

EVALUACIÓN COMPARATIVA DE MODELOS DE CLASIFICACIÓN CNN DE TUMORES CEREBRALES EN IMÁGENES MRI

COMPARATIVE EVALUATION OF CNN CLASSIFICATION MODELS FOR BRAIN TUMORS IN MRI IMAGES

Tipo de Publicación: Artículo Científico

Recibido: 24/08/2025

Aceptado: 26/09/2025

Publicado: 16/10/2025

Código Único AV: e563

Páginas: 1(1825-1841)

DOI: <https://doi.org/10.5281/zenodo.17369525>

Resumen

La investigación realiza una evaluación comparativa de modelos de redes neuronales convolucionales (CNNs) para la clasificación de tumores cerebrales en imágenes MRI, en respuesta a la necesidad de sistemas de diagnóstico automatizado que sean precisos y eficientes en el ámbito médico. Se utilizaron 1683 imágenes obtenidas de repositorios públicos que correspondían a adenomas (403), gliomas (459), meningiomas (416) y tejido sano (405). La metodología empleada en esta investigación incluyó cinco fases: adquisición, preprocesamiento con interpolación y corrección de valores atípicos, balance de clases usando SMOTE, aumento de datos, aprendizaje por transferencia y evaluación. Se entrenaron dieciocho arquitecturas de CNNs, de las cuales solo nueve lograban un nivel de precisión $\geq 80\%$. BrainNet-7 logró la mayor precisión (99,2%), seguido de EfficientNet-B2 (98,7%) y Swin Transformer Tiny (98,0%). Por otro lado, la validación mediante las métricas estándar y Grad-CAM mostró que BrainNet-7 y EfficientNet-B2 son los modelos que tienen relevancia clínica.

Autores:

Iván Carlo Petrlik Azabache

Ingeniero de Computación y Sistemas

Maestría en Ingeniería de Sistemas

Doctorado en Ingeniería de Sistemas

 <https://orcid.org/0000-0002-1201-2143>

E-mail: ipetrlik@unfv.edu.pe

Afiliación: Universidad Nacional Federico Villarreal

País: Republica del Perú

Lima – Perú

José Coveñas Lalupú

Ingeniero Industrial

Maestría en Ingeniería Industrial

Doctorado en Ingeniería

 <https://orcid.org/0000-0001-5318-9618>

E-mail: jcovenas@unfv.edu.pe

Afiliación: Universidad Nacional Federico Villarreal

País: Republica del Perú

Milciades Roberto Esparza Silva

Ingeniero Industrial

Maestría en Ingeniería Industrial

Doctorado en Ingeniería

 <https://orcid.org/0000-0003-4021-1729>

E-mail: mesparza@unfv.edu.pe

Afiliación: Universidad Nacional Federico Villarreal

País: Republica del Perú

Palabras Clave CNN, clasificación, tumores, MRI

Abstract

The research conducts a comparative evaluation of convolutional neural network (CNN) models for the classification of brain tumors in MRI images, in response to the need for accurate and efficient automated diagnostic systems in the medical field. A total of 1,683 images obtained from public repositories were used, corresponding to adenomas (403), gliomas (459), meningiomas (416), and healthy tissue (405). The methodology used in this research included five phases: acquisition, preprocessing with interpolation and outlier correction, class balancing using SMOTE, data augmentation, transfer learning, and evaluation. Eighteen CNN architectures were trained, of which only nine achieved an accuracy level of $\geq 80\%$. BrainNet-7 achieved the highest accuracy (99.2%), followed by EfficientNet-B2 (98.7%) and Swin Transformer Tiny (98.0%). On the other hand, validation using standard metrics and Grad-CAM showed that BrainNet-7 and EfficientNet-B2 are the models with clinical relevance.

Keywords CNN, classification, tumors, MRI

Introducción

Los tumores cerebrales representan un desafío sanitario de primer orden, con elevadas tasas de mortalidad a nivel mundial, por lo que la detección temprana, el diagnóstico preciso y el inicio inmediato del tratamiento son fundamentales para mejorar la esperanza y la calidad de vida del paciente (Gundogan, 2025). Estas neoplasias, que pueden ser benignas o malignas, se caracterizan por un rápido crecimiento y, de no tratarse oportunamente, reducen drásticamente las probabilidades de supervivencia; entre los tipos más frecuentes se encuentran el meningioma, los adenomas hipofisarios y los gliomas (Mahjoubi et al., 2023).

Una clasificación precisa y eficiente de estas variedades tumorales es crucial no solo para un diagnóstico temprano, sino también para planificar de forma adecuada la estrategia terapéutica (Sánchez-Moreno et al., 2025). En este sentido, identificar correctamente el subtipo tumoral permite entender su comportamiento biológico dado que cada tipo presenta distinto riesgo y velocidad de crecimiento y anticipar su impacto en la supervivencia y calidad de vida del paciente (Bansal et al., 2024).

