



Pontificia Universidad
JAVERIANA
Cali

**Facultad de Ingeniería
y Ciencias**

Ingeniería Biomédica

INFORME FINAL DE TRABAJO DE GRADO

Sistema de clasificación de la genética tumoral del
Glioblastoma Multiforme

Juan Pablo Castillo Garzón
Julio Cesar Pineda Toro

Director

Dr. Cristian Alejandro Torres Valencia

12 de junio de 2026

Agradecimientos

Agradecemos a Dios por darnos la fortaleza y la perseverancia necesarias para culminar este proceso. En los momentos de mayor incertidumbre, su presencia fue el refugio que nos permitió continuar.

A nuestros padres, por creer en nosotros desde el primer día. Gracias por cada sacrificio, por sus palabras de aliento en los momentos difíciles y por enseñarnos que el esfuerzo honesto siempre tiene recompensa.

A nuestros hermanos, por su compañía incondicional y por celebrar nuestros avances como si fueran propios.

A nuestro director de tesis, el Dr. Cristian Alejandro Torres Valencia, por compartir su conocimiento con generosidad, por guiarnos con criterio y rigor, y por acompañarnos con paciencia a lo largo de todo este proceso.

A la Pontificia Universidad Javeriana Cali y a sus docentes, por una formación que va más allá de lo técnico y que nos convirtió en mejores personas y profesionales.

Finalmente, nos agradecemos mutuamente. Este proyecto fue el resultado de un trabajo en equipo construido con esfuerzo, respeto y compromiso genuino. Cada obstáculo superado fue posible gracias a la confianza que depositamos el uno en el otro.

Agradecimientos especiales de Julio

En primer lugar, quiero expresar mi más sincero agradecimiento a mis padres, quienes desde siempre han sido mi mayor motivación y el pilar más sólido en mi vida. Gracias por inculcarme el valor del esfuerzo y por enseñarme que los sueños se construyen con dedicación y perseverancia.

A mi pareja Mariangely, quien estuvo presente en cada etapa de este proceso. Su compañía y apoyo incondicional fueron fundamentales para atravesar los momentos de mayor dificultad. Gracias por creer en mí incluso cuando yo dudaba.

Y a mi mascota Tita, por quedarse siempre a mi lado durante las largas noches de trabajo, por su compañía silenciosa y por brindarme calma en los momentos de mayor cansancio.

Agradecimientos especiales de Juan Pablo

Agradezco de manera especial, a mis padres, por ser el motor que impulsa cada uno de mis pasos. Gracias por su amor incondicional, por estar presentes en cada momento importante de mi vida y por enseñarme con su ejemplo que todo lo que se hace con dedicación y honestidad vale la pena. Este logro es tan suyo como mío.

Con mucho cariño, a mi pareja Akemy por creer en mí incluso en los momentos en que yo dudaba de mí mismo. Gracias por cada palabra de aliento, por tu paciencia en los días difíciles y por estar siempre a mi lado. Tu apoyo fue un refugio constante a lo largo de este camino.

Glosario

Símbolos

Símbolos

AUC	Área bajo la curva ROC, valor entre 0 y 1
DSC	Coefficiente de Similitud de Dice
MCC	Coefficiente de Correlación de Matthews
IoU	Intersección sobre Unión
p	Probabilidad de pertenencia a la clase mutado
\hat{y}	Predicción binaria del modelo (0 = wildtype, 1 = mutado)
τ	Umbral de clasificación
N	Número total de casos en el conjunto de datos
k	Número de folds en la validación cruzada

Acrónimos y Abreviaturas

AUC	<i>Area Under the Curve</i>
BraTS	<i>Brain Tumor Segmentation Challenge</i>
CAD	<i>Computer-Aided Diagnosis</i>
CNN	<i>Convolutional Neural Network</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DL	<i>Deep Learning</i>
DSC	<i>Dice Similarity Coefficient</i>
FLAIR	<i>Fluid Attenuated Inversion Recovery</i>
GBM	Glioblastoma Multiforme
GPU	<i>Graphics Processing Unit</i>
HD-BET	<i>High-Definition Brain Extraction Tool</i>
IDH	Isocitrato Deshidrogenasa
IoU	<i>Intersection over Union</i>
LR	Regresión Logística
MCC	<i>Matthews Correlation Coefficient</i>
ML	<i>Machine Learning</i>
MONAI	<i>Medical Open Network for AI</i>
MRI	<i>Magnetic Resonance Imaging</i>
NEC	<i>Not Elsewhere Classified</i>
NIFTI	<i>Neuroimaging Informatics Technology Initiative</i>
nnU-Net	<i>No-New-Net</i>
NOS	<i>Not Otherwise Specified</i>
OMS	Organización Mundial de la Salud
PDF	<i>Portable Document Format</i>
RF	<i>Random Forest</i>
ROC	<i>Receiver Operating Characteristic</i>
RMN	Resonancia Magnética Nuclear
SegResNet	<i>Segmentation Residual Network</i>
SVM	<i>Support Vector Machine</i>
T1GD	T1 con contraste de Gadolinio
TN	Verdadero Negativo
TP	Verdadero Positivo
FN	Falso Negativo
FP	Falso Positivo
VAE	<i>Variational Autoencoder</i>

Términos

Desbalance de clases: Condición en la que las clases de un conjunto de datos presentan una distribución desigual. En este trabajo, la proporción entre casos IDH wildtype e IDH mutado es de 6.6:1.

Fine-tuning: Proceso de ajuste fino de los pesos de un modelo preentrenado sobre un nuevo conjunto de datos, adaptando las representaciones aprendidas a una tarea específica.

Glioblastoma Multiforme (GBM): Tumor cerebral primario maligno de mayor grado (Grado IV según la OMS), caracterizado por crecimiento infiltrativo agresivo y una mediana de supervivencia inferior a 15 meses.

Hold-out: Subconjunto de datos completamente independiente reservado exclusivamente para la evaluación final de los modelos, sin participar en ninguna etapa de entrenamiento ni validación.

IDH mutado: Variante del glioblastoma que presenta mutación en el gen IDH1 o IDH2, asociada a mejor pronóstico y mayor respuesta al tratamiento.

IDH wildtype: Variante del glioblastoma sin mutación en el gen IDH, que representa la forma más agresiva y prevalente de la enfermedad.

Pipeline: Secuencia automatizada de etapas de procesamiento computacional ejecutadas de forma encadenada. En este proyecto integra skull stripping, segmentación tumoral, extracción radiómica y clasificación IDH.

Radiómica: Disciplina que extrae características cuantitativas de alta dimensionalidad a partir de imágenes médicas, incluyendo descriptores de textura, forma e intensidad que codifican información sobre la heterogeneidad del tejido tumoral.

Skull stripping: Proceso de eliminación automática del cráneo y tejidos extracraneales en imágenes de resonancia magnética, conservando únicamente el parénquima cerebral.

Transfer learning: Técnica que reutiliza los pesos de un modelo preentrenado en un dominio diferente, adaptándolo mediante fine-tuning a una nueva tarea con datos limitados.

Umbral de clasificación: Valor de probabilidad a partir del cual el modelo asigna un caso a la clase positiva. En este trabajo se utilizó un umbral de 0.5 para todos los modelos evaluados.

Vóxel: Unidad volumétrica tridimensional equivalente al píxel en imágenes 2D. Representa un pequeño volumen de tejido con un valor de intensidad de señal asociado en imágenes de resonancia magnética.

Whole Tumor: Región tumoral completa que incluye el núcleo necrótico, el realce activo y el edema peritumoral, segmentada automáticamente por SegResNet BraTS23 en este trabajo.

Resumen

El glioblastoma multiforme es el tumor cerebral maligno primario más agresivo en adultos, con una mediana de supervivencia de 15 meses. La distinción entre el glioblastoma con mutación del gen isocitrato deshidrogenasa y el de tipo wildtype tiene implicaciones pronósticas y terapéuticas directas, pero su determinación convencional requiere biopsia tisular invasiva. Este trabajo tuvo como objetivo desarrollar un sistema de aprendizaje automático que clasifique de forma no invasiva el estado mutacional del gen isocitrato deshidrogenasa en glioblastoma multiforme a partir de imágenes de resonancia magnética multiparamétrica, como herramienta de soporte diagnóstico.

Se consolidó una base de datos de 527 casos con etiqueta mutacional validada provenientes de tres repositorios públicos internacionales, con un desbalance de clases de 6.6:1 entre ambas variantes. Sobre este conjunto se desarrolló un pipeline automatizado que integra la eliminación del cráneo mediante redes neuronales con HD-BET, la segmentación tumoral automática con SegResNet BraTS23, y la extracción y selección de características radiómicas sobre la secuencia con gadolinio. A partir de 1130 características extraídas se retuvieron 30 descriptores con mayor poder discriminativo para entrenar un modelo de regresión logística. Paralelamente, se entrenaron dos arquitecturas de aprendizaje profundo volumétrico con transferencia de aprendizaje: SegResNet BraTS23 y MedicalNet ResNet-18. Ambos enfoques incorporaron estrategias explícitas de compensación del desbalance de clases durante el entrenamiento.

Los modelos fueron evaluados sobre un conjunto de prueba independiente de 106 casos mediante métricas robustas frente al desbalance. El modelo de regresión logística obtuvo el mejor rendimiento global con área bajo la curva de 0.9138, especificidad del 90.22 por ciento y coeficiente de correlación de Matthews de 0.5953. MedicalNet ResNet-18 con transferencia de aprendizaje alcanzó un área bajo la curva de 0.9014 con sensibilidad del 85.71 por ciento, sugiriendo que ambos enfoques son competitivos y podrían capturar información complementaria sobre la heterogeneidad tumoral asociada al estado mutacional.

Los resultados fueron integrados en una plataforma interactiva que permite cargar imágenes en formatos nativos de resonancia magnética, ejecutar el pipeline completo de inferencia, visualizar la segmentación tumoral mediante un explorador volumétrico interactivo y exportar un reporte con los resultados. Los hallazgos sugieren la viabilidad técnica de clasificar de forma no invasiva el estado mutacional del gen isocitrato deshidrogenasa mediante resonancia magnética, con potencial para complementar los métodos histopatológicos convencionales en entornos de investigación clínica.

Palabras Clave: Glioblastoma multiforme, mutación IDH, resonancia magnética, inteligencia artificial, segmentación tumoral, imágenes médicas, apoyo diagnóstico.

Abstract

Glioblastoma multiforme (GBM) is the most aggressive primary malignant brain tumor in adults, with a median survival of 15 months. The distinction between GBM harboring an isocitrate dehydrogenase gene mutation and the wildtype variant has direct prognostic and therapeutic implications, yet its conventional determination requires invasive tissue biopsy. This work aimed to develop a machine learning system that non-invasively classifies the isocitrate dehydrogenase mutational status in glioblastoma multiforme from multiparametric magnetic resonance imaging, as a diagnostic support tool.

A dataset of 527 cases with validated mutational labels was consolidated from three international public repositories, with a class imbalance of 6.6:1 between both variants. An automated pipeline was developed integrating skull stripping via neural networks with HD-BET, automatic tumor segmentation with SegResNet BraTS23, and extraction and selection of radiomic features from the gadolinium-enhanced sequence. From 1130 extracted features, 30 descriptors with the highest discriminative power were retained to train a logistic regression model. In parallel, two volumetric deep learning architectures with transfer learning were trained: SegResNet BraTS23 and MedicalNet ResNet-18. Both approaches incorporated explicit class imbalance compensation strategies during training.

The models were evaluated on an independent test set of 106 cases using metrics robust to class imbalance. The logistic regression model achieved the best overall performance with an area under the curve of 0.9138, specificity of 90.22 percent and Matthews correlation coefficient of 0.5953. MedicalNet ResNet-18 with transfer learning reached an area under the curve of 0.9014 with a sensitivity of 85.71 percent, suggesting that both approaches are competitive and may capture complementary information about the tumoral heterogeneity associated with the mutational status.

The results were integrated into an interactive platform that allows loading images in native magnetic resonance imaging formats, executing the complete inference pipeline, visualizing the tumor segmentation through an interactive volumetric explorer, and exporting a report with the results. The findings suggest the technical feasibility of non-invasively classifying the isocitrate dehydrogenase mutational status from magnetic resonance imaging, with potential to complement conventional histopathological methods in clinical research settings.

Keywords: Glioblastoma multiforme, IDH mutation, magnetic resonance imaging, artificial intelligence, tumor segmentation, medical imaging, diagnostic support.

Índice general

1. Introducción	1
2. Planteamiento del Problema	3
2.1. Formulación	5
2.2. Sistematización	5
3. Justificación	7
4. Objetivos	9
4.1. Objetivo General	9
4.2. Objetivos Específicos	9
5. Marco de Referencia	11
5.1. Áreas Temáticas	11
5.2. Marco Teórico	11
5.2.1. Contexto clínico — GBM e IDH	11
5.2.2. Imágenes de resonancia magnética en neurooncología	16
5.2.3. Deep Learning volumétrico con transfer learning	20
5.2.4. Preprocesamiento de imágenes médicas	23
5.2.5. Machine Learning para clasificación molecular	25
5.2.6. Métricas de evaluación para clasificación desbalanceada	32
5.2.7. Herramientas computacionales y plataforma	38
5.3. Trabajos Relacionados	38
6. Materiales y Métodos	41
6.1. Base de datos	41
6.1.1. Descripción del dataset	41
6.1.2. Criterios de selección e inclusión	43
6.1.3. Consolidación del conjunto de datos	44
6.2. Preprocesamiento	44
6.2.1. Skull Stripping	45
6.2.2. Segmentación tumoral	45
6.2.3. División del conjunto de datos	47
6.3. Modelo de Machine Learning	48
6.3.1. Extracción y selección de características radiómicas	48
6.3.2. Modelos de clasificación evaluados	51
6.4. Deep Learning	53
6.4.1. Preprocesamiento y aumentación de datos	54

6.4.2.	Estrategia para el desbalance de clases	55
6.4.3.	Arquitectura DL_01: SegResNet como clasificador	55
6.4.4.	Arquitectura DL_02: MedicalNet ResNet-18	56
6.5.	Plataforma interactiva	59
6.5.1.	Serialización y almacenamiento de modelos	59
6.5.2.	Pipeline de inferencia (<code>gbm_pipeline.py</code>)	60
6.5.3.	Interfaz de usuario (<code>app.py</code>)	62
6.6.	Entorno de desarrollo y recursos computacionales	65
6.6.1.	Tiempos de Cómputo del Preprocesamiento	66
7.	Resultados y Discusión	67
7.1.	Resultados del Skull Stripping	67
7.1.1.	Resultados cuantitativos	67
7.1.2.	Análisis visual	68
7.2.	Resultados de la Segmentación Tumoral	69
7.2.1.	Descripción del proceso	69
7.2.2.	Resultados cuantitativos	69
7.2.3.	Análisis visual	70
7.2.4.	Síntesis de resultados Segmentación Tumoral	70
7.3.	Resultados del Clasificador de Machine Learning	71
7.3.1.	Características radiómicas seleccionadas	71
7.3.2.	Resultados de la validación cruzada	72
7.3.3.	Desempeño en el hold-out independiente	72
7.3.4.	Análisis de las matrices de confusión	73
7.3.5.	Análisis de las curvas ROC	78
7.3.6.	Síntesis de resultados Machine Learning	79
7.4.	Deep Learning	80
7.4.1.	SegResNet BraTS23	80
7.4.2.	MedicalNet ResNet-18	83
7.4.3.	Comparación de los modelos ML y Deep Learning	86
7.4.4.	Análisis comparativo de las curvas ROC	88
7.4.5.	Síntesis de resultados	89
7.5.	Resultados de la Plataforma Interactiva	89
7.5.1.	Descripción general de la interfaz	90
7.5.2.	Módulo de carga de imágenes	91
7.5.3.	Configuración y ejecución del análisis	91
7.5.4.	Módulo de resultados	92
7.5.5.	Exploración volumétrica del tumor	93
7.5.6.	Generación del reporte de análisis	94
8.	Conclusiones	97

Índice general	19
9. Trabajos futuros	99
10. Anexos	101
Anexos	101
10.1. Anexo A: Estructura de la Carpeta GBM_TESIS	101
10.2. Anexo B: Códigos de Preprocesamiento	102
10.3. Anexo C: Códigos de Extracción de Características	102
10.4. Anexo D: Códigos de Clasificación ML	103
10.5. Anexo E: Códigos de Clasificación DL	103
10.6. Anexo F: Código de Interfaz	103
10.7. Anexo G: Repositorio GitHub	104
10.8. Anexo H: Manual de Usuario	104
10.9. Anexo I: Video funcionamiento plataforma interactiva	104
Bibliografía	105

Índice de figuras

5.1. Ruta molecular del gen <i>IDH1</i> mutado. Fuente: [1].	15
5.2. Imagen de MRI (Glioma difuso de línea media). Fuente: [2].	17
5.3. Imagen de resonancia magnética ponderada en T1 del cerebro Fuente: [3].	18
5.4. Imagen de resonancia magnética ponderada en T2 del cerebro, donde las estructuras con líquido como el edema aparecen brillantes. Fuente: [3].	18
5.5. Imagen de resonancia magnética FLAIR del cerebro, donde el líquido cefalorraquídeo aparece oscuro y las lesiones perilesionales destacan con mayor contraste. Fuente: [3].	19
5.6. Imágenes de resonancia magnética T1GD Fuente: [4].	20
5.7. Arquitectura de la red Med3D (MedicalNet) Fuente: [5].	21
5.8. Arquitectura de la red ResNet-18, compuesta por 18 capas que incluyen bloques residuales con conexiones de salto (<i>skip connections</i>) para mitigar el problema del gradiente evanescente. Fuente: [6].	22
5.9. Arquitectura SegResNetVAE. Cada bloque verde es un bloque similar a ResNet con normalización de grupo. La rama VAE reconstruye la imagen de entrada en sí misma y se utiliza solo durante el entrenamiento para regularizar el codificador compartido. Fuente: [7].	23
5.10. Representación gráfica de la función sigmoide utilizada en la regresión logística, que transforma valores de entrada en probabilidades entre 0 y 1. Fuente: [8].	28
5.11. Representación gráfica del proceso de predicción del algoritmo Random Forest, donde múltiples <i>decision trees</i> combinan sus resultados para generar una predicción final. Fuente: [9].	29
5.12. Representación gráfica de una SVM con margen suave, ilustrando el hiperplano óptimo de separación, los vectores de soporte y la variable de holgura. Fuente: [10]. . .	30
5.13. Esquema del algoritmo XGBoost, donde cada árbol aprende sobre los residuos del modelo anterior y la predicción final se obtiene como la suma ponderada de todos los árboles. Fuente: [11].	31
5.14. Estrategia de crecimiento <i>leaf-wise</i> de LightGBM, donde en cada iteración se expande únicamente la hoja que produce la mayor reducción de pérdida. Fuente: [12].	32
5.15. Representación gráfica de la curva ROC Fuente: [13].	33
5.16. Representación de una matriz de confusión para clasificación binaria, mostrando las celdas correspondientes a verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (VN). Fuente: [14].	35
5.17. Representación gráfica del Coeficiente de Similitud de Dice (DSC), donde la región intersectada corresponde al solapamiento entre la segmentación predicha (S_1) y la segmentación de referencia (S_2). Fuente: [15].	36

5.18. Ejemplo de cálculo de IoU para tres modelos (A, B y C) entrenados para predecir aves. La región de referencia se muestra en rojo y las predicciones de los modelos en cian. Fuente: [16].	36
6.1. Diagrama de bloques del pipeline completo de clasificación IDH en GBM. Fuente: Propia.	64
7.1. Resultado del skull stripping aplicado con HD-BET sobre tres casos representativos del dataset. Columna izquierda: imagen original. Columna derecha: imagen tras la eliminación del cráneo.	68
7.2. Resultados de la segmentación tumoral automática con SegResNet BraTS23.	70
7.3. Matriz de confusión del modelo de Regresión Logística sobre el hold-out independiente (106 casos, umbral = 0.5).	74
7.4. Matriz de confusión del modelo Random Forest sobre el hold-out independiente (106 casos, umbral = 0.5).	75
7.5. Matriz de confusión del modelo SVM sobre el hold-out independiente (106 casos, umbral = 0.5).	76
7.6. Matriz de confusión del modelo XGBoost sobre el hold-out independiente (106 casos, umbral = 0.5).	77
7.7. Matriz de confusión del modelo LightGBM sobre el hold-out independiente (106 casos, umbral = 0.5).	78
7.8. Curvas ROC de los cinco modelos evaluados sobre el hold-out independiente (106 casos). La Regresión Logística (LR) obtuvo el mayor AUC-ROC (0.9138).	79
7.9. Curvas de entrenamiento del modelo SegResNet BraTS23. De izquierda a derecha: pérdida BCE, AUC-ROC y exactitud sobre los conjuntos de entrenamiento y validación. 81	
7.10. Evaluación del modelo SegResNet BraTS23 sobre el hold-out independiente. Izquierda: curva ROC (AUC=0.8012). Derecha: matriz de confusión con umbral=0.6915. . .	82
7.11. Curvas de entrenamiento del modelo MedicalNet ResNet-18. De izquierda a derecha: pérdida BCE, AUC-ROC y exactitud sobre los conjuntos de entrenamiento y validación. 84	
7.12. Evaluación del modelo MedicalNet ResNet-18 sobre el hold-out independiente. Izquierda: curva ROC (AUC=0.9014). Derecha: matriz de confusión con umbral=0.5. .	86
7.13. Curvas ROC comparativas de los tres modelos evaluados sobre el hold-out independiente (106 casos). Azul: ML — LR + Radiomics (AUC=0.9138). Verde: DL — MedicalNet ResNet-18 (AUC=0.9014). Naranja: DL — SegResNet BraTS23 (AUC=0.8012). 88	
7.14. Vista general de la plataforma GBM IDH Classifier. El encabezado muestra el título del sistema, su descripción y los componentes tecnológicos integrados mediante etiquetas identificadoras.	90
7.15. Módulo de carga de imágenes en modo NIfTI con detección automática de modalidades. El sistema identifica correctamente T1, T1ce, T2 y FLAIR a partir del nombre de cada archivo cargado.	91

7.16. Módulo de configuración del análisis con selección del modelo de clasificación IDH. Se muestra el modelo Deep Learning — MedicalNet ResNet-18 seleccionado y el análisis completado exitosamente.	92
7.17. Módulo de resultados de la plataforma. Se muestra la predicción IDH Mutado con probabilidad del 98.5 %, indicador de confianza Alta y las métricas del modelo utilizado.	93
7.18. Módulo de exploración volumétrica del tumor. El slider axial permite navegar por los cortes del volumen T1ce mostrando la imagen original (izquierda) y la segmentación tumoral superpuesta en rojo (derecha) generada por SegResNet BraTS23.	94
7.19. Reporte de análisis generado automáticamente en formato PDF. Incluye la predicción IDH Mutado con confianza del 98.5 %, AUC del modelo y visualización del corte axial con segmentación tumoral superpuesta.	95

Índice de tablas

6.1. Distribución general del conjunto de datos consolidado.	42
6.2. Distribución de casos por repositorio y estado mutacional IDH.	42
6.3. Proporción de casos según el estado mutacional del gen IDH.	43
6.4. Parametros de adquisicion de imagen por repositorio.	43
6.5. Transformaciones de preprocesamiento aplicadas previo a la segmentación tumoral .	46
6.6. Distribución del conjunto de datos para el modelo de Machine Learning.	47
6.7. Distribución del conjunto de datos para el modelo de Deep Learning.	48
6.8. Parámetros de configuración para la extracción de características radiómicas.	49
6.9. Top 10 características radiómicas seleccionadas por importancia de Random Forest. .	51
6.10. Hiperparámetros de los modelos de Machine Learning evaluados.	53
6.11. Transformaciones de aumentación de datos aplicadas durante el entrenamiento de los modelos de Deep Learning.	54
6.12. Arquitectura del modelo SegResNet clasificador.	56
6.13. Hiperparámetros de entrenamiento del modelo SegResNet clasificador.	56
6.14. Arquitectura completa del modelo MedicalNet ResNet-18 para clasificación IDH. . .	58
6.15. Hiperparámetros de entrenamiento del modelo MedicalNet ResNet-18.	59
6.16. Archivos serializados del pipeline de inferencia.	60
6.17. Principales librerías y versiones utilizadas en el desarrollo del proyecto.	65
6.18. Tiempos de cómputo de las etapas de preprocesamiento ejecutadas sobre GPU NVI- DIA A100-SXM4-40GB en Google Colab.	66
7.1. Resumen del proceso de skull stripping.	67
7.2. Estadísticas de vóxeles cerebrales retenidos tras el skull stripping (secuencia T1GD). .	67
7.3. Métricas de evaluación de la segmentación tumoral sobre 147 casos con máscara de referencia.	70
7.4. Top 10 características radiómicas seleccionadas por importancia de Random Forest. .	71
7.5. Resultados de la validación cruzada estratificada de 5 pliegues sobre los 421 casos del conjunto de desarrollo.	72
7.6. Métricas de evaluación del modelo de Regresión Logística sobre el hold-out indepen- diente (106 casos, umbral = 0.5).	73
7.7. Métricas de evaluación del modelo SegResNet BraTS23 sobre el hold-out indepen- diente (106 casos).	82
7.8. Métricas de evaluación del modelo MedicalNet ResNet-18 sobre el hold-out indepen- diente (106 casos).	85
7.9. Comparación de métricas de evaluación de los tres modelos sobre el hold-out inde- pendiente (106 casos).	87

7.10. Matrices de confusión de los tres modelos sobre el hold-out independiente (106 casos: 92 wildtype, 14 mutados).	88
--	----

Introducción

El Glioblastoma Multiforme (GBM) es el tumor cerebral primario maligno más frecuente y agresivo en adultos, con una supervivencia media inferior a 15 meses tras el diagnóstico y una tasa de supervivencia a cinco años inferior al 5 % [17]. Su elevada mortalidad, combinada con la complejidad de su biología molecular, lo convierte en uno de los mayores desafíos de la neurooncología moderna.

En este contexto, la clasificación del estado mutacional del gen IDH (*Isocitrato Deshidrogenasa*) representa un marcador molecular de especial relevancia clínica. El gen IDH codifica una enzima del ciclo de Krebs responsable de la conversión del isocitrato en alfaetoglutarato. Cuando este gen presenta una mutación, su función cambia y comienza a producir D-2-hidroxi-glutarato, un oncometabolito que altera los patrones de metilación del ADN y favorece la transformación tumoral [18]. Según la clasificación de la Organización Mundial de la Salud de 2021, el estado IDH es un criterio diagnóstico fundamental para la caracterización de los tumores del sistema nervioso central [19], dado que los pacientes con GBM IDH mutado presentan un mejor pronóstico y una mayor respuesta a los tratamientos convencionales en comparación con los casos IDH wildtype, que representan alrededor del 90 % de los casos y la variante más agresiva de la enfermedad [20].

La determinación del estado IDH se realiza actualmente mediante técnicas histopatológicas y moleculares como la inmunohistoquímica y la secuenciación genética, procedimientos que requieren la obtención de tejido tumoral a través de biopsia o resección quirúrgica [21]. Estos métodos, si bien son el estándar diagnóstico actual, implican riesgos inherentes al procedimiento invasivo, costos elevados y tiempos de espera que pueden retrasar el inicio del tratamiento [22]. En este escenario, el desarrollo de métodos no invasivos capaces de inferir el estado mutacional IDH a partir de imágenes de resonancia magnética multiparamétrica representa una alternativa prometedora con alto potencial de impacto clínico [23].

Los avances recientes en inteligencia artificial, particularmente en Machine Learning y Deep Learning, han demostrado una capacidad creciente para identificar patrones sutiles en imágenes médicas que escapan a la percepción visual humana. Su capacidad para discernir aspectos moleculares a partir de la imagenología no solo permite mejorar la exactitud del diagnóstico, sino que también podría reducir la dependencia de procedimientos invasivos y acelerar el tiempo hasta la obtención de un diagnóstico molecular [24]. En particular, la radiómica permite extraer características cuantitativas de alta dimensionalidad a partir de imágenes médicas, codificando información sobre la heterogeneidad del tejido tumoral no perceptible visualmente [25]. Sin embargo, la aplicación de estas técnicas en el contexto del GBM enfrenta desafíos particulares como la escasez de datos con etiquetas moleculares validadas y el marcado desbalance entre las clases IDH wildtype y mutado [20].

El presente trabajo de grado tiene como objetivo general desarrollar un sistema de aprendizaje

automático que, a partir de imágenes de resonancia magnética cerebral, clasifique el tipo de glioblastoma asociado al estado mutacional del gen IDH1 para el soporte diagnóstico. Para alcanzar este objetivo, se plantearon los siguientes objetivos específicos: estructurar una base de datos a partir de imágenes de resonancia magnética cerebral extraídas de repositorios públicos correspondientes a pacientes con glioblastoma multiforme y etiquetadas según el estado mutacional del gen IDH1 entrenar modelos de aprendizaje automático supervisado para clasificar las imágenes según el estado del gen IDH1 (mutado vs. wildtype); evaluar el rendimiento de los modelos utilizando métricas como precisión, sensibilidad, especificidad y AUC sobre un conjunto de imágenes de prueba previamente definidas; e implementar una plataforma interactiva que permita cargar nuevas imágenes de resonancia magnética y observar los resultados generados por el modelo de clasificación.

Para alcanzar estos objetivos, se desarrolló un pipeline automatizado que integra la eliminación del cráneo con HD-BET, la segmentación tumoral con SegResNet BraTS23, la extracción de características radiómicas con PyRadiomics y dos enfoques complementarios de clasificación: Machine Learning basado en Regresión Logística y Deep Learning mediante MedicalNet ResNet-18. El sistema fue entrenado y evaluado sobre un conjunto de datos consolidado de 527 casos con etiqueta IDH validada, provenientes de tres repositorios públicos internacionales (UPENN-GBM [26], UCSD-PTGBM [27] y UCSF-PDGM [28]), y su funcionalidad fue integrada en una plataforma interactiva desarrollada con Streamlit que permite la inferencia sobre nuevas imágenes en formato NIfTI y DICOM.

El documento se estructura de la siguiente manera: el Capítulo 2 presenta el planteamiento del problema, el Capítulo 3 expone la justificación del trabajo, el Capítulo 4 define los objetivos general y específicos, el Capítulo 5 desarrolla el marco de referencia teórico y los trabajos relacionados, el Capítulo 6 describe los materiales y métodos empleados, el Capítulo 7 presenta los resultados y la discusión, y finalmente los Capítulos 8 y 9 exponen las conclusiones y los trabajos futuros del proyecto.

Planteamiento del Problema

El glioblastoma multiforme (GBM) es el tumor maligno primario más común y agresivo en el sistema nervioso central (SNC) en adultos [29]. Se clasifica principalmente en dos grupos: GBM de tipo salvaje con isocitrato deshidrogenasa (IDH), que anteriormente se denominaba GBM primario y representa alrededor del 90% de los casos, y GBM con mutación de IDH, que se desarrolla a partir de un glioma difuso de grado inferior y representa alrededor del 10% de los casos [20]. La distinción molecular entre ambos tipos de GBM no solo tiene implicaciones pronósticas, sino también funcionales. Fisiológicamente, la enzima IDH convierte el isocitrato en α -cetoglutarato (α -KG); sin embargo, cuando esta enzima presenta una mutación, su función cambia y comienza a reducir α -KG a D-2-hidroxiglutarato (D2-HG), un metabolito oncometabólico que se acumula en niveles elevados en los tumores con IDH mutado [18]. Esta acumulación de D2-HG no solo altera el metabolismo celular, sino que también tiene un impacto significativo en la respuesta inmunitaria antitumoral. D2-HG es fácilmente absorbido por los linfocitos T, células inmunitarias esenciales para combatir el cáncer, y una vez dentro, interfiere con su capacidad de reconocer y destruir las células tumorales [30].

El glioblastoma multiforme (GBM) constituye el 14.5% de todas las neoplasias del sistema nervioso central y el 48.6% de los tumores malignos de este sistema. La mediana de supervivencia global (SG) para los pacientes con GBM es de apenas 15 meses [17]. Dado su agresivo comportamiento y alta incidencia, es fundamental considerar los factores que influyen en su desarrollo, siendo la edad uno de los más relevantes. Muchos estudios confirman que la edad afecta significativamente la incidencia de GBM, donde la gran mayoría de los casos ocurren en personas mayores de 40 años. Esto fue confirmado por Kai et al. [31], donde, para el 47.9% de los sujetos, la edad del diagnóstico de GBM fue ≥ 65 años, y el 46.3% de los sujetos tenían entre 40 y 64 años. Aparte, los gliomas malignos son la causa del 2.5% de las muertes por cáncer y son la tercera causa principal de muerte por cáncer en personas de 15 a 34 años de edad [32].

La incidencia global de GBM es <10 por 100.000 personas [33]. Es importante destacar que la incidencia es mayor en hombres que en mujeres: en el 2020, teniendo en cuenta la base de datos Global Cancer Observatory de la Organización Mundial de la Salud, se registran 168.346 nuevos casos de cáncer en cerebro y sistema nervioso central en hombres y 139.756 en mujeres [31]. A nivel geográfico, el 77.6% de los casos se registraron en Suramérica, siendo 10.864 casos en mujeres (47.43%) y 12.052 en hombres (52.59%) [31]. Respecto a estadísticas locales, en Colombia se registra un comportamiento epidemiológico similar, con una incidencia anual ajustada por edad de 3.4 por cada 100.000 hombres y 2.5 por cada 100.000 mujeres [34][35].

Para determinar el grado de malignidad de los gliomas cerebrales, el método de referencia en el diagnóstico es el análisis histopatológico de las muestras tumorales [36]. Estas muestras pueden

obtenerse mediante biopsia o a través de la resección del tumor, la cual puede ser parcial o total. La resección tumoral permite extraer una porción más amplia de la masa neoplásica, lo que favorece una evaluación más precisa en comparación con la biopsia [37]. Por lo tanto, identificar la región de mayor malignidad dentro de un glioma cerebral para las biopsias es esencial, pero representa un reto significativo. La heterogeneidad característica de estos tumores puede llevar a que una selección inadecuada del sitio de biopsia o una muestra insuficiente resulte en la subestimación del grado tumoral, lo que podría derivar en estrategias de tratamiento inadecuadas [38].

Los tratamientos estándar actuales para el GBM incluyen una combinación de resección quirúrgica, radiación y quimioterapia. En la actualidad, solo hay dos medicamentos aprobados por la FDA para tratar el GBM mediante administración sistemática: temozolomida (TMZ) para el tratamiento del GBM de nuevo diagnóstico (ndGBM) y bevacizumab para el tratamiento del GBM recurrente (rGBM) [39][40].

Es fundamental diferenciar los gliomas según su grado de malignidad, ya que los de alto grado tienen un pronóstico significativamente más desfavorable. Además, el enfoque terapéutico varía según el tipo de tumor: mientras que en los gliomas de bajo grado la cirugía suele ser suficiente, en los de alto grado se requiere complementar el tratamiento con quimioterapia y radioterapia, lo que impacta considerablemente la calidad de vida de los pacientes [36]. Por esta razón, es de suma importancia determinar con precisión el grado de malignidad de los gliomas, ya que esto permite establecer el tratamiento más adecuado para maximizar la supervivencia del paciente, preservando en la mayor medida posible su calidad de vida. Una clasificación errónea de un glioma de alto grado como de bajo grado podría impedir la administración de quimioterapia o radioterapia necesarias, lo que agravaría significativamente el pronóstico e incluso podría derivar en un desenlace fatal que podría haberse evitado con el tratamiento adecuado. De igual manera, clasificar incorrectamente un glioma de bajo grado como de alto grado implicaría informar al paciente y a sus familiares sobre una condición más grave de lo que realmente es, generando un impacto emocional y psicológico considerable. Además, esto podría llevar a la aplicación de un tratamiento excesivamente agresivo, con efectos adversos significativos que afectarían negativamente la calidad de vida del paciente [37].

El diagnóstico inicial de un tumor cerebral suele realizarse mediante una tomografía computarizada en un entorno de urgencias, especialmente cuando el paciente presenta síntomas clínicos agudos, como convulsiones o déficits neurológicos focales. Tras la identificación de una anomalía cerebral en la evaluación inicial, se procede a la adquisición de una resonancia magnética anatómica estándar con la administración de un agente de contraste, con el propósito de caracterizar con mayor precisión la lesión y el tejido cerebral circundante. Además, esta imagen sirve como herramienta fundamental en la planificación preoperatoria para la obtención de un diagnóstico histopatológico definitivo [36]. A pesar de su resolución superior en tejidos blandos, la ausencia de radiación ionizante y su capacidad multiplanar, la resonancia magnética anatómica presenta ciertas limitaciones. Entre ellas, destaca su inespecificidad y su incapacidad para representar cambios fisiológicos o metabólicos, los cuales suelen ser cruciales para evaluar la biología tumoral y la respuesta al tratamiento [36]. En particular, los gliomas difusos y heterogéneos representan un desafío, ya que, aunque la RM permite identificar algunas áreas anormales, en muchos casos no es posible delimitar con precisión los bordes tumorales ni identificar con certeza las regiones de mayor grado de malignidad [41].

