

**Universidad ORT Uruguay**  
**Facultad de Ingeniería**

**Predicción usando Machine Learning del estado del biomarcador genético Metil guanina metil transferasa (MGMT) para el tratamiento de tumores cerebrales.**

Entregado como requisito para la obtención del título de Máster en Big Data.

**Tanya Cohen – 190941**

**Tutor: Sergio Yovine**

**2022**

## **Declaración de autoría**

Yo, Tanya Cohen, declaro que el trabajo que se presenta en esta obra es de mi propia mano. Puedo asegurar que:

- La obra fue producida en su totalidad mientras realizaba la misma;
- Cuando he consultado el trabajo publicado por otros, lo he atribuido con claridad;
- Cuando he citado obras de otros, he indicado las fuentes. Con excepción de estas citas, la obra es enteramente propia;
- En la obra, he acusado recibo de las ayudas recibidas;
- Cuando la obra se basa en trabajo realizado conjuntamente con otros, he explicado claramente qué fue contribuido por otros, y qué fue contribuido por mí;
- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.

A handwritten signature in blue ink, appearing to be 'Tanya Cohen', with a horizontal line extending to the right from the bottom of the signature.

Tanya Cohen

28 de Marzo de 2022

## **Agradecimientos**

- A Sergio Yovine, por su guía y disponibilidad para elaborar este trabajo.
- A Topazium, por su ayuda en brindar conocimientos en el área de oncología.
- A la Universidad ORT por brindar las herramientas necesarias.
- A mi familia y amigos por el apoyo brindado en este tiempo.

## **Abstract**

Un tumor maligno en el cerebro es una afección potencialmente mortal. Conocido como glioblastoma, es tanto la forma más común de cáncer cerebral en adultos como la que tiene el peor pronóstico, con una media de supervivencia de menos de un año. El gen MGMT está localizado en el cromosoma 10q26 y codifica para una enzima reparadora del ADN que remueve grupos alquilo en la posición O de la guanina. Estos grupos O alquilo en la guanina, producen que las hebras de ADN se rompan y las bases no se acoplen, induciendo la apoptosis o muerte celular. El gen MGMT protege a las células normales de carcinógenos.

Actualmente, el análisis genético del cáncer requiere cirugía para extraer una muestra de tejido. Luego puede tomar varias semanas determinar la caracterización genética del tumor. Dependiendo de los resultados y el tipo de terapia inicial elegida, puede ser necesaria una cirugía posterior. Si se pudiera desarrollar un método preciso para predecir la genética del cáncer a través de imágenes (es decir, mediante lo que se conoce como “radiogenómica”) se podría minimizar el número de cirugías y refinar el tipo de terapia requerida. El objetivo general de esta tesis es estudiar el contexto del estado del arte en radiogenómica y su aplicación concreta en la predicción de la metilación del gen MGMT.

## **Palabras clave**

Glioblastoma, Metil guanina metil transferasa (MGMT), Magnetic resonance imaging (MRI), Aprendizaje automático.

# Índice

1. Introducción.....	7
1.1 Glioblastoma.....	7
1.1.1 Diagnóstico.....	9
1.1.2 Tratamiento.....	12
1.2 Promotor MGMT.....	13
1.3 Planteamiento del problema.....	13
1.4 Objetivo general.....	14
2. Estado del arte.....	15
2.1 Radiogenómica.....	15
2.2 Uso del Aprendizaje automático en Radiogenómica.....	17
2.3 Investigación bibliográfica.....	18
3. Datos.....	21
3.1 Descripción general.....	21
3.2 Análisis exploratorio.....	23
4. Predicción del estado de metilación de MGMT.....	26
4.1 Extracción de <i>features</i> radiómicas.....	28
4.2 Construcción de modelos a partir de <i>features</i> .....	29
4.3 Análisis detallado de los resultados obtenidos.....	31
5. Conclusión.....	35
6. Referencias bibliográficas.....	37
7. Anexos.....	40
7.1 <i>Features</i> Radiómicos.....	40

# 1. Introducción

El objetivo de este apartado es introducir la problemática del Glioblastoma, su diagnóstico, tratamiento, el rol del gen MGMT y los objetivos generales que involucran a este trabajo.

## 1.1. Glioblastoma

Un tumor maligno en el cerebro es una afección potencialmente mortal. Conocido como glioblastoma, es tanto la forma más común de cáncer cerebral en adultos como la que tiene el peor pronóstico, con una media de supervivencia de menos de un año. El glioblastoma es el tumor primario maligno más común y agresivo del sistema nervioso central en adultos, con una heterogeneidad intrínseca extrema en apariencia, forma e histología. Los pacientes con glioblastoma tienen un pronóstico muy pobre, y el estándar actual de atención comprende la cirugía, seguida de la radioterapia y la quimioterapia. (1)

La carcinogénesis representa un proceso de múltiples pasos complejos y se asocia con factores biológicos, físicos, químicos y genéticos. La Agencia Internacional para la Investigación del Cáncer (IARC) ha evaluado más de 900 posibles agentes carcinogénicos, incluidas sustancias químicas, mezclas complejas, agentes ocupacionales e infecciosos, pero solo nueve de estas exposiciones se acompañan de informes sobre posibles asociaciones con tumores del Sistema Nervioso Central en seres humanos (berilio, epiclorhidrina, clordano, heptacloro, metil-tiouracilo, propiltiouracilo, plomo, sulfato de diisopropilo y diclorometano). Varios otros productos químicos aumentaron la incidencia de tumores cerebrales en animales de experimentación, pero sin pruebas concomitantes en humanos (aflatoxina B1, sulfato de dietilo, acrilamida, etilnitrosourea, metilnitrosourea, clorhidrato de procarbazona, metanosulfonato de metilo, sulfato de dimetilo, glicidol, dacarbazina, 1,3-propano sulfonato, acrilonitrilo). (3)

Las enfermedades infecciosas causadas por virus, bacterias y parásitos se conocen también como estados asociados o causales, y pueden contribuir a la carcinogénesis cerebral. Cofactores adicionales, tales como la inmunidad sistémica y la inflamación crónica, cumplen un papel importante en este proceso. El análisis de citomegalovirus humano (HCMV) en tejido tumoral y sangre periférica de pacientes con glioblastoma

versus donantes normales (sanguíneos y quirúrgicos), encontró que más del 90% de muestras de tejido tumoral y 80% de muestras de sangre están asociados con el ADN de HCMV, mientras que los donantes normales no mostraron virus detectable. Estos resultados confirman la asociación de HCMV con glioblastoma, y sugieren una reactivación ya sea sistémica del virus en pacientes con glioblastoma o derramamiento de ADN viral a partir de células tumorales infectadas en la periferia. La causa del desarrollo del glioblastoma no es clara. (3) Algunos autores especulan que la infección por citomegalovirus puede generar procesos oncológicos modulando los factores de crecimiento y receptores de expresión de gliomas. (2)

Los glioblastomas pueden comprometer cualquier estructura neuroanatómica, pero en adultos es más común en los hemisferios cerebrales, mientras que en los niños lo es en la fosa posterior. Su crecimiento infiltrativo es extremadamente rápido. Histológicamente está compuesto de células de gran variabilidad morfológica, algunas bizarras, pleomórficas y multinucleadas; con actividad mitótica elevada; proliferación microvascular; severa y característica hiperplasia endotelial; microtrombos intravasculares, y necrosis extensas de carácter isquémico o en forma de pseudoempalizadas. (3)

### 1.1.1. Diagnóstico

La rutina de diagnóstico se basa en resonancia magnética (MRI). La mayoría de los glioblastomas se desarrollan en los hemisferios cerebrales. Sin embargo, su alto grado de infiltración causa que el tumor se disemine en el hemisferio contralateral, con resultados desfavorables para el paciente. Glioblastomas localizados en el cerebelo y corteza cerebral son poco comunes. (6)

La resonancia magnética es uno de los avances más importantes en el diagnóstico y seguimiento de los tumores cerebrales. Describe la morfología del tumor y la reacción del tejido circundante al mismo, aprovechando la característica de espín de los núcleos atómicos. Los núcleos atómicos con espín distinto a cero, aquellos con número impar de neutrones y/o protones, sometidos a un campo magnético externo, son capaces de absorber ondas electromagnéticas en el rango de radiofrecuencias. La liberación de esta energía permite obtener imágenes anatómicas. Las propiedades de la relajación, que viene definida por dos componentes, T1 (espín-entorno) y T2 (espín-espín), proporcionan el contraste entre los diferentes tejidos. (18)

En general, en los tumores del sistema nervioso central, son hipotensos en T1 e hipertensos en T2, debido a un aumento del contenido líquido en el tumor y al edema vasogénico. No obstante, hay excepciones, ya que algunos tumores pueden ser hipertensos en T1 e hipotensos en T2, como los que contienen hemosiderina, melanina, calcificaciones y mucina. Tras el paso de contraste endovenoso (gadolinio) la captación es variable, dependiendo de la vascularización de la lesión y de la integridad de la barrera hematoencefálica. Un aumento de la captación de contraste es sugestivo de lesión agresiva, por rotura de la barrera y por tanto un tumor de alto grado. No obstante, otros factores como la irradiación o el uso de corticoides pueden traducirse en rotura de ésta y por tanto en captación de contraste. (Tabla 1) (18).

La secuencia FLAIR realza imágenes destacadas en T2, anulando el líquido encefaloraquídeo. El FLAIR con contraste es útil para diferenciar al tumor del tejido circundante y para identificar metástasis. (Figura 1) (18)

Intensidad de imagen	Hallazgos compatibles
<b>Hiperintensidad en T1</b>	Metahemoglobina (hemorragia subaguda), melanina (metástasis de melanoma). Alto contenido en proteínas (quiste coloide, craneofaringioma). Grasa (lipoma), colesterol. Agente paramagnético (gadolinio). Magnesio, hierro, cobre, calcio.
<b>Hipointensidad en T2</b>	Hierro (sangre) con necrosis (GB). Hemosiderina (hemorragia crónica), desoxihemoglobina (hemorragia aguda). Calcificación (oligodendroglioma). Elevada relación núcleo-citoplasma (tumor neuroectodérmico primitivo, linfoma). Estroma de fibrocolágeno (meningioma). Alto contenido en proteínas (craneofaringioma). Flujo sanguíneo elevado (hemangioblastoma, GB). Melanina, ferritina, mucina.