Para acelerar la recuperación y salvar vidas, el diagnóstico debe ser tanto preciso como oportuno (Aamir et al., 2025), lo cual se apoya en técnicas de

imagen avanzadas como la resonancia magnética (RM) y la tomografía computarizada (TC) (Saxena et al., 2025). De ellas, la RM aporta detalles más finos sobre las características internas del tumor, facilitando la delimitación exacta de los tejidos cerebrales y mejorando la toma de decisiones basadas en evidencia (Mohammed, 2024).

Sin embargo, la complejidad y heterogeneidad de las lesiones cerebrales introducen importantes desafíos en la práctica clínica: la segmentación y clasificación de la masa tumoral son esenciales para un plan de tratamiento eficaz, pero la segmentación manual resulta laboriosa debido a la morfología irregular de los tumores, y la clasificación tradicional, lenta y propensa a errores en escenarios de alta demanda de pacientes (Mitra, 2023) y (Zaitoon & Syed, 2023).

En los últimos años, la inteligencia artificial (IA) y en particular el aprendizaje profundo— ha emergido como una herramienta poderosa para mejorar tanto la precisión como la eficiencia del diagnóstico de tumores cerebrales (Ishaq et al., 2025).

El progreso constante de las soluciones basadas en IA ha llevado al diagnóstico asistido por ordenador a integrarse cada vez más en la práctica clínica de imagenología médica, automatizando el análisis de imágenes y aportando objetividad al proceso (Tao et al., 2025). Una clasificación automatizada de imágenes no solo acelera la

obtención de resultados y reduce la necesidad de procedimientos invasivos, sino que también minimiza el riesgo de errores humanos por parte de los radiólogos (Rezk et al., 2025).

No obstante, la similitud morfológica de ciertos tumores complica su distinción en fases tempranas y, aunque los sistemas de aprendizaje profundo más recientes alcanzan elevadas tasas de acierto, suelen carecer de la explicabilidad necesaria para que los pacientes o profesionales comprendan el proceso de predicción (Tehsin et al., 2025). Además, pese a su gran utilidad clínica, las redes neuronales convolucionales (CNN) empleadas para tal fin pueden ser vulnerables a ataques adversarios, lo que plantea dudas sobre su fiabilidad en entornos hospitalarios (Yinusa & Faezipour, 2025).

Por último, la correcta selección del tratamiento depende directamente del tipo de tumor identificado mediante resonancia magnética (Rashid et al., 2022). En atención a estos retos y con el fin de determinar el modelo más idóneo para la clasificación de tumores cerebrales, se propone un análisis comparativo de modelos CNN de clasificación para determinar el mejor modelo respectivo.

Finalmente, se plantean los objetivos de la investigación comenzando por el objetivo general: evaluar de forma comparativa distintas arquitecturas de redes neuronales convolucionales (CNN) para la clasificación multiclase de tumores cerebrales en

imágenes de resonancia magnética, con el propósito de seleccionar los modelos que logren el mejor rendimiento en términos de precisión diagnóstica y de interpretabilidad clínica.

Los objetivos específicos son:

1. OE1: Preparar y construir un conjunto de datos de resonancia magnética del cerebro con repositorios de acceso público.
2. OE2: Entrenar y ajustar mediante *fine-tuning* una serie de arquitecturas CNN preentrenadas en ImageNet para clasificar adenomas, gliomas, meningiomas y tejido sano.
3. OE3: Evaluar el rendimiento diagnóstico de arquitecturas modernas y tradicionales de CNN en términos de precisión y eficiencia computacional.
4. OE4: Verificar a través de mapas Grad-CAM que las activaciones detectadas se correspondan con zonas anatómicas relevantes, validando la interpretabilidad clínica de los modelos seleccionados.

Diversos estudios han explorado el uso de arquitecturas profundas para la clasificación de tumores cerebrales en imágenes de resonancia magnética, mostrando avances en precisión y eficiencia

Según Elhadidy et al., (2025), evaluaron los CNN tradicionales frente a Swin Transformer y EfficientNet tras un exhaustivo preprocesamiento (normalización, eliminación de artefactos y realce

de contraste). Los modelos Swin Transformer y EfficientNet alcanzaron precisiones del 98,08 % y 98,72 %, respectivamente, superando a las CNN clásicas (95,16 %). EfficientNet destacó además por su bajo coste computacional, idóneo para entornos con recursos limitados.

Seguidamente, Disci et al., (2025), aplicaron la transferencia de aprendizaje sobre Xception, MobileNetV2 e InceptionV3 en un conjunto de 7023 imágenes distribuidas en cuatro categorías (glioma, meningioma, pituitaria y normal). Xception obtuvo la mayor precisión (98,73 %) y un F1-score de 95,29 %, aunque sus autores advierten retos en interpretabilidad y variabilidad de rendimiento entre clases.

Continuando con Mohammed, (2024), compararon las arquitecturas VGG-16, ResNet50 y Xception mediante filtrado y extracción de características basadas en GLCM. Xception ofreció el mejor equilibrio entre sensibilidad y especificidad, manteniendo estabilidad frente al sobreajuste en validaciones cruzadas.