En este contexto, uno de los desafíos más frecuentemente encontrados en el diagnóstico de tumores cerebrales es el diagnóstico diferencial de un tumor cerebral intraaxial solitario. Determinar si un tumor es benigno o maligno, primario o metastásico, y de bajo o alto grado sigue siendo problemático. A pesar de algunos hallazgos característicos en las imágenes por resonancia magnética, en muchas ocasiones resulta difícil, e incluso imposible, diferenciar entre estos tumores. En tales casos, se hace vital recurrir a una resección macroscópica del tumor o a una biopsia estereotáctica para lograr un diagnóstico preciso [42]. Además, otro problema relevante en el proceso diagnóstico es la existencia de lesiones no neoplásicas del cerebro, como infecciones cerebrales, lesiones desmielinizantes tumefactivas y, con menor frecuencia, infartos, que pueden confundirse con tumores cerebrales. La apariencia de estas lesiones en la resonancia magnética convencional puede ser inespecífica, lo que dificulta su diferenciación de los tumores cerebrales. Incluso con la administración de contraste, la distinción sigue siendo compleja, ya que cualquier proceso que altere la integridad de la barrera hematoencefálica (BHE) puede provocar realce, generando interpretaciones erróneas [36].

2.1. Formulación

¿Cómo puede aprovecharse la información contenida en imágenes de resonancia magnética cerebral para clasificar de manera no invasiva el tipo de GBM asociado al estado mutacional del gen IDH1?

2.2. Sistematización

1. ¿Cómo estructurar una base de datos que combine imágenes de resonancia magnética y datos genéticos para entrenar modelos de aprendizaje profundo?
2. ¿Qué modelos permiten una clasificación efectiva del estado del gen IDH1 a partir de imágenes de resonancia magnética?
3. ¿Cómo se puede medir el desempeño del modelo de clasificación en la predicción del estado mutacional del gen IDH1 utilizando métricas cuantitativas como precisión, sensibilidad, especificidad y AUC?
4. ¿Cómo puede integrarse el modelo de clasificación en una plataforma interactiva que facilite su uso por parte del personal médico en entornos clínicos?

Justificación

El glioblastoma multiforme (GBM) representa la neoplasia maligna primaria más frecuente y agresiva del sistema nervioso central (SNC) en adultos [29]. Este tipo de tumor constituye el 14.5 % de todas las neoplasias que afectan el SNC y el 48.6 % de los tumores malignos dentro de este sistema. A pesar de los avances en su tratamiento, la mediana de supervivencia global (SG) para los pacientes diagnosticados con GBM es de aproximadamente 15 meses [17]. Además, los gliomas malignos en su conjunto representan el 2.5 % de las muertes por cáncer y constituyen la tercera causa principal de mortalidad oncológica en personas de entre 15 y 34 años [32]. Con base en las estadísticas locales, Colombia presenta un comportamiento epidemiológico comparable en relación con los gliomas, con tasas de incidencia anual ajustadas por edad de 3.4 por cada 100.000 hombres y 2.5 por cada 100.000 mujeres [34][35]. Asimismo, la mortalidad estandarizada por edad anual se estima en 2.5 y 1.9 por cada 100.000 hombres y mujeres, respectivamente. En términos de patología tumoral, cerca del 75 % de los casos corresponden a tumores neuroepiteliales, siendo el astrocitoma el tipo histológico más frecuente en ambos sexos, con una prevalencia del 67 % en hombres y 72 % en mujeres [43][44]. Estos datos resaltan la relevancia de los gliomas en la carga oncológica del país y la necesidad de optimizar los enfoques diagnósticos y terapéuticos para mejorar los desenlaces clínicos de los pacientes afectados. En este contexto, es fundamental contar con una infraestructura de salud adecuada para abordar estos desafíos. Para atender casos de glioblastomas, que son un tipo de glioma de alto grado, es crucial que los hospitales o clínicas dispongan de un alto grado de complejidad. Según el Registro Especial de Prestadores de Servicios de Salud (REPS), en Colombia existen 177 instituciones de salud que cuentan con servicios en neurocirugía y oncología clínica, de las cuales 6 están ubicadas en la ciudad de Cali [45]. En el contexto actual del sistema de salud colombiano, se evidencian múltiples barreras que dificultan el acceso oportuno y adecuado a servicios médicos esenciales. Según un informe de la Fundación Retorno Vital, más de 25 millones de personas están afiliadas a Entidades Promotoras de Salud (EPS) que se encuentran intervenidas o en proceso de liquidación. Este panorama ha generado demoras significativas en la atención médica. Se reportan retrasos promedio de 92 días para consultas con especialistas y 36 días para citas de medicina general. Estas deficiencias afectan especialmente a personas con enfermedades oncológicas, autoinmunes y crónicas complejas, comprometiendo la continuidad y eficacia de sus tratamientos [46]. Un caso que ejemplifica esta situación es el cáncer de mama. En un escenario ideal, si una mujer presentara cáncer de mama, este debería ser diagnosticado dentro de los primeros 30 días para iniciar un tratamiento oportuno. No obstante, en Colombia, según la Liga Colombiana contra el Cáncer, el tiempo promedio desde la sospecha hasta la confirmación del diagnóstico es de 50 días, y el inicio del tratamiento suele tardar otros 30 días adicionales [47].

La resonancia magnética (RM) cerebral con contraste es la modalidad de elección cuando se

sospecha un tumor cerebral, idealmente utilizando un protocolo estandarizado de imágenes de tumores cerebrales[48]. Los glioblastomas suelen mostrar realce de contraste en la secuencia ponderada en T1, necrosis central, edema cerebral hiperintenso en T2/FLAIR (recuperación de inversión atenuada por líquido) y signos de efecto de masa[49]. Aun así para un preciso diagnóstico se requiere tejido tumoral. Los pacientes que presentan hallazgos en imágenes que sugieren la presencia de una neoplasia maligna cerebral deben ser remitidos a un centro especializado que cuente con técnicas neuroquirúrgicas avanzadas[22]. Cuando la resección del tumor no es factible, se debe realizar una biopsia para obtener una muestra representativa del tejido tumoral [19]. Esto es fundamental, ya que la gradación precisa de los tumores del SNC permite estimar el pronóstico del paciente y requiere la integración de las características histológicas junto con el análisis de posibles anomalías moleculares[19]. También se requieren pruebas para identificar subtipos de glioma que se caracterizan por alteraciones moleculares específicas, las cuales, además, pueden contribuir a la selección del tratamiento más adecuado [19]. Numerosos centros están adoptando la secuenciación integral del ADN y el análisis del perfil de metilación como métodos para la clasificación de tumores cerebrales [50][21]. Además, se recomienda la remisión a un centro oncológico especializado para realizar un análisis molecular exhaustivo y definir la planificación terapéutica adecuada [21]. En este contexto, la tecnología está desempeñando un papel clave para mejorar estos procesos diagnósticos y terapéuticos. La inteligencia artificial (IA) ha aportado innovaciones transformadoras al manejo de tumores cerebrales, utilizando herramientas de imagenología, histopatológicas y genómicas para una detección, categorización, predicción de resultados y planificación del tratamiento de forma más eficiente [24]. De manera que, los modelos de IA han demostrado superar las evaluaciones humanas en precisión y especificidad. Su capacidad para discernir aspectos moleculares a partir de la imagenología no solo permite mejorar la exactitud del diagnóstico, sino que también podría reducir la dependencia de procedimientos invasivos y acelerar el tiempo hasta la obtención de un diagnóstico molecular [24].

Además, la IA se presenta como una herramienta transformadora en neurooncología, abordando desafíos en diversas etapas de la gestión clínica. En el manejo de tumores cerebrales, su aplicación ha demostrado ser clave en el diagnóstico, pronóstico y planificación del tratamiento, ya que permite acelerar y mejorar la adquisición de imágenes por resonancia magnética, detectar anomalías con mayor precisión, optimizar flujos de trabajo, proporcionar mediciones cuantificables y analizar grandes volúmenes de datos médicos [25]. Asimismo, tiene la capacidad de identificar patrones que no son fácilmente discernibles por los observadores humanos, lo que ha favorecido avances significativos en el análisis detallado de imágenes para diagnósticos más precisos, gradación de tumores, determinación del pronóstico y evaluación de la respuesta al tratamiento [51].

Actualmente, bases de datos (The Cancer Imaging Archive (TCIA) contienen neuroimágenes de pacientes con glioblastoma multiforme, junto con etiquetas que indican el estado mutacional del gen IDH. Estas bases de datos son fundamentales para el desarrollo del presente proyecto, ya que proporcionan información clave para el análisis y la clasificación de los tumores, contribuyendo a una mejor comprensión de la enfermedad y a la optimización de su diagnóstico.

Objetivos

4.1. Objetivo General

Desarrollar un sistema de aprendizaje automático que a partir de imágenes de resonancia magnética cerebral, clasifique el tipo de glioblastoma asociado al estado mutacional del gen IDH1 para el soporte diagnóstico.

4.2. Objetivos Específicos

- Estructurar una base de datos a partir de imágenes de resonancia magnética cerebral extraídas de repositorios públicos, correspondientes a pacientes con glioblastoma multiforme y etiquetadas según el estado mutacional del gen IDH1.
- Entrenar modelos de aprendizaje automático supervisado para clasificar las imágenes según el estado del gen IDH1 (mutado vs. Wild Type).
- Evaluar el rendimiento de los modelos utilizando métricas como precisión, sensibilidad, especificidad y AUC, utilizando un conjunto de imágenes de prueba previamente definidas.
- Implementar una plataforma interactiva que permita cargar nuevas imágenes de resonancia magnética y observar los resultados generados por el modelo de clasificación.

Marco de Referencia

5.1. Áreas Temáticas

Las áreas temáticas que abarca el presente proyecto son las siguientes:

- Computers and Information Processing – Artificial Intelligence – Machine Learning – Supervised Classification Algorithms.
- Computers and Information Processing – Artificial Intelligence – Deep Learning – Transfer Learning and Pretrained Neural Networks.
- Computers and Information Processing – Image Processing and Computer Vision – Volumetric Medical Image Analysis – Brain Tumor Segmentation.
- Biomedical Engineering – Biomedical Signal and Image Processing – Radiomics – Quantitative Feature Extraction from MRI.
- Biomedical Engineering – Clinical Engineering – Molecular Biomarker Identification – Non-Invasive IDH Mutation Classification.
- Life Sciences – Oncology – Neuro-Oncology – Glioblastoma Multiforme Molecular Profiling.
- Health Care – Precision Medicine – Clinical Decision Support Systems – AI-Based Diagnostic Tools.

5.2. Marco Teórico

5.2.1. Contexto clínico — GBM e IDH

- **Tumores:** Masa anormal de tejido que aparece cuando las células se multiplican más de lo debido o no se mueren cuando deberían. Los tumores son benignos (no cancerosos) o malignos (cancerosos). Las masas benignas a veces crecen mucho pero no se diseminan y tampoco invaden los tejidos cercanos ni otras partes del cuerpo. Las masas malignas suelen diseminarse o invadir los tejidos cercanos, y también es posible que se diseminen a otras partes del cuerpo a través de la sangre y el sistema linfático. También se llama neoplasia [52].
- **Tipos de tumores:**

- **Tumores benignos:** Los tumores benignos sólo crecen en una parte del cuerpo, no pueden aparecer ni invadir otras partes del cuerpo (metástasis). Aun así, pueden ser peligrosos si presionan órganos importantes, tales como el cerebro. Los tumores están formados por células extras. En general, las células crecen y se dividen en otras células cuando el cuerpo las necesita. Cuando las células envejecen, mueren y otras células nuevas toman su lugar. Algunas veces, este proceso no resulta como se espera: se forman células nuevas, aun cuando no son necesarias y las células viejas no mueren cuando deberían. Cuando estas células adicionales forman una masa, se les conoce como un tumor. Con frecuencia, el tratamiento incluye cirugía. Los tumores benignos no suelen crecer nuevamente. Diagnóstico y exámenes [53] .
- **Tumores malignos:** El tumor maligno (comúnmente llamado cáncer) es una enfermedad muy complicada caracterizada por el crecimiento descontrolado de células. Las células que normalmente deberían crecer, dividirse y morir pasan por alto este último paso y continúan creciendo, dividiéndose y luego convirtiéndose en un tumor canceroso. Este último, a su vez, puede interferir con los sistemas digestivo, nervioso e incluso circulatorio y, al producir diversas sustancias (hormonas, toxinas y otras), interrumpe el trabajo de todo el cuerpo. Un tumor maligno se caracteriza por que las células se desprenden del tumor primario y se desplazan con la sangre o la linfa a otros lugares del cuerpo, donde, tras su implantación, dan lugar a nuevos tumores. Este proceso se denomina diseminación de la enfermedad y los tumores resultantes se denominan metástasis. Hay más de 100 tipos de cáncer, y el tipo de cáncer está determinado por el tipo de célula afectada. El cáncer puede afectar cualquier parte del cuerpo humano, incluso los ojos o el corazón, lo que suele ser una sorpresa ya que solo se conocen comúnmente los tipos más comunes. Cada tipo de cáncer es diferente, tiene diferentes causas, síntomas y efectos, y se trata de manera diferente. Sin embargo, como con cualquier enfermedad, hay tipos de cáncer que son más comunes que otros [54].
- **Astrocitoma:** Tipo de glioma que empieza en unas células pequeñas en forma de estrella llamadas astrocitos que se encuentran en el encéfalo (tumor cerebral) y la médula espinal. Los astrocitomas pueden ser benignos (no cancerosos) o malignos (cancerosos). Se clasifican en función de su grado, que se determina según cuán anormal es el aspecto de las células tumorales al microscopio y por la rapidez con que el tumor puede crecer, diseminarse y reaparecer después del tratamiento. Los astrocitomas de grado bajo, como los astrocitomas pilocíticos, suelen crecer de forma lenta y son más frecuentes en niños. Los astrocitomas de grado alto, como el glioblastoma, con frecuencia crecen y se diseminan rápido por el encéfalo y la médula espinal, y son más habituales en adultos [55] .
- **Glioblastoma:** El glioblastoma multiforme (GBM) es el subtipo más maligno y extendido de glioma, y el tumor cerebral primario más común en adultos. Sigue siendo una enfermedad incurable con una mediana de supervivencia de 15 meses y se ha clasificado en subtipos de isocitrato deshidrogenasa (IDH) de tipo silvestre y mutante. El único factor causal bien establecido es la exposición a altas dosis de radiación ionizante. El GBM muestra características

de escape inmunitario y una alta heterogeneidad tumoral. Además, el entorno inmunitario relativamente privilegiado, debido a la falta de células presentadoras de antígenos (CPA) y vasos linfáticos en el SNC, contribuye aún más al mal pronóstico en las cohortes con GBM [56].

- **WHO:** Fundada en 1948, la OMS es el organismo de las Naciones Unidas que pone en contacto a naciones, asociados y personas a fin de promover la salud, preservar la seguridad mundial y servir a las poblaciones vulnerables, de modo que todo el mundo, allá donde esté, pueda alcanzar el más alto nivel posible de salud. La OMS lidera los esfuerzos mundiales para ampliar la cobertura sanitaria universal[57].
- **Grados de los astrocitomas (The 2021 WHO Classification of Tumors of the Central Nervous System: clinical implications):** La quinta edición de la Clasificación de Tumores del Sistema Nervioso Central (CNS5) publicada por la Organización Mundial de la Salud (OMS) en 2021 introduce importantes cambios en la categorización de los gliomas, basándose tanto en criterios histológicos como moleculares. En esta nueva clasificación, los glioblastomas se definen exclusivamente como tumores astrocíticos IDH tipo salvaje (IDH wildtype) que presentan características histológicas de alto grado o al menos una de las siguientes alteraciones moleculares: mutación en el promotor de TERT, amplificación del gen EGFR o la combinación de ganancia del cromosoma 7 y pérdida del cromosoma 10 (+7/-10). Incluso si no presentan necrosis o proliferación microvascular, estos tumores pueden ser clasificados como glioblastomas debido a su perfil genético. Por otro lado, los tumores astrocíticos difusos con mutación en IDH se agrupan bajo una única entidad denominada "Astrocitoma, IDH mutado", y se clasifican en grado 2, 3 o 4 dependiendo de criterios histopatológicos y moleculares. En particular, la delección homocigota de CDKN2A/B, detectada mediante técnicas moleculares, se asocia con un peor pronóstico y permite reclasificar un astrocitoma IDH mutado como de grado 4, incluso si no presenta necrosis ni proliferación microvascular [58].
- **Gen IDH1:** Gen que da origen a una proteína que ayuda a descomponer las grasas para producir energía y protege las células de moléculas dañinas. Se han encontrado mutaciones (cambios) en el gen IDH1 en varias afecciones genéticas y en algunos tipos de cáncer, como la leucemia mielógena aguda, los síndromes mielodisplásicos y el cáncer de encéfalo. Es posible que estos cambios hagan que las células anormales, incluso las células cancerosas, se multipliquen y se diseminen por el cuerpo. También se llama gen de la isocitrato deshidrogenasa 1 [59].
- **Ruta molecular del gen IDH1:** Las enzimas isocitrato deshidrogenasa (IDH), de las cuales hay tres isoformas, son enzimas esenciales que participan en varios procesos metabólicos importantes, como el ciclo de Krebs, el metabolismo de la glutamina, la lipogénesis y la regulación redox. IDH1 se encuentra en el citoplasma y los peroxisomas. Los sitios catalíticos de IDH1 e IDH2 exhiben afinidad por el sustrato, isocitrato, junto con nicotinamida adenina dinucleótido fosfato (NADP⁺) y un catión metálico divalente, usualmente magnesio o manganeso, resultando en la formación de α -cetoglutarato (α -KG). IDH3, que también cataliza

la transformación de isocitrato en α -KG, emplea nicotinamida adenina dinucleótido (NAD^+) como su cofactor. La actividad catalítica de la IDH requiere homodimerización junto con una alteración en la conformación enzimática; la unión del isocitrato cambia la estructura de la enzima de una conformación abierta a una cerrada. El reconocimiento del sustrato depende de los residuos de aminoácidos en el sitio activo, mientras que el residuo del sitio activo frecuentemente mutado en el cáncer es la arginina 132 (R132) [1].

- **Gen IDH1 mutado:** Las mutaciones de la isocitrato deshidrogenasa (IDH) son anomalías genéticas comunes en el glioma, que resultan en la acumulación de un “oncometabolito”, el D-2-hidroxisuccinato (D-2-HG). Los niveles anormalmente elevados de D-2-HG resultan en un patrón distintivo en la biología del cáncer, a través de la inhibición competitiva de las dioxigenasas dependientes de α -cetoglutarato (α -KG)/ Fe^{2+} (α -KGDD). Estudios recientes han revelado que la D-2-HG afecta la metilación del ADN/histonas, la señalización de hipoxia, la reparación del ADN y la homeostasis redox, lo cual impacta la oncogénesis de los cánceres con mutación de IDH [60].
- **Ruta molecular del gen IDH1 mutado:** Las mutaciones de IDH que se asocian con el cáncer tienden a localizarse en el residuo de arginina que es crucial para el reconocimiento del isocitrato (R132 para IDH1, R140 o R172 para IDH2). Las mutaciones sin sentido en el gen *IDH1* dan como resultado el reemplazo de un residuo de arginina fuerte y cargado positivamente en la posición 132 con aminoácidos de polaridad más baja, como histidina (H), lisina (K) o cisteína (C), lo que impide la formación de enlaces de hidrógeno con los sitios α -carboxilo y β -carboxilo del isocitrato. Por lo tanto, la enzima IDH mutante exhibe una afinidad disminuida por el isocitrato, junto con una preferencia elevada por NADPH. Sin embargo, solo una copia del gen *IDH* está mutada en tumores y, en células tumorales con mutaciones heterocigotas de *IDH*, se presume que las principales formas de dímeros de IDH son heterodímeros que contienen una versión de IDH1 de tipo silvestre y una versión con la mutación R132H. Como resultado, en células con IDH mutante, el componente IDH1 de tipo silvestre del dímero convierte isocitrato en α -KG para producir NADPH, mientras que la parte mutante del dímero exhibe actividad neomórfica, convirtiendo α -KG en D-2-hidroxisuccinato (D-2-HG) de forma dependiente de NADPH [1]. En la siguiente figura 2.1, se observa la ruta molecular del gen IDH1 mutado

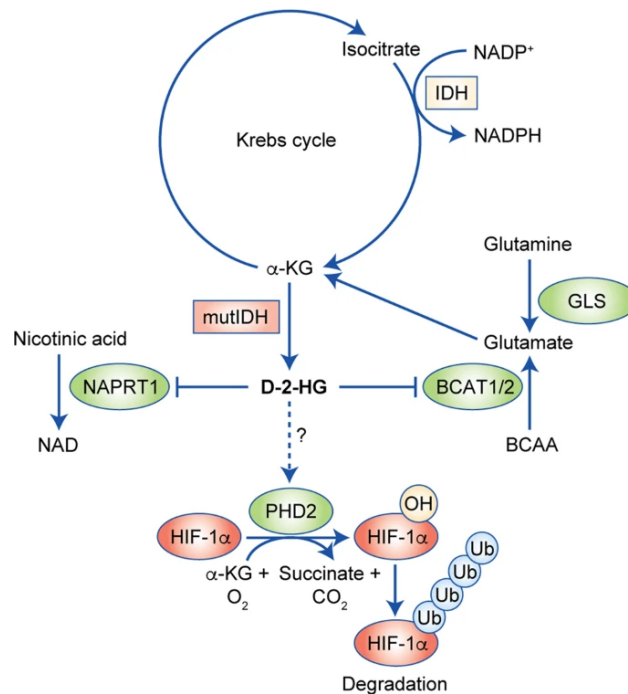


Figura 5.1: Ruta molecular del gen *IDH1* mutado. Fuente: [1].

- **Posibles causas del glioblastoma multiforme:** Las causas subyacentes de estas mutaciones celulares genéticas se desconocen en gran medida. Sin embargo, las investigaciones han revelado que las células de glioblastoma presentan más anomalías genéticas que las células de otros tipos de cáncer cerebral astrocitomaso. Los investigadores creen que diversas mutaciones genéticas intervienen en el desarrollo del glioblastoma. Estas mutaciones genéticas pueden ser causadas por:
 - Defectos hereditarios del ADN.
 - Efectos acumulativos de la exposición a ciertas sustancias química y otros carcinógenos
 - Exposición a dosis altas de radiación ionizante.
 - Desencadenantes adicionales que aún no se han identificado.

El proceso exacto por el cual las células sanas se vuelven cancerosas aún no se comprende por completo [61].

- **Síntomas del glioblastoma multiforme:** Los signos y síntomas del glioblastoma pueden incluir los siguientes [62] :
 - Dolor de cabeza, sobre todo más intenso durante la mañana.
 - Náuseas y vómitos.

- Confusión o deterioro de la función cerebral, como problemas para pensar y comprender información.
 - Pérdida de la memoria.
 - Cambios en la personalidad o irritabilidad.
 - Cambios en la vista, como visión borrosa, visión doble o pérdida de visión periférica.
 - Dificultades con el habla.
 - Dificultad para mantener el equilibrio o la coordinación.
 - Debilidad muscular en la cara, los brazos o las piernas.
 - Reducción de la sensibilidad táctil.
 - Convulsiones, especialmente en personas que nunca las han tenido.
- **Tratamientos del glioblastoma multiforme:** La cirugía es la primera opción de las terapias estandarizadas para GBM. Se recomienda la resección quirúrgica máxima segura para los casos que no son adecuados para la erradicación total, como el GBM invasivo o metastásico. El cuidado postoperatorio estándar también incluye quimioterapia. Solo tres agentes quimioterapéuticos están aprobados por la Administración de Alimentos y Medicamentos (FDA) ahora. La primera clase son nitrosoureas, incluyendo carmustina y lomustina, utilizadas para la quimioterapia de GBM hace más de 40 años. El segundo agente aprobado es la temozolomida (TMZ), que se usa comúnmente en el tratamiento de tumores cerebrales malignos recién diagnosticados. Puede cruzar la barrera hematoencefálica rápidamente, interrumpir la replicación del ADN y causar modificación y reticulación del ADN, lo que principalmente resulta en la apoptosis de células de división rápida ubicadas en el cerebro [63] .

5.2.2. Imágenes de resonancia magnética en neurooncología

- **Imagen por Resonancia Magnética:** La IRM es una tecnología de imágenes no invasiva que produce imágenes anatómicas tridimensionales detalladas, sin el uso de la radiación dañina. Se usa frecuentemente para la detección de enfermedades, el diagnóstico y el monitoreo de tratamientos. Se basa en una tecnología sofisticada que estimula y detecta el cambio en la dirección del eje de rotación de protones que se encuentran en el agua que compone los tejidos vivos. En el cerebro, la IRM puede diferenciar entre la materia blanca y la materia gris, y también puede usarse para diagnosticar aneurismas y tumores. Debido a que la IRM no utiliza rayos X u otra radiación, es la modalidad de imágenes preferida cuando se requieren imágenes frecuentes para el diagnóstico o la terapia, especialmente en el cerebro [64] . En la figura 2.2, se observa un ejemplo de una imagen de MRI.

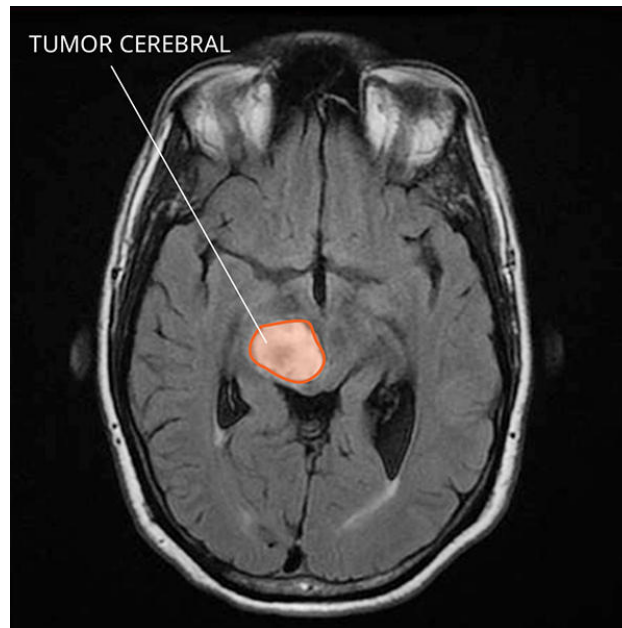


Figura 5.2: Imagen de MRI (Glioma difuso de línea media). Fuente: [2].

- **T1 (*T1-weighted*):** Las imágenes ponderadas en T1 son un tipo de resonancia magnética (RM) comúnmente utilizada en diagnóstico por imagen. En estas imágenes, el contraste está determinado principalmente por el tiempo de relajación longitudinal (T1) de los tejidos, que se relaciona con la rapidez con la que los núcleos de los átomos de hidrógeno regresan a su estado de equilibrio tras ser perturbados por un pulso magnético externo. Los diferentes tejidos presentan distintos tiempos de relajación T1, lo que da lugar a diferentes intensidades de señal en la imagen. Las imágenes ponderadas en T1 generalmente muestran los fluidos oscuros y los tejidos grasos brillantes, lo que las hace útiles para visualizar estructuras anatómicas y ciertas patologías. El contraste se controla mediante un TR corto (400–700 ms) y un TE corto (inferior a 30 ms), parámetros que resaltan los tejidos con tiempos de relajación T1 más cortos, como la grasa, haciéndolos aparecer más brillantes en comparación con otros tejidos [3].

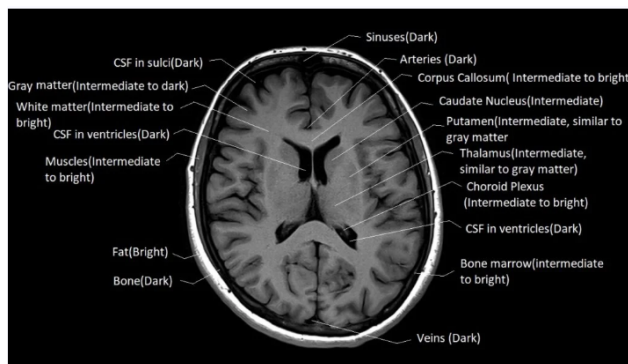


Figura 5.3: Imagen de resonancia magnética ponderada en T1 del cerebro Fuente: [3].

- **T2 (*T2-weighted*):** La resonancia magnética ponderada en T2 es otro tipo común de secuencia de RM utilizada frecuentemente en diagnóstico por imagen. Estas imágenes resaltan el tiempo de relajación T2 de los tejidos, que indica la duración que tardan los núcleos de hidrógeno en perder la coherencia de fase entre espines vecinos tras una perturbación por radiofrecuencia. Los distintos tejidos presentan diferentes tiempos de relajación T2, lo que da lugar a un rango de intensidades de señal en la imagen. Normalmente, las estructuras que contienen líquido, como el edema o el líquido cefalorraquídeo, aparecen brillantes. El contraste se controla mediante un TR prolongado (superior a 2000 ms) y un TE prolongado (60–120 ms), parámetros que proporcionan una visualización clara de los tejidos con tiempos de relajación T2 prolongados, como las regiones llenas de líquido [3].

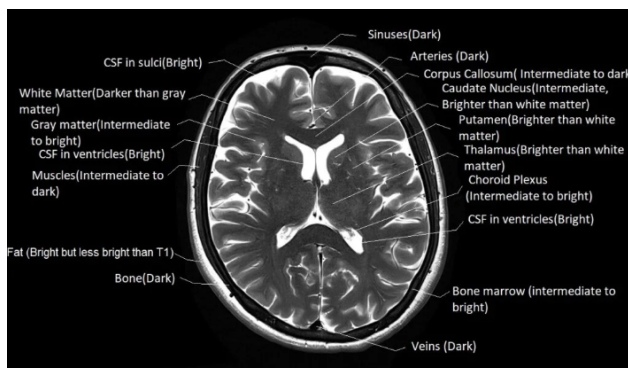


Figura 5.4: Imagen de resonancia magnética ponderada en T2 del cerebro, donde las estructuras con líquido como el edema aparecen brillantes. Fuente: [3].

- **FLAIR (*Fluid Attenuated Inversion Recovery*):** La resonancia magnética FLAIR es una técnica de imagen médica especializada que mejora la detección de anomalías cerebrales y de la médula espinal. Está diseñada para suprimir la señal del líquido cefalorraquídeo (LCR), lo que permite discernir mejor las lesiones cerebrales y medulares. La técnica consiste en un

pulso de recuperación de inversión para anular la señal del LCR, seguido de un retardo y, posteriormente, la adquisición de una imagen ponderada en T2. El resultado es una imagen donde el líquido aparece oscuro, mientras que las alteraciones patológicas en el tejido circundante destacan con mayor contraste, garantizando que las anomalías, a menudo ocultas por la brillante señal del líquido en otras secuencias, se vuelvan más evidentes y fácilmente identificables. Los parámetros clave que determinan el contraste son: un TR largo (superior a 6000 ms), un TE en rango medio-alto (alrededor de 100 ms) y un TI específico (normalmente entre 2000 y 2500 ms) para suprimir eficazmente la señal del LCR [3].

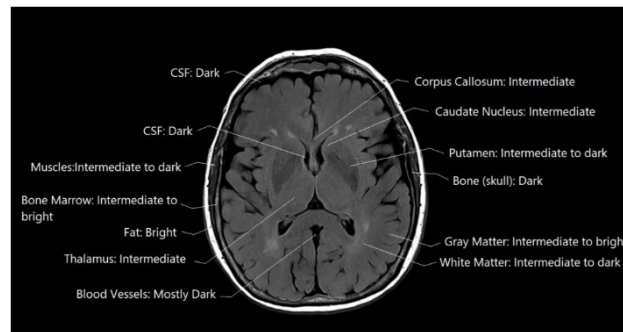


Figura 5.5: Imagen de resonancia magnética FLAIR del cerebro, donde el líquido cefalorraquídeo aparece oscuro y las lesiones perilesionales destacan con mayor contraste. Fuente: [3].

- **T1GD (*T1 con contraste de gadolinio*):** La resonancia magnética T1 con contraste de gadolinio (T1GD) es una variante de la secuencia T1 en la que se administra por vía intravenosa un agente de contraste basado en gadolinio (Gd), un metal de tierra rara paramagnético que acorta significativamente el tiempo de relajación T1 de los tejidos, produciendo un aumento de la intensidad de señal en las imágenes ponderadas en T1 [65]. En condiciones normales, el gadolinio no atraviesa la barrera hematoencefálica (BHE) intacta. Sin embargo, en presencia de tejido tumoral o inflamatorio, la BHE se ve comprometida, permitiendo que el gadolinio se acumule en el espacio extravascular y produzca un realce característico en la región afectada [65]. Esta propiedad hace que la secuencia T1GD sea especialmente útil en neurooncología para identificar la región de realce activo del tumor, delimitar el núcleo necrótico y evaluar la respuesta al tratamiento, ya que el tejido tumoral altamente vascularizado aparece brillante en contraste con el tejido cerebral sano circundante.

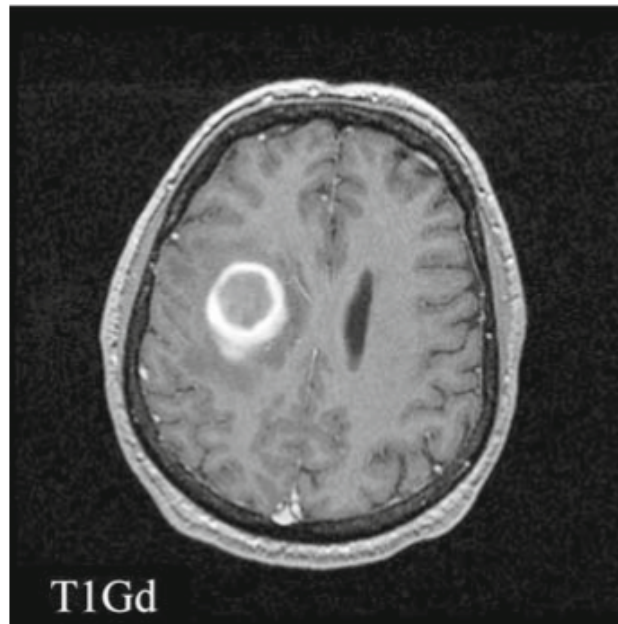


Figura 5.6: Imágenes de resonancia magnética T1GD Fuente: [4].

5.2.3. Deep Learning volumétrico con transfer learning

- **Deep Learning:** El deep learning, por definición, es una subcategoría de la inteligencia artificial que se basa en imitar el funcionamiento del cerebro humano para aprender, razonar y percibir. Se trata de una tecnología basada en redes neuronales artificiales profundas, de ahí el término "deep"(profundo). El deep learning es capaz de procesar grandes cantidades de datos no estructurados y complejos. Cuanto mayor sea el volumen de datos, más precisas serán las conclusiones del modelo de aprendizaje profundo. El deep learning se basa en su capacidad para aprender de forma autónoma a partir de los datos, mejorando su rendimiento con el tiempo y efectuando tareas complicadas que de otro modo serían difíciles o imposibles de realizar con enfoques tradicionales [66] .
- **Transfer Learning (Aprendizaje por Transferencia):** El aprendizaje por transferencia es una técnica en la que el conocimiento adquirido a través de una tarea o conjunto de datos se utiliza para mejorar el rendimiento del modelo en otra tarea relacionada y/o diferente conjunto de datos. En otras palabras, el aprendizaje por transferencia utiliza lo aprendido en un entorno para mejorar la generalización en otro entorno [67].
- **MedicalNet (Med3D):** Para extraer características tridimensionales (3D) médicas generales, Chen et al. diseñaron una red 3D heterogénea llamada Med3D para co-entrenar 3DSeg-8 multidominio y así crear una serie de modelos preentrenados. Los autores transfirieron los modelos preentrenados de Med3D a la segmentación de pulmones en el conjunto de datos LIDC,

la clasificación de nódulos pulmonares en el conjunto de datos LIDC y la segmentación de hígado en el desafío LiTS. Los experimentos demuestran que Med3D puede acelerar la velocidad de convergencia del entrenamiento de tareas médicas 3D objetivo en 2 veces en comparación con un modelo preentrenado en el conjunto de datos Kinetics, y en 10 veces en comparación con el entrenamiento desde cero, además de mejorar la precisión entre un 3% y un 20%. Al transferir el modelo Med3D a la red de segmentación DenseASPP de última generación, en el caso de un solo modelo, se logró un coeficiente Dice del 94.6%, aproximándose al resultado de los algoritmos de mayor rango en el desafío LiTS [5].

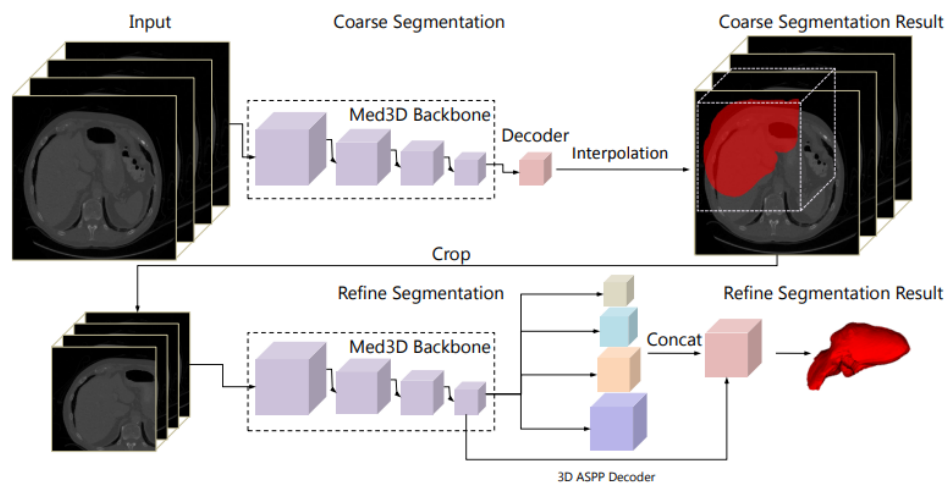


Figura 5.7: Arquitectura de la red Med3D (MedicalNet) Fuente: [5].