Tabla 1: Hallazgos compatibles con hiperintensidad T1 y/o hipointensidad T2. (18)

La mayoría de los pacientes con glioblastoma se someten a una tomografía computarizada del cerebro en la presentación inicial. Una vez que se identifica una masa y se excluye la hemorragia, típicamente se solicita MRI con contraste, con recuperación de inversión estándar ponderada en T2 (T2w), T2-FLAIR atenuada por fluido, secuencias ponderadas en T1 (T1w) y potenciadas por contraste (T1Gd) ponderadas en T1. (18)

El estudio mediante Tomografía de emisión de positrones (PET) también es utilizado para el diagnóstico, visualizar el volumen, heterogeneidad intra tumoral y monitorear la terapia. (2)

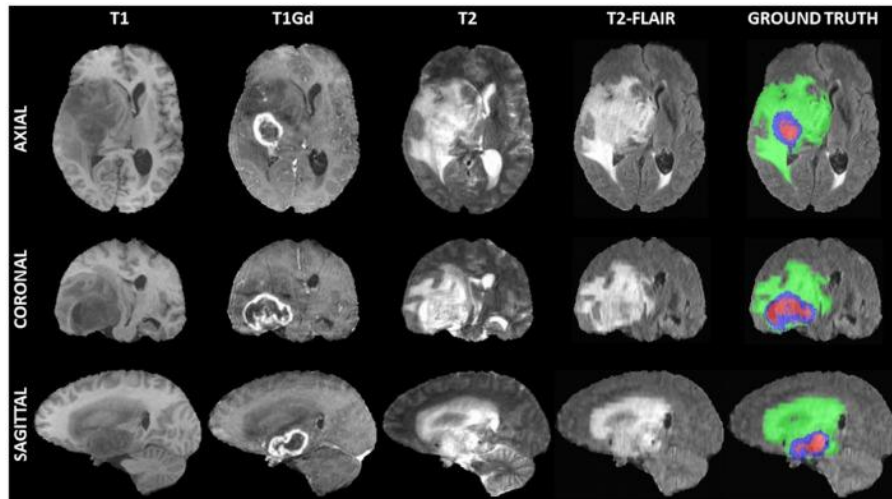


Figura 1: Ejemplo de una exploración de imágenes por resonancia magnética multiparamétrica 3D. De izquierda a derecha, se ilustran las cuatro modalidades de entrada, que incluyen ponderación T1 nativa (T1), poscontraste T1 (T1Gd), ponderación T2 nativa (T2) y recuperación de inversión atenuada por líquido T2 (T2-FLAIR), proporcionada como parte del conjunto de datos BraTS. De arriba a abajo, se representan tres vistas (es decir, axial, coronal, sagital) de estos volúmenes 3D para mostrar la naturaleza tridimensional de estos escans. (25)

### 1.1.2. Tratamiento

Recientemente, se ha progresado en el tratamiento del glioblastoma. El tratamiento adyuvante con radioterapia asociado a quimioterapia, basada en compuestos alquilantes (agentes que transfieren grupos alquil a las bases de guanina, dañan el ADN y causan muerte celular de células tumorales) como la carmustina o la temozolamida, han demostrado que pueden incrementar la supervivencia (3).

La Organización Europea de Investigación y Tratamiento del Cáncer (EORTC) y el Instituto Nacional de Cáncer de Canadá (NCIC) demostraron que el agente alquilante temozolamida en radioterapia mejora la supervivencia en pacientes con diagnóstico de glioblastoma a dos, tres y cinco años, con un aumento del 11% a 27%, 4,4% a 16% y 1,9% a 9,8% respectivamente. La temozolamida es una quimioterapia administrada oralmente que penetra la barrera hematoencefálica. Los efectos citotóxicos de la temozolamida implican la alquilación (metilación) del ADN en varios lugares de la guanina. La formación de O-6-metilguanina inhibe la replicación del ADN y causa la muerte celular a través de la estimulación de la proteína p53. Actualmente, la radioterapia con adyuvancia con temozolamida es lo que se utiliza comúnmente.

En el estudio en fase III AVAglio (24), evaluaron la adición del agente antiangiogénico bevacizumab en combinación con quimioterapia con temozolamida, indicando efectos favorables en supervivencia media de 10.6 meses versus 6.2 meses. (2)

La resección del tumor también representa un tratamiento significativo y es el primer paso. La alta dispersión del glioblastoma en los diferentes tejidos cerebrales hace difícil su resección. Si bien el tratamiento con temozolamida ha mejorado la supervivencia, casi todos los pacientes sufren de recurrencias, debido a que las opciones de tratamiento son limitadas. (2)

## 1.2. Promotor MGMT

La metilación del promotor del gen MGMT, es un factor pronóstico favorable y un fuerte predictor de la capacidad de respuesta a la quimioterapia. Por tanto, la determinación del estado de metilación del promotor MGMT en el glioblastoma recién diagnosticado puede influir en la toma de decisiones sobre el tratamiento. (1)

El gen MGMT está localizado en el cromosoma 10q26 y codifica para una enzima reparadora del ADN que remueve grupos alquilo en la posición O de la guanina, por lo que MGMT protege a las células normales de carcinógenos. Desafortunadamente, también protege células tumorales de efectos normales de la quimioterapia con agentes alquilantes como la temozolamida (2).

## 1.3. Planteamiento del problema

Actualmente, el análisis genético del cáncer requiere cirugía para extraer una muestra de tejido. Luego puede tomar varias semanas determinar la caracterización genética del tumor. Dependiendo de los resultados y el tipo de terapia inicial elegida, puede ser necesaria una cirugía posterior. Si se pudiera desarrollar un método preciso para predecir la genética del cáncer a través de imágenes (es decir, mediante lo que se conoce como “radiogenómica”) se podría minimizar el número de cirugías y refinar el tipo de terapia requerida.

La identificación adecuada de las subregiones del tumor en un MRI es de gran relevancia en muchas aplicaciones clínicas, como ser la planificación de tratamiento por cirugía, intervenciones guiadas por imágenes, monitoreo del crecimiento del tumor, y generación de mapas para radioterapia. Sin embargo, la detección manual y rastreo del tumor es un trabajo tedioso, consumidor y subjetivo, llevado a cabo por radiólogos y con numerosos pacientes. Es de interés automatizar soluciones que contribuyan a mejorar este proceso.

El área de Radiómica y Radiogenómica en Gliomas es un área muy incipiente. El uso de Machine learning y Aprendizaje automático es un campo reciente y muy prometedor para poder predecir propiedades genómicas. (6)

La Sociedad Radiológica de América del Norte (RSNA) se ha asociado con la Sociedad de Computación por Imagen Médica e Intervención Asistida por Computadora (MICCAI) para mejorar el diagnóstico y la planificación del tratamiento para pacientes con glioblastomas, recurriendo a la radiogenómica (5). Para fomentar avances en el desarrollo de estas técnicas, se ha recolectado un banco de imágenes curado que se utiliza anualmente para una competencia internacional. En 2021, el foco estuvo en la predicción del subtipo genético de glioblastoma utilizando escáneres de MRI para entrenar y probar modelos para detectar la presencia de metilación del promotor MGMT.

Si se tiene éxito, se ayudará a los pacientes con cáncer cerebral a recibir diagnósticos y tratamientos menos invasivos. La introducción de estrategias de tratamiento nuevas y personalizadas antes de la cirugía tiene el potencial de mejorar el manejo, la supervivencia y las perspectivas de los pacientes con cáncer cerebral.

#### 1.4 Objetivo general

El objetivo general de esta tesis es estudiar el contexto del estado del arte en radiogenómica y su aplicación concreta en la predicción de la metilación del gen MGMT, utilizando como caso de estudio el propuesto por la Sociedad Radiológica de América del Norte. El caso de estudio fue de apoyo para probar y ejemplificar el problema.

## 2. Estado del Arte

El objetivo de esta sección es presentar el concepto de radiogenómica, así como las aplicaciones en Machine learning, en el contexto basado en literatura relevante, específicamente vinculada con la detección del estado de metilación del gen MGMT de MRI en pacientes con glioblastoma.

### 2.1. Radiogenómica

El comportamiento de gliomas primarios y secundarios, así como la identificación de área de necrosis, progresión tumoral y pseudoprogresión en MRI es complicado. La Radiómica y Radiogenómica prometen ofrecer un diagnóstico preciso, predicción de prognosis y evaluar la respuesta tumoral a quimioterapia, inmunoterapia y radioterapia. Esto se logra mediante extracción de "atributos" o "características" (*features*) morfológicas, texturales y funcionales de imágenes.

La Radiómica refiere a la extracción computarizada cuantificable de datos de imágenes radiológicas, creando bases de datos que pueden ser utilizadas para diagnóstico, caracterización y evaluar o predecir respuesta a cierta terapia. Las mutaciones genéticas usualmente determinan la agresividad del tumor y han demostrado estar asociadas al patrón de lesiones y respuesta a terapia. En el campo de la medicina, la radiómica es un método que extrae un gran número de características de las imágenes médicas utilizando algoritmos de caracterización de datos. Estas características, denominadas características radiómicas o *features*, tienen el potencial de descubrir patrones y características tumorales que no se aprecian a simple vista. La hipótesis de la radiómica es que las características de imagen distintivas entre las formas de la enfermedad pueden ser útiles para predecir el pronóstico y la respuesta terapéutica para varios tipos de cáncer, proporcionando así información valiosa para la terapia personalizada. (10)

Una herramienta muy utilizada en esta área es la biblioteca “*Pyradiomics*” (17), un paquete de Python de código abierto para la extracción de características radiómicas de imágenes médicas. Con este paquete, se establece un estándar de referencia para el análisis radiómico y se proporciona una plataforma de código abierto probada y mantenida para la extracción fácil y reproducible de características radiómicas.

Los datos integrados en escala genómica se denominan Radiogenómica. La extracción de *features* se realiza mediante algoritmos y métodos estadísticos. El propósito del análisis radiogenómico involucra la asociación entre mutaciones genéticas con la distinción de fenotipos de imágenes (Figura 2). (6)

En el contexto de este trabajo, no se cuenta con información de la biopsia, solo con el estado de metilación del gen MGMT. Por lo tanto no se puede realizar un análisis genómico de los datos. Se predice un contexto genómico pero solo a partir de los *features* radiómicos extraídos con *Pyradiomics*.