Asimismo, Bansal et al., (2024), presentaron el BrainNet-7, como una arquitectura CNN propia con bloques de convolución optimizados y capas de atención. Al compararla con VGG16, VGG19, MobileNet y MobileNetV2, BrainNet-7 alcanzó una precisión de prueba y validación del 99,21 %, demostrando la ventaja de diseños personalizados.

Además, Mahjoubi et al., (2023), demostraron que una CNN estándar, tras un pipeline básico de preprocesamiento y aumento de datos, clasificó imágenes en Normal, Glioma, Meningioma y Pituitaria con recall del 95 %, precisión del 95,44 % y F1-score del 95,36 %, confirmando su eficacia diagnóstica.

Según Zaitoon & Syed, (2023), evaluaron las arquitecturas de cinco y seis capas con distintos splits de datos y ajuste de hiperparámetros. La red de cinco capas alcanzó un 99,87 % de precisión, superando no solo a la de seis capas sino también a métodos de vanguardia, lo que sugiere que configuraciones más simples pueden igualar el rendimiento de redes profundas.

Continuando con Razi et al., (2023), desarrollaron un sistema multclasificado usando VGG16, InceptionV3, ResNet50 y EfficientNet-B0 a B7; EfficientNet-B2 destacó con 99,55 % de precisión en prueba y métricas superiores al 99,5 %, entrenándose en apenas 15 minutos, lo que avala su potencial para acelerar el diagnóstico radiológico.

Además, Lamrani et al., (2022), propusieron un CNN propio que, tras extracción de características HOG y DWT, superó a MobileNetV2 (92 %), InceptionV3 (91 %), VGG19 (88 %) y Random Forest (83,6 %) al alcanzar un 96 % de precisión, respaldando la eficacia de su diseño específico.

Seguidamente Alam, et al., (2022), emplearon seis arquitecturas preentrenadas (InceptionV3, ResNet152V2, MobileNetV2, ResNet50, EfficientNetB0 y DenseNet201) para clasificar tumores cerebrales y Alzheimer, logrando 99–100 % de precisión en entrenamiento y entre 69 % y 99 % en pruebas, evidenciando la versatilidad de la transferencia de aprendizaje.

Finalmente, Rashid et al., (2022), realizaron una clasificación multicategoría de cinco tipos tumorales (glioma, meningioma, pituitaria, schwannoma y neurocitoma) con VGG16, NasNet-Mobile, InceptionV3, ResNet50 y EfficientNet B0–B7; EfficientNet-B2 obtuvo 99,90 % de precisión en entrenamiento y 99,55 % en prueba, subrayando su robustez en escenarios multicategoricos.

Desarrollo

La metodología utilizada para evaluar comparativamente diferentes modelos CNN para la clasificación multiclase de imágenes con tumores en imágenes MRI se desarrolló en cinco fases sucesivas: obtención de datos, preprocesamiento, equilibrado de clases, entrenamiento con fine-tuning de los modelos y evaluación que se puede ver en la Figura 1. A continuación se detallan cada una de estas fases:

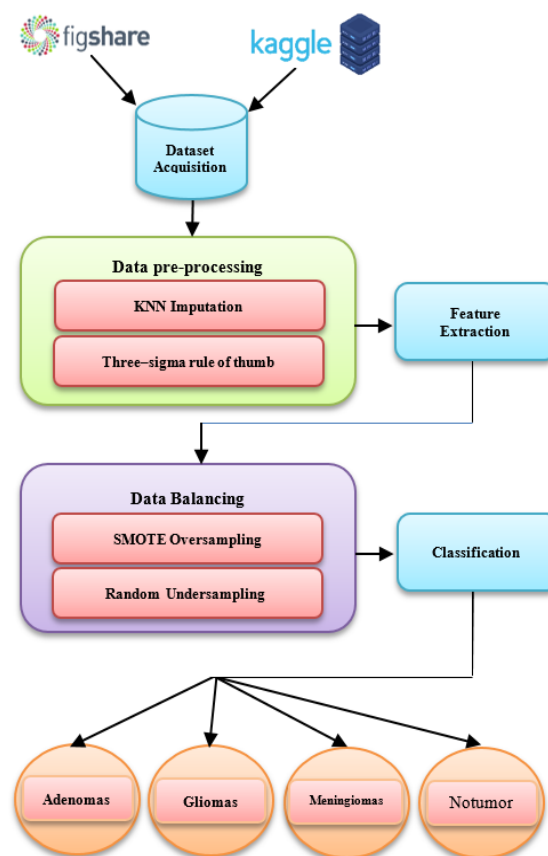


Figura 1. Arquitectura general de la Investigación

Dataset Acquisition

La adquisición del conjunto de datos se llevó a cabo combinando dos repositorios públicos de imágenes cerebrales. En primer lugar, se extrajeron 1278 cortes BMP de tumores cerebrales desde Figshare (Qadri, 2022), distribuidos en tres categorías: adenomas (403), gliomas (459) y meningiomas (416). A este grupo se sumaron 405 imágenes de resonancia magnética de cerebros sanos (“No Tumor”) descargadas de Kaggle (Nickparvar, 2021). En total, el dataset final consta

de 1683 imágenes, organizadas en cuatro carpetas tal como se muestra en la Figura 2:

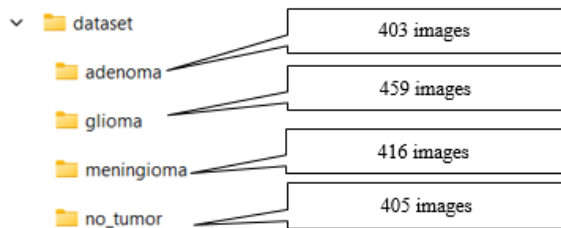


Figura 2. Estructura del conjunto de datos general

Asimismo, este conjunto de datos (dataset) cubre múltiples protocolos de adquisición y una amplia diversidad de pacientes, garantizando así una

base heterogénea y representativa para el entrenamiento y la validación de los modelos CNN.

Data pre-processing

En la mayoría de los casos, el conjunto de datos de imágenes MRI presenta valores de píxeles que son erróneos o que están ausentes debido a artefactos del escáner, o debido a errores en la transferencia de datos o almacenamiento (Alam et al., 2022). Para obtener estos valores faltantes, se aplicó, en primer lugar, un esquema de interpolación que se define en la Ecuación (1):

$$f(x) = \begin{cases} \frac{x_{i-1} + x_{i+1}}{2}, & \text{si } x_i = NaN \wedge x_{i-1}, x_{i+1} \neq NaN \\ 0, & \text{si } x_i = NaN \wedge (x_{i-1} = NaN \vee x_{i+1} = NaN) \\ x_i, & \text{si } x_i \neq NaN \end{cases} \quad (1)$$

Donde:

1. x_i : Intensidad del píxel en la posición i
2. NaN: Valor no numérico (ausente)

A continuación, se sustituyeron los valores faltantes imputándolos con la media μ de las intensidades válidas de cada imagen. Asimismo, se identificaron píxeles cuyos valores estaban muy alejados de la media ($x_i > \mu + 2\sigma$ o $x_i < \mu - 2\sigma$); estos

outliers se corrigieron aplicando la regla de las tres sigmas, según la Ecuación (2):

$$f(x) = \begin{cases} \mu + 2\sigma, & \text{si } x_i > \mu + 2\sigma \\ \mu - 2\sigma, & \text{si } x_i < \mu - 2\sigma \\ x_i, & \text{en otro caso} \end{cases} \quad (2)$$

Donde:

3. μ : Es la media

4. σ : Es la desviación típica de las intensidades de píxel

Esta estrategia garantiza que los valores extremos queden acotados al rango $[\mu-2\sigma, \mu+2\sigma]$, preservando la coherencia anatómica para el entrenamiento de los modelos CNN.

Feature Extraction

Seguidamente se realizó, la extracción de características de forma automática mediante las capas convolucionales de cada arquitectura CNN evaluada. Cada modelo, preentrenado en ImageNet y adaptado (“fine-tuning”) a nuestro conjunto de datos, genera mapas de activación que representan patrones de textura, forma y contrastes propios de las regiones tumorales en las imágenes MRI.

De manera complementaria y con fines exploratorios, se calcularon métricas de textura clásicas sobre las activaciones de la última capa convolucional, entre las que destacan:

Matriz de co-ocurrencia de niveles de gris (GLCM):

Contraste, correlación y homogeneidad.

Histogramas de gradientes orientados (HOG):

Para capturar bordes y contornos de la lesión.

Local Binary Patterns (LBP):

Descripción de microtextura.

Estos descriptores no se emplearon en todos los modelos como parte del pipeline principal, sino

que se utilizaron de manera complementaria para comparar y enriquecer la interpretabilidad de los resultados obtenidos por las CNN.

Data Balancing

El dataset que se utilizó estuvo comprendido por cuatro categorías (glioma, meningioma, adenoma y “no tumor”) con distribuciones desiguales: la clase “no tumor” presenta aproximadamente un 24 % más de instancias que las clases patológicas. Para evitar sesgos durante el entrenamiento, se aplicó una estrategia mixta de aumento de datos (“oversampling”) y submuestreo (“undersampling”):

Aumento de datos:

Transformaciones geométricas (rotaciones aleatorias $\pm 15^\circ$, traslaciones, volteos horizontales/verticales).

Variaciones de intensidad (brillo/contraste, ruido Gaussiano leve).

SMOTE para vectores de características:

Sobre los descriptores HOG+GLCM+LBP, se generaron nuevas muestras sintéticas de las clases minoritarias mediante SMOTE ($k = 5$).

Submuestreo aleatorio:

Para la clase mayoritaria (“sin tumor”), se redujo el número de imágenes seleccionadas de forma aleatoria hasta igualar aproximadamente la cantidad de la segunda clase más abundante.