- **ResNet-18:** El modelo ResNet-18 consta de 18 capas y es una variante de la arquitectura de Red Residual (ResNet). Los bloques residuales son los componentes básicos de ResNet e incluyen conexiones de salto (*skip connections*) que evitan una o más capas. Estos bloques ayudan a mitigar el problema del gradiente evanescente, ya que permiten que la red aprenda mapeos residuales en lugar de los originales, facilitando el flujo del gradiente durante la retropropagación [68]. Un bloque residual en ResNet-18 consta de:

- Dos capas convolucionales.
- Normalización por lotes (*Batch Normalization*).
- Activación ReLU.

La conexión de salto agrega la entrada del bloque a la salida de la segunda capa convolucional, lo que ayuda a la red a aprender características residuales, mejorando la convergencia y el rendimiento general [68].

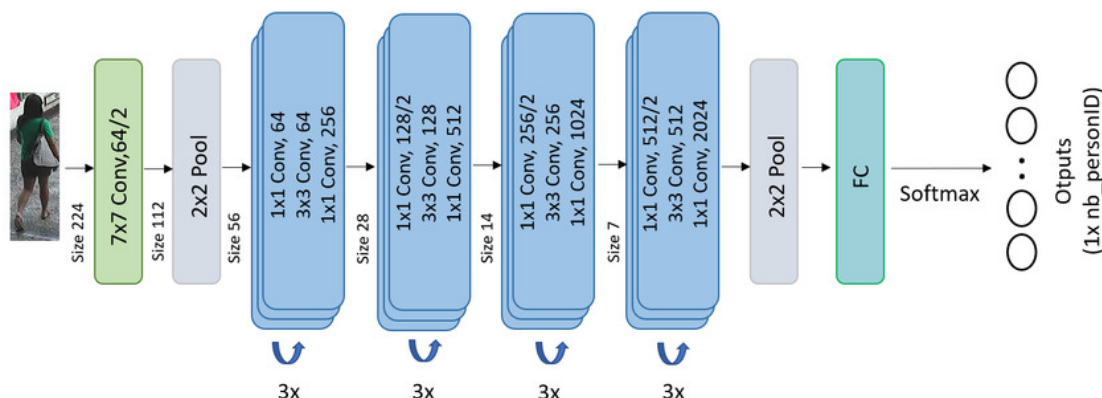


Figura 5.8: Arquitectura de la red ResNet-18, compuesta por 18 capas que incluyen bloques residuales con conexiones de salto (*skip connections*) para mitigar el problema del gradiente evanescente. Fuente: [6].

- SegResNet 3D:** Se trata de una U-Net residual con regularización mediante autoencoder, ganadora del desafío BraTS 2018. Esta U-Net se modifica mediante el diseño de una nueva arquitectura para los bloques del codificador y la adición de una rama de autoencoder variacional (VAE) en el decodificador, que reconstruye la entrada y ejerce un efecto de regularización [7].

El codificador utiliza bloques similares a los de ResNet, donde cada bloque consta de dos convoluciones con normalización de grupo y activación ReLU, seguidas de una conexión de salto de identidad aditiva. La estructura del decodificador es similar a la del codificador, pero con un único bloque por cada nivel espacial. Cada bloque del decodificador comienza reduciendo el número de canales a la mitad (con una convolución de $1 \times 1 \times 1$) y duplicando la dimensión espacial (mediante una convolución bilineal 3D), seguido de una suma con el mapa de características del codificador del nivel espacial equivalente [7].

En la rama VAE del decodificador, primero el mapa de características del cuello de botella se reduce a un espacio de baja dimensión de 256 (128 para representar la media y 128 para representar la desviación estándar). Luego, se extrae una muestra de la distribución gaussiana con la media y la desviación estándar dadas, y se reconstruye en las dimensiones de la imagen de entrada siguiendo la misma arquitectura que el decodificador [7].

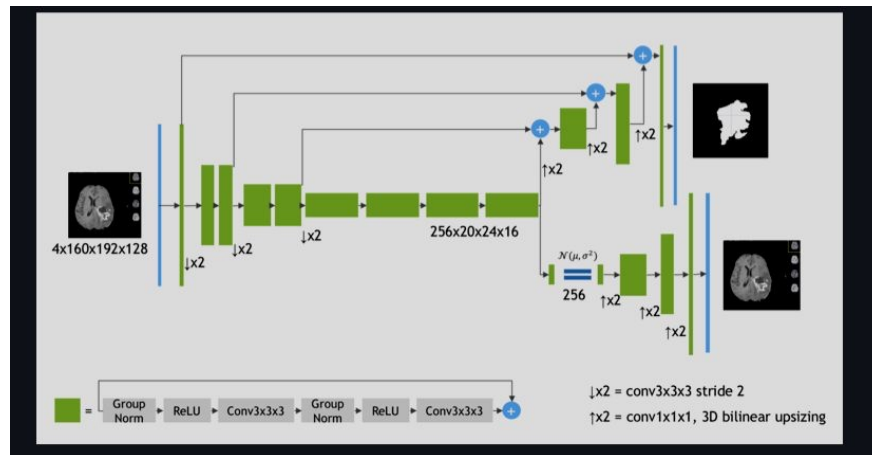


Figura 5.9: Arquitectura SegResNetVAE. Cada bloque verde es un bloque similar a ResNet con normalización de grupo. La rama VAE reconstruye la imagen de entrada en sí misma y se utiliza solo durante el entrenamiento para regularizar el codificador compartido. Fuente: [7].

- Data Augmentation:** El Data Augmentation utiliza datos preexistentes para crear nuevas muestras que pueden mejorar la optimización y la generalización del modelo. En su sentido más general, se refiere a los métodos para complementar conjuntos de datos incompletos mediante la adición de datos faltantes, con el fin de mejorar su capacidad de análisis. Esto se manifiesta en el machine learning mediante la generación de copias modificadas de datos preexistentes para aumentar el tamaño y la diversidad de un conjunto de datos. En otras palabras, puede reducir el sobreajuste y mejorar la robustez del modelo, siendo especialmente útil para corregir conjuntos de datos desequilibrados [69].

5.2.4. Preprocesamiento de imágenes médicas

- Skull Stripping:** La eliminación de la señal no cerebral de los datos de resonancia magnética (RM), conocida como extracción del cráneo, es un componente integral de muchos flujos de análisis de neuroimágenes. A pesar de su abundancia, los métodos clásicos de extracción del cráneo más populares suelen estar adaptados a imágenes con propiedades de adquisición específicas, concretamente resolución casi isotrópica y contraste de RM ponderada en T1 (T1w), que son frecuentes en entornos de investigación. Como resultado, las herramientas existentes tienden a adaptarse mal a otros tipos de imágenes, como pilas de cortes gruesos adquiridos con RM de eco de espín rápido (FSE), comunes en la clínica. Si bien los enfoques basados en aprendizaje para la extracción del cerebro han ganado terreno en los últimos años, estos métodos se enfrentan a una carga similar, ya que solo son efectivos para los tipos de imágenes vistos durante el procedimiento de entrenamiento [70].
- HD-BET:** HD-BET es un algoritmo de extracción cerebral basado en redes neuronales artificiales, desarrollado para superar las limitaciones de las herramientas clásicas de *skull stripping*

en presencia de patología cerebral y datos multicentro [71].

- HD-BET se desarrolló con datos de resonancia magnética (RM) de un amplio ensayo clínico multicéntrico en pacientes adultos con tumores cerebrales, recopilados en 37 instituciones europeas. El estudio incluyó una amplia gama de hardware y parámetros de adquisición de RM, patologías y alteraciones tisulares inducidas por el tratamiento. Se utilizaron dos tercios de los datos para entrenamiento y validación, y un tercio para pruebas. Además, se realizó una prueba independiente de HD-BET en tres conjuntos de datos de referencia públicos (NFBS, LPBA40 y CC-359) [71].
 - HD-BET se entrenó con secuencias T1-w precontraste, T1-w postcontraste, T2-w y FLAIR. Puede realizar la extracción cerebral de forma independiente en diversas secuencias de resonancia magnética y no se limita a las secuencias ponderadas en T1 (T1-w) precontraste [71].
 - HD-BET fue diseñado para ser robusto con respecto a tumores cerebrales, lesiones y cavidades de resección, así como a diferentes equipos de resonancia magnética y parámetros de adquisición [71].
 - HD-BET superó a cinco algoritmos de extracción cerebral disponibles públicamente (FSL BET, AFNI 3DSkullStrip, Brainsuite BSE, ROBEX y BEaST) en todos los conjuntos de datos y produjo mejoras medianas de +1.33 a +2.63 puntos para el coeficiente DICE y de -0.80 a -2.75 mm para la distancia de Hausdorff ($p < 0.001$ ajustado con Bonferroni) [71].
 - HD-BET es muy rápido en GPU, con un tiempo de ejecución inferior a 5 segundos por secuencia de resonancia magnética. Incluso en CPU, no es más lento que otras herramientas de uso común [71].
- **Segmentación tumoral:** La segmentación de tumores cerebrales es una técnica para localizar y etiquetar la posición y los límites precisos de un tumor cerebral en imágenes médicas, particularmente utilizando imágenes de resonancia magnética (IRM). Según el tipo de exploración, como T1, T2, FLAIR o T1-CE, los tumores pueden aparecer claramente al observar estas imágenes precisas del cráneo. En estas exploraciones, la segmentación se refiere a la identificación de la célula maligna o tumoral de la estructura cerebral normal y, ocasionalmente, incluso a la división del tumor en áreas distintas, como el núcleo, el borde activo y la inflamación circundante [72].
- **BraTS (Brain Tumor Segmentation Challenge):** BraTS es una competencia científica internacional organizada en el marco del congreso MICCAI, que reúne a investigadores de todo el mundo para desarrollar y comparar algoritmos de inteligencia artificial aplicados al análisis de tumores cerebrales en imágenes de resonancia magnética [73]. Desde su lanzamiento en 2012, el desafío de Segmentación de Tumores Cerebrales (BraTS) se ha dedicado a evaluar métodos avanzados para la segmentación de tumores cerebrales a partir de imágenes de RMN multiparamétricas (mpMRI). Inicialmente, BraTS tenía como objetivo proporcionar

un conjunto de datos público y establecer un punto de referencia para la comunidad. Utilizó principalmente imágenes de mpMRI preoperatorias procedentes de diversas instituciones, centrándose en la segmentación de tumores cerebrales, especialmente gliomas, que presentan una amplia variabilidad en apariencia, forma e histología. Además, entre 2017 y 2020, BraTS amplió su alcance para incluir predicciones de la supervivencia global (SG) de pacientes con glioblastoma tras resección total macroscópica. Esta ampliación permitió la integración del análisis de características radiómicas y algoritmos de aprendizaje automático para resaltar la importancia clínica de la tarea de segmentación [73].

En las iteraciones de 2021-2022, BraTS continuó concentrándose en la segmentación de subregiones de gliomas, respaldada por un conjunto de datos sustancialmente mayor. El desafío de 2021 también introdujo la tarea clínica crítica de determinar el estado de metilación del promotor de la O6-metilguanina-ADN metiltransferasa (MGMT) en tumores, clasificándolos como metilados o no metilados [73].

BraTS 2023 perfeccionó aún más su marco de evaluación comparativa y amplió su conjunto de datos para abarcar una demografía de pacientes más amplia, incluyendo regiones como África subsahariana, varios tipos de tumores como los meningiomas y nuevos desafíos clínicos como el manejo de datos incompletos. Esta edición introdujo varios subdesafíos que abordan necesidades clínicas específicas y problemas técnicos, como el aumento de datos [73].

La edición 2024 de BraTS amplió aún más su alcance de investigación al introducir nuevos subdesafíos, incluidos aquellos centrados en los resultados del glioma después del tratamiento. El desafío de Patología específicamente tiene como objetivo desarrollar modelos de aprendizaje profundo capaces de identificar características histológicas distintivas dentro de las subregiones del tumor, mejorando así la precisión del diagnóstico y la clasificación de los tumores cerebrales [73].

5.2.5. Machine Learning para clasificación molecular

- **Machine Learning:** El machine learning es la ciencia de desarrollo de algoritmos y modelos estadísticos que utilizan los sistemas de computación con el fin de llevar a cabo tareas sin instrucciones explícitas, en vez de basarse en patrones e inferencias. Los sistemas de computación utilizan algoritmos de machine learning para procesar grandes cantidades de datos históricos e identificar patrones de datos. Esto les permite generar resultados con mayor precisión a partir de un conjunto de datos de entrada [74].
- **Radiómica:** Este marco basado en datos descubre información cuantitativa en las imágenes; se encarga de extraer características más allá de lo visualmente perceptible utilizando estadísticas computacionales para predecir una asociación con un criterio de valoración clínico significativo. Las firmas radiómicas están asociadas con los resultados de procesos biológicos; y pueden utilizarse junto con parámetros clínicos, histológicos y genómicos como parte de un nomograma de características [75].

■ **Características radiómicas de textura:**

- **Matriz de coocurrencia de niveles de gris (GLCM):** Una matriz de coocurrencia de niveles de gris (GLCM) de tamaño $N_g \times N_g$ describe la función de probabilidad conjunta de segundo orden de una región de imagen restringida por la máscara y se define como $P(i, j | \delta, \theta)$. El elemento (i, j) de esta matriz representa el número de veces que la combinación de niveles i y j ocurre en dos píxeles de la imagen, que están separados por una distancia de δ píxeles a lo largo del ángulo θ . La distancia δ desde el vóxel central se define como la distancia según la norma infinito. Para $\delta = 1$, esto resulta en 2 vecinos para cada uno de los 13 ángulos en 3D (conectividad 26) y para $\delta = 2$ una conectividad de 98 (49 ángulos únicos) [76].
- **Matriz de longitud de ejecución de niveles de gris (GLRLM):** El GLRLM proporciona información sobre la distribución espacial de secuencias de píxeles consecutivos con el mismo nivel de gris, en una o más direcciones, en 2 o 3 dimensiones. Las características del GLRLM incluyen la fracción, que evalúa el porcentaje de píxeles o vóxeles dentro de la ROI que son parte de las secuencias y, por lo tanto, reflejan la granularidad; los momentos de énfasis de secuencias largas y cortas, que se ponderan hacia la presencia de números de secuencias largas y cortas, respectivamente; y la no uniformidad del nivel de gris y la longitud de la secuencia, que evalúa la distribución de secuencias en diferentes niveles de gris y longitudes de secuencia, respectivamente [77].
- **Matriz de zonas de tamaño de nivel de gris (GLSZM):** El GLSZM se basa en un principio similar al GLRLM, pero aquí los recuentos del número de grupos de píxeles o vóxeles vecinos interconectados con el mismo nivel de gris forman la base de la matriz. Una textura más homogénea dará como resultado una matriz más ancha y plana. Las características del GLSZM pueden calcularse en 2 dimensiones (8 píxeles vecinos) o 3 dimensiones (26 vóxeles vecinos) e incluyen la fracción, el énfasis en zonas grandes y pequeñas, entre otras [77].
- **Matriz de diferencia de tonos grises del vecindario (NGTDM):** El NGTDM cuantifica la suma de las diferencias entre el nivel de gris de un píxel o vóxel y el nivel de gris medio de sus píxeles o vóxeles vecinos dentro de una distancia predefinida. Las características clave incluyen la rugosidad, la complejidad y la actividad. La rugosidad refleja la diferencia de nivel de gris entre el píxel o vóxel central y su vecindario, capturando la tasa espacial de cambios en las intensidades; mientras que la actividad refleja cambios rápidos en el nivel de gris entre el píxel o vóxel central y sus vecinos [77].
- **Matriz de dependencia de niveles de gris (GLDM):** El GLDM se basa en la relación de nivel de gris entre un píxel o vóxel central y su vecindario. Un píxel o vóxel vecino dentro de una distancia predefinida se considera conectado al píxel o vóxel central si cumple el criterio de dependencia en términos de un rango definido de diferencias de nivel de gris. Las características del GLDM incluyen el énfasis de dependencia grande y pequeña, que reflejan heterogeneidad y homogeneidad, así como la no uniformidad de nivel de gris y la uniformidad de dependencia [77].

- **Características de primer orden (*First Order Statistics*):** Las estadísticas de primer orden describen la distribución de las intensidades de los vóxeles dentro de la región de la imagen definida por la máscara mediante métricas básicas y de uso común. Sea X un conjunto de N_p vóxeles incluidos en la ROI, $P(i)$ el histograma de primer orden con N_g niveles de intensidad discretos, donde N_g es el número de intervalos distintos de cero, igualmente espaciados desde 0 con un ancho definido en el parámetro `binWidth`, y $p(i)$ el histograma de primer orden normalizado e igual a $\frac{P(i)}{N_p}$ [76].
- **Características de forma (*Shape Features 3D*):** Las características de forma tridimensionales son descriptores geométricos que cuantifican el tamaño y la morfología de la región de interés (ROI) independientemente de los valores de intensidad de la imagen [76]. En este grupo de características se incluyen descriptores del tamaño y la forma tridimensionales de la región de interés (ROI). Estas características son independientes de la distribución de intensidad del nivel de gris en la ROI y, por lo tanto, se calculan únicamente sobre la imagen y la máscara no derivadas [76].

Salvo que se especifique lo contrario, las características se derivan de la forma aproximada definida por la malla triangular. Para construir esta malla, primero se definen los vértices como puntos situados a mitad de camino en una arista entre un vóxel incluido en la ROI y otro fuera de ella. Al conectar estos vértices, se obtiene una malla de triángulos conectados, donde cada triángulo está definido por 3 vértices adyacentes y comparte cada lado con exactamente otro triángulo [76].

Esta malla se genera mediante un algoritmo de cubos marchantes. En dicho algoritmo, un cubo de 2×2 se desplaza por el espacio de la máscara. Para cada posición, las esquinas del cubo se marcan como segmentadas (1) o no segmentadas (0). Al tratar las esquinas como bits específicos en un número binario, se obtiene un índice único del cubo (0-255), el cual se utiliza para determinar qué triángulos están presentes en el cubo, definidos en una tabla de consulta [76].

- **Filtro Wavelet:** El filtrado de wavelet produce 8 descomposiciones por nivel, correspondientes a todas las combinaciones posibles de aplicar un filtro de paso alto o de paso bajo en cada una de las tres dimensiones [76].
 - **Filtro LoG (Laplaciano de Gauss):** Es un filtro de realce de bordes que enfatiza las áreas de cambio de nivel de gris, donde el parámetro sigma define cuán gruesa debe ser la textura enfatizada. Un valor bajo de sigma enfatiza las texturas finas (cambio en una distancia corta), mientras que un valor alto de sigma enfatiza las texturas gruesas (cambio de nivel de gris en una distancia grande) [76].
- **Regresión Logística (LR):** La regresión logística es un algoritmo de aprendizaje automático supervisado que predice un resultado discreto o categórico. El modelo predice la probabilidad de que ocurra un resultado específico, con un valor de salida entre 0 y 1. Basándose en un umbral —generalmente 0.5— clasifica el resultado en una de dos categorías. En lugar de ajustar una línea recta a través de los datos, como en la regresión lineal, la regresión logística

ajusta una curva en forma de S para relacionar los valores de entrada con una probabilidad [8].

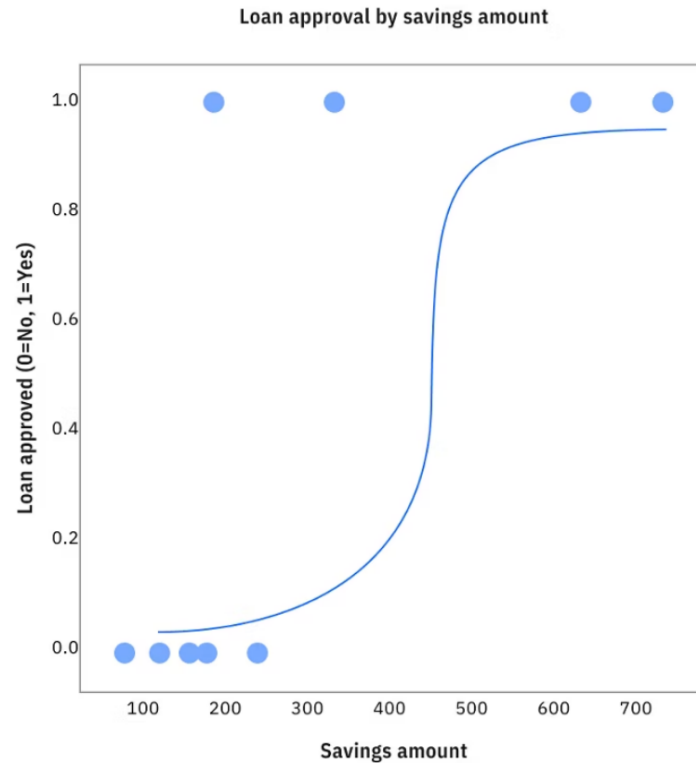


Figura 5.10: Representación gráfica de la función sigmoide utilizada en la regresión logística, que transforma valores de entrada en probabilidades entre 0 y 1. Fuente: [8].

- **Random Forest (RF):** El algoritmo Random Forest es un algoritmo de aprendizaje automático de uso común, desarrollado por Leo Breiman y Adele Cutler, que combina la salida de múltiples *decision trees* para obtener un único resultado. Es una extensión del método de bagging, ya que utiliza tanto bagging como aleatoriedad de características para crear un bosque no correlacionado de *decision trees*. La aleatoriedad de características genera un subconjunto aleatorio de características, garantizando una baja correlación entre los *decision trees*. Esta es una diferencia clave entre los *decision trees* y el Random Forest: mientras que los *decision trees* consideran todas las posibles divisiones de características, el Random Forest solo selecciona un subconjunto de ellas [78].

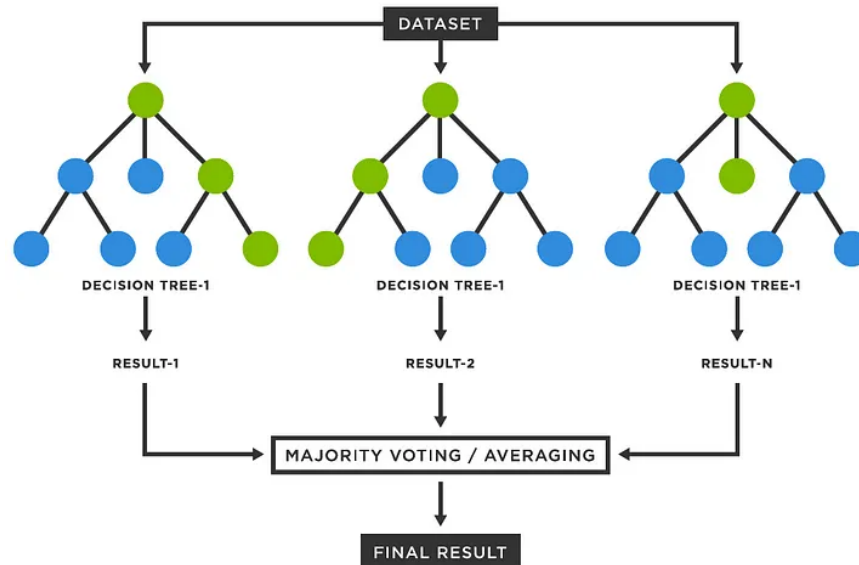


Figura 5.11: Representación gráfica del proceso de predicción del algoritmo Random Forest, donde múltiples *decision trees* combinan sus resultados para generar una predicción final. Fuente: [9].

- Support Vector Machine (SVM):** Una máquina de vectores de soporte (SVM) es un algoritmo de aprendizaje automático supervisado que clasifica los datos encontrando una línea o hiperplano óptimo que maximiza la distancia entre cada clase en un espacio N-dimensional. Las SVM se utilizan comúnmente en problemas de clasificación, distinguiendo entre dos clases al encontrar el hiperplano óptimo que maximiza el margen entre los puntos de datos más cercanos de clases opuestas. El número de características en los datos de entrada determina si el hiperplano es una línea en un espacio bidimensional o un plano en un espacio n-dimensional. Dado que se pueden encontrar múltiples hiperplanos para diferenciar clases, maximizar el margen entre puntos permite al algoritmo encontrar el límite de decisión óptimo entre clases, lo que le permite generalizar bien a nuevos datos y realizar predicciones de clasificación precisas. Las líneas adyacentes al hiperplano óptimo se conocen como vectores de soporte, ya que estos vectores pasan por los puntos de datos que determinan el margen máximo [10].

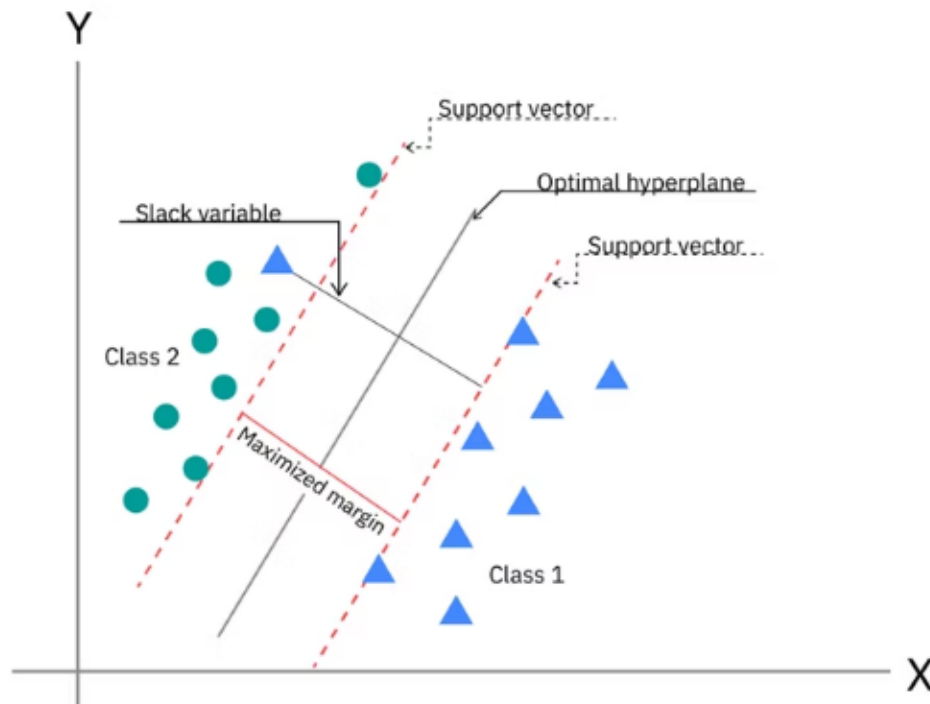


Figura 5.12: Representación gráfica de una SVM con margen suave, ilustrando el hiperplano óptimo de separación, los vectores de soporte y la variable de holgura. Fuente: [10].

- **XGBoost (*eXtreme Gradient Boosting*):** XGBoost es una biblioteca de aprendizaje automático distribuida y de código abierto que utiliza *decision trees* potenciados por gradiente, un algoritmo de potenciación de aprendizaje supervisado que emplea el descenso de gradiente. Se caracteriza por su velocidad, eficiencia y capacidad para escalar bien con grandes conjuntos de datos. Los *decision trees* potenciados por gradiente comienzan con un clasificador débil para realizar predicciones, denominado clasificador base. A continuación, se crean nuevos árboles de forma aditiva a partir de los errores del clasificador base. El algoritmo calcula los residuos de las predicciones de cada árbol para determinar la desviación de las predicciones del modelo con respecto a la realidad, y los residuos se agregan para puntuar el modelo mediante una función de pérdida [79].

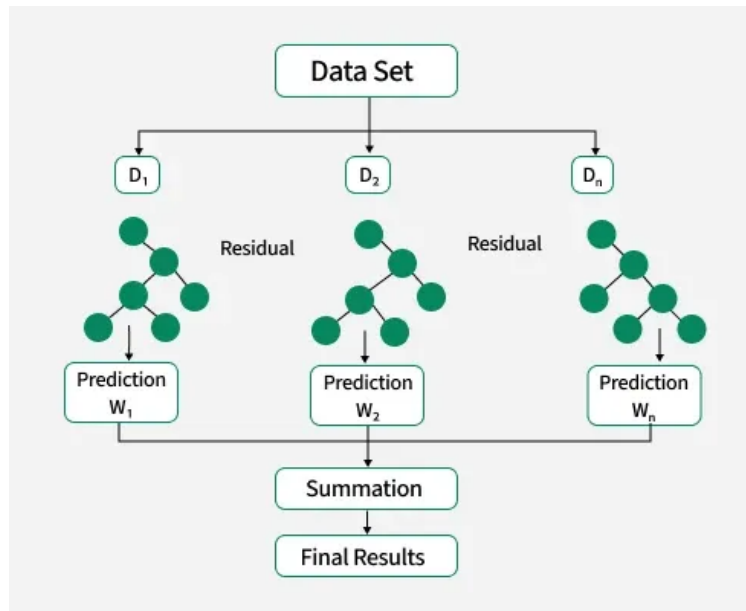


Figura 5.13: Esquema del algoritmo XGBoost, donde cada árbol aprende sobre los residuos del modelo anterior y la predicción final se obtiene como la suma ponderada de todos los árboles. Fuente: [11].

- **LightGBM (*Light Gradient Boosting Machine*):** LightGBM es un marco de trabajo de alto rendimiento de código abierto desarrollado por Microsoft. Se trata de un marco de aprendizaje conjunto que utiliza el método de potenciación de gradiente, el cual construye un modelo robusto añadiendo secuencialmente modelos débiles mediante un descenso de gradiente. Está diseñado para ofrecer eficiencia, escalabilidad y alta precisión, especialmente con grandes conjuntos de datos. Utiliza *decision trees* que crecen de forma eficiente minimizando el uso de memoria y optimizando el tiempo de entrenamiento. Innovaciones clave como el muestreo unilateral basado en gradientes (GOSS), los algoritmos basados en histogramas y el crecimiento de árboles hoja a hoja permiten que LightGBM supere a otros marcos de trabajo tanto en velocidad como en precisión [12].

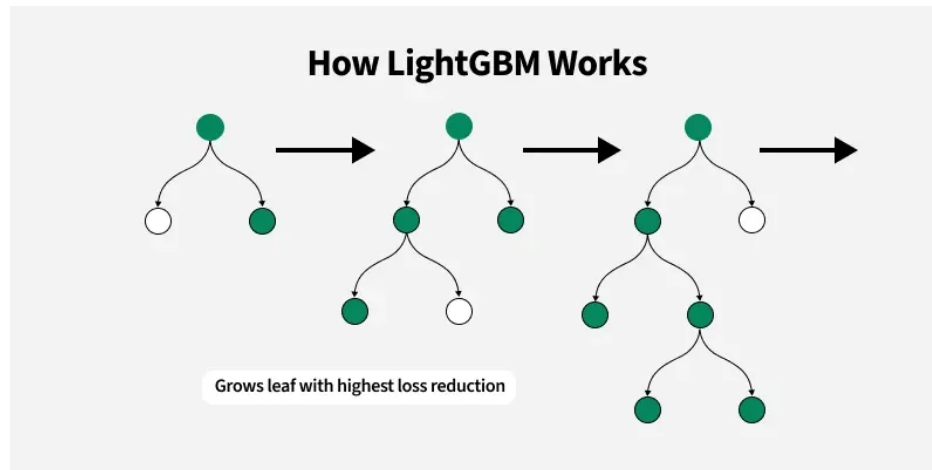


Figura 5.14: Estrategia de crecimiento *leaf-wise* de LightGBM, donde en cada iteración se expande únicamente la hoja que produce la mayor reducción de pérdida. Fuente: [12].

5.2.6. Métricas de evaluación para clasificación desbalanceada

- **AUC-ROC:** Es una métrica de rendimiento utilizada en el aprendizaje automático para evaluar modelos de clasificación binaria. Sus siglas significan Área bajo la curva ROC (*Receiver Operating Characteristic Curve*). Esta métrica ayuda a evaluar la capacidad de un modelo para distinguir entre clases positivas y negativas. AUC-ROC es una curva de probabilidad que representa la tasa de verdaderos positivos (TPR) frente a la tasa de falsos positivos (FPR) en diferentes umbrales. Esta métrica varía de 0 a 1, donde 0 indica un modelo deficiente y 1 un modelo perfecto [80].

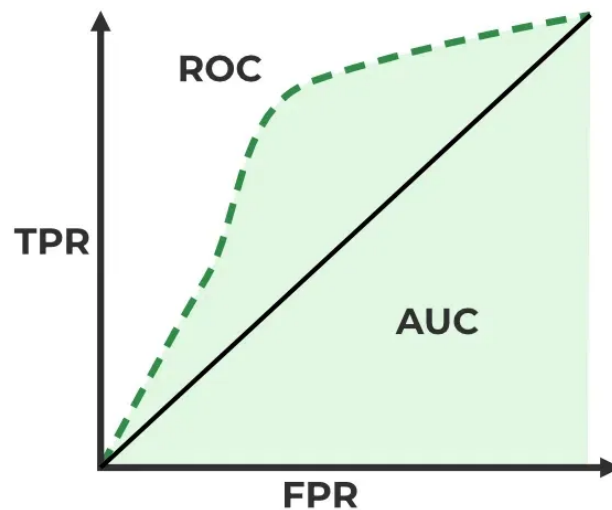


Figura 5.15: Representación gráfica de la curva ROC Fuente: [13].

- **Verdadero Positivo (VP o TP):** Son las instancias en las que el modelo predice correctamente la clase positiva, es decir, el modelo predice que una instancia es positiva y efectivamente lo es [81].
- **Falso Positivo (FP):** Son las instancias de clase negativa que el modelo identifica incorrectamente como positivas. En estadística, también se conocen como errores de tipo I [81].
- **Falso Negativo (FN):** Son las instancias positivas reales que el modelo predice erróneamente como negativas, es decir, casos que deberían haber sido clasificados como positivos pero no lo fueron [81].
- **Verdadero Negativo (VN o TN):** Son las instancias de clase negativa que el modelo predice correctamente como negativas [81].
- **Especificidad:** La especificidad se puede describir como la capacidad del algoritmo/modelo para predecir un verdadero negativo de cada categoría disponible. En la literatura, también se la conoce simplemente como tasa de verdaderos negativos. Formalmente, se puede calcular mediante la siguiente ecuación [82]:

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

- **Sensibilidad:** La sensibilidad en el aprendizaje automático se puede describir como la métrica utilizada para evaluar la capacidad de un modelo para predecir los verdaderos positivos de cada categoría disponible. En la literatura, este término también se conoce como tasa de verdaderos positivos y se puede calcular con la siguiente ecuación [82]:

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

- **Precisión:** La precisión mide la exactitud de un modelo al identificar casos positivos. Se centra en la proporción de objetos positivos identificados correctamente y evita los falsos positivos (identificar erróneamente un objeto negativo como positivo). Esto ayuda a evaluar la capacidad general del modelo para distinguir los objetos relevantes de los irrelevantes. La fórmula para calcular la precisión es [83]:

$$\text{Precisión} = \frac{VP}{VP + FP}$$

- **F1-Score:** La F1 - score es una métrica valiosa para evaluar el rendimiento de un modelo, ya que incorpora el equilibrio entre precisión y exhaustividad. El concepto básico de la puntuación F1 es que mide la media armónica de ambas. Su valor oscila entre 0 y 1, siendo 1 el mejor resultado. La fórmula para la puntuación F1 es [83]:

$$F1 = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}}$$

- **Matriz de confusión:** Una matriz de confusión es una tabla que evalúa el rendimiento de un algoritmo de clasificación. La matriz utiliza valores objetivo para compararlos con los valores predichos por el aprendizaje automático. Cada fila de la matriz representa instancias de una clase predicha y cada columna representa instancias de una clase real, o viceversa [14].

Un ejemplo de matriz de confusión es un algoritmo de clasificación binaria con dos filas y dos columnas. La celda superior izquierda contiene el número de verdaderos positivos, es decir, las instancias clasificadas correctamente como positivas. La celda superior derecha contiene el número de falsos positivos, es decir, las instancias clasificadas incorrectamente como positivas. La celda inferior izquierda contiene el número de falsos negativos, es decir, las instancias clasificadas incorrectamente como negativas. La celda inferior derecha contiene el número de verdaderos negativos, es decir, las instancias clasificadas correctamente como negativas [14].

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 5.16: Representación de una matriz de confusión para clasificación binaria, mostrando las celdas correspondientes a verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (VN). Fuente: [14].

- **Coefficiente de Correlación de Matthews (MCC):** El MCC nos ayuda a representar la matriz de confusión con un solo valor. Si bien no existe una forma perfecta de describir la matriz de confusión de verdaderos y falsos positivos y negativos con un solo número, el coeficiente de correlación de Matthews se considera generalmente uno de los mejores métodos para ello. Es una medida de la calidad de la clasificación binaria y, dado que tiene en cuenta los verdaderos y falsos positivos y negativos, puede utilizarse incluso si las clases tienen tamaños muy diferentes [84].

El MCC es básicamente un coeficiente de correlación entre la serie real y la serie predicha. Devuelve valores entre -1 y +1. Un coeficiente de +1 representa una predicción perfecta, 0 no es mejor que una predicción aleatoria y -1 indica una discrepancia total entre la predicción y la observación. El MCC se puede calcular directamente a partir de la matriz de confusión mediante la siguiente fórmula [84]:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- **Coefficiente de Similitud de Dice (DSC):** El coeficiente de similitud de Dice es un índice de solapamiento espacial y una métrica de validación de la reproducibilidad. El valor de un DSC varía de 0, que indica que no hay solapamiento espacial entre dos conjuntos de resultados de segmentación binaria, a 1, que indica un solapamiento completo. El coeficiente de similitud de Dice se ha adoptado para validar la segmentación de lesiones de sustancia blanca en resonancias magnéticas y la zona periférica de la glándula prostática en braquiterapia de próstata [15].