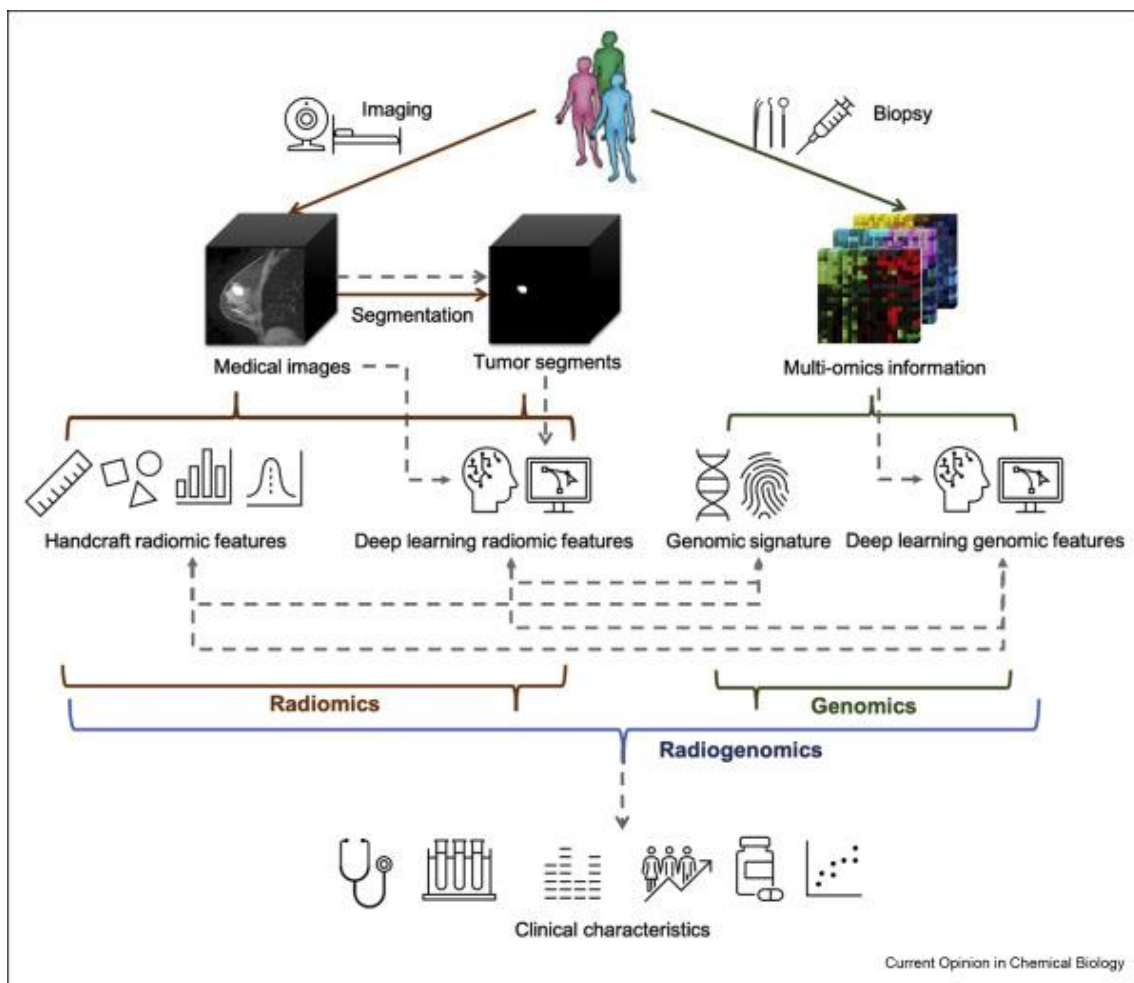


Figura 2: Relación entre Radiogenómica, Radiómica y Genómica asociado a mutaciones genéticas. (21)

## 2.2. Uso de Aprendizaje automático en Radiogenómica

Machine learning es una disciplina del campo de la inteligencia artificial que a través de algoritmos de aprendizaje busca identificar patrones en datos masivos y elaborar predicciones.

Este aprendizaje se divide en dos categorías; supervisado y no supervisado. El aprendizaje supervisado trabaja con datos etiquetados, es decir, datos para los que ya conoce la respuesta objetivo. El algoritmo trata de buscar patrones relacionándolos con el objetivo, dadas unas variables predictoras de entrada. A través de dicho histórico de datos, el algoritmo aprende una función que le permita predecir el atributo objetivo para nuevos datos. (12)

El aprendizaje no supervisado no cuenta con un conjunto de datos etiquetados para las pruebas de entrenamiento. Este modelo suele utilizarse como método exploratorio o de análisis. Los usos más comunes de este tipo de aprendizaje son: problemas de clustering, agrupamientos de co-ocurrencia o profiling de datos.

Deep Learning (aprendizaje profundo) es un tipo de aprendizaje en el que se utilizan redes neuronales artificiales que permiten descubrir patrones complejos en grandes volúmenes de datos. Estos modelos han realizado contribuciones en el área de computer vision, lenguaje y robótica. En el área de la salud, su utilización es prometedora dado que se genera y captura gran cantidad de datos conteniendo información valiosa, de una manera que sobrepasa a los métodos tradicionales. (12)

Las redes neuronales convolucionales (CNN) pueden ser utilizadas para mejorar la eficiencia en radiología. En cuanto a MRI, Deep learning se aplica en problemas de segmentación, predicción de enfermedades, reconstrucción de imágenes, entre otros. En un problema de segmentación, se particiona una imagen en múltiples regiones que comparten atributos similares, permitiendo localización y cuantificación. (12)

### 2.3. Investigación bibliográfica

En esta sección se presenta una investigación realizada del estado del arte en cuanto al análisis de imágenes del problema en cuestión. Dentro de la búsqueda realizada con los distintos criterios y palabras claves, se realiza una preselección de artículos relacionados al problema y se compilan para su posterior lectura.

Una vez realizada esta selección se procede a revisarlos para analizar las problemáticas tratadas, las técnicas que fueron usadas y los resultados obtenidos en cada uno de ellos. Con los resultados de este análisis, se selecciona el enfoque a seguir en esta tesis.

El primer artículo corresponde a “*MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks*”. En este estudio, se utilizan imágenes MRI de cerebro de pacientes con glioblastoma, del Cancer Imaging Archive (TCIA), combinado con datos de metilación del Cancer Genome Atlas (TCGA) para predecir el estado de metilación del gen MGMT en dichos pacientes. Se utiliza una red convolucional recurrente bidireccional (CRNN) que tiene como ventaja trabajar con los scans MRI tridimensionales. Se obtiene una *accuracy* de 67% en datos de validación y 62% en datos de test, se debería ver *precisión* y *recall* ambos de 67%. (8)

El segundo artículo corresponde a “*An overview of deep learning in medical imaging focusing on MRI*”. El mismo describe como *deep learning* ha sido aplicado al procesamiento de MRI, desde la adquisición hasta la segmentación y predicción de enfermedades. (7)

Otro artículo de interés trata sobre “*The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification*”. La competencia BraTS 2021 fue organizada por la Sociedad radiológica de Norte América (RSNA), la Sociedad americana de Neuroradiología (ASNR) y Medical Image Computing and Computer Assisted Interventions (MICCAI). Esta competencia se centra en evaluar diferentes algoritmos en 2000 pacientes diagnosticados con glioblastoma en 2 tareas: segmentación de las subregiones cerebrales y clasificación de la metilación del promotor MGMT. (5)

Otro artículo que se investigó es “*Is it Possible to Predict MGMT Promoter Methylation from Brain Tumor MRI Scans using Deep Learning Models?*” en donde explica que un

par de publicaciones recientes proponen una conexión entre el estado de metilación del promotor MGMT y los escáneres de MRI de los tumores, sugiriendo que el uso de *Deep learning* para este propósito es prometedor, incluyendo modelos CNN 2D y 3D. (16)

El artículo “*Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features*” se enfoca en el etiquetado de las subregiones cerebrales de scans de MRI, disponibles del Cancer Genome Atlas (TCGA), publicadas en The Cancer Imaging Archive (TCIA). (10)

En el trabajo “*Novel Local Radiomic Bayesian Classifiers for Non-Invasive Prediction of MGMT Methylation Status in Glioblastoma*” se presenta un clasificador Bayesiano que indica la probabilidad de predecir el estado de metilación del promotor MGMT basado en las *features* radiómicas de imágenes FLAIR. (19)

En cuanto a herramientas disponibles en Deep learning, el artículo “*GaNDLF: A Generally Nuanced Deep Learning Framework for Scalable End-to-End Clinical Workflows in Medical Imaging*” describe el impacto del *machine learning* optimizado en el área de la ciencia y clínica. Las librerías open source como TensorFlow (Google), PyTorch (Facebook), y MXNet (Apache), contribuyen al desarrollo de algoritmos y la variabilidad de su implementación llevan al desarrollo más estable, reproducible y escalable. Generally Nuanced Deep Learning Framework (GaNDLF) permite trabajar con *k-fold cross-validation*, *data augmentation*, múltiples modalidades y con imágenes radiológicas e histológicas. (11)

Otro artículo relevante es el de “*Automatic Prediction of MGMT Status in Glioblastoma via Deep Learning-Based MR Image Analysis*”. La metilación del gen MGMT se correlaciona con la salud de los pacientes. En este estudio, se diseña un algoritmo de *Deep learning* para predecir el estado del MGMT en 87 pacientes con imágenes T1Gd y 66 pacientes FLAIR. Las tareas involucraron la segmentación y la clasificación. En este artículo se dice haber utilizado la segmentación, se utilizó un modelo para MRI 2D del siguiente, una red neuronal basada a la red neuronal para MRI 3D que fue la ganadora de la competencia BraTS 2018 a continuación (13). Si bien se hicieron experimentos con este enfoque, no arrojaron los resultados del artículo, lo cual generó dudas sobre los experimentos publicados dado que no fue posible replicarlos.