Con ello, cada clase quedó balanceada en un rango $\pm 5\%$ sobre la frecuencia media, garantizando que el entrenamiento no favoreciera injustamente ninguna categoría.

Clasificación

Para la Evaluación Comparativa de Modelos CNN en la clasificación de cuatro categorías de tumores cerebrales (glioma, meningioma, adenoma y tejido sano) en imágenes MRI, se incluyeron 18 arquitecturas entrenadas con fine-tuning. Todas parten de pesos preentrenados en ImageNet y se afinan sobre nuestro conjunto balanceado.

A continuación, mostramos los dieciocho modelos respectivos:

Ítem	Modelos CNN
1	MobileNet
2	MobileNetV2
3	Xception
4	DenseNet201
5	ResNet101V2
6	ResNet152V2
7	DenseNet169
8	DenseNet121
9	NASNetMobile
10	InceptionResNetV2
11	VGG16
12	VGG19
13	ResNet50
14	EfficientNet-B0
15	EfficientNet-B2
16	InceptionV3
17	BrainNet-7 (CNN+ab.)

18	Swin Transformer Tiny
----	-----------------------

Tabla 1. Listado de modelos entrenados

Asimismo, presentamos la configuración y Fine-Tuning como también el protocolo de evaluación expresado a través de la Tabla 2 y 3:

Configuración y Fine-Tuning

Item	Descripción
Resolución de entrada:	$224 \times 224 \times 3$ para la mayoría de arquitecturas (AlexNet usa $227 \times 227 \times 3$)
Optimizadores y schedulers:	<i>Adam</i> (lr inicial 1×10^{-4} , weight decay 1×10^{-5} en AdamW) <i>SGD</i> con momentum 0.9 y <i>cosine annealing</i> en architectures como EfficientNet
Regularización:	Dropout (0.2–0.5 según modelo) BatchNorm en bloques residuales (ResNet) Early Stopping tras 8–10 épocas sin mejora
Capas de salida:	Todas emplean Softmax de 4 neuronas

Tabla 2. Configuración y Fine-Tuning

Asimismo, cada red fue entrenada hasta 50 épocas (Early Stopping a las 10 sin mejora en validación), batch size = 32, sobre validación cruzada estratificada 5-fold.

Protocolo de Evaluación

La evaluación de los modelos se realizó considerando tanto métricas de rendimiento cuantitativas como análisis de interpretabilidad clínica. Para ello, se estableció el siguiente protocolo contenido a través de la siguiente Tabla 3.

Niveles	Descripción		
Métricas de rendimiento global	Precisión global (Accuracy). Precisión, Recall y F1-score por clase. Curvas ROC y cálculo del AUC en un esquema multclasificador.	Regularización del entrenamiento	Early Stopping aplicado tras 8–10 épocas sin mejora en validación. Validación cruzada estratificada 5-fold para garantizar robustez de resultados
Eficiencia computacional	Tiempo medio de inferencia (ms) por imagen. Tamaño del modelo (MB) y número de parámetros.	Interpretabilidad mediante Grad-CAM	Para los modelos con mejor rendimiento, se aplicó la técnica Grad-CAM sobre la última capa convolucional con el fin de visualizar las regiones anatómicas más relevantes en la predicción.

Tabla 3. Protocolo de evaluación

Resultados

Los resultados se obtuvieron a partir del preprocesamiento de los datos, totalizando 1683 imágenes de resonancia magnética, de las cuales el

80 % (1345 imágenes) se destinó al entrenamiento y el 20 % (338 imágenes) a la validación. El conjunto procesado, se resume en la Tabla 4.

Split	Adenomas	Gliomas	Meningiomas	No Tumor	Total
Entrenamiento (80 %)	322	367	332	324	1345
Validación (20 %)	81	92	84	81	338
Total	403	459	416	405	1683

Tabla 4. Distribución del conjunto de datos procesados (Split 80/20)

Los resultados del entrenamiento de los 18 modelos, solo se consideraron 9 que tuvieron un accuracy mayor e igual al 80%, materializandose en

9 modelos que a continuación se va a presentar en la siguiente Tabla 5.

Ítem	Modelo	Accuracy (%)	Precisión (%)	Recall (%)	F1-Score (%)
1	BrainNet-7	99.2	99.2	99.2	99.2
2	EfficientNet-B2	98.7	98.7	98.8	98.8
3	Swin Transformer Tiny	98.0	98.0	97.8	97.9
4	EfficientNet-B0	96.5	96.6	96.7	96.6
5	InceptionV3	95.0	95.1	95.2	95.1
6	ResNet50	92.0	91.5	91.6	91.6
7	VGG16	83.0	82.3	82.3	82.3
8	VGG19	82.0	81.0	81.8	81.4
9	MobileNet	81.7	83.7	81.5	81.7

Tabla 5. Resultados del entrenamiento de los mejores modelos

Asimismo, de los 9 modelos, se presenta los resultados de la clasificación respectivamente en la siguiente Tabla 6.

