenso-de-gradiente-los-par%C3%A1metros-del-modelo-para-minimizar-la-funci%C3%B3n-objetivo-y-qu%C3%A9-papel-juega-la-tasa-de-aprendizaje-en-este-proceso%3F/>
35. Brain MRI Images for Brain Tumor Detection [Internet]. [cited 2024 Dec 30]. Available from: <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection>

36. Brain Tumor Classification (MRI) [Internet]. [cited 2024 Dec 30]. Available from: <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri?select=Training>
37. Bakas S, Sako C, Akbari H, Bilello M, Sotiras A, Shukla G, et al. The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics. *Sci Data*. 2022 Jul 29;9(1):453.
38. UPENN-GBM [Internet]. The Cancer Imaging Archive (TCIA). [cited 2024 Dec 30]. Available from: <https://www.cancerimagingarchive.net/collection/upenn-gbm/>
39. CPTAC-GBM [Internet]. The Cancer Imaging Archive (TCIA). [cited 2025 Jan 13]. Available from: <https://www.cancerimagingarchive.net/collection/cptac-gbm/>
40. Liu J, Cao S, Imbach KJ, Gritsenko MA, Lih TSM, Kyle JE, et al. Multi-scale signaling and tumor evolution in high-grade gliomas. *Cancer Cell*. 2024 Jul 8;42(7):1217-1238.e19.
41. REMBRANDT [Internet]. The Cancer Imaging Archive (TCIA). [cited 2025 Jan 9]. Available from: <https://www.cancerimagingarchive.net/collection/rembrandt/>
42. A Hybrid Feature Extraction Method With Regularized Extreme Learning Machine for Brain Tumor Classification | IEEE Journals & Magazine | IEEE Xplore [Internet]. [cited 2024 Dec 29]. Available from: <https://ieeexplore.ieee.org/document/8664160>
43. mri\_deface - Free Surfer Wiki [Internet]. [cited 2024 Dec 30]. Available from: [https://surfer.nmr.mgh.harvard.edu/fswiki/mri\\_deface](https://surfer.nmr.mgh.harvard.edu/fswiki/mri_deface)
44. CBICA/BrainMaGe [Internet]. Center for Biomedical Image Computing & Analytics (CBICA); 2024 [cited 2024 Dec 30]. Available from: <https://github.com/CBICA/BrainMaGe>
45. Alsaif H, Guesmi R, Alshammari BM, Hamrouni T, Guesmi T, Alzamil A, et al. A Novel Data Augmentation-Based Brain Tumor Detection Using Convolutional Neural Network. *Appl Sci-BASEL*. 2022 Apr;12(8):3773.
46. Musallam AS, Sherif AS, Hussein MK. A New Convolutional Neural Network Architecture for Automatic Detection of Brain Tumors in Magnetic Resonance Imaging Images. *IEEE ACCESS*. 2022;10:2775–82.
47. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [Internet]. *arXiv*; 2015 [cited 2025 Jan 9]. Available from: <http://arxiv.org/abs/1409.1556>
48. Nguyen TH, Nguyen TN, Ngo BV. A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease. *AgriEngineering*. 2022 Dec;4(4):871–87.
49. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.
50. Khazaee Z, Langarizadeh M, Ahmadabadi MES. Developing an Artificial Intelligence Model for Tumor Grading and Classification, Based on MRI Sequences of Human Brain Gliomas. *Int J Cancer Manag* [Internet]. 2022 [cited



2024 Dec 29];15(1). Available from: <https://brieflands.com/articles/ijcm-120638#abstract>

51. Adithya V, Williams B, Czanner S, Kavitha S, Friedman D, Willoughby C, et al. EffUnet-SpaGen: An Efficient and Spatial Generative Approach to Glaucoma Detection. *J Imaging*. 2021 May 30;7:92.
52. Mock A, Geisenberger C, Orlik C, Warta R, Schwager C, Jungk C, et al. LOC283731 promoter hypermethylation prognosticates survival after radiochemotherapy in IDH1 wild-type glioblastoma patients. *Int J Cancer*. 2016;139(2):424–32.
53. Hao J, Kim Y, Mallavarapu T, Oh JH, Kang M. Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. *BMC Med Genomics*. 2019 Dec 23;12(10):189.
54. Yang Q, Xiong Y, Jiang N, Zeng F, Huang C, Li X. Integrating Genomic Data with Transcriptomic Data for Improved Survival Prediction for Adult Diffuse Glioma. *J CANCER*. 2020;11(13):3794–802.
55. Shboul ZA, Alam M, Vidyaratne L, Pei L, Elbakary MI, Iftekharruddin KM. Feature-Guided Deep Radiomics for Glioblastoma Patient Survival Prediction. *Front Neurosci* [Internet]. 2019 Sep 20 [cited 2025 Jan 13];13. Available from: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2019.00966/full>
56. CNN Architecture - Detailed Explanation [Internet]. InterviewBit. 2022 [cited 2025 Jan 14]. Available from: <https://www.interviewbit.com/blog/cnn-architecture/>
57. 3D Slicer image computing platform [Internet]. 3D Slicer. [cited 2025 Jan 14]. Available from: <https://slicer.org/>

## 8. Anexos

En vez de introducir anexos de código, se ha preferido añadir este a un repositorio de GitHub público, para que el código quede disponible para quien le pueda interesar.

<https://github.com/VaroLG/Barin-Tumor-Classification.git>