El siguiente artículo fue ganador de la competencia BraTS 2018; “3D MRI brain tumor segmentation using autoencoder regularization”. La automatización de la segmentación de MRI 3D de tumores cerebrales es necesaria para el diagnóstico, monitoreo y tratamiento de la enfermedad. La segmentación manual requiere de conocimiento de anatomía, es de alto costo, consume tiempo y puede llevar a errores. Aquí se describe una red para la segmentación de la subregión de diferentes MRI basado en la arquitectura *encoder-decoder*. Se basa en la arquitectura CNN con *encoder* para extraer las *features* de las imágenes y *decoder* para reconstruir la máscara. El *encoder* utiliza bloques de ResNet, con normalización y ReLU. El *decoder* tiene una estructura similar, pero reduce el número de *features* (Figura 3). (14)

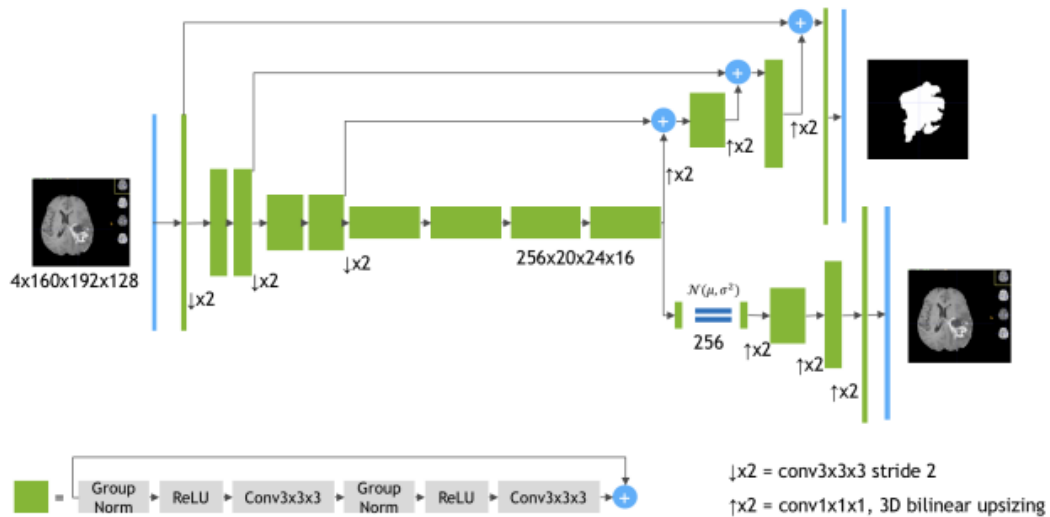


Figura 3: Arquitectura *encoder-decoder* (14)

Un interés particular de este último artículo es que utiliza los *features* de características radiómicas de las imágenes y además las compara con las que se extraen de las imágenes (auto *encoders*).

### 3. Datos

En esta sección se introduce el *dataset* que se utiliza para experimentar e ilustrar los resultados de diversas técnicas basadas en la bibliografía descrita en la sección anterior.

#### 3.1. Descripción general

Los datos para este trabajo se obtuvieron de la competencia RSNA-ASNR-MICCAI BraTS 2021. (5) El *dataset* describe una colección de scans MRI adquiridos de diferentes instituciones bajo estándares clínicos, pero con diferentes equipos y protocolos, resultando en una gran heterogeneidad de la calidad de las imágenes. Dentro de los criterios de inclusión se confirma el diagnóstico patológico y estado disponible de la metilación del promotor MGMT. Se tienen aproximadamente 2000 casos.

RSNA es un conjunto de datos de pacientes que tienen dos versiones de las resonancias. Por un lado, se encuentran las resonancias con la segmentación (3D) y por otro lado imágenes 2D en formato DCM extraídas de las resonancias, junto con el estatus de metilación. Cada uno de estos *dataset* son utilizados para cada competencia. Debido a que el *dataset* en formato DCM es variable en tamaño, con mala calidad y con imposibilidad de reconstruir en 3D, se prosigue a cruzar los IDs y trabajar con las resonancias de los pacientes para los cuales se conoce la segmentación y estatus de metilación, sumando en total 577 casos.

Las imágenes incluyen diferentes MRI *scans* a) Native (T1), b) post-contrast T1 weighted (T1Gd – Gadolinium), c) T2 weighted (T2) and d) T2 Fluid Attenuates Inversion Recovery (T2 FLAIR). Las imágenes se encuentran en formato DCM y al aplicar un preprocesamiento se convierten en formato NIFTI. El preprocesamiento se lleva a cabo con la herramienta Cancer Imaging Phenomics Toolkit (CaPTk). Las subregiones tumorales se determinan a través de observaciones que involucran Gd- enhancing tumor (ET – label 4), edema peritumoral (ED – label 2) y tumor necrótico (NC – label 1). ET es la porción del tumor en aumento compuesto por áreas T1Gd MRI. NC es el área necrótica del tumor, hipotensa en T1Gd MRI. ED es el edema peritumoral y tejido infiltrado definido por T2 FLAIR (Figura 4).

También se utiliza el software SimpleITK, es una interfaz simplificada de código abierto para *Insight Segmentation and Registration Toolkit* (ITK). Las imágenes fueron analizadas con su segmentación y su máscara. (Figuras 4 y 5)

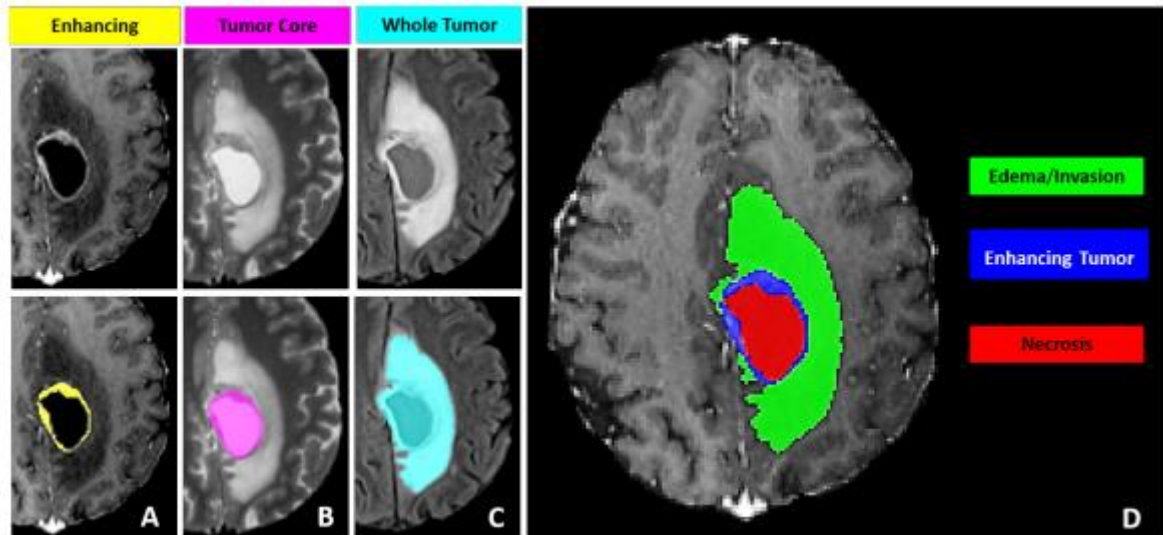


Figura 4: Subregiones en MRI scans. A - Enhancing tumor (ET) visible en un T1Gd scan, B – Tumor core (TC), C- Whole tumor (WT) visible en T2 y T2 FLAIR. D- Edema, Enhancing tumor y Necrosis. (5)

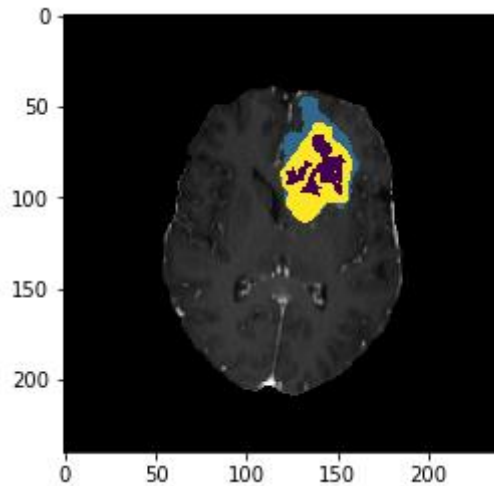


Figura 5: Visualización de la imagen “BraTS2021\_00000\_t1ce.nii.gz” de las tres regiones Edema, Enhancing tumor y Necrosis, pero en el *dataset* utilizando la herramienta *SimpleITK*. Esta imagen tiene valor de metilación 1.

El estatus de metilación del promotor MGMT es binario (0: no metilado, 1: metilado) y previsto en un .csv junto con los identificadores de los MRI. El estatus de metilación fue determinado por secuenciación. Nos encontramos en un contexto de radiogenómica donde el objetivo es predecir información genómica a partir de información radiómica.

### 3.2. Análisis exploratorio

Mediante la biblioteca pandas, se lee el archivo “train-labels.csv”, el cual posee los identificadores de las imágenes y su valor de metilación. Se realizó un histograma para analizar las distribuciones de metilación y no metilación en el conjunto de datos. (Figura 5).

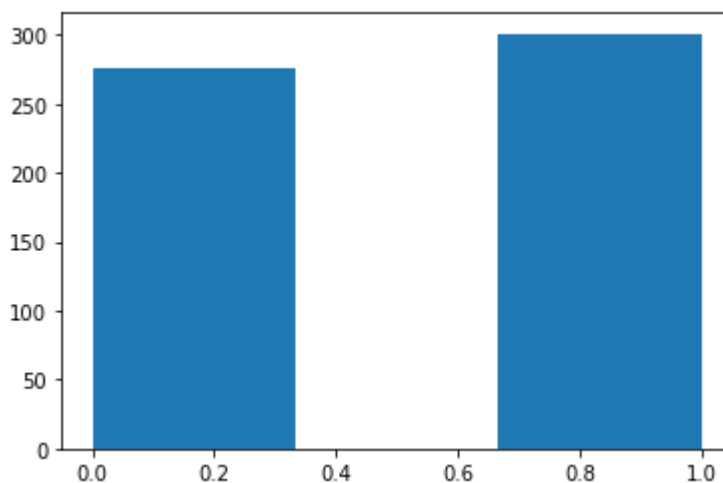


Figura 5: Distribución de metilación (derecha) y no metilación (izquierda). Se observa un *dataset* equilibrado.

También se realizó un análisis de la correlación entre Enhancing tumor (ET), Edema (ED), Necrosis (NC) y MGMT. Se observa que existe poca correlación entre las mismas. (Figura 6) A su vez se analizó la distribución de estas en cuanto a la metilación o no metilación de MGMT (Figura 7 a y b).

	nc	ed	et	mgmt
nc	1	0.11	0.51	0.03
ed	0.11	1	0.29	-0.014
et	0.51	0.29	1	0.013
mgmt	0.03	-0.014	0.013	1

Figura 6: Estudio de la correlación entre variables MGMT, NC, ED y ET.