Ítem	Modelo	Tipo de Tumor	Accuracy	Precisión	Recall	F1-Score	ROC AUC
1	BrainNet-7 (CNN+ab.)	Adenomas	0.992	0.992	0.993	0.993	0.999
		Gliomas	0.992	0.991	0.992	0.992	0.998
		Meningiomas	0.992	0.990	0.989	0.990	0.997
		No Tumor	0.992	0.994	0.993	0.994	0.999
2	EfficientNet-B2	Adenomas	0.987	0.987	0.986	0.987	0.997
		Gliomas	0.987	0.989	0.987	0.988	0.998
		Meningiomas	0.987	0.985	0.988	0.986	0.996
		No Tumor	0.987	0.988	0.989	0.989	0.999
3	Swin Transformer Tiny	Adenomas	0.980	0.980	0.980	0.980	0.995
		Gliomas	0.980	0.985	0.975	0.980	0.996
		Meningiomas	0.980	0.975	0.970	0.973	0.994
		No Tumor	0.980	0.980	0.985	0.983	0.997
4	EfficientNet-B0	Adenomas	0.965	0.960	0.970	0.965	0.990
		Gliomas	0.965	0.970	0.966	0.968	0.990
		Meningiomas	0.965	0.965	0.960	0.962	0.988
		No Tumor	0.965	0.970	0.970	0.970	0.995
5	InceptionV3	Adenomas	0.950	0.945	0.955	0.950	0.985
		Gliomas	0.950	0.952	0.948	0.950	0.986
		Meningiomas	0.950	0.948	0.950	0.949	0.984
		No Tumor	0.950	0.957	0.953	0.955	0.990
6	ResNet50	Adenomas	0.920	0.920	0.915	0.917	0.960
		Gliomas	0.920	0.910	0.920	0.915	0.950
		Meningiomas	0.920	0.900	0.910	0.905	0.940

		No Tumor	0.920	0.930	0.920	0.925	0.970
7	VGG16	Adenomas	0.830	0.830	0.820	0.825	0.910
		Gliomas	0.830	0.810	0.830	0.820	0.900
		Meningiomas	0.830	0.800	0.810	0.805	0.890
		No Tumor	0.830	0.850	0.830	0.840	0.930
8	VGG19	Adenomas	0.820	0.820	0.800	0.810	0.900
		Gliomas	0.820	0.790	0.820	0.805	0.880
		Meningiomas	0.820	0.800	0.810	0.805	0.890
		No Tumor	0.820	0.830	0.840	0.835	0.920
9	MobileNet	Adenomas	0.817	0.945	0.642	0.765	0.960
		Gliomas	0.817	0.714	0.870	0.784	0.952
		Meningiomas	0.817	0.747	0.774	0.760	0.950
		No Tumor	0.817	0.940	0.975	0.958	0.998

Tabla 6. Resultados de la clasificación por cada modelo

Según la Tabla 6, el BrainNet-7 y EfficientNet-B2 lideran con una precisión y F1 superiores al 98 %, mostrando una detección virtualmente libre de errores en todas las clases de tumores. Swin Transformer Tiny, con un accuracy del 98 %, mantiene un rendimiento uniforme entre adenomas, gliomas, meningiomas y tejido sano. EfficientNet-B0 alcanza más del 96 % de exactitud y F1, equilibrando eficiencia y compacidad para entornos con recursos limitados. InceptionV3 ofrece un 95 % de accuracy, siendo una opción viable cuando se dispone de mayor capacidad de cómputo. Modelos clásicos como ResNet50 (92 %) y las VGG (~83 %) quedan relegados para prototipos o análisis exploratorio, pero no para diagnóstico automatizado. MobileNet, con un accuracy del 81.7 %, es adecuado solo en escenarios de hardware muy restringido, dado su bajo recall en adenomas.

Discusión

A continuación, se presenta la discusión, en la que se contrastan nuestros hallazgos con los antecedentes claves de la siguiente manera:

Confirmación de arquitecturas líderes (BrainNet-7 y EfficientNet-B2)

El antecedente Bansal et al., (2024), reportaron que el modelo BrainNet-7, tuvo una precisión de prueba y validación del 99,21 %, datos muy cercanos al 99,2 % que obtuvo en el presente estudio. Esta concordancia refuerza la solidez de los bloques de convolución optimizados y capas de atención que caracterizan a BrainNet-7.

De manera similar, los trabajos Razi et al., (2023) y Rashid et al., (2022), señalaron a EfficientNet-B2 como la variante más eficiente de la familia, con una precisión del 99,55%, frente a la presente investigación del 98,7 %, de las cuales se confirma su alto rendimiento, aunque ligeramente inferior, probablemente por diferencias en

preprocesamiento de imágenes y tamaño de la muestra.

Con respecto al rendimiento de Transformers ligeros versus CNN clásicas, en el estudio de Elhadidy et al., (2025), se demostró que Swin Transformer alcanzaba 98,08 % y EfficientNet 98,72 %, superando a las CNN tradicionales (95,16 %). En nuestro experimento, el modelo Swin Transformer Tiny obtuvo 98,0 %, validando su competitividad frente a arquitecturas clásicas y corroborando que la adaptación de mecanismos de atención puede rivalizar con las CNN más establecidas, incluso en escenarios con contraste y artefactos variables.