La fórmula del coeficiente de Dice es:

$$DSC = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}$$

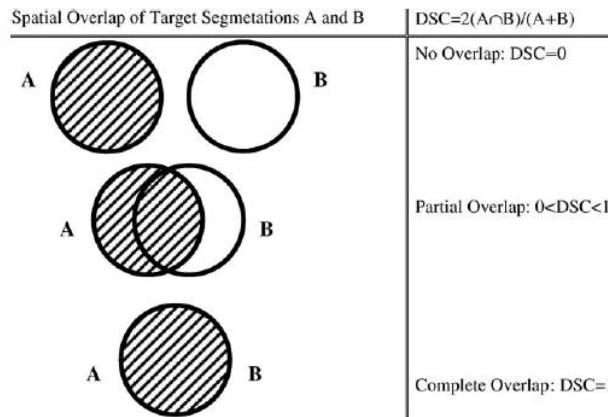


Figura 5.17: Representación gráfica del Coeficiente de Similitud de Dice (DSC), donde la región interseccionada corresponde al solapamiento entre la segmentación predicha (S_1) y la segmentación de referencia (S_2). Fuente: [15].

- Intersección sobre Unión (IoU):** La intersección sobre unión (IoU) es un número que cuantifica el grado de superposición entre dos regiones. En el caso de la detección y segmentación de objetos, la IoU evalúa la superposición entre la región de referencia y la región de predicción [16].

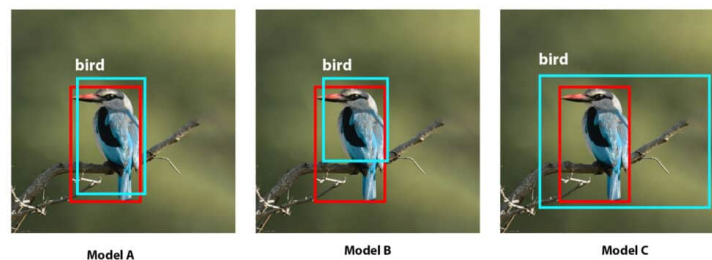


Figura 5.18: Ejemplo de cálculo de IoU para tres modelos (A, B y C) entrenados para predecir aves. La región de referencia se muestra en rojo y las predicciones de los modelos en cian. Fuente: [16].

En la detección de objetos, la IoU es una métrica auxiliar. Sin embargo, en la segmentación de imágenes, la IoU es la métrica principal para evaluar la precisión del modelo [16]. Formalmente se calcula como:

$$IoU = \frac{TP}{TP + FP + FN}$$

- **Índice de Youden (*Youden's J statistic*):** El estadístico J de Youden es un estadístico único que refleja el rendimiento de una prueba diagnóstica dicotómica. Fue sugerido por W.J. Youden en 1950 como una forma de resumir el rendimiento de una prueba diagnóstica [85]. Se define como:

$$J = \text{Sensibilidad} + \text{Especificidad} - 1 = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1 \quad (5.1)$$

donde TP es el número de verdaderos positivos, TN el número de verdaderos negativos, FP el número de falsos positivos y FN el número de falsos negativos. Su valor oscila entre -1 y 1 , donde un valor de 0 indica que la prueba no tiene capacidad diagnóstica y un valor de 1 indica que la prueba es perfecta, sin falsos positivos ni falsos negativos. El índice otorga el mismo peso a los valores falsos positivos y falsos negativos [85].

- **SMOTE (*Synthetic Minority Over-sampling Technique*):** SMOTE es una técnica de sobremuestreo propuesta por primera vez en 2002 que sintetiza nuevos puntos de datos a partir de los puntos existentes en la clase minoritaria. Consiste en el siguiente proceso [86]:
 1. Buscar los K vecinos más cercanos para todos los puntos de datos de clase minoritaria, donde K suele ser 5.
 2. Para cada punto de datos de clase minoritaria, elegir uno de sus K vecinos más cercanos.
 3. Seleccionar un punto aleatorio en el segmento de línea que conecta estos dos puntos en el espacio de características para generar una nueva muestra mediante interpolación.
 4. Repetir los pasos anteriores según la cantidad de sobremuestreo deseada, utilizando un vecino más cercano diferente en cada iteración.

SMOTE resuelve el problema del sobreajuste en el sobremuestreo aleatorio al añadir nuevos datos nunca antes vistos al conjunto de datos, en lugar de simplemente duplicar los datos preexistentes. Sin embargo, la generación artificial de puntos de datos puede añadir ruido adicional al conjunto y provocar superposiciones entre las clases minoritarias y mayoritarias que no reflejan la realidad, fenómeno conocido como sobregeneralización [86].

- **Validación Cruzada (*Cross-Validation*):** La validación cruzada es una técnica que se utiliza para comprobar el rendimiento de un modelo de machine learning con datos no vistos, evitando al mismo tiempo el sobreajuste. Funciona de la siguiente manera [87]:
 - Dividir el conjunto de datos en varias partes.
 - Entrenar el modelo con algunas partes y probarlo con la parte restante.

- Repetir este proceso de remuestreo varias veces seleccionando diferentes partes del conjunto de datos.
- Promediar los resultados de cada paso de validación para obtener el rendimiento final.

5.2.7. Herramientas computacionales y plataforma

- **GPU (*Graphics Processing Unit*):** Una unidad de procesamiento gráfico (GPU) es un circuito electrónico diseñado para acelerar el procesamiento informático de gráficos e imágenes en diversos dispositivos, incluyendo tarjetas de vídeo, placas de sistema, teléfonos móviles y ordenadores personales. Una GPU dispone de su propia memoria de acceso rápido (RAM), utilizada para almacenar código y datos a los que el chip puede acceder y modificar según sus necesidades. Las GPU avanzadas suelen contar con RAM diseñada específicamente para almacenar los grandes volúmenes de datos necesarios para tareas de cálculo intensivo, como la edición de gráficos, los videojuegos o los casos de uso de inteligencia artificial y machine learning [88].
- **DICOM (*Digital Imaging and Communications in Medicine*):** El estándar DICOM especifica un protocolo de intercambio de datos, un formato de imagen digital y una estructura de archivos no propietarios para imágenes biomédicas e información relacionada con imágenes. DICOM abarca cinco áreas de aplicación generales: (1) gestión de imágenes en red, (2) gestión de interpretación de imágenes en red, (3) gestión de impresión en red, (4) gestión de procedimientos de imagen y (5) gestión de medios de almacenamiento fuera de línea. DICOM es una especificación completa de los elementos necesarios para lograr un nivel práctico de interoperabilidad automática entre sistemas informáticos de imágenes biomédicas, desde la capa de aplicación hasta la codificación de flujo de bits [89].
- **Streamlit:** Streamlit es una biblioteca de Python de código abierto para crear aplicaciones web interactivas utilizando únicamente Python. Es ideal para crear paneles de control, aplicaciones web basadas en datos, herramientas de generación de informes e interfaces de usuario interactivas sin necesidad de HTML, CSS ni JavaScript. Streamlit ofrece funciones integradas sencillas que permiten mostrar texto, añadir widgets, visualizar datos y gestionar la entrada del usuario [90].

5.3. Trabajos Relacionados

- **Clasificación de gliomas basada en IA para un diagnóstico confiable: un proceso analítico para una mayor confiabilidad:** Este estudio aborda la clasificación precisa de los gliomas, los tumores cerebrales malignos más comunes y agresivos en adultos, un desafío significativo para los radiólogos. Se explora la aplicación de técnicas de aprendizaje profundo para facilitar la clasificación tumoral mediante imágenes de resonancia magnética (IRM). Utilizando una base de datos de gliomas proveniente de múltiples conjuntos de datos públicos

y comparando diferentes enfoques, el objetivo es desarrollar un sistema de clasificación robusto y confiable. Los resultados muestran que, al enfocarse en la región tumoral de interés y emplear técnicas de aumento de datos, se logra una mejora sustancial tanto en la precisión como en la confianza de las clasificaciones de los grados tumorales. Aunque se logró diferenciar con éxito los gliomas de bajo grado de los de alto grado, la clasificación precisa de los grados 2, 3 y 4 sigue siendo un reto persistente [91] .

- **Un marco CNN guiado por la atención para la segmentación y clasificación del glioma mediante exploraciones de resonancia magnética 3D:** En este estudio, se propone un enfoque basado en redes neuronales convolucionales (CNN) para la gradación no invasiva de tumores utilizando resonancias magnéticas 3D. El marco desarrollado integra dos arquitecturas CNN innovadoras. La primera arquitectura se encarga de la segmentación de los tumores a partir de volúmenes de resonancia magnética multimodal. Esta red de segmentación emplea módulos de atención espacial y de canal, los cuales permiten recalibrar los mapas de características en las capas intermedias. La segunda red adopta una estrategia de aprendizaje multitarea para clasificar los gliomas en función de tres características clave: la distinción entre tumores de bajo y alto grado, la identificación de la alteración genética 1p19q y el estado mutacional de la isocitrato deshidrogenasa (IDH). A través de varios experimentos, se ha demostrado que el método propuesto supera a otros enfoques de vanguardia en términos de rendimiento. Además, se realizó la prueba t de Welch para validar la significancia estadística de los resultados obtenidos [92] .
- **Robust Deep Learning–based Segmentation of Glioblastoma on Routine Clinical MRI Scans Using Sparsified Training:** Eijgelaar et al. (2020) propusieron un modelo robusto de segmentación automática de glioblastoma basado en aprendizaje profundo, específicamente utilizando la arquitectura DeepMedic. Su estudio abordó una problemática común en la práctica clínica: la variabilidad en la calidad y completitud de las imágenes de resonancia magnética (MRI) adquiridas en distintos centros hospitalarios. Para enfrentar este reto, los autores desarrollaron una estrategia de entrenamiento "sparsificado", que consiste en simular de forma aleatoria la ausencia de secuencias secundarias durante el entrenamiento del modelo. Esta técnica permitió aumentar la robustez del modelo ante datos incompletos, mejorando su rendimiento en imágenes clínicas reales que no siempre incluyen todas las secuencias estándar. A través del uso combinado de datos públicos del desafío BraTS y de un conjunto clínico propio con 634 pacientes, se demostró que el modelo entrenado con esta estrategia logró un Dice score comparable al de modelos entrenados con datos completos, incluso cuando se enfrentaba a secuencias faltantes. Este trabajo representa una contribución relevante hacia la implementación de segmentadores automáticos de tumores cerebrales en entornos clínicos reales [93] .
- **Fusión de aprendizaje automático para la detección de tumores de glioma:** Un estudio propone un sistema automatizado para la detección y clasificación de tumores cerebrales, enfocado especialmente en gliomas. El proceso inicia con la adquisición de imágenes de

resonancia magnética cerebral, de las cuales se extraen características relevantes asociadas a la presencia de tumores. Estas características son analizadas y clasificadas individualmente, y posteriormente se utiliza un modelo de aprendizaje profundo para identificar el tipo de tumor. El sistema distingue entre tres clases principales: meningioma, tumor pituitario y glioma. Los resultados del modelo revelan un alto desempeño, alcanzando una precisión del 99.21 %, una especificidad del 98.3 % y una sensibilidad del 97.83 %. No obstante, los autores resaltan la necesidad de realizar investigaciones adicionales para validar su efectividad en entornos clínicos reales [94] .

- **Una herramienta de aprendizaje profundo accesible para la clasificación voxel por vóxel de neoplasias malignas cerebrales a partir de resonancia magnética de perfusión:** Alonso-García et al. (2024) desarrollaron una herramienta de aprendizaje profundo denominada Diagnosis In Susceptibility Contrast Enhancing Regions for Neuro-oncology (DISCERN), orientada a la clasificación no invasiva de tumores cerebrales. Utilizando imágenes de resonancia magnética con contraste de susceptibilidad dinámica (DSC-MRI), entrenaron una red neuronal convolucional con aproximadamente 50.000 vóxeles provenientes de 40 pacientes con diagnóstico histológico confirmado de glioblastoma, metástasis o linfoma. El modelo generó mapas de probabilidad intratumoral que permitieron una clasificación precisa de las lesiones. La validación del modelo se realizó en una cohorte adicional de 400 casos y una cohorte externa de 128 pacientes, alcanzando una precisión del 78 % en la clasificación de los tres tipos de tumores. Este rendimiento superó al de métricas convencionales de MRI, como el volumen sanguíneo cerebral (55 %) y el porcentaje de recuperación de la señal (59 %), evidenciando el potencial de DISCERN como herramienta de apoyo en el diagnóstico clínico de tumores cerebrales mediante imágenes de resonancia magnética estándar [95] .

Materiales y Métodos

6.1. Base de datos

6.1.1. Descripción del dataset

Para el desarrollo de este trabajo se utilizaron imágenes de resonancia magnética provenientes del dataset University of Pennsylvania Glioblastoma [26], el dataset University of California San Diego Pediatric and Adult Brain Tumor [27] y el dataset University of California San Francisco Preoperative Diffuse Glioma [28].

Este conjunto de datos es una colección retrospectiva que reúne imágenes de resonancia magnética (MRI) multiparamétrica de 681 pacientes diagnosticados con Glioblastoma Multiforme (GBM) de novo (Grado IV, según la OMS). En el repositorio, cada paciente cuenta con cuatro tipos de imágenes estructurales clave para el análisis neurorradiológico:

- **T1:** Resonancia basal, hecha sin contraste [3].
- **T1GD:** Secuencia con realce de contraste, que ayuda a definir el núcleo del tumor y las áreas donde la barrera hematoencefálica está dañada [65].
- **T2:** Útil para ver la presencia de líquido y la estructura básica del tejido [3].
- **FLAIR:** Indispensable para detectar el edema alrededor del tumor y la infiltración de células cancerosas en el tejido cercano [3].

Las imágenes vienen en formato NIfTI (.nii.gz), que es lo que suele usarse en investigación neurocientífica, ya que conserva la información espacial y la intensidad de los voxels tal como se capturó, lo que permite que se apliquen preprocesamientos [96]. Además, el dataset trae datos genómicos importantes, en particular el estado mutacional del gen IDH, que se usa como referencia para entrenar el modelo de clasificación incluido en este proyecto.

6.1.1.1. Organización y distribución del dataset

De los 722 casos disponibles, 527 contaban con etiqueta validada del estado mutacional del gen IDH y fueron utilizados para el entrenamiento y la evaluación de los modelos, distribuidos en 458 casos IDH wildtype y 69 casos IDH mutado. Los pacientes restantes del dataset UPENN-GBM fueron excluidos del análisis por dos razones: 106 por presentar clasificación IDH de tipo NOS/NEC, sin confirmación molecular definitiva, y 89 como parte de una estrategia de submuestreo controlado de

la clase wildtype con el fin de reducir el desbalance de clases de 7.9:1 a 6.6:1, proporción consistente con la epidemiología reportada en la literatura donde la variante IDH wildtype representa aproximadamente el 90% de los casos [17]. Adicionalmente, el dataset UPENN-GBM incluye máscaras de segmentación tumoral delimitadas manualmente y aprobadas por neuroradiólogos expertos [26]. De los casos disponibles en este trabajo, 147 contaban con máscara de segmentación validada, las cuales fueron empleadas para la evaluación cuantitativa del modelo de segmentación. El Cuadro 6.1 resume la distribución general del conjunto de datos.

Cuadro 6.1: Distribución general del conjunto de datos consolidado.

Característica	N° de casos
Total de casos procesados	722
UPENN-GBM	671
UCSD-PTGBM + UCSF-PDGM	51
Casos con etiqueta IDH	527
IDH Wildtype	458
IDH Mutado	69
Casos excluidos del análisis	195
NOS/NEC	106
Submuestreo controlado	89
Máscaras de segmentación	147

Nota: El total de casos (722) es superior al total de pacientes (681) debido a que 41 pacientes del dataset UPENN-GBM cuentan con dos sesiones de adquisición, cada una tratada como un caso independiente en el pipeline.

El dataset UPENN-GBM es la fuente principal, con 671 casos de los cuales 458 presentan IDH wildtype y 18 IDH mutado. Dado que esta proporción genera un desbalance de clase considerable, se incorporaron casos IDH mutado provenientes del UCSD-PTGBM (20 casos)[27] y del UCSF-PDGM (31 casos)[28] con el objetivo de enriquecer la representación de la clase minoritaria. Cabe destacar que las imágenes de ambos datasets externos fueron incorporadas directamente al pipeline sin requerir la etapa de skull stripping, dado que ya se encontraban sin cráneo en sus repositorios de origen. El Cuadro 6.2 detalla la distribución por repositorio.

Cuadro 6.2: Distribución de casos por repositorio y estado mutacional IDH.

Dataset	Total casos	IDH Wildtype	IDH Mutado	Excluidos
UPENN-GBM	671	458	18	195
UCSD-PTGBM	20	0	20	0
UCSF-PDGM	31	0	31	0
Total	722	458	69	195

Nota: Los casos del UCSD-PTGBM y UCSF-PDGM fueron seleccionados específicamente por contar con etiqueta IDH mutado y registros completos de las cuatro modalidades de imagen requeridas.

El Cuadro 6.3 muestra la proporción final de clases sobre los 527 casos con etiqueta IDH, evidenciando un desbalance de aproximadamente 6.6:1 entre la clase wildtype y la clase mutada.

Cuadro 6.3: Proporción de casos según el estado mutacional del gen IDH.

Estado IDH	N° de casos	Porcentaje
IDH Wildtype	458	86.9 %
IDH Mutado	69	13.1 %
Total	527	100 %

6.1.1.2. Parámetros de adquisición

El Cuadro 6.4 resume los parámetros de adquisición de las secuencias estructurales para cada repositorio. Para el UPENN-GBM, los valores reportados corresponden a los parámetros modales del conjunto, dado que el dataset incluye imágenes adquiridas en múltiples escaneros institucionales a lo largo de varios años [26].

Cuadro 6.4: Parámetros de adquisición de imagen por repositorio.

Dataset	Secuencia	Campo (T)	TR (ms)	TE (ms)	TI (ms)	Resolución (mm)
UPENN-GBM	T1	3.0 ^a	~1621	~4.1	~939	~1.0
	T1GD	3.0 ^a	~1665	~3.6	~927	~1.0
	T2	3.0 ^a	~4700	~148	–	~1.0
	FLAIR	3.0 ^a	~9557	~132	~2484	~1.0
UCSD-PTGBM	T1/T1GD	3.0	6.5	2.8	450	0.93 × 0.93 × 1.2
	FLAIR	3.0	6000	126	1863	0.93 × 0.93 × 1.2
UCSF-PDGM	T1/T1GD	3.0	N/D	N/D	N/D	N/D
	T2/FLAIR	3.0	N/D	N/D	N/D	N/D

^a El 88.8 % de los casos del UPENN-GBM fueron adquiridos a 3.0 T y el 11.1 % restante a 1.5 T, todos en escanadores Siemens (98.8 %) con bobinas de cabeza estándar [26]. El UCSD-PTGBM utilizó escanadores GE de 3.0 T con bobina de 8 canales [27]. El UCSF-PDGM utilizó un escaner GE Discovery 750 de 3.0 T con bobina de 8 canales [28].

6.1.2. Criterios de selección e inclusión

Con el fin de garantizar la calidad y consistencia del conjunto de datos utilizado para el entrenamiento y evaluación de los modelos, se definieron criterios de inclusión y exclusión para la selección de casos a partir de los tres repositorios.

Criterios de inclusión

- Disponibilidad de las cuatro secuencias estructurales requeridas: T1, T1GD, T2 y FLAIR.
- Disponibilidad de etiqueta validada del estado mutacional del gen IDH1, obtenida mediante prueba inmunohistoquímica y que esté reportada en el repositorio.

Criterios de exclusión

- Clasificación IDH de tipo NOS/NEC o ausencia de etiqueta confirmada en los metadatos clínicos del repositorio (106 casos excluidos).
- Indisponibilidad de una o más de las cuatro modalidades de imagen requeridas.

La aplicación de estos criterios resultó en un conjunto final de 527 casos válidos sobre un total de 722 casos disponibles, provenientes de los tres repositorios

6.1.3. Consolidación del conjunto de datos

Una vez identificados los repositorios y aplicados los criterios de selección, se llevó a cabo un proceso de consolidación para unificar los casos provenientes de las tres fuentes en una estructura común de trabajo. Las imágenes de cada paciente, correspondientes a las cuatro modalidades estructurales, fueron organizadas en carpetas individuales por sujeto en formato NIfTI comprimido (.nii.gz) y almacenadas en Google Drive, lo que facilitó el acceso desde el entorno de desarrollo Google Colaboratory, donde se ejecutaron todos los pipelines computacionales del proyecto. El dataset consolidado, compuesto por los 722 casos con sus cuatro modalidades de imagen, ocupa un total aproximado de 37 GB en disco, volumen que refleja la alta resolución volumétrica de las imágenes multiparamétricas utilizadas.

Adicionalmente, se construyó un archivo de metadatos denominado `labels_idh.csv`, que registró para cada uno de los 527 casos válidos el identificador del sujeto y la etiqueta del estado mutacional IDH (wildtype o mutado).

6.2. Preprocesamiento

En este apartado se describen las decisiones metodológicas adoptadas durante el preprocesamiento de las imágenes de resonancia magnética. Este proceso comprendió una serie de transformaciones secuenciales orientadas a estandarizar la representación espacial y de intensidades de los volúmenes, eliminar estructuras anatómicas no relevantes para el análisis tumoral, y garantizar que los datos de entrada cumplieran con los requisitos específicos de cada modelo implementado. La correcta ejecución de estas etapas es fundamental para asegurar la calidad, consistencia y comparabilidad de los datos a lo largo de todo el pipeline de clasificación.

6.2.1. Skull Stripping

Como primera etapa del pipeline de preprocesamiento, se llevó a cabo la eliminación del cráneo (*skull stripping*) sobre las imágenes de resonancia magnética. Este proceso consiste en la identificación y eliminación automática del cráneo, y los tejidos extracraneales, permitiendo que los modelos de segmentación y clasificación se enfoquen exclusivamente en el parénquima cerebral y el tejido tumoral.

Para este propósito se utilizó la herramienta HD-BET (*High-Definition Brain Extraction Tool*) [71], basada en redes neuronales profundas con una arquitectura nnU-Net [97]. HD-BET fue entrenada con datos provenientes de un ensayo clínico multicéntrico que abarcó 37 instituciones en Europa, lo que le confiere una alta robustez frente a variaciones en los parámetros de adquisición y en la presencia de patologías cerebrales como tumores, resecciones quirúrgicas y alteraciones del tejido inducidas por tratamiento. HD-BET genera automáticamente una máscara binaria del tejido cerebral, que es aplicada sobre cada modalidad para suprimir los vóxeles extracerebrales y conservar únicamente el parénquima cerebral y las estructuras intracraneales de interés.

El procedimiento consistió en aplicar HD-BET sobre la secuencia T1 de cada caso para generar una máscara cerebral binaria, la cual fue posteriormente propagada a las modalidades T1GD, T2 y FLAIR mediante una operación de multiplicación vóxel a vóxel, garantizando la consistencia espacial entre las cuatro modalidades. El proceso se ejecutó sobre los 722 casos disponibles, produciendo un total de 2.888 volúmenes stripped listos para las etapas posteriores del pipeline.

6.2.2. Segmentación tumoral

Una vez completado el skull stripping, se procedió a la segmentación automática del tumor sobre los 722 casos. El objetivo de esta etapa fue obtener una máscara binaria volumétrica de la región tumoral (*whole tumor*), que delimita el área de interés para la extracción de características radiómicas y la generación de parches de entrada al modelo de Deep Learning.

Para este propósito se utilizó el modelo SegResNet preentrenado distribuido por MONAI como parte del reto BraTS 2023 [98], cargado directamente desde su bundle oficial mediante la interfaz `monai.bundle`. Este modelo recibe como entrada las cuatro modalidades de imagen simultáneamente y produce como salida logits para tres regiones tumorales: núcleo necrótico, tumor realzado y edema peritumoral. Para los fines de este trabajo se utilizó únicamente la región *whole tumor*, obtenida aplicando una función sigmoide sobre los logits y umbralizando en 0.5.

Previo a la inferencia del modelo, todas las imágenes fueron sometidas a una secuencia de transformaciones estandarizadas implementadas mediante la librería MONAI [99], con el fin de garantizar la consistencia espacial y de intensidad entre los casos del dataset. El Cuadro 6.5 resume las transformaciones aplicadas en orden secuencial. En primer lugar, las cuatro modalidades de imagen fueron cargadas en formato NIfTI mediante `LoadImaged` y reorganizadas en el formato de tensor requerido por el modelo mediante `EnsureChannelFirstd`. A continuación, se aplicó una reorientación espacial al estándar RAS (*Right-Anterior-Superior*) mediante `Orientatiod`, garantizando que todas las imágenes presentaran la misma orientación independientemente del protocolo de adquisición original. Posteriormente, se realizó un resampling isotrópico a una resolución de $1 \times 1 \times 1$ mm mediante

`Spacingd`, utilizando interpolación bilineal para preservar las transiciones graduales de intensidad características de las imágenes de resonancia magnética. Finalmente, se aplicó una normalización de intensidad por modalidad mediante `NormalizeIntensityd`, con los parámetros `nonzero=True` para operar exclusivamente sobre los vóxeles con señal cerebral, y `channel_wise=True` para normalizar cada modalidad de forma independiente, eliminando la variabilidad en los rangos de intensidad entre diferentes escáneres y protocolos de adquisición.

Cuadro 6.5: Transformaciones de preprocesamiento aplicadas previo a la segmentación tumoral

Orden	Transformación	Descripción
1	<code>LoadImaged</code>	Carga las 4 modalidades en formato NIFTI
2	<code>EnsureChannelFirstd</code>	Reorganiza las dimensiones del tensor de entrada
3	<code>Orientationd</code>	Reorienta todas las imágenes al estándar RAS
4	<code>Spacingd</code>	Resampling isotrópico a $1 \times 1 \times 1$ mm mediante interpolación bilineal
5	<code>NormalizeIntensityd</code>	Normalización de intensidad por modalidad (<code>nonzero=True</code> , <code>channel_wise=True</code>)

Para la inferencia, las cuatro modalidades preprocesadas (T1, T1GD, T2 y FLAIR) fueron apiladas como un tensor de cuatro canales con dimensiones $(1, 4, H, W, D)$, donde el primer eje corresponde al tamaño del lote (*batch size*), el segundo a las cuatro modalidades y los tres últimos a las dimensiones espaciales del volumen. Dado que los volúmenes pueden tener dimensiones variables tras el resampling, la segmentación se realizó mediante inferencia por ventana deslizante (*sliding window inference*) con un tamaño de región de interés (*ROI size*) de $128 \times 128 \times 128$ vóxeles, un solapamiento (*overlap*) de 0.5 y ponderación gaussiana (`mode='gaussian'`) para suavizar las transiciones entre ventanas adyacentes y reducir los artefactos en los bordes. Las probabilidades resultantes fueron binarizadas con un umbral de 0.5 para obtener la máscara binaria del tumor, correspondiente a la región Whole Tumor (WT). Finalmente, la máscara generada fue reorientada al espacio original de la imagen T1GD de referencia para garantizar la correspondencia espacial con las imágenes originales del paciente.

Para la validación cuantitativa, las máscaras de referencia del UPENN-GBM fueron binarizadas previamente, dado que en su formato original contienen etiquetas múltiples correspondientes a distintas subregiones tumorales: necrosis (etiqueta 1), tumor realzado (etiqueta 2) y edema peritumoral (etiqueta 4), siguiendo la convención del reto BraTS [98]. La binarización consistió en asignar valor 1 a cualquier vóxel con etiqueta distinta de cero, obteniendo así la región *whole tumor* como referencia para el cálculo de la métrica de similitud de Dice. Las máscaras de segmentación generadas fueron validadas cuantitativamente utilizando estas 147 máscaras de referencia disponibles en el dataset UPENN-GBM [26]. Los resultados de esta evaluación se presentan en el Capítulo 7.

6.2.3. División del conjunto de datos

6.2.3.1. Machine Learning

Para el modelo de Machine Learning, la división del conjunto de datos se realizó sobre los 527 casos con etiqueta IDH validada. Se implementó una estrategia de tipo *hold-out*, particionando el conjunto en un subconjunto de desarrollo (80 %) y un subconjunto de evaluación independiente (20 %), mediante la función `train_test_split` de la librería `scikit-learn`, la cual es una herramienta de código abierto ampliamente utilizada para el aprendizaje automático en Python que integra algoritmos de preprocesamiento, clasificación y regresión [100]. La división fue estratificada (`stratify=y`) para preservar la proporción original de clases en ambos subconjuntos, garantizando que el desbalance de aproximadamente 6.6:1 entre la clase *wildtype* y la clase *mutada* se mantuviera representado de forma proporcional. El Cuadro 6.6 resume la distribución resultante.

Cuadro 6.6: Distribución del conjunto de datos para el modelo de Machine Learning.

Conjunto	Total	Wildtype	Mutado	Porcentaje
Desarrollo	421	366	55	80 %
Hold-out	106	92	14	20 %
Total	527	458	69	100 %

El subconjunto de *hold-out* fue apartado desde el inicio del proceso y no participó en ninguna etapa de entrenamiento, validación ni selección de modelos, siendo utilizado exclusivamente para la evaluación final del modelo seleccionado.

6.2.3.2. Deep Learning

Para el modelo de Deep Learning se utilizaron los mismos identificadores de casos generados en la división del modelo de Machine Learning, garantizando que el subconjunto de *hold-out* fuera idéntico en ambos enfoques. Esta decisión asegura que la comparación de rendimiento entre los modelos de Machine Learning y Deep Learning se realice sobre exactamente los mismos 106 casos de evaluación, eliminando la variabilidad introducida por diferentes particiones.

A diferencia del modelo de Machine Learning, el entrenamiento de redes neuronales requiere un conjunto de validación independiente para monitorear la convergencia del modelo durante el entrenamiento. Por esta razón, el subconjunto de desarrollo de 421 casos fue dividido nuevamente en un conjunto de entrenamiento (80 %) y un conjunto de validación (20 %), aplicando nuevamente estratificación por clase. El Cuadro 6.7 resume la distribución final.

Cuadro 6.7: Distribución del conjunto de datos para el modelo de Deep Learning.

Conjunto	Total	Wildtype	Mutado	Porcentaje
Entrenamiento	336	292	44	64.8 %
Validación	85	74	11	15.2 %
Hold-out	106	92	14	20.0 %
Total	527	458	69	100 %

La información de cada subconjunto fue almacenada en tres archivos de metadatos independientes: `dl_train_ids.csv`, `dl_val_ids.csv` y `dl_holdout_ids.csv`. Cada archivo registró, para cada caso, el identificador del sujeto, las rutas a las cuatro modalidades de imagen sin cráneo, la ruta a la máscara de segmentación tumoral generada en la etapa anterior y la etiqueta del estado mutacional IDH.

6.3. Modelo de Machine Learning

El modelo de Machine Learning desarrollado en este trabajo tiene como objetivo clasificar el estado mutacional del gen IDH a partir de características radiómicas extraídas de las imágenes de resonancia magnética. El modelo comprende cinco etapas secuenciales: extracción y selección de características radiómicas, estrategia para el desbalance de clases, configuración de los modelos de clasificación y protocolo de entrenamiento y validación.

6.3.1. Extracción y selección de características radiómicas

Radiomics es un enfoque que permite extraer un gran número de características cuantitativas a partir de imágenes médicas, con el objetivo de capturar información sobre la heterogeneidad del tejido que no es perceptible a simple vista [101]. En el contexto de este trabajo, la extracción de características radiómicas constituye la etapa de representación cuantitativa del tejido tumoral en el modelo de Machine Learning, generando un vector de características numéricas que sirve como entrada al clasificador.

6.3.1.1. Preprocesamiento previo a la extracción

Previo a la extracción de características, cada imagen T1GD fue sometida a las mismas transformaciones de estandarización descritas en el Cuadro 6.5: reorientación al estándar RAS, resampling isótropo a $1 \times 1 \times 1$ mm mediante interpolación bilineal y normalización de intensidades sobre los vóxeles no nulos. Paralelamente, la máscara tumoral binaria fue ajustada al mismo espacio que la imagen normalizada mediante interpolación por vecino más cercano (*nearest neighbor*), preservando los valores discretos de la máscara [102]. Los casos cuya máscara tumoral presentara menos de 50 vóxeles fueron excluidos del análisis por considerarse insuficientes para una caracterización representativa del tejido tumoral.

6.3.1.2. Configuración de la extracción

La extracción se realizó exclusivamente sobre la secuencia T1GD, dado que esta modalidad refleja directamente las zonas de ruptura de la barrera hematoencefálica, siendo la más informativa para caracterizar el comportamiento del tumor realizado en GBM [103][65]. Se utilizó la librería PyRadiomics [102], que implementa el estándar de extracción de características definido por la Sociedad de Imagen Cuantitativa en Medicina (*QIBA - Quantitative Imaging Biomarkers Alliance*). La extracción fue configurada mediante un archivo de parámetros `radiomics_params.yaml`, cuyos valores siguieron las recomendaciones establecidas en la literatura de radiomics para neuroimagen [102, 104].

Los parámetros de configuración determinan directamente qué características se calculan y sobre qué representaciones de la imagen, por lo que su definición constituye una decisión metodológica fundamental que afecta el espacio de características resultante. El Cuadro 6.8 resume los parámetros utilizados.

Cuadro 6.8: Parámetros de configuración para la extracción de características radiómicas.

Parámetro	Valor
Modalidad de entrada	T1GD
Tipos de imagen	Original, LoG ($\sigma = 1.0, 2.0, 3.0$ mm), Wavelet
Clases de características	<i>firstorder</i> , <i>shape</i> , GLCM, GLRLM, GLSZM, NGTDM, GLDM
Resampling interno	$1 \times 1 \times 1$ mm (BSpline)
Escala de normalización	100
Ancho de bin (<i>binWidth</i>)	25
Tamaño mínimo de ROI	50 vóxeles

Los **tipos de imagen** definen las representaciones de la imagen T1GD sobre las cuales se calculan las características [102]:

- **Original:** la imagen T1GD preprocesada sin transformaciones adicionales.
- **LoG (*Laplacian of Gaussian*):** filtro que resalta regiones de cambio de intensidad a diferentes escalas espaciales [76]. Se aplicó con tres valores de sigma ($\sigma = 1.0, 2.0$ y 3.0 mm), siendo el sigma pequeño sensible a cambios finos de intensidad y el sigma mayor a estructuras más gruesas dentro del tumor.
- **Wavelet:** descomposición de la imagen en componentes de alta y baja frecuencia en las tres dimensiones espaciales [102], capturando información sobre la heterogeneidad intratumoral a múltiples resoluciones.

Las **clases de características** definen los descriptores matemáticos calculados sobre cada tipo de imagen [102]:

- **Firstorder:** estadísticas de la distribución de intensidades dentro del tumor (media, percentiles, entropía, entre otros).

- **Shape:** descriptores geométricos de la máscara tumoral (volumen, esfericidad, diámetros máximos). Se calcula únicamente sobre la imagen original.
- **GLCM** (*Gray Level Co-occurrence Matrix*): mide la relación espacial entre pares de vóxeles con niveles de gris similares, describiendo la textura del tejido.
- **GLRLM** (*Gray Level Run Length Matrix*): cuantifica la longitud de secuencias de vóxeles con el mismo nivel de intensidad en diferentes direcciones.
- **GLSZM** (*Gray Level Size Zone Matrix*): describe el tamaño de zonas homogéneas de intensidad dentro del tumor.
- **NGTDM** (*Neighbourhood Gray Tone Difference Matrix*): captura la diferencia de intensidad entre un vóxel y sus vecinos inmediatos.
- **GLDM** (*Gray Level Dependence Matrix*): mide cuántos vóxeles vecinos tienen una intensidad similar al vóxel central.

El proceso de extracción se ejecutó sobre los 527 casos con etiqueta IDH, produciendo una matriz de características de dimensiones 527×1.130 , sin valores nulos ni características con varianza cero.

6.3.1.3. Selección de características

Dado el alto número de características extraídas en relación con el tamaño del conjunto de datos, se implementó un proceso de selección en dos etapas para reducir la dimensionalidad y retener únicamente las características con mayor poder discriminativo para la clasificación IDH [105].

En la primera etapa se aplicó el test estadístico ANOVA (*Analysis of Variance*) [106] mediante `SelectKBest` de `scikit-learn` [100], seleccionando las 100 características con mayor estadístico F, es decir, aquellas que presentaron mayor diferencia estadísticamente significativa entre las distribuciones de las clases IDH mutado e IDH wildtype.

En la segunda etapa, un clasificador Random Forest [107] con 200 árboles y pesos de clase balanceados fue entrenado sobre las 100 características preseleccionadas. Se retuvieron las 30 características con mayor importancia según el criterio de impureza de Gini. Este enfoque en dos etapas combina la eficiencia del filtrado estadístico con la capacidad del Random Forest para capturar relaciones no lineales entre características, siguiendo estrategias reportadas en estudios de radiomics en gliomas [108, 109].

El número de características finales fue establecido en 30 considerando la relación entre el tamaño de la clase minoritaria en el conjunto de desarrollo (55 casos IDH mutado) y la dimensionalidad del espacio de características, con el objetivo de reducir el riesgo de sobreajuste. Con 30 características la relación es de aproximadamente 1.8 casos por característica para la clase minoritaria. Adicionalmente, las 10 características de mayor importancia acumularon el 37.59 % de la capacidad discriminativa total del modelo de selección, con un decaimiento progresivo en las posiciones posteriores, lo que indica que el conjunto de 30 características concentra la mayor parte de la información discriminativa disponible en el espacio original de 1.130 descriptores. Este criterio es consistente

con estudios de radiomics en imágenes de resonancia magnética que reportan modelos competitivos construidos con subconjuntos reducidos de características seleccionadas, como el trabajo de Wong et al. [110], que obtuvo un AUC de 0.85 utilizando únicamente 17 características seleccionadas de un pool inicial de 422, lo que respalda que una dimensionalidad reducida puede ser suficiente para capturar la información discriminativa relevante en tareas de clasificación médica basadas en radiomics.