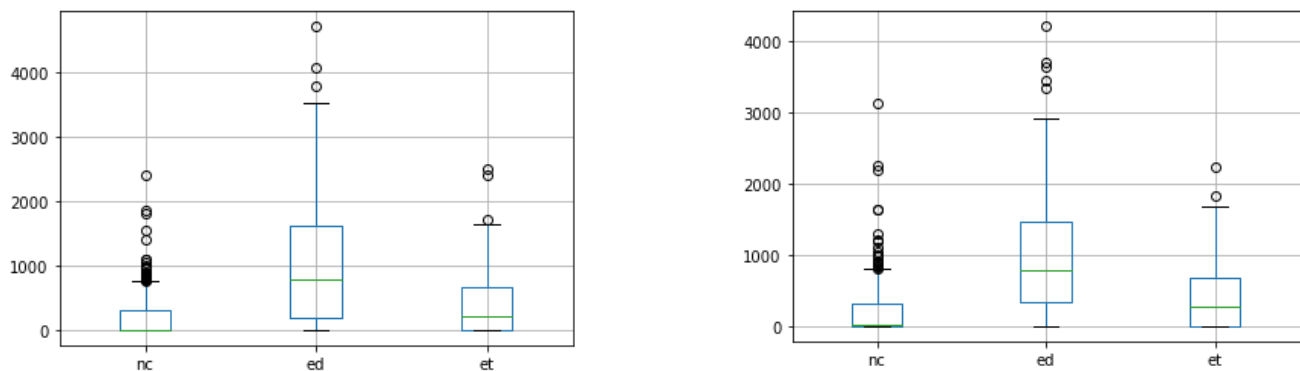


Figura 7: A – Boxplot de la distribución del área medida en píxeles para NC, ED y ET para no metilación. B- Boxplot de la distribución del área medida en píxeles para NC, ED y ET para metilación.

Se observa en el boxplot que son prácticamente iguales ambas comparaciones con la misma media. El área necrótica es alrededor de 0, el edema es alrededor de 800 y el enhancing tumor es un poco menor para la no metilación.

Luego de este análisis exploratorio se decidió trabajar con imágenes en formato NIFTI y no en formato DCM. Las imágenes DCM están desordenadas y hay información faltante, además los artículos relevantes leídos sugieren que tener la segmentación es muy importante para la predicción. En el otro *dataset* en formato NIFTI estaba presente la segmentación y teniendo en cuenta esto es que aporta a elegir porque no era el objetivo la segmentación sino detectar la metilación.

Como primer abordaje al problema de predicción se intentó hacer una clasificación basada solamente en la segmentación. Esta sola no sirve a pesar de ser importante, se necesita un análisis más profundo. (Figura 8)

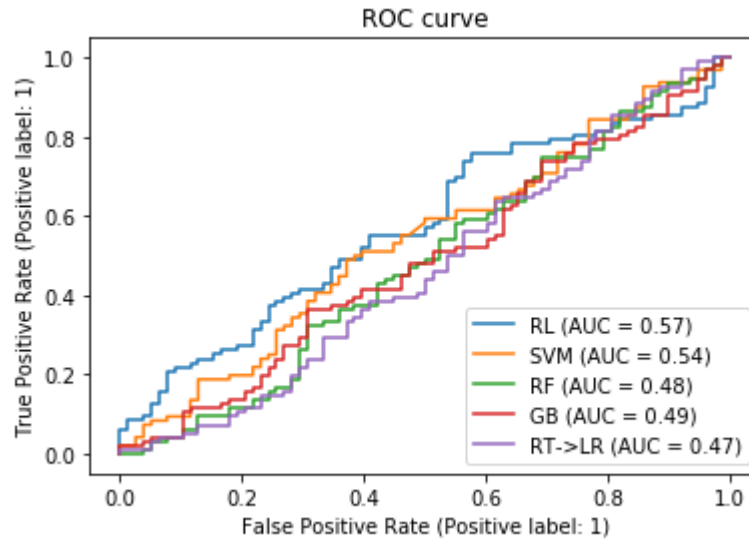


Figura 8: Clasificación basada solamente en segmentación. Valores de AUC por debajo de 0,5 no son lo esperado.

## 4. Predicción del estado de metilación de MGMT

Este experimento se basa en el artículo de “*Prediction of MGMT Methylation Status of Glioblastoma using Radiomics and Latent Space Shape Features*”.(15) Es el primer abordaje con Radiogenómica y que testea y evalúa modelos sencillos con resultados similares. El mismo corresponde a la competencia de BraTS 2021, donde se propone un método para predecir el estatus de MGMT en gliomas de alto grado. A partir de los MRI, se segmenta el tumor utilizando redes neuronales convolucionales y se extraen las *features* por un “*variational autoencoder*”.

Se parte de las imágenes en formato el NIFTI y su segmentación. Las radiomic *features* se extraen de las regiones tumorales segmentadas para cada modalidad T1, T2, Flair, T1Gd y para cada región (whole tumor, enhancing, non enhancing y necrotic). Para cada uno se generaron las radiomic *features* con el *feature extractor*. Esto genera 2 tipos de *features*: diagnostics y original. La partición consiste en 597 imágenes, 90% train y 10% test. El *feature extractor* toma el csv de train y busca en todas las imágenes NIFTI de las cuales tienen segmentación y lo hace con el 90% de train (Figura 9 y Tabla 2)

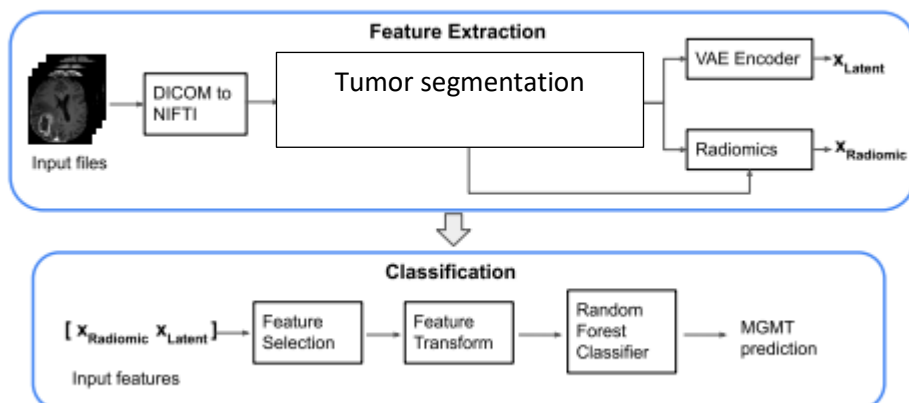


Figura 9: Arquitectura y Feature extraction adaptada del artículo “*Prediction of MGMT Methylation Status of Glioblastoma using Radiomics and Latent Space Shape Features*”

(8)

**Table 1.** List of selected radiomic features.

Category	Feature name	Modality	Region
Shape	Maximum 3D Diameter	-	Enh-core
First order	Interquartile Range	T1-ce	Core
First order	Mean Absolute Deviation	T1-ce	Core
First order	Mean	T1-ce	Core
First order	Median	T1-ce	Core
First order	Median	T1-ce	Whole
First order	Variance	T1-ce	Core
First order	10Percentile	FLAIR	Core
GLRLM	Graylevel non-uniformity normalized	FLAIR	Whole
GLRLM	Graylevel variance	FLAIR	Whole
GLSZM	Small area emphasis	FLAIR	Whole
GLSZM	Small area high graylevel emphasis	FLAIR	Whole
GLSZM	Small area low graylevel emphasis	FLAIR	Whole
NGTDM	Busyness	FLAIR	Whole
GLSZM	Small area high graylevel emphasis	T2	Whole

Tabla 2: Lista de *features* radiómicas (8)

#### 4.1 Extracción de *features* radiómicas

La primera parte de este experimento fue la extracción de *features* radiómicas. Las *features* se explican de la biblioteca “*Pyradiomics*”, un paquete de Python de código abierto para la extracción de características de Radiomics de imágenes médicas. Con este paquete, se establece un estándar de referencia para el análisis radiómico y proporcionar una plataforma de código abierto probada y mantenida para la extracción fácil y reproducible de características radiómicas. La plataforma admite la extracción de características en 2D y 3D y se puede utilizar para calcular valores únicos por característica para una región de interés ("basada en segmentos") o para generar mapas de características ("basada en vóxeles"). (18)

A continuación, se muestra a modo de ejemplo para una de las imágenes, los valores de algunos *features* radiómicos para el paciente 132 correspondiente a la imagen T1. La plataforma permite la extracción en 2D y 3D pero se trabaja con 3D. Se extraen 100 *features*, en adición a MGMT y feature ID. (Figura 10)

```
id                                00132
original_shape_Elongation         0.649015
original_shape_Flatness           0.42152
original_shape_LeastAxisLength    40.6597
original_shape_MajorAxisLength    96.4598
...
original_gldm_LowGrayLevelEmphasis 0.01376954050833343
original_gldm_SmallDependenceEmphasis 0.013278215062626886
original_gldm_SmallDependenceHighGrayLevelEmphasis 1.03026572774992
original_gldm_SmallDependenceLowGrayLevelEmphasis 0.00022327547014408788
mgmt                               0
Name: 85, Length: 102, dtype: object
```

Figura 10: Featrures radiómicos para el paciente 132 en la notebook

*RadiomicFeatureAnalysisML.*

## 4.2 Construcción de modelos a partir de *features*

Se continúa con la construcción de modelos a partir de *features*. Se generan 4x4 *datasets* porque se generan *features* para ED, ET, NC y WT para T1, T2, T1Gd y FLAIR.

Se probó con los *radiomic features* de T2 en el *whole tumor* y el resultado fue 0,57 comparable a la competencia BraTS que dio 0,66. Se probaron todas las variantes y se obtuvo un resultado similar. T2 con *whole tumor* fue el de mejor área bajo la curva con una Regresión Logística (RL). (Figura 11) Se usó RL porque fue el modelo que mejores resultados arrojó con la segmentación en el análisis preliminar (ver Figura 9).