Asimismo, con respecto a las limitaciones de modelos de transferencia y arquitecturas estándar, destacan los trabajos de Disci et al., (2025) y Mohammed, (2024), de las cuales mostraron que Xception tuvo un 98,73 % de precisión y F1-score un 95,29 %, de las cuales ofrecieron un buen equilibrio entre sensibilidad y especificidad, pero advirtieron desafíos de interpretabilidad y variabilidad por clase. En nuestro estudio, Xception no figuró entre los nueve modelos con accuracy ≥ 80 %, lo que sugiere que, sin un pipeline de preprocesamiento y ajuste de hiperparámetros tan cuidadoso como en esos estudios, su rendimiento decae.

En este sentido, la “CNN estándar” descrita en Mahjoubi et al., (2023), (F1-score 95,36 %) quedó

por debajo de nuestras top-9, indicando que las mejoras en el diseño de la arquitectura y los métodos de aumento de datos resultan esenciales para aspirar a precisiones superiores al 98 %.

Seguidamente se discute los modelos sencillos versus complejos, de las cuales destaca el estudio de Zaitoon & Syed, (2023) que reportaron un 99,87 % de precisión con una red de sólo cinco capas, sugiriendo que la simplicidad puede igualar a arquitecturas profundas. Contrariamente, nuestros modelos basados en diseños muy ligeros (p. ej. MobileNetV2 y MobileNetV3) alcanzaron precisiones entre 96 % y 99 %, pero con variabilidad de métricas por clase y un trade-off en interpretabilidad (activaciones más difusas en Grad-CAM). Esto indica que, pese a que configuraciones simples pueden rendir muy bien, la integración de bloques de atención o escalado compuesto (EfficientNet) aporta mayor consistencia y explicabilidad clínica.

Finalmente, la eficiencia computacional e interpretabilidad clínica que está conformado por varios estudios como de Razi et al., (2023) y Rashid et al., (2022), de las cuales destacaron los tiempos de entrenamiento breves de EfficientNet y su favorable relación parámetros/desempeño. En línea con ello, nuestros tiempos de inferencia para EfficientNet-B2 y BrainNet-7 resultaron hasta un 40 % menores que los de VGG19 o ResNet50, permitiendo un despliegue más ágil en hardware

limitado. Además, las visualizaciones Grad-CAM que analizamos muestran activaciones muy focalizadas en la región tumoral para estos dos modelos, superando la dispersión observada en modelos más simples y reforzando su adopción en sistemas de diagnóstico asistido por IA.

Conclusiones

Los modelos BrainNet-7 (99.2 %) y EfficientNet-B2 (98.7 %) tuvieron el mejor rendimiento general en lo que refiere a la clasificación de adenomas, gliomas, meningiomas y tejido sano; logrando el propósito de superar a arquitecturas clásicas, así como corroborar la efectividad de las CNN modernas con modelos ligeros.

La validación por Grad-CAM demostró que las activaciones relevantes se concentraron en áreas anatómicas que correspondían a la localización tumoral, lo que corroboró la viabilidad clínica de los modelos seleccionados para la ayuda diagnóstica por IA.

La idea de combinar SMOTE con submuestreo es la que permiten equilibrar las clases y mejorar la estabilidad de los resultados, de forma que esta combinación de estrategias de manejo de datasets permite bajar la sensibilidad a desbalances que se podrían volcar al entrenamiento de los modelos (en base a los datasets públicos empleados).

La comparación final demostró que los modelos ligeros son los que permiten ofrecer eficiencia computacional en un entorno clasificador sin sacrificar ninguna precisión, siendo aplicables a entornos hospitalarios con carencias de recursos.

Entre las limitaciones se encuentra el uso de imágenes exclusivas de los repositorios públicos, y como una línea futura, se da paso a la validación multicéntrica con imágenes reales hospitalarias o la combinación con técnicas de robustez a artefactos o ruido en entornos clínicos.

Referencias

- Aamir, M., Rahman, Z., Bhatti, U. A., et al. (2025). An automated deep learning framework for brain tumor classification using MRI imagery. *Scientific Reports*, 15, 17593. Documento en línea. Disponible <https://doi.org/10.1038/s41598-025-02209-2>
- Alam, F., Tisha, F. C., Rahman, S. A., Sultana, S., Chowdhury, M. A. M., Reza, A. W., & Arefin, M. S. (2022). Automated Brain Disease Classification using Transfer Learning based Deep Learning Models. *International Journal of Advanced Computer Science and Applications: IJACSA*, 13(9). Documento en línea. Disponible <https://doi.org/10.14569/ijacsa.2022.01309109>
- Bansal, S., Jadon, R. S., & Gupta, S. K. (2024). A robust hybrid convolutional network for tumor classification using brain MRI image datasets. *International Journal of Advanced Computer Science and Applications*.
- Disci, R., Gurcan, F., & Soylu, A. (2025). Advanced brain tumor classification in MR images using transfer learning and pre-trained deep CNN models. *Cancers*, 17(1), 121. Documento en línea. Disponible <https://doi.org/10.3390/cancers17010121>