El Cuadro 7.4 presenta las 10 características de mayor importancia del conjunto final seleccionado.

Cuadro 6.9: Top 10 características radiómicas seleccionadas por importancia de Random Forest.

Orden	Característica	Importancia
1	log-sigma-2-0-mm-3D_firstorder_Mean	0.0616
2	log-sigma-3-0-mm-3D_firstorder_Mean	0.0581
3	wavelet-LHL_glcM_Correlation	0.0423
4	log-sigma-1-0-mm-3D_firstorder_Mean	0.0412
5	wavelet-LLL_firstorder_90Percentile	0.0403
6	original_firstorder_90Percentile	0.0287
7	wavelet-LLL_firstorder_10Percentile	0.0286
8	original_firstorder_10Percentile	0.0281
9	log-sigma-3-0-mm-3D_firstorder_10Percentile	0.0264
10	wavelet-HLL_firstorder_Mean	0.0206

Las características seleccionadas están dominadas por descriptores de primer orden calculados sobre imágenes filtradas con LoG y Wavelet, lo que indica que la distribución de intensidades a diferentes escalas espaciales dentro de la región tumoral contiene información discriminativa para la clasificación del estado mutacional IDH [108].

6.3.2. Modelos de clasificación evaluados

Con el fin de identificar el clasificador de mayor rendimiento para la tarea de clasificación del estado mutacional IDH, se evaluaron cinco modelos de Machine Learning supervisado ampliamente utilizados en problemas de clasificación binaria con datos tabulares en el contexto biomédico. Cada modelo fue implementado como un pipeline secuencial mediante la clase `ImbPipeline` de la librería `imbalanced-learn` [100], integrando en un único objeto las etapas de balanceo de clases, normalización y clasificación. La entrada a cada pipeline es un vector de dimensión $(N, 30)$, donde N corresponde al número de casos y 30 a las características radiómicas seleccionadas.

6.3.2.1. Estrategia para el desbalance de clases

El desbalance de aproximadamente 6.6:1 entre la clase `wildtype` y la clase `mutada` representa un desafío para el entrenamiento de clasificadores, ya que los modelos tienden a favorecer la clase mayoritaria. Para mitigar este efecto se adoptaron dos estrategias complementarias según el modelo:

La primera estrategia fue el sobremuestreo sintético mediante SMOTE (*Synthetic Minority Over-sampling Technique*) [86], aplicado como primera etapa de cada pipeline. SMOTE genera casos sintéticos de la clase minoritaria interpolando linealmente entre casos reales existentes en el espacio de características, aumentando la representación de la clase mutada en el conjunto de entrenamiento sin duplicar instancias existentes. De forma crítica, SMOTE fue aplicado exclusivamente dentro de cada subconjunto de entrenamiento durante la validación cruzada, sin contaminar los subconjuntos de validación ni el conjunto de hold-out. La segunda estrategia fue el ajuste de pesos de clase mediante el parámetro `class_weight='balanced'` en los modelos de Regresión Logística, Random Forest y SVM, y mediante `scale_pos_weight=6` en XGBoost y LightGBM. Este valor fue calculado como la razón entre el número de casos wildtype y mutados en el conjunto de desarrollo ($366/55 \approx 6$), indicando al modelo que los errores sobre la clase mutada tienen un costo seis veces mayor que los errores sobre la clase wildtype.

6.3.2.2. Descripción de los modelos

Los cinco modelos evaluados y sus configuraciones son los siguientes:

Regresión Logística (LR): modelo lineal que estima la probabilidad de pertenencia a cada clase mediante una función sigmoide aplicada sobre una combinación lineal de las características de entrada [100]. Se configuró con regularización L2 con parámetro $C = 0,1$, favoreciendo soluciones más regularizadas adecuadas para datasets de tamaño moderado, y con un máximo de 2000 iteraciones para garantizar la convergencia. Requiere normalización previa de las características, por lo que incluye un paso de `StandardScaler` en el pipeline.

Random Forest (RF): conjunto de 300 árboles de decisión entrenados sobre submuestras aleatorias del conjunto de entrenamiento y subconjuntos aleatorios de características en cada nodo [107]. La predicción final se obtiene por votación mayoritaria entre los árboles. Al estar basado en árboles de decisión, no requiere normalización de características.

Máquina de Vectores de Soporte (SVM): clasificador que busca el hiperplano de máximo margen que separa las clases en un espacio de alta dimensión [100]. Se utilizó el kernel RBF (*Radial Basis Function*) con $C = 1,0$, que permite capturar relaciones no lineales entre las características. Se habilitó la estimación de probabilidades mediante `probability=True` para obtener scores continuos requeridos por la métrica AUC. Requiere normalización previa mediante `StandardScaler`.

XGBoost: implementación optimizada de *Gradient Boosting* que construye secuencialmente 300 árboles de decisión, donde cada árbol corrige los errores del anterior mediante descenso por gradiente [111]. Se utilizó `eval_metric= 'logloss'` como función de pérdida durante el entrenamiento. No requiere normalización de características.

LightGBM: implementación de Gradient Boosting basada en histogramas que reduce el costo computacional mediante muestreo de instancias y agrupamiento de características [112]. Se configuró con 300 estimadores. Al igual que XGBoost, no requiere normalización de características.

El Cuadro 6.10 resume los hiperparámetros de cada modelo.

Cuadro 6.10: Hiperparámetros de los modelos de Machine Learning evaluados.

Modelo	Hiperparámetro	Valor	Justificación
LR	C	0.1	Regularización moderada para dataset pequeño
	max_iter	2000	Garantiza convergencia
	class_weight	balanced	Penalización proporcional al desbalance
RF	n_estimators	300	Balance entre rendimiento y costo computacional
	class_weight	balanced	Penalización proporcional al desbalance
SVM	kernel	rbf	Captura relaciones no lineales
	C	1.0	Valor estándar de regularización
	class_weight	balanced	Penalización proporcional al desbalance
XGBoost	n_estimators	300	Balance entre rendimiento y costo computacional
	scale_pos_weight	6	Razón wildtype/mutado en desarrollo (366/55)
LightGBM	n_estimators	300	Balance entre rendimiento y costo computacional
	scale_pos_weight	6	Razón wildtype/mutado en desarrollo (366/55)

Los hiperparámetros fueron definidos con base en valores estándar reportados en la literatura para problemas de clasificación binaria con datasets de tamaño moderado y desbalance de clases [105, 109], sin recurrir a búsqueda exhaustiva de hiperparámetros dado el tamaño limitado del conjunto de datos.

6.3.2.3. Protocolo de entrenamiento y validación

El entrenamiento y la selección del modelo final se realizaron mediante validación cruzada estratificada de cinco pliegues (*Stratified 5-Fold Cross-Validation*) sobre los 421 casos del conjunto de desarrollo. En cada uno, el pipeline completo fue ajustado sobre los datos de entrenamiento, incluyendo la aplicación de SMOTE y la normalización cuando correspondía, y evaluado sobre el subconjunto de validación. Las métricas reportadas corresponden al promedio y la desviación estándar obtenidos sobre los cinco pliegues.

El modelo con mayor AUC promedio en validación cruzada fue reentrenado sobre la totalidad de los 421 casos del conjunto de desarrollo y evaluado una única vez sobre los 106 casos del hold-out independiente, con un umbral de clasificación de 0.5. Los resultados de esta evaluación se presentan en el Capítulo 7.

6.4. Deep Learning

El modelo de Deep Learning desarrollado en este trabajo tiene como objetivo clasificar el estado mutacional del gen IDH a partir de volúmenes de resonancia magnética cerebral. A diferencia del modelo de Machine Learning, que opera sobre un vector de características radiómicas extraídas manualmente, el enfoque de Deep Learning aprende representaciones directamente desde los vóxeles de la imagen, capturando patrones espaciales volumétricos de la región tumoral de forma automática.

Dado el tamaño limitado del conjunto de datos disponible (527 casos, 69 mutados), entrenar una red neuronal convolucional tridimensional desde cero no es viable, ya que este tipo de arquitecturas requieren grandes volúmenes de datos etiquetados para aprender representaciones útiles sin incurrir en sobreajuste [113]. Por esta razón, se adoptó una estrategia de Transfer Learning, que consiste en reutilizar los pesos aprendidos por una red entrenada en un conjunto de datos grande y adaptarlos a la tarea específica de clasificación IDH [114]. Esta estrategia ha demostrado ser especialmente efectiva en imágenes médicas volumétricas, donde la disponibilidad de datos etiquetados es estructuralmente limitada por restricciones técnicas y económicas [113].

Se evaluaron dos arquitecturas de Deep Learning con Transfer Learning bajo condiciones experimentales idénticas, garantizando una comparación justa entre ambas. Ambas arquitecturas recibieron exactamente el mismo conjunto de entrenamiento, validación y hold-out, el mismo tamaño de crop, el mismo esquema de aumentación de datos, el mismo mecanismo de balanceo de clases y el mismo scheduler de tasa de aprendizaje. El hold-out independiente de 106 casos es idéntico al utilizado en el modelo de Machine Learning, lo que permite una comparación directa y sin sesgos entre los tres enfoques evaluados en este trabajo.

6.4.1. Preprocesamiento y aumentación de datos

El preprocesamiento aplicado fue idéntico para ambas arquitecturas. Cada volumen de las cuatro modalidades fue normalizado de forma independiente sobre los vóxeles no nulos, sustrayendo la media e dividiendo por la desviación estándar del tejido cerebral. Posteriormente, se extrajo un parche volumétrico centrado en el centroide de la máscara tumoral generada en la etapa de segmentación, produciendo un crop de dimensiones $64 \times 64 \times 64$ vóxeles. Si la máscara no contenía vóxeles positivos, el crop se centró en el volumen. Las regiones que excedieran los límites del volumen fueron completadas con valor cero (*zero-padding*).

Para mitigar el sobreajuste dado el tamaño reducido del conjunto de entrenamiento, se aplicó aumentación de datos diferencial: aumentación fuerte sobre los casos IDH mutado, cuya baja frecuencia en el dataset genera mayor riesgo de sobreajuste, y aumentación leve sobre los casos wildtype. El Cuadro 6.11 resume las transformaciones aplicadas.

Cuadro 6.11: Transformaciones de aumentación de datos aplicadas durante el entrenamiento de los modelos de Deep Learning.

Transformación	Parámetros	P (mutado)	P (wildtype)
RandFlipd (eje 0)	–	0.5	0.5
RandFlipd (eje 1)	–	0.5	0.5
RandFlipd (eje 2)	–	0.5	0.5
RandRotate90d	max_k=3, ejes (0.1)	0.5	0.5
RandGaussianNoised	$\sigma = 0,1$ (fuerte)	0.3	–
RandGaussianNoised	$\sigma = 0,05$ (leve)	–	0.2
RandZoomd	zoom $\in [0.85, 1.15]$	0.3	–

6.4.2. Estrategia para el desbalance de clases

Ambas arquitecturas emplearon dos mecanismos complementarios para compensar el desbalance de clases. El primero fue un `WeightedRandomSampler` que asignó a cada caso una probabilidad de muestreo inversamente proporcional a la frecuencia de su clase, garantizando que cada batch contuviera una representación aproximadamente balanceada de ambas clases durante el entrenamiento. El segundo fue el parámetro `pos_weight` de la función de pérdida `BCEWithLogitsLoss` [115], calculado como la razón entre el número de casos wildtype y mutados en el conjunto de entrenamiento, limitado a un valor máximo de 3.0 para evitar una penalización excesiva sobre la clase minoritaria.

6.4.3. Arquitectura DL_01: SegResNet como clasificador

6.4.3.1. Justificación

El modelo SegResNet fue propuesto originalmente por Myronenko [7] para la segmentación de tumores cerebrales en el reto BraTS, donde demostró un rendimiento de estado del arte. La hipótesis que motivó su evaluación como clasificador fue que su encoder, preentrenado con imágenes de GBM multiparamétricas, podría haber aprendido representaciones tumorales que contuvieran información discriminativa sobre el estado mutacional IDH, sin necesidad de reentrenar el backbone completo. Esta estrategia de reutilizar encoders de redes de segmentación como extractores de características para clasificación ha sido explorada en otros contextos de análisis de imágenes médicas [5].

6.4.3.2. Arquitectura del modelo

El modelo se construyó reutilizando el encoder del bundle oficial de MONAI BraTS 2023, compuesto por la capa de convolución inicial `convInit` y las capas descendentes `down_layers`. El encoder fue mantenido completamente congelado durante todo el entrenamiento, es decir, ninguno de sus parámetros fue actualizado durante la retropropagación. Esta decisión se fundamentó en la hipótesis de que las representaciones preaprendidas en la tarea de segmentación eran suficientemente informativas para clasificación sin necesidad de adaptación.

La entrada al modelo es un tensor de dimensiones $(B, 4, 64, 64, 64)$, donde B es el tamaño del lote y 4 corresponde a las cuatro modalidades de imagen (T1, T1GD, T2 y FLAIR). Tras pasar por las capas descendentes del encoder, se produce un mapa de características de dimensiones $(B, 128, 8, 8, 8)$. Una operación de Global Average Pooling 3D [116] reduce este tensor a un vector de características de dimensión $(B, 128)$, que constituye la entrada a la cabeza de clasificación.

La cabeza de clasificación es un perceptrón multicapa (MLP) mínimo, diseñado para evitar el sobreajuste dado el reducido número de parámetros entrenables. La salida final es un logit escalar al que se aplica una función sigmoide para obtener la probabilidad de pertenencia a la clase IDH mutado, $P(\text{mutado}) \in [0, 1]$. La clasificación binaria se obtiene aplicando un umbral de 0.5 sobre esta probabilidad. El Cuadro 6.12 describe la arquitectura completa con las dimensiones del tensor en cada etapa.

Cuadro 6.12: Arquitectura del modelo SegResNet clasificador.

Componente	Operación	Dimensión de salida	Entrenable
Encoder (congelado)	convInit	$(B, 32, 64, 64, 64)$	No
	down_layers	$(B, 128, 8, 8, 8)$	No
	AdaptiveAvgPool3d(1)	$(B, 128)$	–
Cabeza MLP (entrenable)	Linear(128→32) + ReLU	$(B, 32)$	Sí
	Dropout(0.5)	$(B, 32)$	–
	Linear(32→1)	$(B, 1)$	Sí

6.4.3.3. Protocolo de entrenamiento

El modelo fue entrenado con un optimizador Adam [117] durante un máximo de 100 épocas. El scheduler CosineAnnealingLR [118] redujo la tasa de aprendizaje de forma suave desde el valor inicial hasta η_{min} siguiendo una función coseno, lo que permite una exploración gradual del espacio de parámetros sin oscilaciones bruscas. Los mejores pesos fueron guardados en cada época en que el AUC de validación mejoraba. El Cuadro 6.13 resume los hiperparámetros utilizados.

Cuadro 6.13: Hiperparámetros de entrenamiento del modelo SegResNet clasificador.

Hiperparámetro	Valor
Épocas máximas	100
Tamaño de lote (<i>batch size</i>)	8
Optimizador	Adam [117]
Tasa de aprendizaje inicial	1×10^{-3}
Weight decay	1×10^{-2}
Función de pérdida	BCEWithLogitsLoss
pos_weight	$\min(n_{wt}/n_{mut}, 3, 0)$
Scheduler	CosineAnnealingLR [118]
T_{max}	100 épocas
η_{min}	1×10^{-5}
Crop size	$64 \times 64 \times 64$ vóxeles
Semilla aleatoria	42

Los resultados obtenidos con esta arquitectura se presentan y discuten en el Capítulo 7.

6.4.4. Arquitectura DL_02: MedicalNet ResNet-18

6.4.4.1. Justificación

MedicalNet [5] es un framework de Transfer Learning para imágenes médicas volumétricas que proporciona pesos preentrenados para múltiples variantes de la arquitectura ResNet 3D [119]. El preentrenamiento se realizó sobre 23 datasets médicos 3D que abarcan diferentes modalidades y

regiones anatómicas, con un total de aproximadamente 1.7 millones de segmentos volumétricos [5]. Este preentrenamiento extenso permite al modelo aprender representaciones volumétricas generales del tejido biológico, transferibles a tareas de clasificación con conjuntos de datos pequeños, lo que lo convierte en una alternativa más adecuada para el presente trabajo que arquitecturas entrenadas exclusivamente para segmentación [113].

Se seleccionó la variante ResNet-18 por su balance entre capacidad representacional y costo computacional, siendo la arquitectura más ligera del framework MedicalNet y adecuada para el tamaño del dataset disponible. La arquitectura ResNet se basa en conexiones residuales que permiten entrenar redes profundas de forma estable al mitigar el problema del desvanecimiento del gradiente [119].

6.4.4.2. Modificaciones al backbone preentrenado

El backbone ResNet-18 de MedicalNet fue diseñado originalmente para recibir imágenes de un único canal. Para aceptar las cuatro modalidades de imagen simultáneamente, se modificó la primera capa convolucional `conv1` expandiendo su dimensión de entrada de 1 a 4 canales. La modificación consistió en replicar los pesos originales cuatro veces a lo largo de la dimensión de canales y dividirlos por 4, preservando la magnitud de la respuesta inicial de la capa y garantizando una inicialización informada para los cuatro canales [5].

6.4.4.3. Estrategia de congelamiento selectivo

Se adoptó una estrategia de congelamiento parcial del encoder que permite al modelo adaptarse a la tarea de clasificación IDH. Las capas `bn1`, `layer1` y `layer2` fueron congeladas por aprender características de bajo y medio nivel (bordes, texturas, gradientes de intensidad) que son universales entre tareas de imagen médica y no requieren adaptación [120]. Las capas `conv1` (modificada), `layer3` y `layer4` fueron mantenidas entrenables, permitiendo que las representaciones de alto nivel específicas de la tarea se especializaran en los patrones relevantes para la clasificación IDH. Esta estrategia de congelamiento selectivo sigue las recomendaciones establecidas para Transfer Learning en imágenes médicas con datasets pequeños [113].

6.4.4.4. Arquitectura del modelo

La entrada al modelo es un tensor de dimensiones $(B, 4, 64, 64, 64)$. El encoder procesa este tensor a través de una serie de bloques residuales que incrementan progresivamente la profundidad de los mapas de características mientras reducen las dimensiones espaciales, extrayendo representaciones de complejidad creciente [119]. La salida del encoder es reducida mediante Global Average Pooling 3D a un vector de dimensión $(B, 512)$, que captura la representación global del parche tumoral y es la entrada a la cabeza de clasificación.

La cabeza de clasificación es más profunda que la del SegResNet e incorpora Batch Normalization [121] para estabilizar el entrenamiento y Dropout con tasa 0.3 para regularización. La salida final es un logit escalar al que se aplica una función sigmoide para obtener $P(\text{mutado}) \in [0, 1]$.

La clasificación binaria se obtiene con un umbral de 0.5. El Cuadro 6.14 describe la arquitectura completa.

Cuadro 6.14: Arquitectura completa del modelo MedicalNet ResNet-18 para clasificación IDH.

Componente	Operación	Dimensión de salida	Entrenable
Encoder	conv1 ($7 \times 7 \times 7$, stride=2)	$(B, 64, 32, 32, 32)$	Sí
	bn1 + ReLU + MaxPool3d	$(B, 64, 16, 16, 16)$	No
	layer1 (2 bloques res.)	$(B, 64, 16, 16, 16)$	No
	layer2 (2 bloques res.)	$(B, 128, 8, 8, 8)$	No
	layer3 (2 bloques res.)	$(B, 256, 4, 4, 4)$	Sí
	layer4 (2 bloques res.)	$(B, 512, 2, 2, 2)$	Sí
	AdaptiveAvgPool3d(1)	$(B, 512)$	–
Cabeza MLP	Linear(512→128)	$(B, 128)$	Sí
	BatchNorm1d(128) + ReLU	$(B, 128)$	Sí
	Dropout(0.3)	$(B, 128)$	–
	Linear(128→32) + ReLU	$(B, 32)$	Sí
	Dropout(0.3)	$(B, 32)$	–
	Linear(32→1)	$(B, 1)$	Sí

6.4.4.5. Protocolo de entrenamiento

El modelo fue entrenado con un optimizador Adam [117] durante un máximo de 100 épocas. La tasa de aprendizaje de 1×10^{-4} es un orden de magnitud inferior a la del SegResNet, dado que MedicalNet tiene más parámetros entrenables en el encoder y una tasa más alta podría deteriorar las representaciones preaprendidas durante el fine-tuning [113]. Se aplicó gradient clipping con norma máxima de 1.0 para prevenir el problema de gradientes explosivos durante la retropropagación [122]. El scheduler CosineAnnealingLR [118] redujo la tasa de aprendizaje desde 1×10^{-4} hasta 1×10^{-6} a lo largo de las 100 épocas. Los mejores pesos fueron guardados en cada época en que el AUC de validación mejoraba respecto al mejor valor registrado hasta ese momento. El Cuadro 6.15 resume los hiperparámetros utilizados.

Cuadro 6.15: Hiperparámetros de entrenamiento del modelo MedicalNet ResNet-18.

Hiperparámetro	Valor
Épocas máximas	100
Tamaño de lote (<i>batch size</i>)	8
Optimizador	Adam [117]
Tasa de aprendizaje inicial	1×10^{-4}
Weight decay	1×10^{-3}
Función de pérdida	BCEWithLogitsLoss
pos_weight	$\min(n_{wt}/n_{mut}, 3, 0)$
Scheduler	CosineAnnealingLR [118]
T_{max}	100 épocas
η_{min}	1×10^{-6}
Gradient clipping	max_norm = 1.0
Crop size	$64 \times 64 \times 64$ vóxeles
Semilla aleatoria	42

Los resultados de la evaluación en el hold-out independiente se presentan en el Capítulo 7.

6.5. Plataforma interactiva

Como parte del cuarto objetivo específico de este trabajo, se desarrolló una plataforma interactiva que permite cargar imágenes de resonancia magnética cerebral y obtener de forma automática la clasificación del estado mutacional IDH mediante los modelos entrenados. La plataforma integra todas las etapas del pipeline computacional en una interfaz web accesible, orientada a facilitar la exploración y validación de los modelos desarrollados en un contexto de investigación.

6.5.1. Serialización y almacenamiento de modelos

Con el fin de desacoplar el entrenamiento de la inferencia, todos los componentes del pipeline fueron serializados y almacenados en Google Drive al finalizar el desarrollo de cada etapa. El Cuadro 6.16 resume los archivos clave del sistema y su función dentro del pipeline de inferencia.

Cuadro 6.16: Archivos serializados del pipeline de inferencia.

Archivo	Descripción	Tamaño
medicalnet_best.pth	Pesos del modelo MedicalNet ResNet-18 entrenado	133.3 MB
resnet18.pth	Pesos preentrenados MedicalNet base [5]	132.0 MB
best_model.joblib	Modelo LR serializado con pipeline SMOTE + Scaler	<1 MB
pipeline_config.json	Nombres de features, métricas y umbral ML	<1 MB
medicalnet_config_v2.json	Umbral, métricas y rutas del modelo DL	<1 MB
radiomics_params.yaml	Parámetros de configuración PyRadiomics	<1 MB
hdbet_weights/	Pesos HD-BET para skull stripping [71]	Carpeta
brats_mri_segmentation/	Bundle SegResNet BraTS23 para segmentación tumoral	Carpeta

Este esquema garantiza que la plataforma opera completamente de forma local sin requerir conexión a servicios externos ni claves de API durante la inferencia. Todos los pesos y configuraciones se cargan desde Google Drive al inicio de cada sesión.

6.5.2. Pipeline de inferencia (`gbm_pipeline.py`)

El módulo `gbm_pipeline.py` implementa la clase `GBMPipeline`, que encapsula la lógica completa del pipeline de inferencia desde la imagen de entrada hasta la predicción final. Su diseño modular permite reutilizar cada etapa de forma independiente y facilita el mantenimiento del sistema. El pipeline ejecuta las siguientes etapas en orden secuencial:

6.5.2.1. Conversión DICOM a NIFTI

Dado que los modelos fueron entrenados con imágenes en formato NIFTI, pero las imágenes clínicas se adquieren y almacenan típicamente en formato DICOM, se implementó un módulo de conversión automática mediante la librería SimpleITK [123]. El método `analizar_zip_dicom` acepta uno o varios archivos ZIP con imágenes DICOM, extrae los archivos, lee los headers de cada serie y sugiere automáticamente la modalidad correspondiente (T1, T1GD, T2 o FLAIR) basándose en los campos `SeriesDescription` y `SequenceName`. El usuario puede confirmar o corregir la asignación antes de proceder. La conversión a NIFTI se realiza mediante el método `convertir_serie_dicom`, que utiliza el lector DICOM de SimpleITK para manejar correctamente las orientaciones oblicuas y los metadatos de adquisición originales.

6.5.2.2. Reorientación espacial

Una vez en formato NIfTI, todas las modalidades son reorientadas al estándar LPS (*Left-Posterior-Superior*) mediante la librería NiBabel [124], garantizando una orientación consistente independientemente del sistema de coordenadas original de cada escáner.

6.5.2.3. Registro rígido entre modalidades

En el contexto clínico, las cuatro modalidades de un mismo paciente pueden haber sido adquiridas con diferentes campos de visión, resoluciones o en sesiones separadas, lo que resulta en volúmenes con dimensiones físicas incompatibles. Para detectar y corregir esta situación, el método `necesita_registro` compara las dimensiones físicas en milímetros de cada modalidad respecto a T1GD. Si la diferencia supera 10 mm en cualquier eje, se activa automáticamente un registro rígido mediante el método `registrar_al_t1ce`, que utiliza la métrica de información mutua de Mattes con optimización gradiente descendente en tres escalas de resolución, implementado con SimpleITK [123]. Este paso es especialmente relevante para imágenes clínicas donde T1GD es una secuencia 3D isótropa de alta resolución y T2 o FLAIR son secuencias 2D con mayor grosor de corte.

6.5.2.4. Skull stripping

Se aplica HD-BET sobre la secuencia T1 para generar la máscara cerebral, que es propagada a las demás modalidades mediante multiplicación vóxel a vóxel, tal como se describe en la Sección 6.2.1. El pipeline detecta automáticamente si las imágenes ya han sido sometidas a skull stripping previamente, omitiendo este paso en ese caso.

6.5.2.5. Segmentación tumoral

Se ejecuta la inferencia del modelo SegResNet BraTS23 mediante sliding window sobre las cuatro modalidades stripped, produciendo la máscara binaria del tumor necesaria tanto para la extracción de características radiómicas como para el crop centrado en el tumor del modelo de Deep Learning, tal como se describe en la Sección 6.2.2.

6.5.2.6. Clasificación IDH

Dependiendo del modelo seleccionado por el usuario, el pipeline ejecuta uno de dos clasificadores:

- **Machine Learning (LR + Radiomics):** aplica el preprocesamiento descrito en la Sección 6.3.1.1, extrae las 30 características radiómicas con PyRadiomics [102] sobre T1GD, y las pasa al modelo LR serializado en `best_model.joblib`.
- **Deep Learning (MedicalNet ResNet-18):** normaliza los volúmenes, extrae el crop centrado en el tumor de $64 \times 64 \times 64$ vóxeles sobre las cuatro modalidades, y realiza la inferencia con el modelo cargado desde `medicalnet_best.pth`.

Ambos clasificadores retornan un diccionario con la predicción binaria, las probabilidades $P(\text{mutado})$ y $P(\text{wildtype})$, el AUC del modelo sobre el hold-out, la sensibilidad y la especificidad, permitiendo que la interfaz muestre un resultado clínicamente interpretable.

6.5.3. Interfaz de usuario (app.py)

La interfaz de usuario fue implementada mediante el framework Streamlit [125], que permite construir aplicaciones web interactivas directamente en Python sin necesidad de desarrollo frontend separado. La aplicación se organiza en dos pestañas principales: carga de imágenes y resultados.

6.5.3.1. Entorno de ejecución y acceso remoto

La aplicación se ejecuta sobre Google Colaboratory con GPU Tesla T4, requerida por tres componentes del pipeline que demandan procesamiento paralelo en GPU: HD-BET para el skull stripping, el modelo SegResNet para la segmentación tumoral, y el modelo MedicalNet para la clasificación DL.

Sin GPU, el tiempo de inferencia por caso aumentaría considerablemente: el skull stripping con HD-BET, que en GPU procesa las 4 modalidades de un caso en aproximadamente 26.53 segundos, requiere entre 2 y 5 minutos por secuencia en CPU según lo reportado por sus autores [71], lo que elevaría el tiempo de esta única etapa a un rango de 8 a 20 minutos por caso. Sumado a los tiempos adicionales de segmentación e inferencia DL en CPU, el tiempo total de procesamiento por caso superaría ampliamente los tiempos esperados para una herramienta de uso interactivo, en la que se espera una respuesta en el orden de minutos y no de decenas de minutos.

El servidor Streamlit se lanza en el puerto 8501 del entorno Colab, que es un entorno de ejecución remoto sin interfaz gráfica local. Para exponer la aplicación públicamente se utiliza un túnel seguro

6.5.3.2. Soporte de formatos de entrada

La interfaz acepta dos formatos de entrada:

- **NIfTI (.nii / .nii.gz):** el usuario carga las cuatro modalidades directamente. El sistema detecta automáticamente la modalidad de cada archivo por su nombre.
- **DICOM (archivos ZIP):** el usuario puede subir uno o varios archivos ZIP con imágenes DICOM. El sistema analiza los headers de todas las series encontradas, sugiere la modalidad de cada una y presenta una tabla de confirmación antes de proceder con la conversión y el análisis.

6.5.3.3. Visualización de resultados

Una vez completado el pipeline, la pestaña de resultados muestra la predicción del estado IDH con su probabilidad asociada, un indicador de confianza basado en la distancia de la probabilidad al umbral de decisión, las métricas del modelo utilizado (AUC, sensibilidad y especificidad) como

referencia para la interpretación clínica, y un slicer axial interactivo que permite navegar por los slices del volumen T1GD con la segmentación tumoral superpuesta.

6.5.3.4. Generación de reportes

La plataforma incluye un módulo de generación de reportes en formato PDF mediante la librería Matplotlib [126]. El reporte incluye los datos del análisis, la predicción con sus probabilidades, el AUC del modelo y una figura con el slice de mayor área tumoral con la segmentación superpuesta. El reporte incluye un aviso explícito de que el sistema ha sido desarrollado con fines exclusivamente investigativos y que los resultados no constituyen un diagnóstico médico, en cumplimiento de los principios éticos aplicables al desarrollo de herramientas de apoyo diagnóstico basadas en inteligencia artificial [127].

La Figura 6.1 presenta el diagrama de bloques del pipeline completo desarrollado en este trabajo, desde la carga de las imágenes de entrada hasta la predicción final del estado mutacional IDH. El pipeline integra las etapas de preprocesamiento descritas en la Sección 6.2 con los dos enfoques de clasificación implementados: Machine Learning basado en características radiómicas y Deep Learning mediante Transfer Learning volumétrico. Ambas ramas comparten la misma etapa de preprocesamiento y convergen en una interfaz web interactiva que presenta el resultado al usuario de forma accesible.

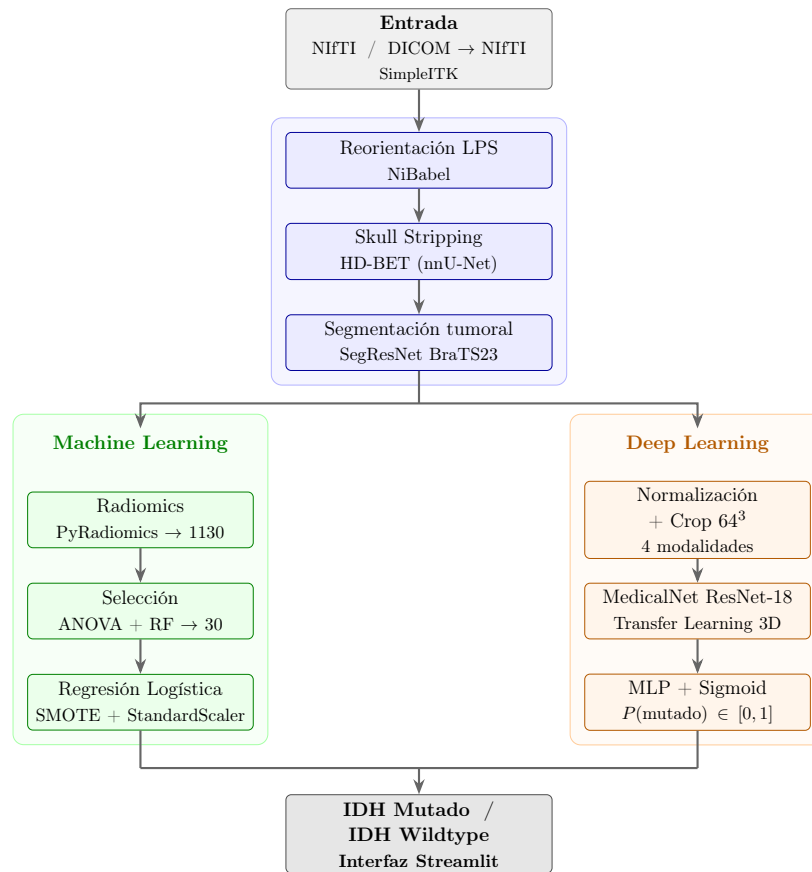


Figura 6.1: Diagrama de bloques del pipeline completo de clasificación IDH en GBM. Fuente: Propia.

6.6. Entorno de desarrollo y recursos computacionales

Todo el desarrollo computacional del proyecto se realizó sobre Google Colaboratory [128], una plataforma de cómputo en la nube que proporciona acceso a entornos de ejecución con aceleración por GPU. Esta plataforma fue seleccionada por su capacidad de integración directa con Google Drive, donde se almacenaron todos los datos, modelos serializados y resultados del proyecto, eliminando la necesidad de infraestructura local de alto rendimiento. Esta decisión es consistente con la práctica estándar en investigación académica de Deep Learning, donde el acceso a recursos computacionales de alto desempeño mediante plataformas en la nube es ampliamente adoptado [128].

El entrenamiento de los modelos de Deep Learning se realizó sobre una GPU NVIDIA A100-SXM4-40GB (42.4 GB de VRAM), disponible mediante Google Colaboratory Pro, lo que permitió manejar los tensores volumétricos de entrada y los pesos de las redes neuronales 3D sin restricciones de memoria. La plataforma interactiva de inferencia se ejecuta sobre una GPU NVIDIA Tesla T4 (15.6 GB de VRAM), suficiente para el pipeline completo de inferencia incluyendo HD-BET, SegResNet y MedicalNet. El uso de GPU es indispensable en ambos contextos dado el costo computacional del skull stripping, la segmentación volumétrica y la inferencia del modelo de Deep Learning, cuyo tiempo de procesamiento por caso sería prohibitivo en CPU.

El entorno de ejecución utiliza Python 3.12 sobre Ubuntu 22.04. El Cuadro 6.17 resume las principales librerías utilizadas y sus versiones.

Cuadro 6.17: Principales librerías y versiones utilizadas en el desarrollo del proyecto.

Librería	Versión	Uso principal
PyTorch	2.10.0	Framework de Deep Learning
MONAI	1.5.2	Segmentación y preprocesamiento de imágenes médicas
NiBabel	5.4.2	Lectura y escritura de archivos NIfTI
SimpleITK	2.5.3	Conversión DICOM y registro rígido
PyRadiomics	3.1.1	Extracción de características radiómicas
Scikit-learn	1.6.1	Modelos de Machine Learning y métricas
Imbalanced-learn	0.14.1	Pipeline con SMOTE para desbalance de clases
HD-BET	[71]	Skull stripping basado en nnU-Net
Joblib	1.5.3	Serialización del modelo ML
PyDICOM	3.0.2	Lectura de headers DICOM
Matplotlib	3.10.0	Generación de figuras y reporte PDF
SciPy	1.16.3	Operaciones sobre máscaras volumétricas
Streamlit	1.56.0	Interfaz web interactiva
Cloudflared	–	Túnel seguro para acceso remoto a la interfaz

6.6.1. Tiempos de Cómputo del Preprocesamiento

Cuadro 6.18: Tiempos de cómputo de las etapas de preprocesamiento ejecutadas sobre GPU NVIDIA A100-SXM4-40GB en Google Colab.

Etapas	Casos	Tiempo total	Tiempo por caso
Skull Stripping (HD-BET)	722	5 h 19 min 14 s	26.53 s/caso
Segmentación tumoral (SegResNet)	722	20 min 26 s	1.70 s/caso
Total preprocesamiento	722	5 h 39 min 40 s	28.23 s/caso

Skull Stripping procesa 4 modalidades por caso (T1, T1GD, T2, FLAIR), lo que explica la diferencia en tiempo respecto a la segmentación, que opera sobre las 4 modalidades de forma conjunta como un único volumen multicanal. La ejecución en CPU no es viable para el volumen de datos procesado, dado que HD-BET requiere entre 2 y 5 minutos por volumen en CPU [71], lo que implicaría aproximadamente 240 horas para el dataset completo frente a las 5 horas 19 minutos obtenidas con GPU.

Resultados y Discusión

7.1. Resultados del Skull Stripping

Como primera etapa del pipeline de preprocesamiento, se aplicó el algoritmo HD-BET sobre las imágenes de resonancia magnética de los 722 casos disponibles, incluyendo los 671 casos del dataset UPENN-GBM y los 51 casos provenientes de los datasets externos UCSD-PTGBM y UCSF-PDGM, correspondientes a las cuatro modalidades de imagen (T1, T1GD, T2 y FLAIR). El proceso se ejecutó de forma completamente automatizada, completándose sin errores en ninguno de los casos y alcanzando una tasa de éxito del 100 %, generando un total de 2.888 volúmenes stripped.