Se intentó hacer un *majority voting*, el cual no mejora el resultado y hay una razón debido a que hay correlación entre los votantes porque votan igual y eso destruye el ensemble. Se podría sacar uno de los votantes, y al sacar FLAIR el resultado es 67%, el cual es comparable y mejor que T2 solo ya que rompe la correlación y mejora el resultado. Este *majority voting* no se hizo con el mejor umbral. Habría que calcular el mejor umbral y luego hacer el voto y luego este análisis. Pero parece que se entendería mejor si se toman los casos en los que T2 que es el mejor predictor individual, tiene mucha confianza en su predicción (alto porcentaje mayor a 80%). Se podría considerar un modelo entrenado con los datos de los *features* radiómicos de todos los tipos de MRI juntos, es decir, un modelo entrenado con 400 *features* en vez de 4 modelos entrenados con 100 *features*.

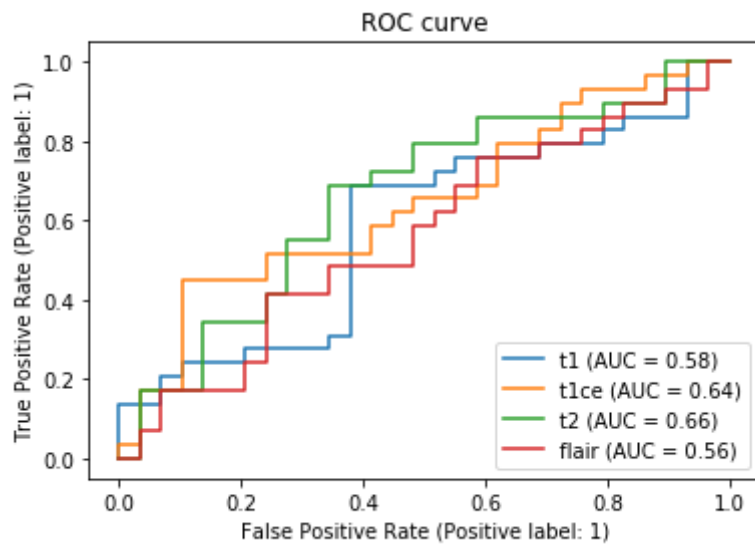


Figura 11: Predicción utilizando *features* radiómicos. Curva ROC para las modalidades T1, T2, T1ce y FLAIR con el WT. El mejor resultado obtenido de 0,66 es comparable con la competencia y con el artículo en cuestión.

### 4.3 Análisis detallado de los resultados obtenidos

A partir de los resultados del apartado anterior, se hizo un relevamiento de las imágenes de test y se seleccionaron algunas para evaluar si existe algún patrón entre el estado de metilación real y predicho. Para ello se evalúa la realidad con las predicciones de las imágenes MRI T2 y el voto mayoritario del ensemble T1, T2 y T1CE juntos. (Tabla 3)

ID	Imagen	Realidad	T2	Majority voting
513	342	1	1	1
810	549	0	0	0
611	417	1	1	0
452	305	0	0	1
716	485	1	0	1
575	388	0	1	0
808	547	1	0	0
555	373	0	1	1

Tabla 3: Relación entre el estado de metilación real y predicho para 8 imágenes.

Primero se comparan imágenes cuya realidad y predicción por T2 y majority voting es igual. Se observa que en la imagen con ID 513, cuya realidad es metilada, el whole tumor presenta una zona necrótica clara. Al contrario, la imagen de ID 810, cuya realidad es no metilada, se observa un zona necrótica indefinida.

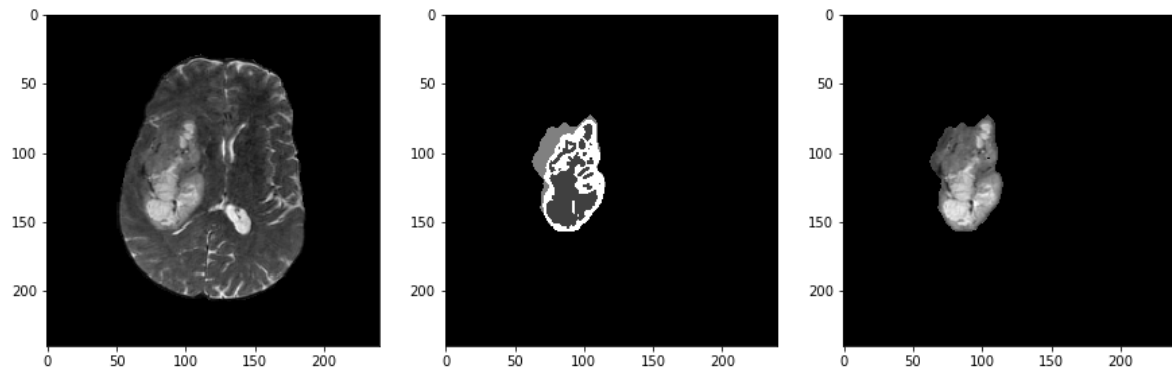
Luego se comparan imágenes donde la realidad y la predicción de T2 es igual, pero difiere del majority voting. Se observa el mismo patrón anterior. Se observa que en la imagen con ID 611, cuya realidad es metilada y coincide con T2, el whole tumor presenta una zona necrótica clara. Al contrario, la imagen de ID 452, cuya realidad es no metilada y coincide con T2, se observa una zona necrótica indefinida. Un punto a investigar es la relación entre zona necrótica y metilación. Esto no se menciona en la literatura relevada.

A continuación, se comparan imágenes donde la realidad y la predicción de T2 es diferente pero es igual a la del majority voting. El patrón que se venía observando no se cumple con los IDs 716 y 575. Ambos presentan zona necrótica y la misma es grande en proporción al whole tumor, más de lo que se observó en casos anteriores.

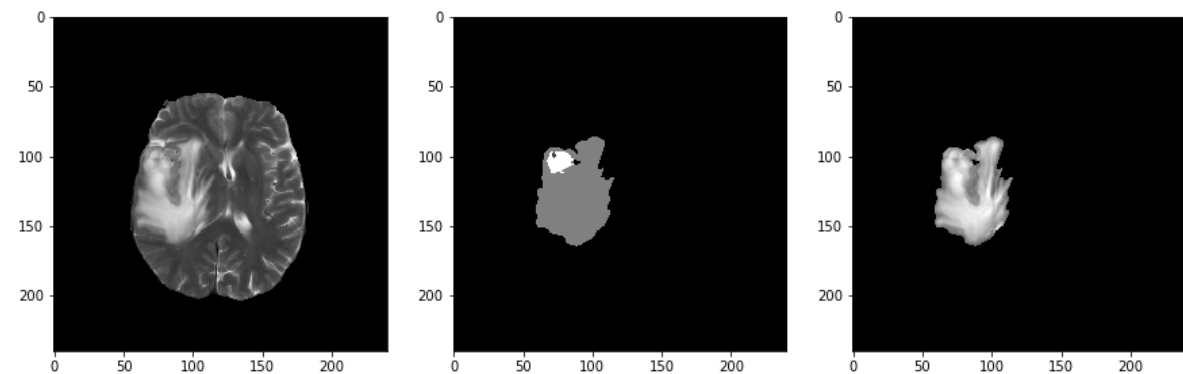
Por último, se comparan imágenes donde la realidad difiere a la predicción de T2 y majority voting, siendo estos dos iguales. Las imágenes cuyos ID son 808 y 555 son irregulares, y de formas particulares. No se observa un patrón.

Por otra parte, en la bibliografía analizada no se encontró ninguna mención a una posible relación entre la metilación y el tamaño de la zona necrótica.