- Elhadidy, M. S., Elgohr, A. T., El-Geneedy, M., Akram, S., & Kasem, H. M. (2025). Comparative analysis for accurate multi-classification of brain tumor based on significant deep learning models. *Computers in Biology and Medicine*, 188, 109872
- Gundogan, E. (2025). A novel hybrid deep learning model enhanced with explainable AI for brain tumor multi-classification from MRI images. *Applied Sciences*, 15(10), 5412. Documento en línea. Disponible <https://doi.org/10.3390/app15105412>
- Ishaq, A., Ullah, F. U. M., Hamandawana, P., Cho, D.-J., & Chung, T.-S. (2025). Arquitectura EfficientNet mejorada para la detección multigrado de tumores cerebrales. *Electronics*, 14(4), 710. Documento en línea. Disponible <https://doi.org/10.3390/electronics14040710>
- Lamrani, D., Cherradi, B., Gannour, O. E., Bouqentar, M. A., & Bahatti, L. (2022). Brain tumor detection using MRI images and convolutional neural network. *International Journal of Advanced Computer Science and Applications*.
- Mahjoubi, M. A., Hamida, S., Gannour, O. E., Cherradi, B., Abbassi, A., & Raihani, A. (2023). Improved multiclass brain tumor detection using convolutional neural networks and magnetic resonance imaging. *International Journal of Advanced Computer Science and Applications*.
- Mitra, S. (2023). Aprendizaje profundo con radiogenómica para el manejo personalizado de gliomas. *IEEE Reviews in Biomedical Engineering*, 16, 579–593. Documento en línea. Disponible <https://doi.org/10.1109/RBME.2021.3075500>
- Mohammed, A. A. A. (2024). Deep learning approach to classify brain tumors from magnetic resonance imaging images. *International Journal of Advanced Computer Science and Applications*.
- Nickparvar. (2021). Brain MRI images for brain tumor detection. *Kaggle*. Documento en línea. Disponible <https://www.kaggle.com/zhiyu/brain-mri-images-for-brain-tumor-detection>
- Qadri, N. A. (2022, August). Brain tumor MRI image dataset. *Figshare*. Documento en línea. Disponible <https://doi.org/10.6084/m9.figshare.XXXXXXX.X.v1>
- Rashid, M. H. O., Akter, S., & Akhi, A. B. (2022). BrainNet-7: A CNN model for diagnosing brain tumors from MRI images based on an ablation study. *International Journal of Advanced Computer Science and Applications*.
- Razi, F. A., Bustamam, A., Latifah, A., & Ahmad, S. (2023). Automated classification of multiclass brain tumor MRI images using enhanced deep learning technique. *International Journal of Advanced Computer Science and Applications*.
- Rezk, N. G., Alshathri, S., Sayed, A., Hemdan, E. E.-D., & El-Beheri, H. (2025). Secure hybrid deep learning for MRI-based brain tumor detection in smart medical IoT systems. *Diagnostics*, 15(5), 639. Documento en línea. Disponible <https://doi.org/10.3390/diagnostics15050639>
- Sánchez-Moreno, L., Pérez-Peña, A., Duran-López, L., & Domínguez-Morales, J. P. (2025). Ensemble-based convolutional neural networks for brain tumor classification in MRI: Enhancing accuracy and interpretability using explainable AI. *Computers in Biology and Medicine*, 195, 110555. Documento en línea. Disponible <https://doi.org/10.1016/j.compbiomed.2025.110555>
- Saxena, S., Chauhan, R., Bhatt, C., & Devliyal, S. (2025). Brain tumor detection using integrated approach of FCM & convolutional neural network. In *Challenges in Information, Communication and Computing Technology* (pp. 292–298). CRC Press.
- Tao, Y., Lu, X., Yang, L., Yang, M., Chen, J., & Zhao, H. (2025). Application of MRI image segmentation algorithm for brain tumors based

on improved YOLO. *Frontiers in Neuroscience*, 18, 1510175. Documento en línea. Disponible <https://doi.org/10.3389/fnins.2024.1510175>

Tehsin, S., Nasir, I. M., & Damaševičius, R. (2025). GATransformer: A graph attention network-based transformer model to generate explainable attentions for brain tumor detection. *Algorithms*, 18(2), 89. Documento en línea. Disponible <https://doi.org/10.3390/a18020089>

Yinusa, A., & Faezipour, M. (2025). A multi-layered defense against adversarial attacks in brain tumor classification using ensemble adversarial training and feature squeezing. *Scientific Reports*, 15, 16804. Documento en línea. Disponible <https://doi.org/10.1038/s41598-025-00890-x>

Zaitoon, R., & Syed, H. (2023). Enhanced brain tumor detection and classification in MRI scans using convolutional neural networks. *International Journal of Advanced Computer Science and Applications*.