7.1.1. Resultados cuantitativos

El Cuadro 7.1 resume los resultados generales del proceso de skull stripping sobre el dataset consolidado.

Cuadro 7.1: Resumen del proceso de skull stripping.

Parámetro	Valor
Total de casos procesados	722
UPENN-GBM	671
UCSD-PTGBM + UCSF-PDGM	51
Casos procesados exitosamente	722
Casos con error	0
Tasa de éxito	100 %
Modalidades procesadas por caso	4
Volúmenes generados	2.888

Cuadro 7.2: Estadísticas de vóxeles cerebrales retenidos tras el skull stripping (secuencia T1GD).

Estadístico	Vóxeles
Media	1.516.567
Mediana	1.517.178
Mínimo	1.047.445
Máximo	1.951.930

El Cuadro 7.2 resume las estadísticas del volumen cerebral retenido tras el skull stripping sobre la secuencia T1GD. La reducida diferencia entre la media y la mediana indica una distribución simétrica del volumen cerebral entre los casos. La ausencia de valores extremos inusuales sugiere que no se produjeron eliminaciones excesivas ni insuficientes del tejido cerebral en ningún caso.

7.1.2. Análisis visual

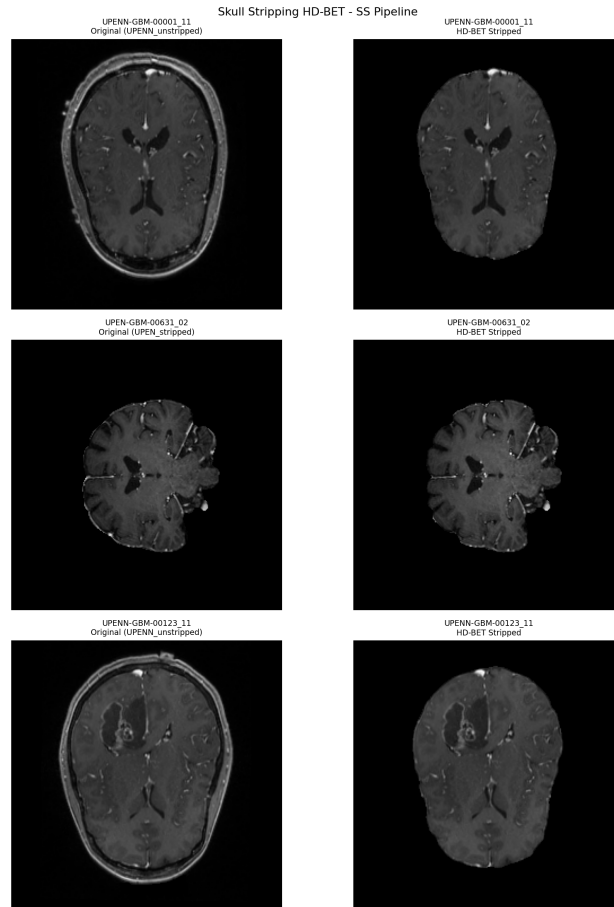


Figura 7.1: Resultado del skull stripping aplicado con HD-BET sobre tres casos representativos del dataset. Columna izquierda: imagen original. Columna derecha: imagen tras la eliminación del cráneo.

La Figura 7.1 ilustra el resultado del proceso de skull stripping sobre tres casos representativos del dataset. El primer caso (UPENN-GBM-00001_11) corresponde a un caso del dataset UPENN-GBM con cráneo, donde se observa la presencia del cuero cabelludo y tejidos extracraneales en la imagen original, los cuales fueron eliminados correctamente por HD-BET conservando íntegramente el parénquima cerebral. El segundo caso (UPEN-GBM-00631_02) corresponde a un caso del dataset

externo que ya se encontraba sin cráneo en el repositorio de origen, por lo que la imagen original y la procesada son prácticamente idénticas. El tercer caso (UPENN-GBM-00123_11) ilustra un caso con masa tumoral visible como región hiperintensa en la imagen T1GD, donde el proceso de skull stripping eliminó correctamente las estructuras extracraneales preservando la región tumoral de interés.

7.1.2.1. Verificación de calidad

El proceso de skull stripping se completó sin errores de ejecución en los 722 casos procesados. La verificación del proceso se realizó mediante tres criterios automáticos: ausencia de errores en la ejecución del programa, generación exitosa del archivo de máscara cerebral para cada caso, y presencia de vóxeles cerebrales en todas las máscaras generadas. Todos los casos produjeron una máscara cerebral con un número de vóxeles anatómicamente plausible. Dado que el resampling isótropo estandarizó todos los volúmenes a $1 \times 1 \times 1$ mm, cada vóxel equivale a 1 mm^3 de tejido. Los valores obtenidos (media: 1.517 cm^3 , mediana: 1.517 cm^3 , rango: $1.047 - 1.952 \text{ cm}^3$) son coherentes con el volumen cerebral típico en adultos, que se encuentra aproximadamente entre 1.000 y 1.600 cm^3 según mediciones por resonancia magnética [129], con valores superiores atribuibles a la masa tumoral adicional característica del GBM. La ausencia de máscaras vacías o con volúmenes anormalmente pequeños descarta fallos totales de extracción cerebral en cualquier caso del dataset.

7.2. Resultados de la Segmentación Tumoral

7.2.1. Descripción del proceso

Como segunda etapa del pipeline, se aplicó el modelo SegResNet preentrenado en el reto BraTS 2023 para la segmentación automática de la región tumoral (*whole tumor*) sobre los 722 casos disponibles. El modelo recibió como entrada las cuatro modalidades de imagen sin cráneo (T1, T1GD, T2 y FLAIR) y generó una máscara binaria volumétrica correspondiente a la región *whole tumor*, que incluye el tumor realzado, el núcleo tumoral necrótico y el edema peritumoral. La validación cuantitativa se realizó sobre las 147 máscaras de referencia delimitadas manualmente disponibles en el dataset UPENN-GBM.

7.2.2. Resultados cuantitativos

El Cuadro 7.3 resume las métricas de evaluación obtenidas sobre las 147 máscaras de referencia disponibles. El modelo alcanzó un Dice medio de 0.836, con una mediana de 0.870, lo que indica que la mayoría de los casos presentaron una superposición alta entre la segmentación automática y la referencia manual. La precisión media fue de 0.885 y el recall medio de 0.811, lo que sugiere que el modelo tiende a ser más conservador en la delimitación del tumor, generando menos falsos positivos a costa de algunos falsos negativos. El IoU medio de 0.730 confirma una suficiente superposición espacial entre las máscaras generadas y las de referencia.

Cuadro 7.3: Métricas de evaluación de la segmentación tumoral sobre 147 casos con máscara de referencia.

Estadístico	Dice	Precisión	Recall	IoU
Media	0.836	0.885	0.811	0.730
Mediana	0.870	0.904	0.847	0.770
Mínimo	0.247	0.574	0.155	0.141
Máximo	0.961	0.993	0.990	0.925
Desv. estándar	0.102	0.084	0.148	0.131

7.2.3. Análisis visual

La Figura 7.2 ilustra los resultados de segmentación obtenidos por el pipeline SS sobre tres casos representativos del dataset UPENN-GBM, seleccionados aleatoriamente. En estos escenarios, que abarcan desde masas compactas hasta lesiones heterogéneas con complejos patrones de necrosis y realce, las máscaras generadas delimitan con precisión la región tumoral. En los tres casos, las máscaras generadas delimitan correctamente la región tumoral mostrando una buena correspondencia espacial con las máscaras de referencia y demostrando la capacidad del modelo para segmentar lesiones con distintas morfologías, tamaños y patrones de realce.

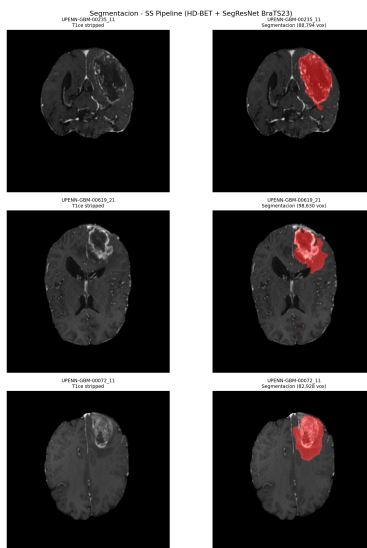


Figura 7.2: Resultados de la segmentación tumoral automática con SegResNet BraTS23.

7.2.4. Síntesis de resultados Segmentación Tumoral

El modelo SegResNet BraTS23 procesó los 722 casos del dataset consolidado sin errores de ejecución, generando máscaras binarias de la región whole tumor para cada uno de ellos. El volumen

tumoral promedio obtenido fue de 72.346 vóxeles, con un valor mínimo de 168 y un valor máximo de 200.092, lo que refleja la alta variabilidad morfológica característica del GBM.

La validación cuantitativa sobre las 147 máscaras de referencia disponibles arrojó un Dice promedio de 0.836 y una mediana de 0.870, de acuerdo con el rendimiento esperado para modelos preentrenados en el desafío BraTS 2023 aplicado a datos de características similares [98], siendo comparable con el rango de equipos participantes en dicho reto, donde el ganador alcanzó un Dice de 0.9005 para la región whole tumor [130]. No obstante, la validación se limitó a los 147 casos con máscara de referencia disponible, lo que constituye una limitación para generalizar este resultado al conjunto completo de 527 casos. Los resultados obtenidos respaldan el uso de SegResNet BraTS23 como etapa de preprocesamiento, dado que las máscaras generadas proporcionan una delimitación tumoral suficientemente consistente para la extracción de características radiómicas y la definición del crop volumétrico en los modelos de Deep Learning.

7.3. Resultados del Clasificador de Machine Learning

7.3.1. Características radiómicas seleccionadas

La extracción de características radiómicas sobre la secuencia T1GD de los 527 casos produjo una matriz de dimensiones 527×1.130 , sin valores nulos ni características con varianza cero. Tras el proceso de selección en dos etapas descrito en la Sección 6.3.1.3 (ANOVA + Random Forest), se retuvieron 30 características con mayor poder discriminativo para la clasificación IDH. El Cuadro 7.4 presenta las 10 características de mayor importancia del conjunto final seleccionado.

Cuadro 7.4: Top 10 características radiómicas seleccionadas por importancia de Random Forest.

Orden	Característica	Importancia
1	log-sigma-2-0-mm-3D_firstorder_Mean	0.0616
2	log-sigma-3-0-mm-3D_firstorder_Mean	0.0581
3	wavelet-LHL_glcm_Correlation	0.0423
4	log-sigma-1-0-mm-3D_firstorder_Mean	0.0412
5	wavelet-LLL_firstorder_90Percentile	0.0403
6	original_firstorder_90Percentile	0.0287
7	wavelet-LLL_firstorder_10Percentile	0.0286
8	original_firstorder_10Percentile	0.0281
9	log-sigma-3-0-mm-3D_firstorder_10Percentile	0.0264
10	wavelet-HLL_firstorder_Mean	0.0206

Las características derivadas de la descomposición Wavelet, como `wavelet-LHL_glcm_Correlation` y los percentiles sobre componentes de baja frecuencia (`wavelet-LLL`), capturan la textura del tejido tumoral a múltiples resoluciones. La correlación GLCM sobre la componente Wavelet-LHL mide la regularidad de las relaciones espaciales entre vóxeles vecinos dentro del tumor, reflejando el grado de homogeneidad tisular, mientras que

los percentiles 10 y 90 sobre la componente de baja frecuencia capturan el rango dinámico de intensidades en su representación de menor resolución.

En conjunto, el predominio de descriptores de primer orden sobre imágenes filtradas con LoG y Wavelet indica que la información discriminativa para la clasificación IDH reside principalmente en la distribución de intensidades a diferentes escalas espaciales dentro del tumor. Estos hallazgos son consistentes con la literatura, que reporta que los tumores IDH wildtype exhiben un realce periférico heterogéneo con zonas necróticas centrales y bordes irregulares en T1GD, mientras que los tumores IDH mutado presentan márgenes tumorales más definidos, menor heterogeneidad intratumoral y realce más homogéneo [108, 23].

Finalmente, los percentiles 10 y 90 de la imagen original (`original_firstorder`) cuantifican directamente el rango de intensidades de los vóxeles tumorales en la imagen T1GD sin transformaciones adicionales, siendo sensibles a las diferencias en la captación máxima y mínima de contraste dentro de la región tumoral. En conjunto, el predominio de descriptores de primer orden sobre imágenes filtradas con LoG y Wavelet indica que la información discriminativa para la clasificación IDH reside principalmente en la distribución de intensidades a diferentes escalas espaciales dentro del tumor, más que en descriptores geométricos o de textura de orden superior.

7.3.2. Resultados de la validación cruzada

El Cuadro 7.5 presenta los resultados de la validación cruzada estratificada de 5 pliegues sobre los 421 casos del conjunto de desarrollo. El modelo de Regresión Logística (LR) obtuvo el mayor AUC promedio (0.9157 ± 0.0621), seguido por LightGBM (0.9105 ± 0.0647) y XGBoost (0.9064 ± 0.0760). Los modelos Random Forest y SVM presentaron AUC promedio de 0.9026 y 0.8847, respectivamente. Con base en el mayor AUC promedio en validación cruzada, la Regresión Logística fue seleccionada como el modelo final para su evaluación sobre el hold-out independiente.

Cuadro 7.5: Resultados de la validación cruzada estratificada de 5 pliegues sobre los 421 casos del conjunto de desarrollo.

Modelo	AUC	Desv. AUC	F1	Sensitivity
LR	0.9157	0.0621	0.6442	0.7818
LightGBM	0.9105	0.0647	0.6125	0.6545
XGBoost	0.9064	0.0760	0.5967	0.6727
RF	0.9026	0.0765	0.6168	0.6182
SVM	0.8847	0.0735	0.6248	0.7091

7.3.3. Desempeño en el hold-out independiente

El modelo de Regresión Logística fue reentrenado sobre la totalidad de los 421 casos del conjunto de desarrollo y evaluado una única vez sobre los 106 casos del hold-out independiente, compuesto por 92 casos IDH wildtype y 14 casos IDH mutado, con un umbral de clasificación de 0.5. El Cuadro 7.6 presenta las métricas obtenidas.

Cuadro 7.6: Métricas de evaluación del modelo de Regresión Logística sobre el hold-out independiente (106 casos, umbral = 0.5).

Métrica	Valor
AUC-ROC	0.9138
F1-Score	0.6471
Sensitivity	0.7857
Specificity	0.9022
MCC	0.5953
TN	83
FP	9
FN	3
TP	11

El modelo alcanzó un AUC-ROC de 0.9138, consistente con el AUC promedio obtenido en validación cruzada (0.9157), lo que sugiere una buena capacidad de generalización sin evidencia de sobreajuste. La sensitivity de 0.7857 indica que el modelo identificó correctamente 11 de los 14 casos IDH mutado del hold-out, mientras que la specificity de 0.9022 refleja que 83 de los 92 casos IDH wildtype fueron correctamente clasificados. El coeficiente de correlación de Matthews (MCC) de 0.5953 confirma un rendimiento moderadamente alto considerando el desbalance de clases de 6.6:1 presente en el dataset.

7.3.4. Análisis de las matrices de confusión

La evaluación sobre el hold-out independiente de 106 casos (92 wildtype, 14 mutados) con umbral de clasificación de 0.5 permitió analizar el comportamiento de clasificación de cada modelo de forma individual.

7.3.4.1. Regresión Logística (LR)

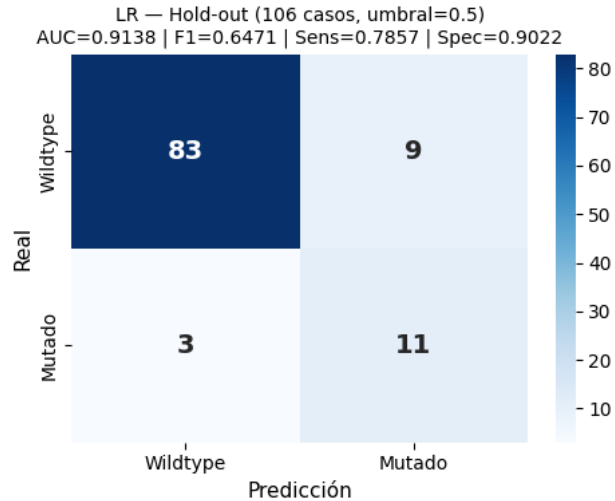


Figura 7.3: Matriz de confusión del modelo de Regresión Logística sobre el hold-out independiente (106 casos, umbral = 0.5).

El modelo de Regresión Logística obtuvo el mejor balance entre sensibilidad y especificidad entre los cinco modelos evaluados. Identificó correctamente 11 de los 14 casos IDH mutado (sensitivity = 0.7857) y 83 de los 92 casos IDH wildtype (specificity = 0.9022), con un total de 3 falsos negativos y 9 falsos positivos. Este comportamiento refleja la capacidad del modelo para capturar la mayoría de los casos mutados a costa de una tasa moderada de falsos positivos, lo cual es deseable en un contexto clínico donde la no detección de una mutación IDH tiene mayor impacto diagnóstico que una clasificación falsa positiva.

7.3.4.2. Random Forest (RF)

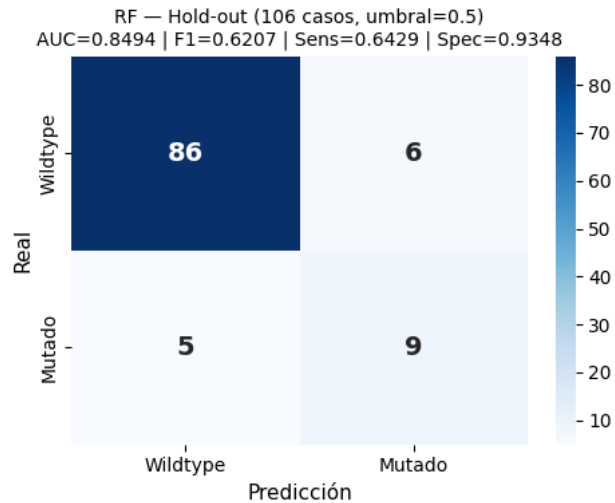


Figura 7.4: Matriz de confusión del modelo Random Forest sobre el hold-out independiente (106 casos, umbral = 0.5).

El modelo de Random Forest presentó la mayor especificidad entre los cinco clasificadores evaluados (0.9348), clasificando correctamente 86 de los 92 casos wildtype del conjunto de hold-out. Sin embargo, esta precisión sobre la clase mayoritaria tuvo un costo en la detección de casos mutados: el modelo identificó correctamente solo 9 de los 14 casos mutados, resultando en una sensibilidad de 0.6429 y 5 falsos negativos.

Este comportamiento refleja una tendencia conservadora del modelo hacia la clase mayoritaria, priorizando la clasificación correcta de los casos wildtype sobre la detección de los casos mutados. Dicha tendencia es característica de los modelos basados en árboles de decisión en presencia de datasets con marcado desequilibrio de clases, donde las funciones objetivo tienden a favorecer la clase mayoritaria [131], como ocurre en el presente trabajo, donde la proporción wildtype/mutado es de 6.6:1.

7.3.4.3. Máquina de Vectores de Soporte (SVM)

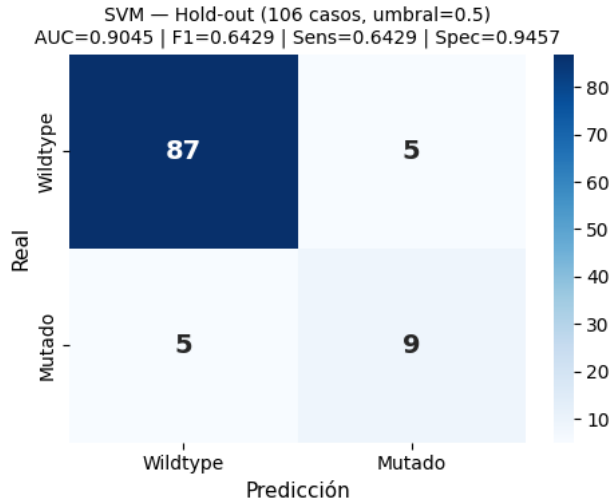


Figura 7.5: Matriz de confusión del modelo SVM sobre el hold-out independiente (106 casos, umbral = 0.5).

El modelo SVM obtuvo la mayor especificidad entre los cinco modelos evaluados (0.9457), con solo 5 falsos positivos y 87 casos wildtype correctamente clasificados. Al igual que Random Forest, detectó 9 de los 14 casos mutados (sensitivity = 0.6429) con 5 falsos negativos. La combinación de alta especificidad y sensibilidad moderada posiciona a SVM como el modelo más conservador del conjunto evaluado, adecuado para escenarios donde minimizar los falsos positivos sea prioritario, aunque a costa de una menor capacidad de detección de la clase mutada.

7.3.4.4. XGBoost

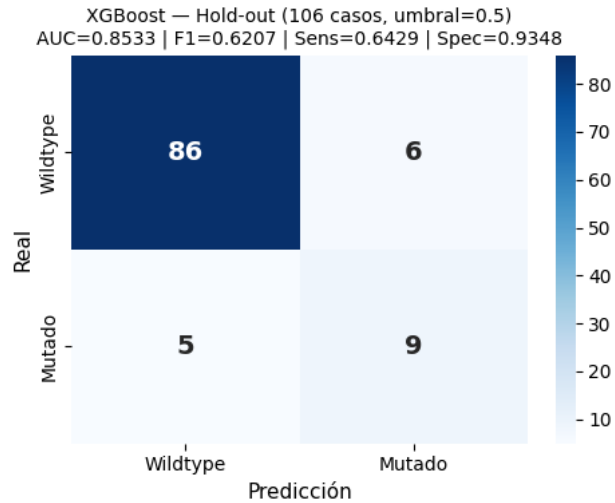


Figura 7.6: Matriz de confusión del modelo XGBoost sobre el hold-out independiente (106 casos, umbral = 0.5).

El modelo XGBoost presentó un patrón de clasificación idéntico al de Random Forest, con 86 casos wildtype correctamente clasificados (TN=86, FP=6) y 9 casos mutados detectados (TP=9, FN=5). A pesar de compartir la misma distribución en la matriz de confusión con Random Forest, XGBoost obtuvo un AUC-ROC de 0.8533, lo que indica diferencias en la distribución de probabilidades predichas aunque el comportamiento a umbral fijo de 0.5 sea equivalente. Este resultado refleja que ambos modelos, aunque con diferente capacidad discriminativa global, convergen al mismo comportamiento de clasificación bajo el umbral estándar de 0.5.

7.3.4.5. LightGBM

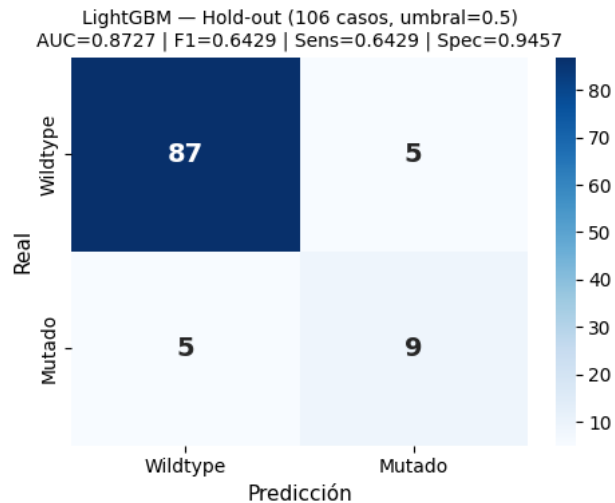


Figura 7.7: Matriz de confusión del modelo LightGBM sobre el hold-out independiente (106 casos, umbral = 0.5).

El modelo LightGBM presentó el mismo patrón de clasificación que SVM, con 87 casos wildtype correctamente clasificados (TN=87, FP=5) y 9 casos mutados detectados (TP=9, FN=5). Su AUC-ROC de 0.8727 fue superior al de Random Forest y XGBoost, lo que sugiere que LightGBM genera probabilidades mejor calibradas en términos de separabilidad entre clases, aunque a umbral fijo de 0.5 produzca el mismo resultado de clasificación que SVM. En conjunto con SVM, LightGBM representa la segunda mejor opción tras la Regresión Logística en términos de balance entre sensibilidad y especificidad.

7.3.5. Análisis de las curvas ROC

La Figura 7.8 presenta las curvas ROC de los cinco modelos evaluados sobre el hold-out independiente. La Regresión Logística obtuvo el mayor AUC-ROC (0.9138), seguida por LightGBM, XGBoost, RF y SVM. La consistencia entre los AUC de validación cruzada y hold-out en todos los modelos evaluados sugiere que los pipelines son estables y no presentan sobreajuste al conjunto de desarrollo.

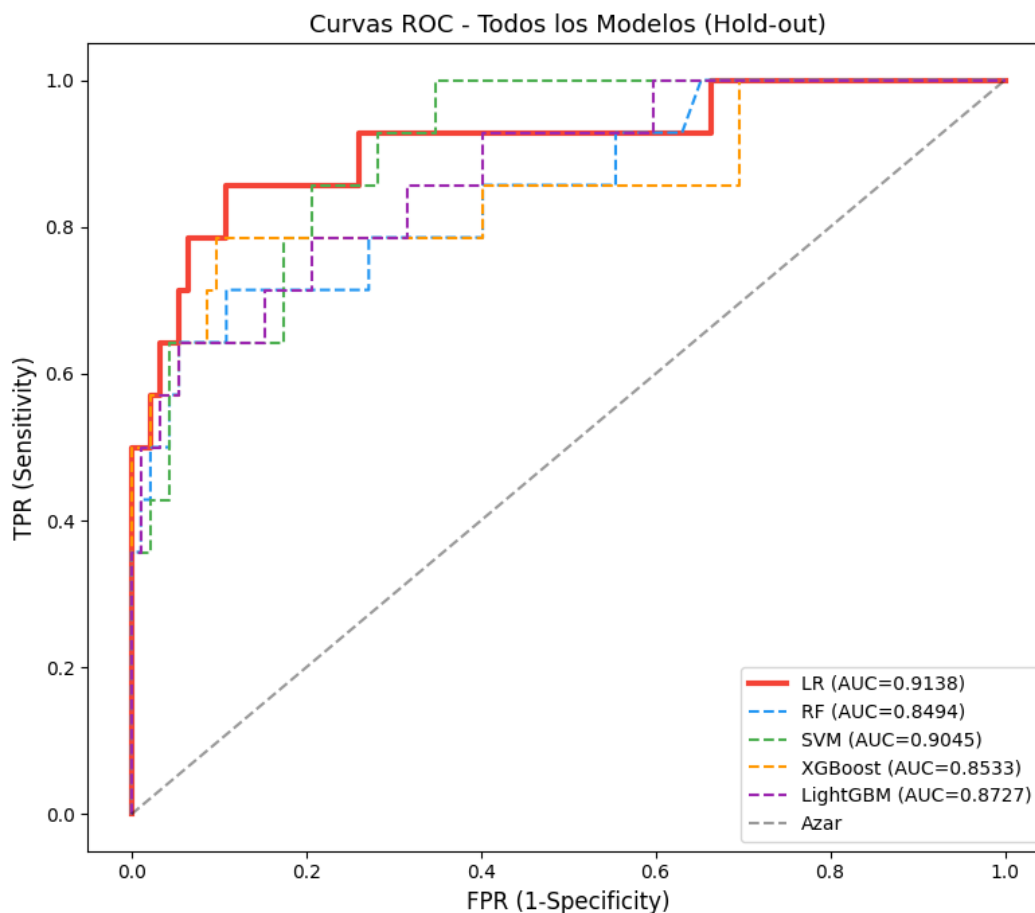


Figura 7.8: Curvas ROC de los cinco modelos evaluados sobre el hold-out independiente (106 casos). La Regresión Logística (LR) obtuvo el mayor AUC-ROC (0.9138).

7.3.6. Síntesis de resultados Machine Learning

En términos generales, los cinco modelos presentan un patrón común: alta especificidad en la identificación de casos IDH wildtype y sensibilidad moderada en la detección de casos IDH mutado, atribuible al desbalance de clases de 6.6:1 presente en el dataset. De entre todos los modelos evaluados, la Regresión Logística se destaca como el modelo con mejor balance entre ambas métricas, siendo el único que detecta 11 de los 14 casos mutados del hold-out, confirmando su selección como modelo final del pipeline de clasificación ML. Este clasificador, entrenado sobre características radiómicas de la secuencia T1GD, alcanzó un AUC-ROC de 0.9138 sobre el conjunto de hold-out independiente de 106 casos, respaldado por la consistencia entre el AUC de validación cruzada de 0.9157 y el AUC de hold-out de 0.9138. Si bien el F1-Score de 0.6471 refleja las limitaciones inherentes al desbalance de clases de 6.6:1, el MCC de 0.5953 confirma que el modelo logra un equilibrio real entre la detección de casos mutados y la correcta clasificación de casos wildtype.

Estos resultados servirán como referencia para los modelos de aprendizaje profundo evaluados en las secciones siguientes.

7.4. Deep Learning

Como se describió en la Sección 6.4, se evaluaron dos arquitecturas de Deep Learning con Transfer Learning sobre el mismo hold-out independiente de 106 casos utilizado en el modelo de Machine Learning, garantizando una comparación directa entre los tres enfoques. La primera arquitectura corresponde al modelo SegResNet BraTS23, que reutiliza el encoder preentrenado para segmentación tumoral como extractor de características. La segunda corresponde al modelo MedicalNet ResNet-18, preentrenado sobre 23 datasets médicos volumétricos. Ambos modelos fueron entrenados sobre los mismos 336 casos de entrenamiento y evaluados sobre los mismos 85 casos de validación y 106 casos de hold-out, bajo condiciones experimentales idénticas descritas en el Cuadro 6.13 y el Cuadro 6.15.

7.4.1. SegResNet BraTS23

El modelo SegResNet BraTS23 fue evaluado como primera arquitectura de Deep Learning para la clasificación del estado mutacional IDH. A diferencia del modelo MedicalNet, cuyo preentrenamiento se realizó sobre una amplia variedad de tareas médicas volumétricas, el SegResNet fue preentrenado exclusivamente para la segmentación de tumores cerebrales en el reto BraTS 2023 [98], lo que lo convierte en un extractor de características especializado en la morfología y heterogeneidad del tejido tumoral cerebral. La hipótesis que motivó su evaluación como clasificador fue que su encoder, entrenado con imágenes multiparamétricas de GBM, podría haber aprendido representaciones tumorales con información discriminativa sobre el estado mutacional IDH, sin necesidad de reentrenar el backbone completo.

7.4.1.1. Curvas de entrenamiento

La Figura 7.9 presenta las curvas de pérdida, AUC-ROC y exactitud durante el entrenamiento del modelo SegResNet BraTS23 sobre los 336 casos del conjunto de entrenamiento, monitoreadas sobre los 85 casos del conjunto de validación a lo largo de 100 épocas.

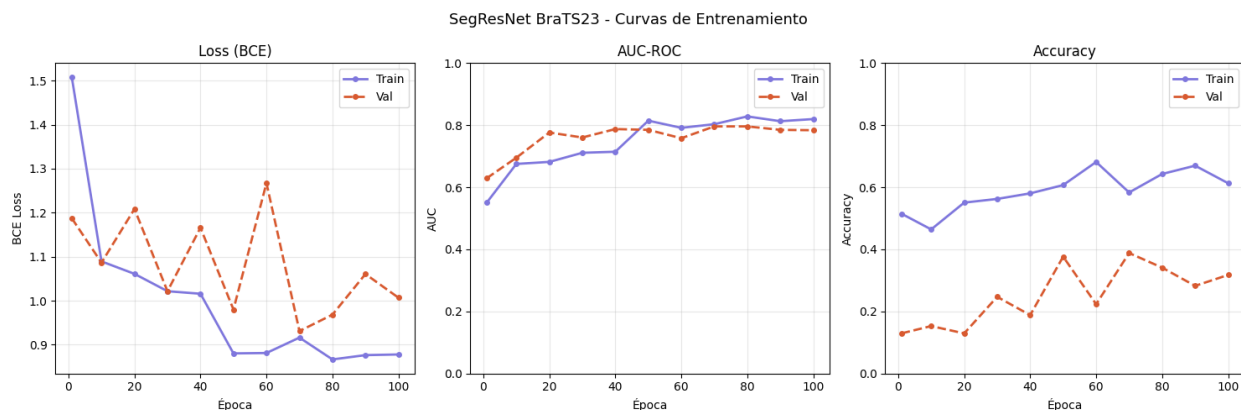


Figura 7.9: Curvas de entrenamiento del modelo SegResNet BraTS23. De izquierda a derecha: pérdida BCE, AUC-ROC y exactitud sobre los conjuntos de entrenamiento y validación.

El análisis de las curvas de entrenamiento evidencia una convergencia progresiva del modelo a lo largo de las 100 épocas. En la fase inicial del entrenamiento, el AUC sobre el conjunto de entrenamiento partió de un valor cercano al azar (0.5502 en la época 1), incrementándose de forma sostenida hasta alcanzar valores superiores a 0.82 en las épocas finales. Este comportamiento confirma que el encoder del SegResNet, aunque mantenido completamente congelado, proporcionó representaciones suficientemente informativas para que la cabeza de clasificación aprendiera a discriminar entre los subtipos moleculares IDH.

La curva de validación presentó una mayor variabilidad respecto a la de entrenamiento, con fluctuaciones pronunciadas a lo largo del proceso, particularmente entre las épocas 20 y 70. Esta inestabilidad es atribuible al tamaño reducido del conjunto de validación (85 casos) y al desbalance de clases de 6.6:1, factores que amplifican la sensibilidad de las métricas ante variaciones en la composición de los mini-batches. El mejor AUC de validación (0.8133) se alcanzó en torno a las épocas 70 y 80, identificándose este punto como el mejor compromiso entre aprendizaje y generalización. Los pesos correspondientes a este máximo fueron guardados y utilizados para la evaluación en el hold-out independiente.

La brecha entre las curvas de entrenamiento y validación en las épocas finales sugiere un inicio de sobreajuste moderado, esperable dado el tamaño limitado del dataset y el número de parámetros entrenables de la cabeza de clasificación. No obstante, la evaluación definitiva de la capacidad de generalización del modelo se sustenta en el desempeño obtenido sobre el hold-out independiente.

7.4.1.2. Desempeño en el hold-out independiente

El modelo SegResNet BraTS23 fue evaluado sobre los 106 casos del hold-out independiente, compuesto por 92 casos IDH wildtype y 14 casos IDH mutado, utilizando el umbral óptimo de clasificación calculado mediante el índice de Youden sobre el conjunto de validación (umbral = 0.6915). El Cuadro 7.7 presenta las métricas obtenidas.

Cuadro 7.7: Métricas de evaluación del modelo SegResNet BraTS23 sobre el hold-out independiente (106 casos).

Métrica	Valor
AUC-ROC	0.8012
F1-Score	0.4706
Sensitivity	0.8571
Specificity	0.7283
MCC	0.4158
Umbral	0.6915
TN	67
FP	25
FN	2
TP	12

El modelo alcanzó un AUC-ROC de 0.8012 sobre el hold-out independiente, evidenciando una capacidad discriminativa buena entre los subtipos IDH mutado y wildtype. La sensitivity de 0.8571 indica que el modelo identificó correctamente 12 de los 14 casos IDH mutado, con únicamente 2 falsos negativos. La specificity de 0.7283 refleja que 67 de los 92 casos wildtype fueron correctamente clasificados, con 25 falsos positivos. El coeficiente de correlación de Matthews (MCC) de 0.4158 confirma un rendimiento moderado considerando el desbalance de clases presente en el hold-out.

7.4.1.3. Análisis de la matriz de confusión y curva ROC

La Figura 7.10 presenta la curva ROC y la matriz de confusión del modelo SegResNet BraTS23 sobre el hold-out independiente.

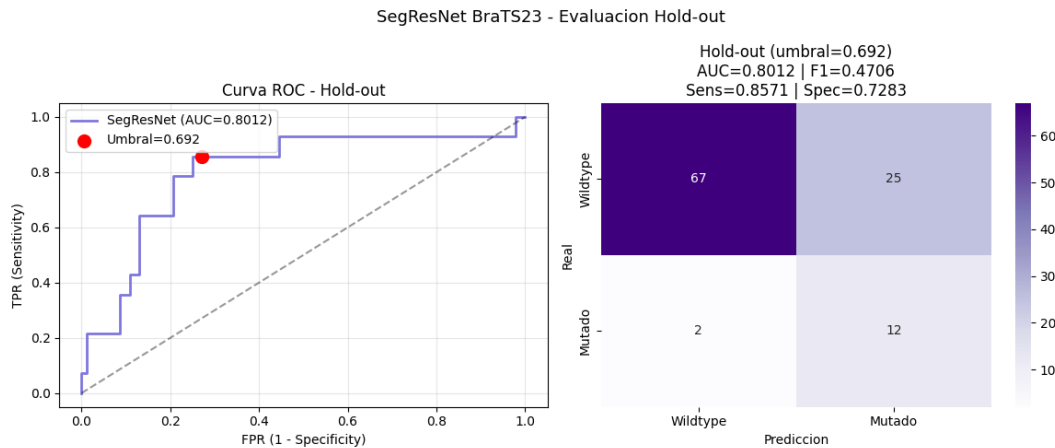


Figura 7.10: Evaluación del modelo SegResNet BraTS23 sobre el hold-out independiente. Izquierda: curva ROC (AUC=0.8012). Derecha: matriz de confusión con umbral=0.6915.