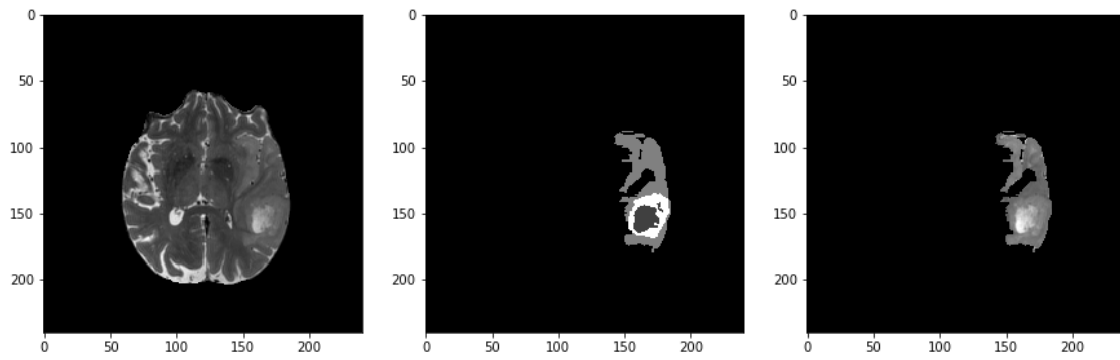
### **ID 513**



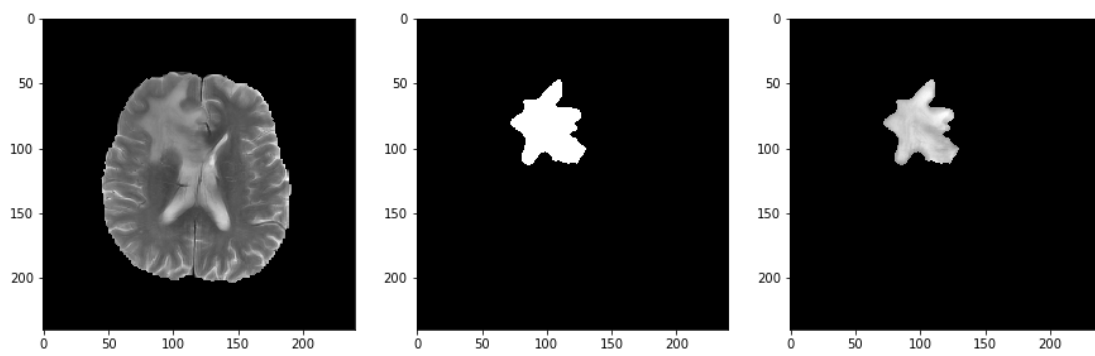
### **ID 810**



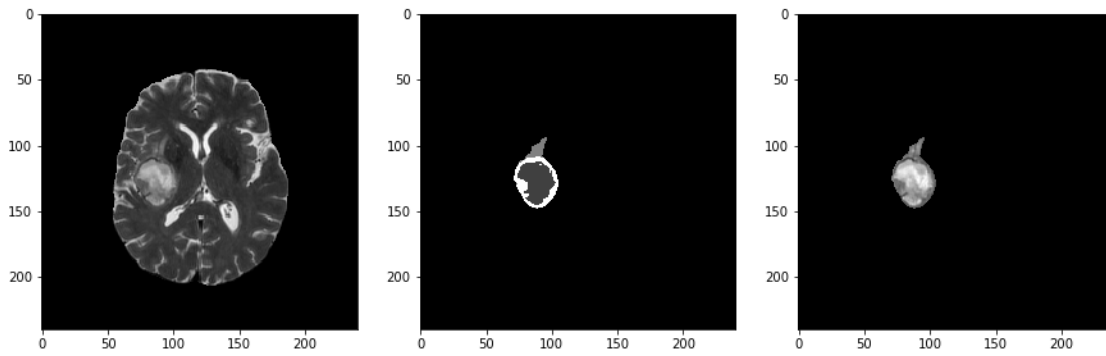
**ID 611**



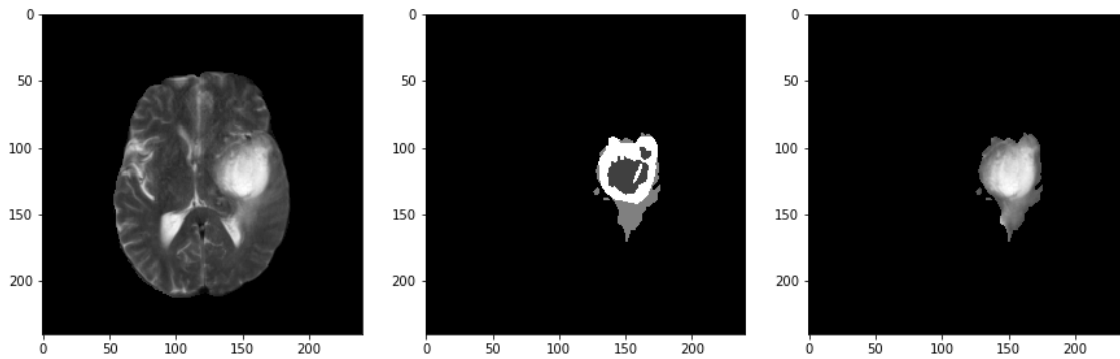
**ID 452**



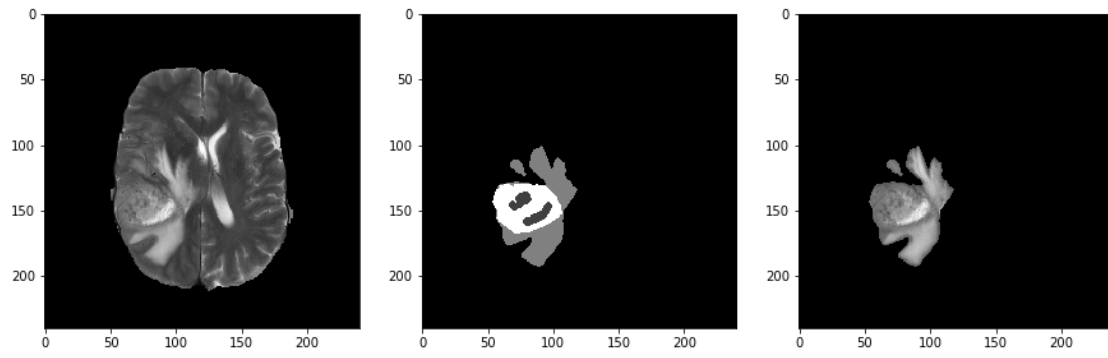
**ID 716**



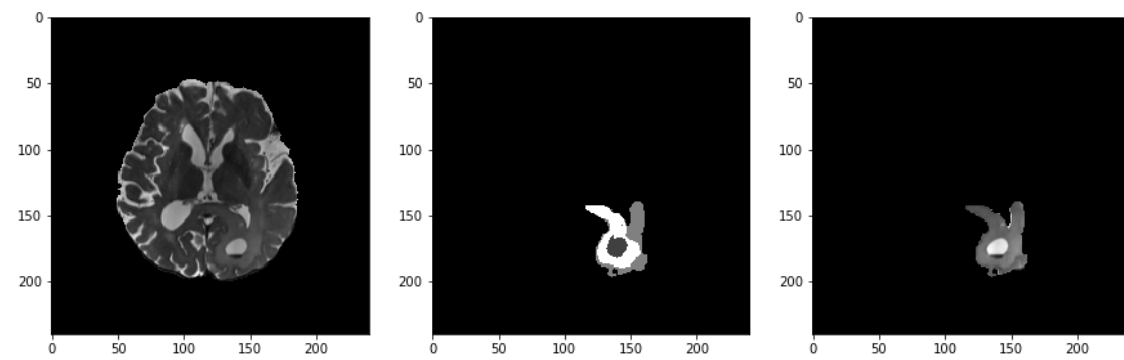
**ID 575**



**ID 808**



**ID 555**



## 5. Conclusión

Tiene sentido que T2 sea el mejor resultado porque es el que muestra el whole tumor, con medio de contraste, es esencial para detectar y caracterizar lesiones. Esta secuencia deja ver tejido como una señal de baja intensidad y, el líquido, como una señal de alta intensidad, por lo que es útil en la identificación de lesiones patológicas que suelen caracterizarse por un aumento en el contenido de agua.

Se obtuvieron resultados favorables utilizando *features* radiómicos, y con un modelo sencillo a diferencia de la bibliografía de base, como lo es la Regresión logística.

Tal vez no estamos teniendo en cuenta la ubicación del tumor y parece ser relevante. Según bibliografía, el patrón de metilación depende de la ubicación en el hemisferio izquierdo o derecho. (Figura 12 a y b, Figura 13). Parece ser relevante en bibliografía cuyos resultados observados sugieren que el patrón de metilación tiene relación con la localización tumoral. (22, 23). Sin embargo, los experimentos detallados en el capítulo 4 no son concluyentes con respecto a la posición tal como lo sugieren estos trabajos.

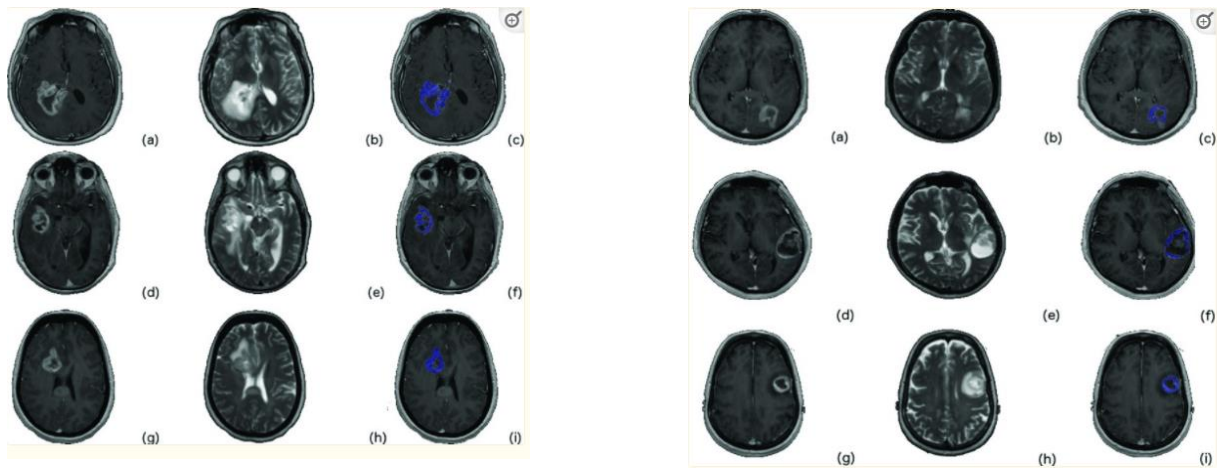


Figura 12: A- Casos de glioblastoma con gen MGMT metilado. Se muestran el T1 posterior al contraste [(a), (d) y (g)], el T2 [(b), (e) y (h)] y el ROI del tumor superpuesto (azul) en el T1 posterior al contraste [(c), (f) y (i)].

B- Casos de glioblastoma con gen MGMT no metilado. Se muestran el T1 posterior al contraste [(a), (d) y (g)], el T2 [(b), (e) y (h)] y el ROI del tumor superpuesto (azul) en el T1 posterior al contraste [(c), (f) y (i)]. (22)

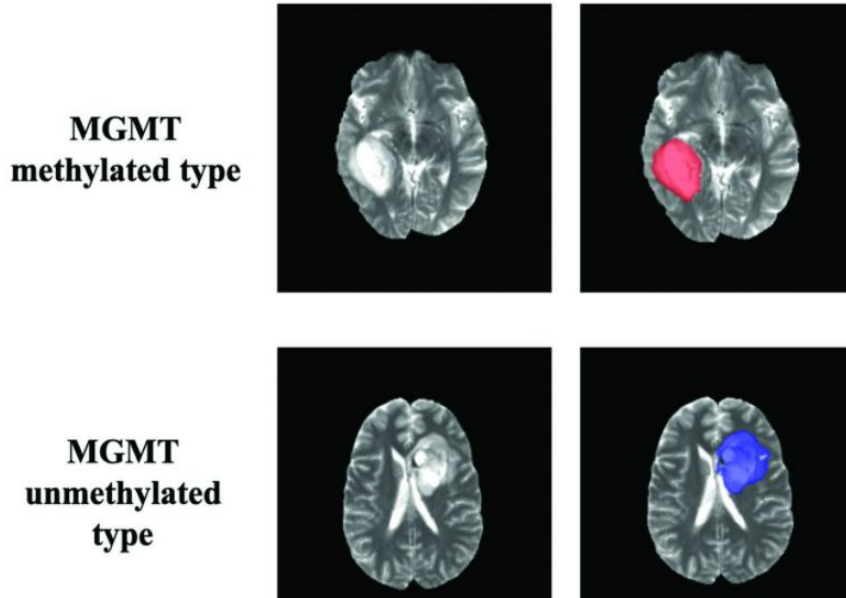


Figura 13: Los vóxeles rojos representan el estado del promotor de MGMT metilado y los vóxeles azules representan el estado del promotor de MGMT no metilado. (23)

A partir de los experimentos realizados y otros resultados en otras bibliografías, como trabajo a futuro, se podría tener en cuenta la posición del tumor en los *features*, realizar *data augmentation* mediante la ubicación del tumor o utilizando VAE, usar análisis de imágenes para el aprendizaje de *features* radiómicos con redes neuronales. También se podría trabajar con información genética que no sea de la biopsia (ej. sangre) y otros datos del paciente para disponer de más variables.

Queda pendiente un contacto con médicos instancia de interacción con expertos oncólogos y radiólogos para analizar estos resultados desde el punto de vista médico, en particular la observación de una posible relación entre metilación y zona necrótica.

## 6. Referencias bibliográficas

- (1) RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge 2021, CBICA, Perelman School of Medicine at the University of Pennsylvania (upenn.edu)
- (2) Niklas Thon, Simone Kreth, Friedrich-Wilhelm Kreth. Personalized treatment strategies in glioblastoma: MGMT promoter methylation status. *OncoTargets and Therapy*. Hospital of the University of Munich, Campus Grosshadern, Munich, Germany, 2013.
- (3) Carlos A Castañeda, Sandro Casavilca. Glioblastoma: Análisis molecular y sus implicancias clínicas. Instituto Nacional de Enfermedades Neoplásicas. Lima, Perú, 2015.
- (4) Gaurav Shukla, Gregory S. Alexander. Advanced magnetic resonance imaging in glioblastoma: a review. *Chinese clinical oncology*. Vol 6, No 4, August 2017.
- (5) Ujjwal Baid, Satyam Ghodasara. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification.
- (6) Singh, G., Manjila, S., Sakla, N. *et al*. Radiomics and radiogenomics in gliomas: a contemporary update. *Br J Cancer* **125**, 641–657 (2021)
- (7) Alexander Selvikvåg Lundervold, Arvid Lundervold. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, Volume 29, Issue 2, Pages 102-127, 2019.
- (8) Han, Lichy, and Maulik R Kamdar. “MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks.” *Pacific Symposium on Biocomputing*. *Pacific Symposium on Biocomputing* vol. 23 (2018): 331-342.
- (9) Rivera, Andreana L et al. “MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma.” *Neuro-oncology* vol. 12,2 (2010): 116-21.
- (10) Bakas, S., Akbari, H., Sotiras, A. et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic *features*. *Sci Data* 4, 170117 (2017)
- (11) Pati, Sarthak & Thakur, Siddhesh & Bhalerao, Megh & Baid, Ujjwal & Grenko, Caleb & Edwards, Brandon & Sheller, Micah & Agraz, Jose & Baheti, Bhakti &

- Bashyam, Vishnu & Sharma, Parth & Haghighi, Babak & Gastounioti, Aimilia & Bergman, Mark & Menze, Bjoern & Kontos, Despina & Davatzikos, Christos & Bakas, Spyridon. (2021). GaNDLF: A Generally Nuanced Deep Learning Framework for Scalable End-to-End Clinical Workflows in Medical Imaging.
- (12) Khan, Adnan A et al. “Deep learning applications in neuro-oncology.” *Surgical neurology international* vol. 12 435. 30 Aug. 2021
- (13) Chen, Xin et al. “Automatic Prediction of MGMT Status in Glioblastoma via Deep Learning-Based MR Image Analysis.” *BioMed research international* vol. 2020 9258649. 23 Sep. 2020
- (14) Andriy Myronenko. 3D MRI brain tumor segmentation using autoencoder regularization. NVIDIA, Santa Clara, CA. 2018
- (15) Sveinn Pálsson , Stefano Cerri, Koen Van Leemput. Prediction of MGMT Methylation Status of Glioblastoma using Radiomics and Latent Space Shape *Features*. Department of Health Technology, Technical University of Denmark, Denmark, 2021.
- (16) Saeed, N. ‘‘Is it Possible to Predict MGMT Promoter Methylation from Brain Tumor MRI Scans using Deep Learning Models?’’. Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
- (17) Van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21), e104–e107.
- (18) Anna Estival Gonzale. Analisis de la metilacion del promotor del gen MGMT en muestras de tejido y sangre de pacientes con glioblastoma, mediante pirosecuenciacion. UAB, Barcelona, 2019
- (19) Rhao, M. Novel Local Radiomic Bayesian Classifiers for NonInvasive Prediction of MGMT Methylation Status in Glioblastoma. Chatham High School Chatham, NJ November 2021.
- (20) Smedley, N. F., & Hsu, W. (2018). USING DEEP NEURAL NETWORKS FOR RADIOGENOMIC ANALYSIS. *Proceedings. IEEE International Symposium on Biomedical Imaging, 2018*, 1529–1533.

- (21) Qian Liu, Pingzhao Hu, Extendable and explainable deep learning for pan-cancer radiogenomics research. *Current Opinion in Chemical Biology*, Volume 66, 2022.
- (22) Korfiatis, P., Kline, T. L., Coufalova, L., Lachance, D. H., Parney, I. F., Carter, R. E., Buckner, J. C., & Erickson, B. J. (2016). MRI texture *features* as biomarkers to predict MGMT methylation status in glioblastomas. *Medical physics*, 43(6), 2835–2844.
- (23) Yogananda, B.R. Shah, S.S. Nalawade, G.K. Murugesan, F.F. Yu, M.C. Pinho, B.C. Wagner, B. Mickey, T.R. Patel, B. Fei, A.J. Madhuranthakam, J.A. Maldjian MRI-Based Deep-Learning Method for Determining Glioma *MGMT* Promoter Methylation Status C.G.B. . *American Journal of Neuroradiology*. Mar 2021
- (24) Chinot, O L et al. “AVAglio: Phase 3 trial of bevacizumab plus temozolomide and radiotherapy in newly diagnosed glioblastoma multiforme.” *Advances in therapy* vol. 28,4 (2011): 334-40. doi:10.1007/s12325-011-0007-3
- (25) Reina GA, Panchumarthy R, Thakur SP, Bastidas A, Bakas S. Systematic Evaluation of Image Tiling Adverse Effects on Deep Learning Semantic Segmentation. *Frontiers in Neuroscience*. 2020 ;14:65

## 7. Anexos

### 7.1 Features Radiómicos

- 1) **original\_shape\_Elongation:** La elongación muestra la relación entre los dos componentes principales más grandes en ROI. Por razones de cálculo, esta característica se define como la inversa del alargamiento verdadero, teniendo en cuenta las longitudes de los ejes del componente principal más grande y el segundo más grande. Los valores oscilan entre 1 (donde la sección transversal a través del primer y segundo momento principal más grande es circular (no alargada) y 0 (donde el objeto tiene una elongación máxima: es decir, una línea de 1 dimensión).
- 2) **original\_shape\_Flatness:** La llanura muestra la relación entre los componentes principales más grandes y más pequeños en ROI. Se tienen en cuenta las longitudes de los ejes de los componentes principales más grande y más pequeño. Los valores oscilan entre 1 (no plano, con forma de esfera) y 0 (un objeto plano o segmentación de un solo corte).
- 3) **original\_shape\_LeastAxisLength:** La longitud mínima del eje produce la longitud de eje más pequeña del elipsoide que encierra el ROI y se calcula utilizando el componente principal más grande. En caso de una segmentación 2D, este valor será 0.
- 4) **original\_shape\_MajorAxisLength:** La longitud del eje principal produce la longitud de eje más grande del elipsoide que encierra el ROI y se calcula utilizando el componente principal más grande.
- 5) **original\_shape\_Maximum2DDiameterColumn:** El diámetro 2D máximo (columna) se define como la distancia euclidiana por pares más grande entre los vértices de malla de la superficie del tumor en el plano de corte por fila.

- 6) **original\_shape\_Maximum2DDiameterRow**: El diámetro 2D máximo (fila) se define como la distancia euclidiana por pares más grande entre los vértices de malla de la superficie del tumor en el plano de corte de columna.
- 7) **original\_shape\_Maximum2DDiameterSlice**: El diámetro 2D máximo (rebanada) se define como la distancia euclidiana por pares más grande entre los vértices de la malla de la superficie del tumor en el plano fila-columna.
- 8) **original\_shape\_Maximum3DDiameter**: El diámetro 3D máximo se define como la distancia euclidiana por pares más grande entre los vértices de la malla de la superficie del tumor. También conocido como diámetro de Feret.
- 9) **original\_shape\_MeshVolume**: El volumen de la malla ROI  $V$  se calcula a partir de la malla triangular del ROI. para cada cara  $i$  en la malla, definida por puntos  $a_i$ ,  $b_i$ ,  $c_i$ , el volumen  $V_i$  del tetraedro definido por esa cara y el origen de la imagen ( $O$ ) es calculado. El signo del volumen está determinado por el signo de la normal, que debe definirse consistentemente ya sea mirando hacia afuera o hacia adentro de la ROI. Luego tomando la suma de todos  $V_i$ , se obtiene el volumen total del ROI.
- 10) **original\_gldm\_GrayLevelNonUniformity**: Falta de uniformidad del nivel de gris mide la similitud de los valores de intensidad de nivel de gris en la imagen, donde un valor de GLN más bajo se correlaciona con una mayor similitud en los valores de intensidad.
- 11) **original\_gldm\_GrayLevelVariance**: Mide la variación en el nivel de gris de la imagen.
- 12) **original\_gldm\_HighGrayLevelEmphasis**: mide la distribución de los valores de nivel de gris más altos; un valor más alto indica una mayor proporción de valores de nivel de gris más altos y zonas de tamaño en la imagen.

- 13) original\_gldm\_LargeDependenceEmphasis:** Una medida de la distribución de grandes dependencias, con un mayor valor indicativo de mayor dependencia y texturas más homogéneas.
- 14) original\_gldm\_LargeDependenceHighGrayLevelEmphasis:** Mide la distribución conjunta de gran dependencia con valores de nivel de gris más altos.
- 15) original\_gldm\_LargeDependenceLowGrayLevelEmphasis:** Mide la distribución conjunta de gran dependencia con valores de nivel de gris más bajos.
- 16) original\_gldm\_LowGrayLevelEmphasis:** Mide la distribución conjunta de tiradas largas con valores de nivel de gris más bajos.
- 17) original\_gldm\_SmallDependenceEmphasis:** Una medida de la distribución de pequeñas dependencias, con un mayor valor indicativo de menor dependencia y texturas menos homogéneas.
- 18) original\_gldm\_SmallDependenceHighGrayLevelEmphasis:** Mide la distribución conjunta de dependencia pequeña con valores de nivel de gris más altos.
- 19) original\_gldm\_SmallDependenceLowGrayLevelEmphasis:** Mide la distribución conjunta de pequeña dependencia con valores de nivel de gris más bajos.