La curva ROC muestra una clara distancia respecto al clasificador aleatorio y tiene un ascenso marcado en la zona de baja tasa de falsos positivos. Esto indica que el modelo se puede configurar para detectar la mayoría de los casos mutados mientras mantiene una especificidad aceptable. El punto de operación óptimo, indicado en rojo en la curva, es el umbral de Youden de 0.6915. Este umbral maximiza la suma de sensitivity y specificity en el grupo de validación. La matriz de confusión confirma el patrón visto en las métricas. El modelo da prioridad a la detección de casos IDH mutado (alta sensitivity) a cambio de una tasa de falsos positivos moderada. Este comportamiento es clínicamente razonable en el contexto de la clasificación IDH porque no detectar una mutación (falso negativo) tiene un mayor impacto en diagnóstico y pronóstico que una clasificación falsa positiva.

7.4.1.4. Síntesis de resultados Deep Learning Modelo SegResNet BraTS23

El modelo SegResNet BraTS23 que funciona como extractor de características congelado junto con una cabeza de clasificación entrenable mostró una capacidad discriminativa media para clasificar el estado mutacional IDH. Alcanzó un AUC-ROC de 0.8012 sobre el conjunto reservado independiente. La alta sensitivity lograda (0.8571) indica que las representaciones que el encoder SegResNet aprende durante la segmentación incluyen información parcialmente relevante para la clasificación molecular. Pero la especificidad media (0.7283) muestra que estas representaciones no son suficientemente discriminativas para diferenciar con precisión ambos subtipos moleculares.

7.4.2. MedicalNet ResNet-18

El modelo MedicalNet ResNet-18 fue evaluado como segunda arquitectura de Deep Learning para la clasificación del estado mutacional IDH. A diferencia del SegResNet, cuyo preentrenamiento se orientó exclusivamente a la segmentación tumoral cerebral, MedicalNet fue preentrenado sobre 23 datasets médicos volumétricos de diversas modalidades y regiones anatómicas [5], proporcionando representaciones volumétricas generales del tejido biológico transferibles a tareas de clasificación con conjuntos de datos pequeños. Se adoptó una estrategia de congelamiento selectivo del encoder, manteniendo entrenables las capas `conv1`, `layer3` y `layer4`, que capturan representaciones de alto nivel específicas de la tarea, mientras que `bn1`, `layer1` y `layer2` fueron congeladas por aprender características de bajo nivel universales entre tareas [113]. El umbral de clasificación fue fijado en 0.5.

7.4.2.1. Curvas de entrenamiento

La Figura 7.11 presenta las curvas de pérdida, AUC-ROC y exactitud durante el entrenamiento del modelo MedicalNet ResNet-18 sobre los 336 casos del conjunto de entrenamiento, monitoreadas sobre los 85 casos del conjunto de validación a lo largo de 100 épocas.

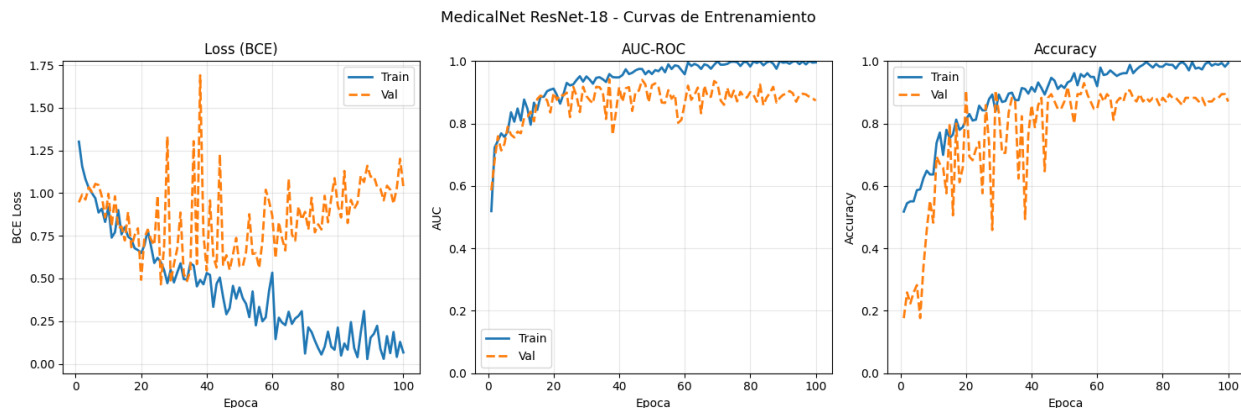


Figura 7.11: Curvas de entrenamiento del modelo MedicalNet ResNet-18. De izquierda a derecha: pérdida BCE, AUC-ROC y exactitud sobre los conjuntos de entrenamiento y validación.

El estudio de las curvas de entrenamiento muestra una convergencia más rápida y estable que la que se ve en el modelo SegResNet. Esto indica el beneficio del amplio preentrenamiento de MedicalNet en 23 conjuntos de datos médicos. Al principio, el AUC en el conjunto de entrenamiento comenzó con un valor cercano al azar (0.5195 en la época 1) y subió notablemente, pasando 0.90 antes de la época 20 y superando 0.99 en las épocas finales. Este comportamiento demuestra que descongelar las capas `layer3` y `layer4` permitió que el modelo se adaptara de manera eficiente a la tarea de clasificación IDH usando representaciones previamente aprendidas.

La curva de validación mostró un aumento sostenido durante las primeras 50 épocas y llegó a su punto máximo en la época 37 con un AUC de 0.9423. Los mejores pesos del modelo se guardaron en ese momento. Después de la época 37, el AUC de validación empezó a bajar con algunas fluctuaciones, mientras que el AUC de entrenamiento siguió subiendo, acercándose a 1.0. Esta creciente diferencia entre las curvas de entrenamiento y validación desde la época 37 indica un sobreajuste progresivo. Esto se debe al tamaño limitado del conjunto de datos y al mayor número de parámetros entrenables comparado con SegResNet. Pero la evaluación final de la capacidad de generalización del modelo se basa en su desempeño en el hold-out independiente, el cual no se usó en el entrenamiento ni en la selección del umbral.

7.4.2.2. Desempeño en el hold-out independiente

El modelo MedicalNet ResNet-18 fue evaluado sobre los 106 casos del hold-out independiente, compuesto por 92 casos IDH wildtype y 14 casos IDH mutado, con un umbral de clasificación de 0.5. El Cuadro 7.8 presenta las métricas obtenidas.

Cuadro 7.8: Métricas de evaluación del modelo MedicalNet ResNet-18 sobre el hold-out independiente (106 casos).

Métrica	Valor
AUC-ROC	0.9014
F1-Score	0.4364
Sensitivity	0.8571
Specificity	0.6848
MCC	0.3767
Umbral	0.5000
TN	63
FP	29
FN	2
TP	12

El modelo alcanzó un AUC-ROC de 0.9014 sobre el hold-out independiente, evidenciando una capacidad discriminativa alta entre los subtipos IDH mutado y wildtype. La sensitivity de 0.8571 indica que el modelo identificó correctamente 12 de los 14 casos IDH mutado, con únicamente 2 falsos negativos, resultado idéntico al obtenido por el SegResNet. La specificity de 0.6848 refleja que 63 de los 92 casos wildtype fueron correctamente clasificados, con 29 falsos positivos, valor inferior al del SegResNet (0.7283) debido al uso del umbral estándar de 0.5 en lugar del umbral optimizado mediante el índice de Youden. El MCC de 0.3767 confirma un rendimiento moderado, afectado principalmente por la tasa de falsos positivos.

7.4.2.3. Análisis de la matriz de confusión y curva ROC

La Figura 7.12 presenta la curva ROC y la matriz de confusión del modelo MedicalNet ResNet-18 sobre el hold-out independiente.

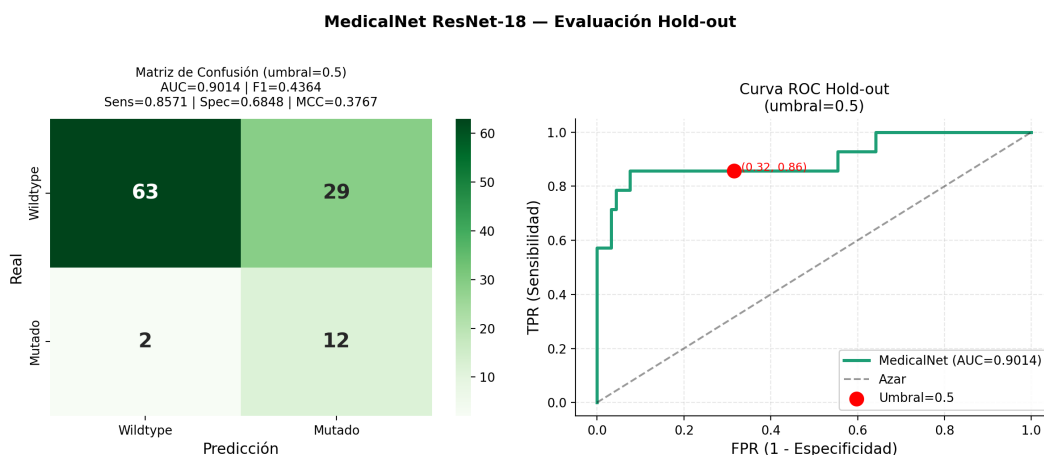


Figura 7.12: Evaluación del modelo MedicalNet ResNet-18 sobre el hold-out independiente. Izquierda: curva ROC (AUC=0.9014). Derecha: matriz de confusión con umbral=0.5.

La curva ROC muestra una separación clara respecto al clasificador aleatorio, con un AUC de 0.9014 que posiciona al modelo MedicalNet como el de mayor capacidad discriminativa entre las dos arquitecturas de Deep Learning evaluadas. La matriz de confusión confirma el patrón observado: alta sensitivity en la detección de casos IDH mutado a costa de una tasa moderada de falsos positivos, comportamiento consistente con el uso del umbral estándar de 0.5 sobre un modelo cuyas probabilidades predichas presentan una distribución sesgada hacia valores bajos debido al desbalance de clases.

7.4.2.4. Síntesis de resultados Deep Learning Modelo MedicalNet ResNet-18

El modelo MedicalNet ResNet-18 demostró la mayor capacidad discriminativa entre las dos arquitecturas de Deep Learning evaluadas, alcanzando un AUC-ROC de 0.9014 sobre el hold-out independiente, valor comparable al obtenido por el modelo de Machine Learning (AUC=0.9138). La convergencia más rápida durante el entrenamiento y el mayor AUC de validación (0.9423) respecto al SegResNet (0.8133) sugieren que el preentrenamiento extenso sobre 23 datasets médicos volumétricos proporciona representaciones más transferibles para la tarea de clasificación IDH que el preentrenamiento exclusivo en segmentación tumoral.

7.4.3. Comparación de los modelos ML y Deep Learning

El Cuadro 7.9 presenta una comparación directa de los tres modelos evaluados sobre el mismo hold-out independiente de 106 casos (92 wildtype, 14 mutados), permitiendo una evaluación objetiva y sin sesgos del rendimiento relativo de cada enfoque.

Cuadro 7.9: Comparación de métricas de evaluación de los tres modelos sobre el hold-out independiente (106 casos).

Modelo	AUC	F1	Sens.	Spec.	MCC
ML — LR + Radiomics	0.9138	0.6471	0.7857	0.9022	0.5953
DL — SegResNet BraTS23	0.8012	0.4706	0.8571	0.7283	0.4158
DL — MedicalNet ResNet-18	0.9014	0.4364	0.8571	0.6848	0.3767

El análisis comparativo revela patrones diferenciados entre los enfoques de Machine Learning y Deep Learning. En términos de capacidad discriminativa global, el modelo de Machine Learning basado en características radiómicas (LR) obtuvo el mayor AUC-ROC (0.9138), seguido por MedicalNet ResNet-18 (0.9014) y SegResNet BraTS23 (0.8012). La diferencia de 0.0124 puntos entre LR y MedicalNet es marginal, sugiriendo que ambos enfoques son comparables en términos de capacidad discriminativa global.

En cuanto a la sensitivity, los dos modelos de Deep Learning obtuvieron valores idénticos (0.8571), superiores al modelo de Machine Learning (0.7857), identificando correctamente 12 de los 14 casos IDH mutado frente a los 11 identificados por LR. Este resultado sugiere que los modelos de Deep Learning, al operar directamente sobre los vóxeles de la imagen, capturan información espacial volumétrica complementaria a las características radiómicas extraídas manualmente, lo que les confiere una mayor capacidad para detectar casos mutados.

Sin embargo, esta mayor sensitivity de los modelos DL se obtuvo a costa de una menor specificity. El modelo de Machine Learning alcanzó la mayor especificidad (0.9022), clasificando correctamente 83 de los 92 casos wildtype con solo 9 falsos positivos. En contraste, SegResNet y MedicalNet presentaron 25 y 29 falsos positivos respectivamente, lo que se refleja en valores de specificity considerablemente menores (0.7283 y 0.6848). Esta diferencia es atribuible a la mayor capacidad de regularización implícita de los clasificadores de Machine Learning frente a las redes neuronales profundas en datasets de tamaño limitado con desbalance de clases.

El coeficiente de correlación de Matthews (MCC), que constituye la métrica más robusta para evaluar clasificadores binarios en presencia de desbalance de clases, confirma la superioridad global del modelo de Machine Learning (MCC=0.5953) frente a SegResNet (MCC=0.4158) y MedicalNet (MCC=0.3767). Esta diferencia refleja que, aunque los modelos DL presentan mayor sensibilidad, su mayor tasa de falsos positivos penaliza su rendimiento global en el contexto del desbalance de clases presente en el dataset.

En términos de F1-Score, el modelo de Machine Learning también obtuvo el valor más alto (0.6471), seguido por SegResNet (0.4706) y MedicalNet (0.4364). La diferencia en F1 entre ML y DL es más pronunciada que la diferencia en AUC, lo que indica que los modelos DL tienen dificultades para mantener un balance adecuado entre precisión y sensibilidad con el umbral de clasificación utilizado.

El Cuadro 7.10 presenta los valores de la matriz de confusión de los tres modelos para facilitar la comparación directa de los patrones de clasificación.

Cuadro 7.10: Matrices de confusión de los tres modelos sobre el hold-out independiente (106 casos: 92 wildtype, 14 mutados).

Modelo	TP	FP	TN	FN
ML — LR + Radiomics	11	9	83	3
DL — SegResNet BraTS23	12	25	67	2
DL — MedicalNet ResNet-18	12	29	63	2

7.4.4. Análisis comparativo de las curvas ROC

La Figura 7.13 presenta las curvas ROC de los tres modelos evaluados sobre el hold-out independiente de 106 casos, permitiendo una comparación visual directa de su capacidad discriminativa a lo largo de todos los umbrales de decisión posibles.

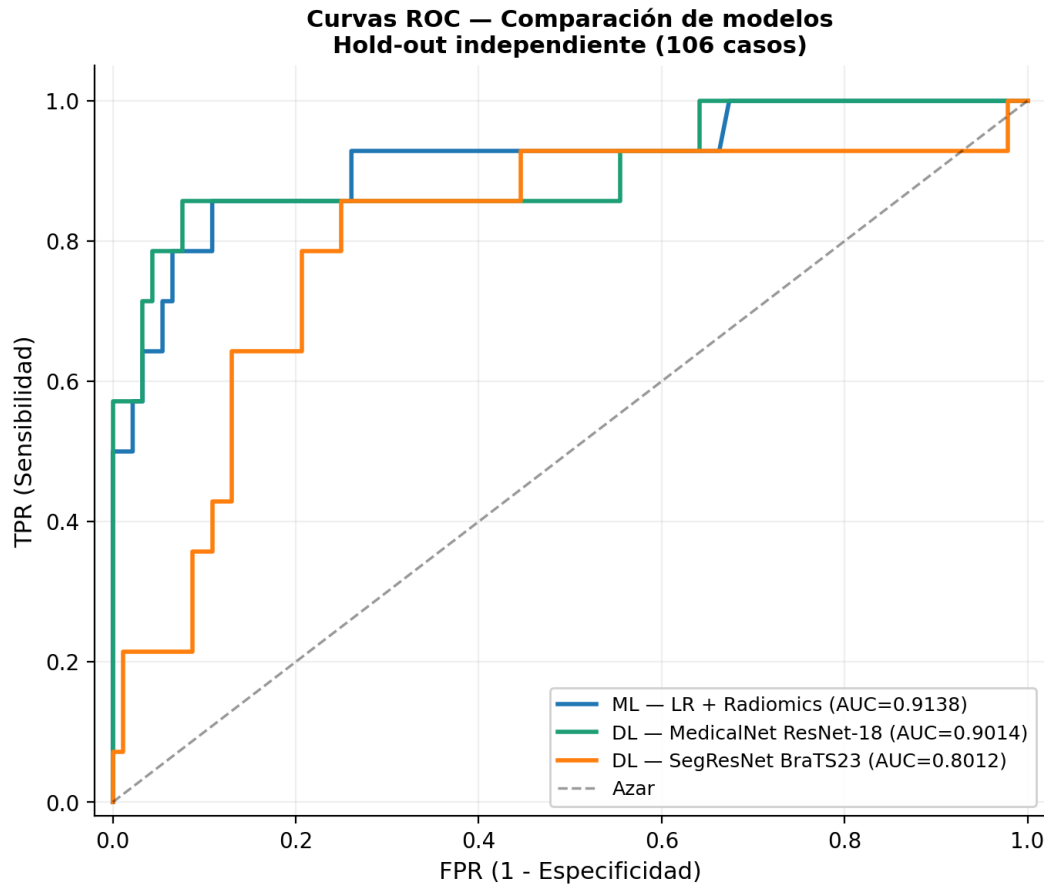


Figura 7.13: Curvas ROC comparativas de los tres modelos evaluados sobre el hold-out independiente (106 casos). Azul: ML — LR + Radiomics (AUC=0.9138). Verde: DL — MedicalNet ResNet-18 (AUC=0.9014). Naranja: DL — SegResNet BraTS23 (AUC=0.8012).

El análisis de la figura muestra patrones diferentes entre los tres enfoques. Las curvas del modelo de Machine Learning en azul y del modelo MedicalNet ResNet-18 en verde siguen trayectorias parecidas en la zona donde la tasa de falsos positivos es baja ($FPR < 0.1$). Aquí, muestran sensibilidades cercanas al 57 % y 79 % respectivamente con menos de un 5 % de falsos positivos. Esta variación inicial favorece a MedicalNet en zonas de alta especificidad, relevante para casos donde es prioritario reducir los falsos positivos. Pero la curva del modelo de Machine Learning cubre más área en la parte media del espacio ROC. Esto se traduce en un AUC-ROC global más alto (0.9138 comparado con 0.9014). La curva del modelo SegResNet BraTS23 en naranja tiene una trayectoria diferente. Sube de manera más gradual y menos pronunciada al inicio, lo que concuerda con su menor AUC-ROC de 0.8012. Sin embargo, las tres curvas se separan claramente de la línea del clasificador aleatorio en todo el espacio de umbrales, lo que sugiere que los tres modelos presentan capacidad discriminativa real para clasificar el estado mutacional IDH en GBM, aunque con diferencias en su rendimiento global reflejadas en los valores de AUC-ROC obtenidos.

7.4.5. Síntesis de resultados

En síntesis, el modelo de Machine Learning basado en características radiómicas demostró el mejor rendimiento global entre los tres enfoques evaluados, alcanzando el mayor AUC-ROC (0.9138), la mayor especificidad (0.9022), el mayor F1-Score (0.6471) y el mayor MCC (0.5953) sobre el hold-out independiente. Este resultado es consistente con estudios previos que reportan que los modelos de radiomics con Machine Learning son competitivos con enfoques de Deep Learning en datasets de tamaño limitado, donde la extracción manual de características actúa como un regularizador implícito que reduce el riesgo de sobreajuste [105, 101].

El modelo MedicalNet ResNet-18 obtuvo el segundo mejor AUC-ROC (0.9014), comparable al del modelo de Machine Learning, lo que sugiere que el preentrenamiento extenso sobre 23 datasets médicos volumétricos proporciona representaciones transferibles para la tarea de clasificación IDH. El modelo SegResNet BraTS23, aunque preentrenado específicamente en imágenes de GBM, obtuvo el menor AUC-ROC (0.8012), posiblemente debido a que su encoder congelado no fue suficientemente adaptado a la tarea de clasificación molecular.

Los tres modelos coincidieron en identificar entre 11 y 12 de los 14 casos IDH mutado del hold-out, con 2 o 3 falsos negativos, lo que refleja la dificultad inherente de la tarea de clasificación IDH en GBM de alto grado, donde la variante mutada representa apenas el 13.1 % de los casos y presenta características radiológicas frecuentemente solapadas con la variante wildtype.

7.5. Resultados de la Plataforma Interactiva

La plataforma interactiva desarrollada como parte del cuarto objetivo específico de este trabajo fue evaluada funcionalmente mediante la ejecución del pipeline completo de inferencia sobre un caso representativo del dataset. A continuación se describen los módulos principales de la interfaz y los resultados obtenidos durante la prueba de funcionamiento.

7.5.1. Descripción general de la interfaz

La Figura 7.14 presenta la vista general de la plataforma al momento de acceder. El encabezado identifica el sistema como *GBM IDH Classifier* e incluye una descripción concisa de su propósito: la clasificación automática del estado IDH en Glioblastoma Multiforme mediante segmentación volumétrica y modelos de aprendizaje profundo. Un conjunto de etiquetas resume los componentes tecnológicos integrados: SegResNet BraTS23 para la segmentación tumoral, LR + Radiomics con AUC=0.914 para el modelo de Machine Learning, MedicalNet ResNet-18 con AUC=0.901 para el modelo de Deep Learning, HD-BET para el skull stripping y soporte de formatos DICOM y NIfTI. La interfaz se organiza en dos pestañas principales: *Cargar Imágenes* y *Resultados*, que guían al usuario a través del flujo de análisis de forma secuencial.

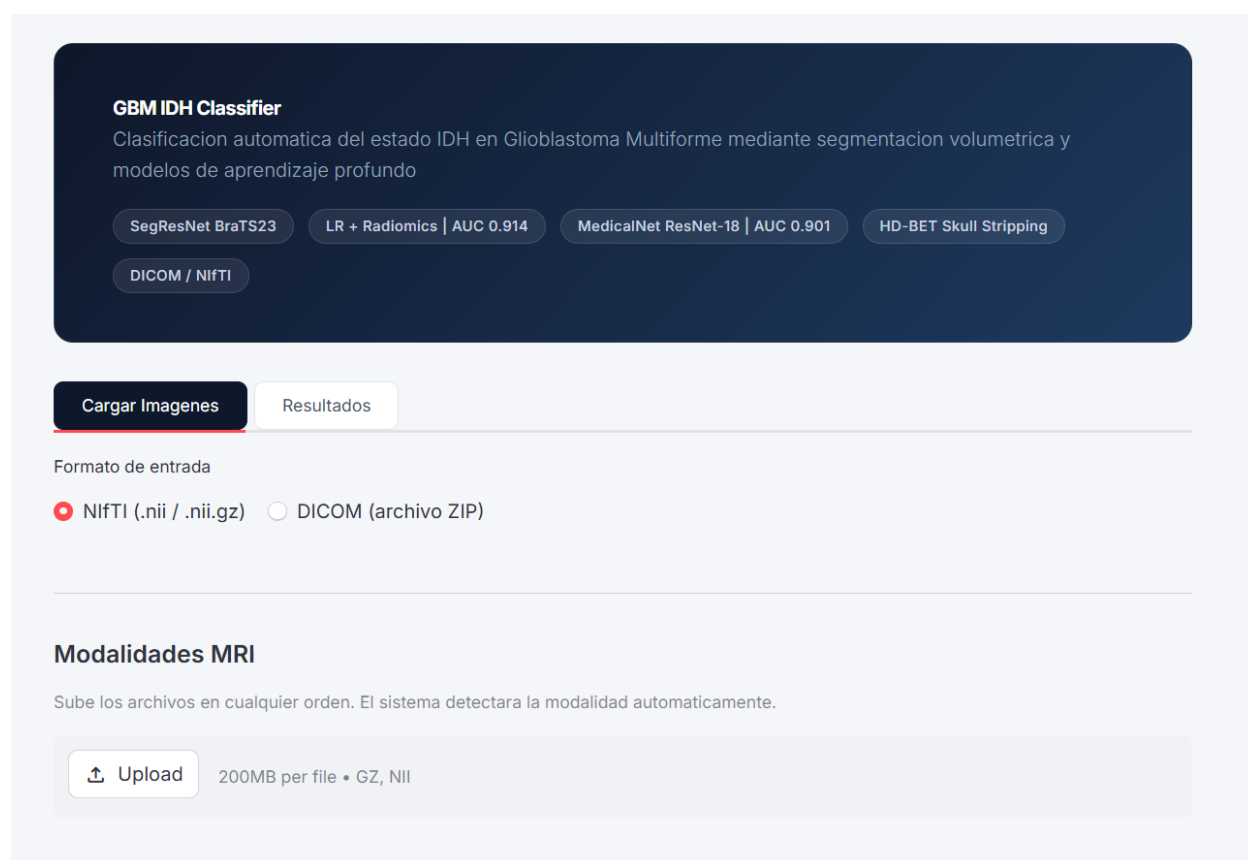


Figura 7.14: Vista general de la plataforma GBM IDH Classifier. El encabezado muestra el título del sistema, su descripción y los componentes tecnológicos integrados mediante etiquetas identificadoras.

7.5.2. Módulo de carga de imágenes

La Figura 7.15 ilustra el módulo de carga de imágenes en modo NIfTI. El usuario selecciona el formato de entrada y carga las cuatro modalidades en cualquier orden. El sistema detecta automáticamente la modalidad correspondiente a cada archivo a partir de su nombre, identificando T1, T1ce, T2 y FLAIR de forma independiente. En el caso presentado, el sistema detectó correctamente las cuatro modalidades del caso UPEN-GBM-00658_01, confirmando el funcionamiento del módulo de detección automática implementado en `gbm_pipeline.py`.

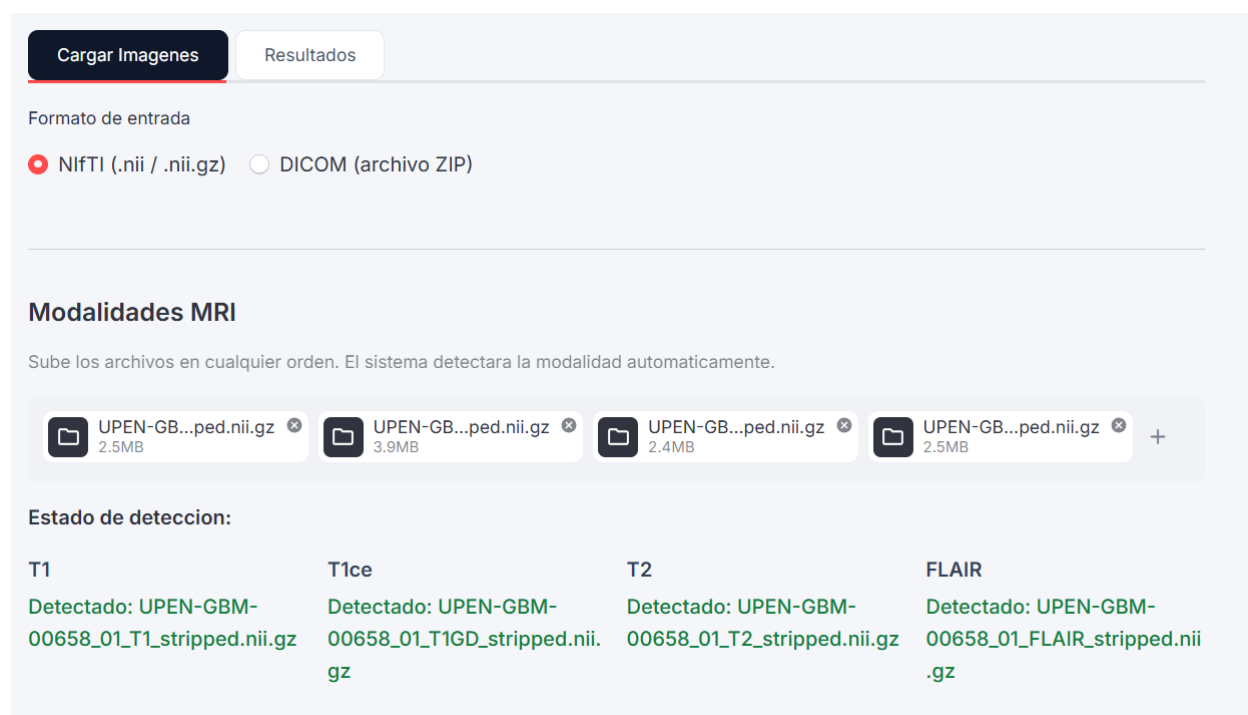


Figura 7.15: Módulo de carga de imágenes en modo NIfTI con detección automática de modalidades. El sistema identifica correctamente T1, T1ce, T2 y FLAIR a partir del nombre de cada archivo cargado.

7.5.3. Configuración y ejecución del análisis

La Figura 7.16 presenta el módulo de configuración del análisis. El usuario selecciona el modelo de clasificación IDH de entre dos opciones disponibles: *Machine Learning — LR + Radiomics* o *Deep Learning — MedicalNet ResNet-18*. En el caso ilustrado se seleccionó el modelo de Deep Learning. Una vez configurado, el usuario ejecuta el pipeline mediante el botón *Ejecutar análisis*, que desencadena secuencialmente las etapas de reorientación, skull stripping, segmentación tumoral y clasificación IDH. Al completarse el proceso, la interfaz muestra un mensaje de confirmación y

un aviso clínico que recuerda que los resultados no constituyen un diagnóstico médico y han sido generados con fines exclusivamente investigativos.



Figura 7.16: Módulo de configuración del análisis con selección del modelo de clasificación IDH. Se muestra el modelo Deep Learning — MedicalNet ResNet-18 seleccionado y el análisis completado exitosamente.

7.5.4. Módulo de resultados

La Figura 7.17 presenta la pestaña de resultados tras la ejecución del pipeline. La interfaz muestra la predicción del estado mutacional IDH en un panel destacado visualmente, indicando *IDH Mutado* con una probabilidad de 98.5 %. Bajo el resultado principal se presenta el indicador de confianza, clasificado como *Alta*, acompañado de la descripción: “*El modelo clasifica este caso como mutado con alta certeza (P=98.5 %)*”. Este indicador comunica al usuario no solo la predicción binaria sino también la certeza del modelo en dicha predicción. Cinco tarjetas de métricas presentan $P(\text{mutado})=98.5 \%$, $P(\text{wildtype})=1.5 \%$, $\text{AUC del modelo}=0.9014$, $\text{Sensibilidad}=85.7 \%$ y $\text{Especificidad}=65.2 \%$, permitiendo contextualizar el resultado individual dentro del rendimiento general del modelo sobre el hold-out de evaluación.



Figura 7.17: Módulo de resultados de la plataforma. Se muestra la predicción IDH Mutado con probabilidad del 98.5 %, indicador de confianza Alta y las métricas del modelo utilizado.

7.5.5. Exploración volumétrica del tumor

La Figura 7.18 presenta el módulo de exploración volumétrica disponible en la pestaña de resultados. Un control deslizante permite navegar a través de los slices axiales del volumen T1ce, mostrando simultáneamente la imagen original y la segmentación tumoral superpuesta en rojo generada por el modelo SegResNet BraTS23. Esta funcionalidad permite al usuario inspeccionar la correspondencia entre la máscara automática y la compleja morfología del tumor, facilitando la validación visual de la segmentación incluso en regiones con patrones de realce periférico o núcleos necróticos. De este modo, el módulo se constituye como una herramienta esencial para verificar con precisión la calidad y consistencia del pipeline sobre cada caso analizado

Exploracion volumetrica del tumor

NAVEGACION DE SLICES AXIALES

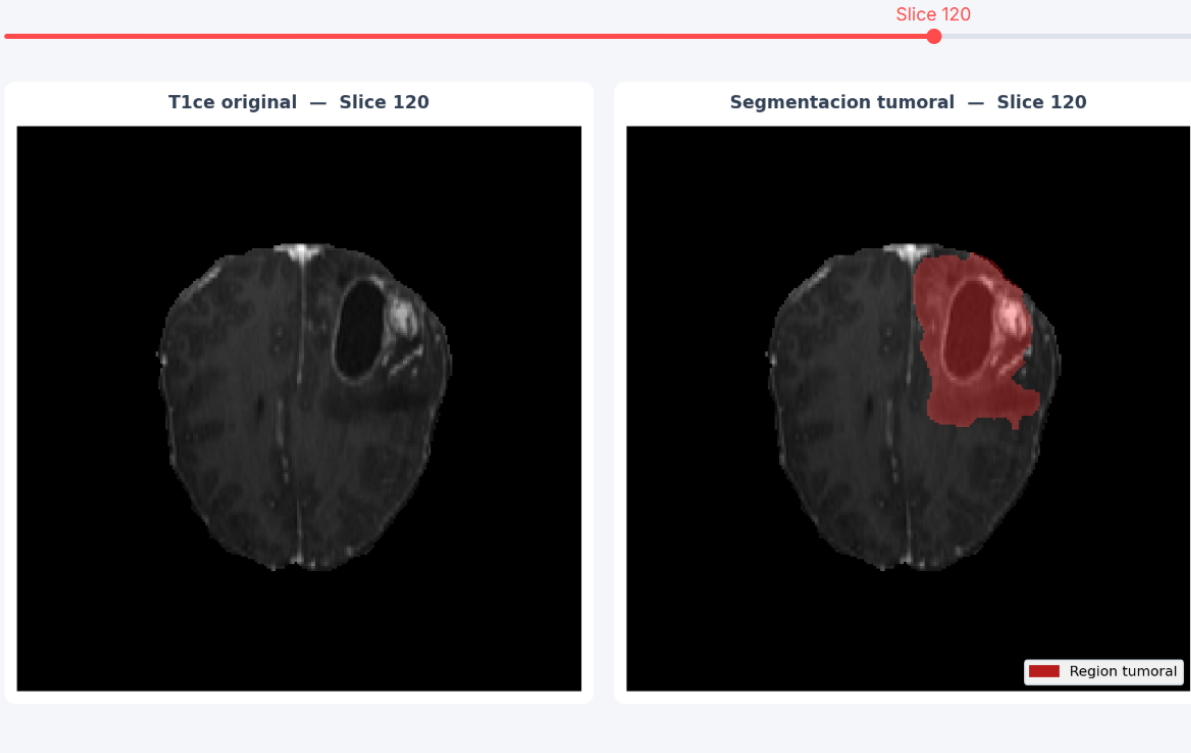


Figura 7.18: Módulo de exploración volumétrica del tumor. El slider axial permite navegar por los cortes del volumen T1ce mostrando la imagen original (izquierda) y la segmentación tumoral superpuesta en rojo (derecha) generada por SegResNet BraTS23.

7.5.6. Generación del reporte de análisis

La Figura 7.19 presenta el reporte de análisis generado automáticamente en formato PDF. El reporte incluye el título del sistema, la fecha y hora de generación, el modelo utilizado, la predicción del estado IDH con su nivel de confianza y las probabilidades asociadas, el AUC del modelo como referencia de rendimiento, y una representación visual del slice axial de mayor área tumoral con la segmentación superpuesta en rojo. El documento incluye al pie un aviso explícito que indica que el reporte ha sido generado por una herramienta de investigación y no constituye un diagnóstico médico, en cumplimiento de los principios éticos aplicables al desarrollo de herramientas de apoyo diagnóstico basadas en inteligencia artificial [127].



Figura 7.19: Reporte de análisis generado automáticamente en formato PDF. Incluye la predicción IDH Mutado con confianza del 98.5 %, AUC del modelo y visualización del corte axial con segmentación tumoral superpuesta.

Conclusiones

- Se desarrolló un sistema de aprendizaje automático capaz de clasificar el estado mutacional del gen IDH en pacientes con Glioblastoma Multiforme a partir de imágenes de resonancia magnética multiparamétrica, integrando un pipeline completo que abarca desde el preprocesamiento de las imágenes hasta la presentación de resultados en una plataforma interactiva.
- Se consolidó una base de datos de 722 casos provenientes de tres repositorios públicos (UPENN-GBM, UCSD-PTGBM y UCSF-PDGM), de los cuales 527 contaban con etiqueta IDH validada. La integración de casos IDH mutado de fuentes externas permitió reducir el desbalance de clases de 7.9:1 a 6.6:1, proporción consistente con la epidemiología reportada en la literatura.
- El flujo de preprocesamiento automatizado demostró ser consistente y confiable. El uso de HD-BET permitió procesar exitosamente los 722 casos disponibles, sin generar errores de ejecución y produciendo en todos los casos máscaras cerebrales con volúmenes anatómicamente plausibles, incluso en presencia de patologías complejas y variaciones en los parámetros de adquisición entre los distintos repositorios. Esto permitió que las etapas posteriores de segmentación y clasificación operaran sobre el parénquima cerebral con una interferencia mínima de tejidos extracraneales.
- A partir de 1130 características radiómicas extraídas de la secuencia T1GD mediante ANOVA y Random Forest, se identificaron 30 descriptores que, en el contexto de este trabajo, mostraron capacidad para capturar información discriminativa asociada al estado mutacional IDH, respaldado por el AUC-ROC de 0.9138 y MCC de 0.5953 obtenidos por el modelo de Regresión Logística sobre el hold-out independiente de 106 casos. Si bien el F1-Score de 0.6471 refleja las limitaciones inherentes al desbalance de clases de 6.6:1, estos resultados sugieren que la radiómica constituye un enfoque interpretable y computacionalmente eficiente frente a las arquitecturas de Deep Learning en escenarios con datos limitados y desbalance de clases.
- Se entrenaron cinco modelos de Machine Learning y dos arquitecturas de Deep Learning para la clasificación del estado IDH. Los resultados sugieren que el enfoque de Machine Learning basado en características radiómicas de la secuencia T1GD superó ligeramente a las arquitecturas de Deep Learning en rendimiento global. El modelo de Regresión Logística obtuvo el mejor desempeño con un AUC-ROC de 0.9138 y una especificidad del 90.22% sobre el conjunto de hold-out independiente de 106 casos.

- La evaluación cuantitativa mediante métricas como AUC-ROC, F1-Score, sensibilidad, especificidad y MCC permitió una comparación rigurosa y objetiva entre los modelos. El coeficiente MCC resultó especialmente informativo en presencia del desbalance de clases, evidenciando que la Regresión Logística logra el mejor equilibrio entre la detección de casos mutados y la correcta clasificación de casos wildtype. Por su parte, el modelo MedicalNet ResNet-18 demostró ser altamente competitivo con un AUC-ROC de 0.9014 y, junto con SegResNet, alcanzó una sensibilidad superior del 85.71 %, identificando correctamente a la mayoría de los pacientes con la mutación IDH.
- En condiciones de un conjunto de datos limitado y un marcado desbalance de clases (6.6:1), el enfoque de Machine Learning basado en radiómica ofrece una capacidad de regularización superior a las arquitecturas de Deep Learning volumétricas. Mientras que los modelos MedicalNet y SegResNet tendieron a sobreajustarse a la clase minoritaria (IDH mutado) a costa de una alta tasa de falsos positivos, la Regresión Logística logró el desempeño más equilibrado (MCC = 0.5953). Este resultado sugiere que, en condiciones de dataset limitado y marcado desbalance, la extracción dirigida de características cuantitativas puede ofrecer mayor estabilidad, mitigando el riesgo de interpretaciones erróneas derivadas del desbalance epidemiológico del glioblastoma.
- Los resultados del preentrenamiento de MedicalNet sobre 23 conjuntos de datos médicos confirman que las características anatómicas generales aprendidas en múltiples dominios son más transferibles y proporcionan una convergencia más rápida y estable durante el entrenamiento en tareas de clasificación con muestras reducidas y desbalanceadas, superando al modelo SegResNet adaptado para clasificación IDH (AUC = 0.8012 vs AUC = 0.9014).
- Se implementó una plataforma interactiva desarrollada con Streamlit que integra el pipeline completo de inferencia. La plataforma permite cargar imágenes en formato NIfTI y DICOM, ejecutar el preprocesamiento automático, seleccionar el modelo de clasificación y visualizar los resultados con un slicer volumétrico interactivo y un reporte exportable en PDF, facilitando su uso por parte de investigadores y profesionales del área de la salud.
- Los resultados obtenidos en conjunto sugieren la viabilidad técnica de clasificar de forma no invasiva el estado mutacional IDH en Glioblastoma Multiforme mediante imágenes de resonancia magnética, con potencial para complementar los métodos histopatológicos convencionales en entornos de investigación clínica, si bien una validación prospectiva sobre datos externos sería necesaria antes de considerar cualquier aplicación asistencial.

Trabajos futuros

El presente trabajo abre diversas líneas de investigación y desarrollo que podrían fortalecer y extender los resultados obtenidos.

- **Modelo híbrido ensemble ML + DL.** El análisis de los resultados sobre el hold-out independiente evidenció que el modelo de Regresión Logística y MedicalNet ResNet-18 fallaron en casos diferentes, detectando 11 y 12 de los 14 casos IDH mutado respectivamente. Esta complementariedad sugiere que un modelo ensemble que combine las probabilidades de salida de ambos enfoques mediante stacking o votación ponderada podría superar el rendimiento individual de cada modelo, constituyendo una línea de trabajo con alto potencial de mejora.
- **Validación externa prospectiva.** Los modelos desarrollados fueron entrenados y evaluados exclusivamente sobre datos de tres repositorios públicos con características de adquisición similares. Una validación prospectiva sobre imágenes provenientes de instituciones clínicas con equipos de MRI distintos, protocolos de adquisición diferentes y poblaciones de pacientes no vistas constituye un paso necesario para demostrar la generalización real del sistema antes de cualquier aplicación asistencial.
- **Incorporación de múltiples modalidades para extracción de características radiómicas.** En este trabajo la extracción de características radiómicas se realizó exclusivamente sobre la secuencia T1GD. La incorporación de las secuencias T2 y FLAIR permitiría capturar información complementaria sobre el edema peritumoral y la heterogeneidad intratumoral que T1GD no representa, con potencial de incrementar el poder discriminativo del modelo de Machine Learning.
- **Ampliación de la base de datos mediante repositorios adicionales y convenios institucionales.** El tamaño limitado del dataset y el marcado desbalance de clases (6.6:1) constituyeron las principales restricciones del trabajo. La incorporación de repositorios como TCGA-GBM y TCIA, así como el establecimiento de convenios con instituciones clínicas que dispongan de imágenes de MRI con etiqueta IDH confirmada, permitiría aumentar la representación de la clase minoritaria y mejorar la capacidad de generalización de los modelos.
- **Segmentación tumoral multi-clase.** La segmentación empleada en este trabajo consideró la región *whole tumor* como una única máscara binaria. El uso de las subregiones tumorales de forma independiente — núcleo activo, zona necrótica y edema peritumoral — para la extracción de características radiómicas y el crop volumétrico de los modelos de Deep Learning podría capturar patrones de heterogeneidad intratumoral más específicos asociados al estado mutacional IDH.

- **Optimización del pipeline para entornos sin GPU.** El pipeline desarrollado depende de infraestructura de cómputo en la nube con GPU para ejecutarse en tiempos razonables. El desarrollo de una versión optimizada para CPU, empaquetada como aplicación standalone, ampliaría la accesibilidad del sistema a instituciones que no disponen de infraestructura especializada.
- **Segmentación tumoral interactiva con reclasificación automática.** La plataforma actual presenta al usuario la segmentación generada por SegResNet BraTS23 sin posibilidad de corrección. La incorporación de un módulo de edición interactiva de la máscara tumoral, que permita al especialista ajustar manualmente la segmentación y desencadene automáticamente una reclasificación IDH basada en la nueva región de interés, cerraría el ciclo clínico entre el modelo y el experto, mejorando la confiabilidad del resultado final.
- **Integración con sistemas PACS.** La plataforma actual requiere que el usuario cargue manualmente las imágenes en formato NIfTI o DICOM. La integración con sistemas de archivo y comunicación de imágenes (PACS) hospitalarios permitiría que las imágenes lleguen directamente desde el equipo de MRI al pipeline de análisis, reduciendo la fricción operativa y facilitando su adopción en entornos clínicos reales.
- **Incorporación de variables clínicas y biomarcadores moleculares adicionales.** El sistema desarrollado clasifica el estado IDH exclusivamente a partir de características de imagen. La incorporación de variables clínicas como edad, sexo y grado histológico, junto con biomarcadores moleculares relevantes en el contexto del GBM como el estado de metilación del promotor MGMT, la mutación TERT y la pérdida de expresión ATRX, como features complementarias al modelo podría incrementar el poder predictivo del sistema y acercarlo a un perfil molecular más completo del tumor.

Los siguientes anexos forman parte integral de este trabajo de grado. Debido a la extensión de los códigos desarrollados, todos los archivos fuente se encuentran almacenados en el repositorio GitHub y en Google Drive indicados en los Anexos G y H.

10.1. Anexo A: Estructura de la Carpeta GBM_TESIS

La carpeta principal del proyecto alojada en Google Drive contiene la siguiente estructura de directorios y archivos:

```
GBM_TESIS/  
|-- Dataset/  
|   |-- labels_idh.csv  
|-- SS_PIPELINE/  
|   |-- ss_radiomics_config.json  
|   |-- radiomics/  
|       |-- params/radiomics_params.yaml  
|       |-- features/features_t1ce.csv  
|       |-- selected_features/features_selected.csv  
|   |-- clasificador_v2/  
|       |-- models/  
|       |-- results/  
|       |-- figures/  
|   |-- inference_pipeline_v2/  
|       |-- best_model.joblib  
|       |-- pipeline_config.json  
|   |-- dl_clasificador_v2/  
|       |-- medicalnet/  
|           |-- models/medicalnet_best.pth  
|           |-- results/  
|           |-- figures/  
|       |-- segresnet/  
|           |-- models/segresnet_best.pth  
|           |-- results/  
|           |-- figures/  
|   |-- metrics/
```

```

| |-- figures/
|-- MODELO PREENTRENADO/
| |-- models/brats_mri_segmentation/
|-- hdbet_weights/
| |-- release_2.0.0/fold_all/checkpoint_final.pth
|-- medicalnet_weights/
| |-- resnet18.pth
|-- notebooks /
| |-- SS_01_skull_stripping.ipynb
| |-- SS_02_segmentacion.ipynb
| |-- SS_03_radiomics.ipynb
| |-- SS_04_clasificador.ipynb
| |-- DL_01_segresnet.ipynb
| |-- DL_02_medicalnet.ipynb
|-- Interfaz/
| |-- SS_06_interfaz.ipynb
| |-- app.py
| |-- gbm_pipeline.py

```

A continuación se detalla el contenido de cada directorio y archivo relevante.

10.2. Anexo B: Códigos de Preprocesamiento

Ubicación: notebooks/

Contiene los notebooks de preprocesamiento de las imágenes MRI multiparamétricas del dataset UPENN-GBM.

- `SS_01_skull_stripping.ipynb`: Pipeline de extracción de cráneo mediante HD-BET sobre las 722 imágenes del dataset. Incluye detección automática del tipo de imagen, procesamiento por lotes y generación de métricas de calidad.
- `SS_02_segmentacion.ipynb`: Segmentación tumoral automática utilizando el modelo preentrenado SegResNet BraTS23 mediante el framework MONAI. Genera las máscaras de Whole Tumor para cada caso y calcula métricas Dice e IoU contra las 147 máscaras de referencia disponibles.

10.3. Anexo C: Códigos de Extracción de Características

Ubicación: notebooks/

Contiene el notebook de extracción de características radiómicas a partir de las imágenes preprocesadas.

- `SS_03_radiomics.ipynb`: Extracción de características radiómicas mediante PyRadiomics sobre la imagen T1GD de cada caso. Genera 1130 características por caso (722 casos en total) y aplica selección de características mediante ANOVA y Random Forest, obteniendo las 30 características más discriminativas para la clasificación IDH.

10.4. Anexo D: Códigos de Clasificación ML

Ubicación: `notebooks/`

Contiene el notebook de entrenamiento y evaluación de los modelos de Machine Learning.

- `SS_04_clasificador.ipynb`: Entrenamiento y evaluación de cinco clasificadores de Machine Learning (Regresión Logística, Random Forest, SVM, XGBoost y LightGBM) sobre las 30 características radiómicas seleccionadas. Implementa validación cruzada estratificada 5-fold, optimización de hiperparámetros y evaluación en el conjunto de hold-out independiente de 106 casos.

10.5. Anexo E: Códigos de Clasificación DL

Ubicación: `notebooks/`

Contiene los notebooks de entrenamiento y evaluación de los modelos de Deep Learning.

- `DL_01_segresnet.ipynb`: Entrenamiento y evaluación del clasificador basado en SegResNet BraTS23 para la clasificación IDH. Incluye el pipeline completo de preprocesamiento, entrenamiento con datos aumentados y evaluación en el hold-out independiente.
- `DL_02_medicalnet.ipynb`: Entrenamiento y evaluación del clasificador MedicalNet ResNet-18 con pesos preentrenados en imágenes médicas 3D. El modelo recibe las 4 modalidades MRI simultáneamente y clasifica el estado IDH. Incluye fine-tuning, curvas de entrenamiento y evaluación en hold-out con umbral de clasificación 0.5.

10.6. Anexo F: Código de Interfaz

Ubicación: `Interfaz/`

Contiene el código completo de la interfaz web desarrollada con Streamlit para la inferencia clínica del sistema de clasificación IDH.

- `gbm_pipeline.py`: Módulo principal del pipeline de inferencia. Implementa skull stripping con HD-BET, segmentación tumoral con SegResNet BraTS23, extracción de características radiómicas con PyRadiomics y clasificación IDH con los modelos ML y DL entrenados.
- `app.py`: Interfaz gráfica Streamlit que integra el pipeline completo. Permite cargar imágenes en formato NIfTI o DICOM, seleccionar el modelo de clasificación, visualizar resultados con el slicer interactivo y exportar el reporte en formato PDF.

- `SS_06_interfaz.ipynb`: Notebook de Google Colab para desplegar la interfaz. Instala las dependencias, monta Google Drive, restaura los pesos preentrenados y lanza la aplicación Streamlit con URL pública mediante cloudflared.

10.7. Anexo G: Repositorio GitHub

Ubicación: <https://github.com/juliopinedatoro2004-tech/gbm-idh-classifier>

El repositorio GitHub contiene todos los notebooks, scripts y archivos de configuración del proyecto, incluyendo el código completo del pipeline de preprocesamiento, extracción de características radiómicas, clasificadores ML y DL, y la interfaz Streamlit.

Nota: Todos los códigos incluidos están documentados internamente con comentarios explicativos y siguen las buenas prácticas de programación en Python.

10.8. Anexo H: Manual de Usuario

Ubicación: https://drive.google.com/file/d/1p_jv8GS5w0_opz-xxB76gu0yU7fdgQ8K/view?usp=sharing

Documentación completa que describe el uso de la interfaz de clasificación desarrollada. El manual incluye los requisitos previos, el proceso de inicio de la interfaz, la carga de imágenes MRI en formato NIFTI y DICOM, la interpretación de resultados y los errores frecuentes con sus respectivas soluciones.

Nota: La estructura presentada refleja la organización real de archivos en el momento de la entrega del trabajo de grado. Se recomienda consultar el Manual de Usuario para información específica sobre el uso de la interfaz de clasificación.

10.9. Anexo I: Video funcionamiento plataforma interactiva

Ubicación: https://drive.google.com/file/d/14lioX0PkZD_s1Yv2UMsoQd9JgTAa-T8l/view?usp=drive_link

El anterior enlace corresponde a un video demostrativo que ilustra el funcionamiento completo de la plataforma interactiva, desde la carga de imágenes en formato NIFTI hasta la visualización de la segmentación tumoral y la generación del reporte de análisis en formato PDF.

Bibliografía

- [1] S. Han, Y. Liu, S. J. Cai, M. Qian, J. Ding, M. Larion, M. R. Gilbert, and C. Yang, “Mutación de idh en gliomas: mecanismos moleculares y posibles dianas terapéuticas,” *British Journal of Cancer*, vol. 122, no. 11, pp. 1580–1589, 2020.
- [2] Instituto Nacional del Cáncer, “Glioma difuso de línea media,” <https://www.cancer.gov/rare-brain-spine-tumor/espanol/tumores/gliomas-difuso-linea-media>, 2020.
- [3] MRI Master, “T1 vs t2 vs pd vs flair mri,” <https://mrimaster.com/t1-vs-t2-vs-pd-vs-flair-mri/>, 2024, accedido: 2025.
- [4] ResearchGate, “T1gd and t2-weighted mri images of a patient diagnosed with glioblastoma multiforme (gbm),” https://www.researchgate.net/figure/T1Gd-and-T2-weighted-MRI-images-of-a-patient-diagnosed-with-glioblastoma-multiforme-GBM_fig3_23278845, 2024, accedido: 2025.
- [5] S. Chen, K. Ma, and Y. Zheng, “Med3D: Transfer learning for 3D medical image analysis,” *arXiv preprint arXiv:1904.00625*, 2019. [Online]. Available: <https://arxiv.org/abs/1904.00625>
- [6] ResearchGate, “Resnet-18 architecture,” https://www.researchgate.net/figure/ResNet-18-Architecture_fig2_382725662, 2024, accedido: 2025.
- [7] A. Myronenko, “3D MRI brain tumor segmentation using autoencoder regularization,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018*, ser. Lecture Notes in Computer Science, vol. 11384. Springer, 2019, pp. 311–320. [Online]. Available: <https://arxiv.org/abs/1810.11654>
- [8] IBM, “What is logistic regression?” <https://www.ibm.com/think/topics/logistic-regression>, 2024, accedido: 2025.
- [9] Data Science Dojo. (2024) Understanding the random forest algorithm – a comprehensive guide. Consultado el 8 de mayo de 2026. [Online]. Available: <https://datasciencedojo.com/blog/random-forest-algorithm/>
- [10] IBM, “What is a support vector machine?” <https://www.ibm.com/think/topics/support-vector-machine>, 2024, accedido: 2025.
- [11] GeeksforGeeks, “Implementation of xgboost,” <https://www.geeksforgeeks.org/machine-learning/implementation-of-xgboost-extreme-gradient-boosting/>, 2024, accedido: 2025.
- [12] —, “Lightgbm (light gradient boosting machine),” <https://www.geeksforgeeks.org/machine-learning/lightgbm-light-gradient-boosting-machine/>, 2024, accedido: 2025.

- [13] —, “AUC-ROC curve,” <https://www.geeksforgeeks.org/machine-learning/auc-roc-curve/>, 2024, accedido: 2025.
- [14] H2O.ai, “Confusion matrix,” <https://h2o.ai/wiki/confusion-matrix/>, 2024, accedido: 2025.
- [15] K. H. Zou *et al.*, “Statistical validation of image segmentation quality based on a spatial overlap index,” *Academic Radiology*, vol. 11, no. 2, pp. 178–189, 2004. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1415224/>
- [16] LearnOpenCV, “Intersection over IoU in object detection and segmentation,” <https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/>, 2024, accedido: 2025.
- [17] S. Grochans, A. M. Cybulska, D. Simińska, J. Korbecki, K. Kojder, D. Chlubek, and I. Baranowska-Bosiacka, “Epidemiology of glioblastoma multiforme—literature review,” *Cancers (Basel)*, vol. 14, no. 10, p. 2412, May 2022.
- [18] Y. Zhang and et al., “Emerging role of ferroptosis in glioblastoma: Potential mechanisms and therapeutic targeting strategies,” *Frontiers in Oncology*, vol. 13, p. 10252079, 2023.
- [19] D. N. Louis, A. Perry, P. Wesseling *et al.*, “Clasificación de la oms de tumores del sistema nervioso central de 2021,” *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.
- [20] B. Huang, X. Li, Y. Li, J. Zhang, Z. Zong, and H. Zhang, “Current immunotherapies for glioblastoma multiforme,” *Frontiers in Immunology*, vol. 11, p. 603911, Mar 2021.
- [21] D. J. Brat, K. Aldape, J. A. Bridge *et al.*, “Pruebas de biomarcadores moleculares para el diagnóstico de gliomas difusos,” *Archives of Pathology & Laboratory Medicine*, vol. 146, no. 5, pp. 547–574, 2022.
- [22] N. Sanai and M. S. Berger, “Surgical oncology for gliomas: The state of the art,” *Nature Reviews Clinical Oncology*, vol. 15, no. 2, pp. 112–125, 2018.
- [23] M. Smits and M. J. van den Bent, “Imaging correlates of adult glioma genotypes,” *Radiology*, vol. 284, no. 2, pp. 316–331, 2017.
- [24] J. Smith *et al.* (2024) Recent advances in glioma research. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10980741/>
- [25] A. K. Philip, B. A. Samuel, S. Bhatia, S. A. M. Khalifa, and H. R. El-Seedi, “Artificial intelligence and precision medicine: A new frontier for brain tumor treatment,” *Life*, vol. 13, no. 1, p. 24, 2023.
- [26] S. Bakas, C. Sako, H. Akbari, M. Bilello, A. Sotiras, G. Shukla *et al.*, “The university of pennsylvania glioblastoma (UPenn-GBM) cohort: Advanced MRI, clinical, genomics, & radiomics,” *Scientific Data*, vol. 9, p. 453, 2022.

- [27] L. Gagnon, D. Gupta, U. Nguyen *et al.*, “The university of california san diego post-treatment glioblastoma (UCSD-PTGBM) annotated multimodal MRI dataset,” *Scientific Data*, vol. 13, p. 183, 2026.
- [28] E. Calabrese, J. E. Villanueva-Meyer, J. D. Rudie, A. M. Rauschecker, U. Baid, S. Bakas, S. Cha, J. T. Mongan, and C. P. Hess, “The UCSF preoperative diffuse glioma MRI (UCSF-PDGM) dataset,” *Radiology: Artificial Intelligence*, 2022.
- [29] M. E. Davis, “Epidemiology and overview of gliomas,” *Seminars in Oncology Nursing*, vol. 34, no. 5, pp. 420–429, Dec 2018.
- [30] N. Staff, “How a cancer-causing metabolite blocks immune cells,” Online, Aug 2022, national Cancer Institute. [Online]. Available: <https://www.cancer.gov/news-events/cancer-current-s-blog/2022/idh1-cancer-metabolite-blocks-immune-cells>
- [31] K. Li, D. Lu, Y. Guo, C. Wang, X. Liu, Y. Liu, and D. Liu, “Trends and patterns of incidence of diffuse glioma in adults in the united states, 1973–2014,” *Cancer Medicine*, vol. 7, no. 10, pp. 5281–5290, Oct 2018.
- [32] M. Salcman, “Epidemiology and factors affecting survival,” in *Malignant Cerebral Glioma*, M. L. J. Apuzzo, Ed. Park Ridge, IL: American Association of Neurological Surgeons, 1990, pp. 95–110.
- [33] G. Zada and *et al.*, “Advances in glioma surgery,” Online, Sep 2019, *frontiers in Oncology*. [Online]. Available: <https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2019.00963/full>
- [34] M. I. O. Navia, J. C. G. Vega, and O. H. F. Lee, “Epidemiología y caracterización general de los tumores cerebrales primarios en el adulto,” 2018, online. [Online]. Available: [https://revistas.javeriana.edu.co/filesarticulos/UMED/60-1%20\(2019-I\)/231057460010/](https://revistas.javeriana.edu.co/filesarticulos/UMED/60-1%20(2019-I)/231057460010/)
- [35] A. Miranda-Filho, M. Piñeros, I. Soerjomataram, I. Deltour, and F. Bray, “Cancers of the brain and cns: Global patterns and trends in incidence,” *Neuro-Oncology*, vol. 19, no. 2, pp. 270–280, Feb 2017.
- [36] S. Cha, “Perfusion mr imaging of brain tumors,” *Topics in Magnetic Resonance Imaging*, vol. 15, no. 5, pp. 279–289, Oct 2004.
- [37] J. J. C. E. de los Monteros, R. V. Portero, A. L. Méndez, R. R. Romero, and J. V. García, *Resonancia magnética perfusión en la diferenciación de los gliomas cerebrales*. Sevilla, Spain: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, 2008.
- [38] E. M. Driscoll, R. E. Anderson, and R. E. Latchaw, “Percutaneous biopsy of the brain: A safe and effective technique,” *American Journal of Roentgenology*, vol. 179, no. 3, pp. 783–786, Sep 2002.

- [39] R. Stupp *et al.*, “Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma,” *New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, Mar 2005.
- [40] M. H. Cohen, Y. L. Shen, P. Keegan, and R. Pazdur, “Fda drug approval summary: Bevacizumab (avastin) as treatment of recurrent glioblastoma multiforme,” *The Oncologist*, vol. 17, no. 11, pp. 1482–1482, 2012.
- [41] J. Arbizu, P. D. Domínguez, R. Díez-Valle, C. Vigil, R. García-Eulate, J. L. Zubieta, and J. A. Richter, “Neuroimagen de los tumores cerebrales,” *Revista Española de Medicina Nuclear*, vol. 30, no. 1, pp. 47–65, 2011.
- [42] H. Zhang *et al.*, “Glioma treatment advances,” *PMC*, Mar 2021.
- [43] C. Pardo and R. Cendales, *Incidencia, mortalidad y prevalencia de cáncer en Colombia 2007–2011*. Bogotá, Colombia: Instituto Nacional de Cancerología, 2015, vol. 1.
- [44] A. Suárez, M. Castellanos, A. Simbaqueba, and Gamboa, “Aspectos clínicos y demora para el diagnóstico en niños con tumores del sistema nervioso central,” *Revista Colombiana de Cancerología*, vol. 15, no. 3, pp. 127–134, 2011.
- [45] Ministerio de Salud y Protección Social. (n.d.) Registro de prestadores habilitados. [Online]. Available: https://prestadores.minsalud.gov.co/habilitacion/consultas/habilitados_reps.aspx
- [46] D. Beltrán. (2025, May) Panorama del servicio de salud en colombia. [Online]. Available: <https://www.infobae.com/colombia/2025/05/04/panorama-del-servicio-de-salud-en-colombia-a-crisis-en-eps-pone-en-riesgo-atencion-de-millones-de-ciudadanos/>
- [47] M. C. B. Castro and L. C. L. Chaparro. (2023) Entre la burocracia y la negligencia: cuando hay fallas en la atención médica. [Online]. Available: <https://www.javeriana.edu.co/pesquisa/acceso-salud-colombia-derecho-petro/>
- [48] D. R. Louis *et al.*, “Molecular classification of gliomas,” *Neuro-Oncology*, vol. 17, no. 9, pp. 1188–1190, 2015.
- [49] A. B. Smith *et al.* (2024) Glioma research updates. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11445779/>
- [50] D. Capper, D. T. W. Jones, M. Sill *et al.*, “Clasificación de tumores del sistema nervioso central basada en la metilación del adn,” *Nature*, vol. 555, no. 7697, pp. 469–474, 2018.
- [51] T. T. Dunder *et al.*, “Machine learning-based surgical planning for neurosurgery,” *Frontiers in Surgery*, vol. 9, p. 863633, 2022.
- [52] Instituto Nacional del Cáncer, “Diccionario de cáncer: Tumor,” <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/tumor>, 2025.

- [53] Top Doctors, “Tumores benignos,” <https://www.topdoctors.com.co/diccionario-medico/tumores-benignos/>, 2023.
- [54] Alivia, “Tumores malignos: ¿por qué son así?” <https://alivia.es/sobre-el-cancer/tumores-malignos-por-que-son-asi/>, 2025.
- [55] Instituto Nacional del Cáncer, “Diccionario de cáncer: Astrocitoma,” <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/astrocitoma>, 2025.
- [56] H. Khurshid and S. A. Hassan, “Glioblastoma multiforme,” in *StatPearls*. Treasure Island, FL: StatPearls Publishing, 2023.
- [57] Organización Mundial de la Salud, “Acerca de la oms,” <https://www.who.int/es/about>, 2023.
- [58] P. Y. Wen and R. J. Packer, “The 2021 who classification of tumors of the central nervous system: clinical implications,” *Neuro-Oncology*, vol. 23, no. 8, pp. 1215–1217, 2021.
- [59] Instituto Nacional del Cáncer, “Diccionario de cáncer: Gen idh1,” <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/gen-idh1>, 2025.
- [60] F.-J. Chou, Y. Liu, F. Lang, and C. Yang, “D-2-hidroxiglutarato en la biología del glioma,” *Cells*, vol. 10, no. 9, p. 2345, 2021.
- [61] Moffitt Cancer Center, “Glioblastoma causes & risk factors,” <https://www.moffitt.org/cancers/glioblastoma/diagnosis/causes/>, 2025.
- [62] Mayo Clinic, “Glioblastoma: síntomas y causas,” <https://www.mayoclinic.org/es/diseases-conditions/glioblastoma/symptoms-causes/syc-20569077>, 2025.
- [63] J. Li, L. Feng, and Y. Lu, “Glioblastoma multiforme: Diagnosis, treatment, and invasion,” *Journal of Biomedical Research*, vol. 37, no. 1, pp. 47–58, 2022.
- [64] Instituto Nacional de Imágenes Biomédicas y Bioingeniería, “Imagen por resonancia magnética (irm),” <https://www.nibib.nih.gov/espanol/temas-cientificos/imagen-por-resonancia-magn%C3%A9tica-irm>, 2025.
- [65] E. Kanal and M. F. Tweedle, “Gadolinium magnetic resonance imaging,” In: *StatPearls*. <https://www.ncbi.nlm.nih.gov/books/NBK482487/>, 2023, accedido: 2025.
- [66] Salesforce, “¿qué es el deep learning?” <https://www.salesforce.com/es/resources/definition/deep-learning/>, 2025.
- [67] IBM, “Transfer learning,” <https://www.ibm.com/mx-es/think/topics/transfer-learning>, 2024, accedido: 2025.
- [68] GeeksforGeeks, “Resnet18 from scratch using pytorch,” <https://www.geeksforgeeks.org/deep-learning/resnet18-from-scratch-using-pytorch/>, 2024, accedido: 2025.

- [69] IBM, “What is data augmentation?” <https://www.ibm.com/think/topics/data-augmentation>, 2024, accedido: 2025.
- [70] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, “Synthstrip: skull-stripping for any brain image,” *NeuroImage*, vol. 260, p. 119474, 2022. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9465771/>
- [71] F. Isensee, M. Schell, I. Pflueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H. P. Schlemmer, S. Heiland, W. Wick, M. Bendszus, K. H. Maier-Hein, and P. Kickingereder, “Automated brain extraction of multisequence MRI using artificial neural networks,” *Human Brain Mapping*, vol. 40, no. 17, pp. 4952–4964, 2019. [Online]. Available: <https://github.com/MIC-DKFZ/HD-BET>
- [72] M. Iqbal *et al.*, “Enhanced brain tumor segmentation in medical imaging using multi-modal multi-scale contextual aggregation and attention fusion,” *Scientific Reports*, vol. 15, 2025. [Online]. Available: <https://www.nature.com/articles/s41598-025-21255-4>
- [73] B. Bonato, L. Nanni, and A. Bertoldo, “Advancing precision: A comprehensive review of MRI segmentation datasets from BraTS challenges (2012–2025),” *Sensors*, vol. 25, no. 6, p. 1838, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11945730/>
- [74] Amazon Web Services, “¿qué es el machine learning?” <https://aws.amazon.com/es/what-is/machine-learning/>, 2025.
- [75] NUBIX, “Radiómica, un camino a mejorar la medicina de precisión,” <https://nubix.cloud/radiologia/radiomica-un-camino-a-mejorar-la-medicina-de-precision>, 2021.
- [76] PyRadiomics Development Team, “Radiomic features — PyRadiomics documentation,” <https://pyradiomics.readthedocs.io/en/latest/features.html>, 2024, accedido: 2025.
- [77] T. Furbaa *et al.*, “Introduction to radiomics,” *Journal of Nuclear Medicine*, vol. 61, no. 4, pp. 488–495, 2020. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9374044/>
- [78] IBM, “What is random forest?” <https://www.ibm.com/think/topics/random-forest>, 2024, accedido: 2025.
- [79] —, “What is xgboost?” <https://www.ibm.com/think/topics/xgboost>, 2024, accedido: 2025.
- [80] H2O.ai, “AUC-ROC,” <https://h2o.ai/wiki/auc-roc/>, 2024, accedido: 2025.
- [81] IBM, “What is a confusion matrix?” <https://www.ibm.com/think/topics/confusion-matrix>, 2024, accedido: 2025.
- [82] Deepchecks, “Sensitivity and specificity in machine learning,” <https://deepchecks.com/glossary/sensitivity-and-specificity-of-machine-learning/>, 2024, accedido: 2025.
- [83] Roboflow, “F1 score,” <https://blog.roboflow.com/f1-score/>, 2024, accedido: 2025.

- [84] Quantdare, “Confusion matrix MCC statistic,” <https://quantdare.com/confusion-matrix-mcc-statistic/>, 2024, accedido: 2025.
- [85] Wikipedia, “Youden’s j statistic,” https://en.wikipedia.org/wiki/Youden%27s_J_statistic, 2024, accedido: 2025.
- [86] IBM, “What is upsampling?” <https://www.ibm.com/think/topics/upsampling>, 2024, accedido: 2025.
- [87] GeeksforGeeks, “Cross validation in machine learning,” <https://www.geeksforgeeks.org/machine-learning/cross-validation-machine-learning/>, 2024, accedido: 2025.
- [88] IBM, “¿qué es una gpu?” <https://www.ibm.com/es-es/think/topics/gpu>, 2024, accedido: 2025.
- [89] W. D. Bidgood, S. C. Horii, F. W. Prior, and D. E. V. Syckle, “Understanding and using DICOM, the data interchange standard for biomedical imaging,” *Journal of the American Medical Informatics Association*, vol. 4, no. 3, pp. 199–212, 1997. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC61235/>
- [90] GeeksforGeeks, “A beginners guide to streamlit,” <https://www.geeksforgeeks.org/python/a-begginners-guide-to-streamlit/>, 2024, accedido: 2025.
- [91] C. Pitarch, V. Ribas, and A. Vellido, “Ai-based glioma grading for a trustworthy diagnosis,” *Cancers*, vol. 15, no. 13, p. 3369, 2023.
- [92] P. C. Tripathi and S. Bag, “An attention-guided cnn framework for segmentation and grading of glioma using 3d mri scans,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 3, pp. 1890–1904, 2023.
- [93] R. S. Eijgelaar *et al.*, “Robust deep learning–based segmentation of glioblastoma on routine clinical mri scans using sparsified training,” *Radiology: Artificial Intelligence*, vol. 2, no. 5, p. e190103, 2020.
- [94] C. Gunasundari and K. S. Bhuvaneshwari, “Machine learning fusion for glioma tumor detection,” *Scientific Reports*, vol. 15, p. 11236, 2025.
- [95] A. García-Ruiz *et al.*, “Una herramienta de aprendizaje profundo accesible para la clasificación voxel por voxel de neoplasias malignas cerebrales,” *Cell Reports Medicine*, vol. 5, no. 3, p. 101464, 2024.
- [96] Neuroimaging Informatics Technology Initiative, “NIfTI: Neuroimaging informatics technology initiative,” <https://nifti.nimh.nih.gov/>, 2004, accedido: 2025.
- [97] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, pp. 203–211, 2021.

- [98] U. Baid *et al.*, “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification,” in *arXiv preprint arXiv:2107.02314*, 2021. [Online]. Available: <https://arxiv.org/abs/2107.02314>
- [99] MONAI Consortium, “MONAI: Medical open network for AI,” <https://monai.io>, 2020.
- [100] F. P. et al., “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [101] P. Lambin *et al.*, “Radiomics: the bridge between medical imaging and personalized medicine,” *Nature Reviews Clinical Oncology*, vol. 14, pp. 749–762, 2017.
- [102] J. J. M. van Griethuysen *et al.*, “Computational radiomics system to decode the radiographic phenotype,” *Cancer Research*, vol. 77, no. 21, pp. e104–e107, 2017.
- [103] W. B. Pope *et al.*, “Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy,” *Journal of Neuro-Oncology*, vol. 107, no. 1, pp. 197–205, 2012.
- [104] A. Zwanenburg *et al.*, “The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping,” *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.
- [105] R. Parmar *et al.*, “Radiomic machine-learning classifiers for prognostic biomarkers of HPV oropharyngeal cancer,” *Frontiers in Oncology*, vol. 5, p. 272, 2015.
- [106] R. A. Fisher, “Statistical methods for research workers,” 1925.
- [107] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [108] H. Zhou *et al.*, “MRI features predict survival and molecular markers in diffuse lower-grade gliomas,” *Neuro-Oncology*, vol. 19, no. 6, pp. 862–870, 2017.
- [109] S. Bakas *et al.*, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Scientific Data*, vol. 4, p. 170117, 2017.
- [110] L. M. Wong, Q. Y. H. Ai, R. Zhang, F. Mo, and A. D. King, “Radiomics for discrimination between early-stage nasopharyngeal carcinoma and benign hyperplasia with stable feature selection on MRI,” *Cancers*, vol. 14, no. 14, p. 3433, 2022. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9324280/>
- [111] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

- [112] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: <https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>
- [113] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [114] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [115] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, “PyTorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 8026–8037. [Online]. Available: <https://papers.nips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>
- [116] M. Lin, Q. Chen, and S. Yan, “Network in network,” in *International Conference on Learning Representations*, 2014. [Online]. Available: <https://arxiv.org/abs/1312.4400>
- [117] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *International Conference on Learning Representations*, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [118] I. Loshchilov and F. Hutter, “SGDR: Stochastic gradient descent with warm restarts,” in *International Conference on Learning Representations*, 2017. [Online]. Available: <https://arxiv.org/abs/1608.03983>
- [119] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [120] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in *Advances in Neural Information Processing Systems*, vol. 27, 2014. [Online]. Available: <https://papers.nips.cc/paper/2014/hash/375c71349b295f5be2dcdca9206f20a06-Abstract.html>
- [121] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 448–456. [Online]. Available: <https://proceedings.mlr.press/v37/ioffe15.html>
- [122] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *Proceedings of the 30th International Conference on Machine Learning*, 2013, pp. 1310–1318. [Online]. Available: <https://proceedings.mlr.press/v28/pascanu13.html>

- [123] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The design of SimpleITK,” *Frontiers in Neuroinformatics*, vol. 7, p. 45, 2013.
- [124] M. Brett *et al.*, “NiBabel: Access a cacophony of neuro-imaging file formats,” <https://zenodo.org/record/3668070>, 2020.
- [125] Streamlit Inc., “Streamlit,” <https://streamlit.io>, 2019.
- [126] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [127] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, pp. 44–56, 2019.
- [128] Google Research, “Google colab,” <https://colab.research.google.com>, 2017.
- [129] Y. Taki, B. Thyreau, S. Kinomura, K. Sato, R. Goto, R. Kawashima, and H. Fukuda, “Correlations among brain gray matter volumes, age, gender, and hemisphere in healthy individuals,” *PLOS ONE*, vol. 6, no. 7, p. e22734, Jul 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144937/>
- [130] R. A. Zeineldin, M. E. Karar, O. Burgert, and F. Mathis-Ullrich, “How we won BraTS 2023 adult glioma challenge? Just faking it! Enhanced synthetic data augmentation and model ensemble for brain tumour segmentation,” *arXiv preprint arXiv:2402.17317*, 2024. [Online]. Available: <https://arxiv.org/abs/2402.17317>
- [131] D. A. Cieslak and N. V. Chawla, “Learning decision trees for unbalanced data,” in *Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*, 2008, pp. 241–256.