Universidad CENFOTEC

Maestría en Tecnologías de Bases de Datos

Documento final de Proyecto de Investigación Aplicada 2

Tema:

Desarrollo de una solución en la nube basada en arquitectura Big Data y redes neuronales para la clasificación de tumores cerebrales en imágenes de resonancia magnética (MRI)

Garita Matamoros, Harold Haydn

Abril, 2026

Declaratoria de derecho de autor

Yo, Harold Garita Matamoros, declaro que soy el autor original del presente documento titulado “Documento final de Proyecto de Investigación Aplicada 2”, elaborado como parte del Proyecto de Investigación Aplicada 1 y 2 de la Maestría en Tecnologías de Bases de Datos. Afirmo que el contenido de este trabajo es producto de mi propio esfuerzo intelectual, análisis académico y desarrollo técnico, y que todas las fuentes externas utilizadas han sido citadas y referenciadas conforme a las normas APA, respetando los derechos de propiedad intelectual correspondientes.

Asimismo, manifiesto que este trabajo es inédito, no ha sido presentado anteriormente para la obtención de otro grado académico ni ha sido sometido a evaluación en ninguna otra institución. Autorizo a la Universidad a utilizar este documento para fines académicos, investigativos y de preservación digital, manteniendo siempre mi reconocimiento como autor.

Dedicatoria

Quiero dedicar esta tesis, en primer lugar, a mi abuela Grace. Ella fue médica y siempre despertó en mí un gran interés por la medicina. Además, la quiero muchísimo, y por eso este trabajo tiene un significado tan especial para mí. Decidí enfocar mi tesis en este tema como una forma de honrar su memoria, especialmente después de todo lo que significó para mí y por su lucha contra un tumor cerebral. Espero que este trabajo pueda aportar, aunque sea un poco, a ayudar a otras personas que enfrentan esta enfermedad.

También quiero dedicársela a mi novia Andrea, por haberme acompañado y apoyado durante todo este proceso. No solo estuvo conmigo durante el desarrollo de la tesis, sino también a lo largo de toda la maestría. Siempre me motivó, creyó en mí y estuvo a mi lado en los momentos difíciles. Incluso me prestó su computadora para hacer los trabajos de la universidad cuando yo no tenía una, y ese gesto significa muchísimo para mí.

A mis papás, por apoyarme siempre en mis estudios, por darme la mejor educación posible y por estar conmigo en cada una de mis ideas y proyectos. Les agradezco profundamente su amor, su confianza y todo lo que han hecho por mí. Gracias a ustedes he podido crecer y formarme como profesional.

Al profesor Jorge Arturo Garnier, por su apoyo y guía durante el desarrollo de esta tesis, y por haberme inspirado a llevar a cabo este proyecto de machine learning. Es un excelente profesor, muy apasionado por lo que hace, y eso fue una gran motivación para mí.

Y, finalmente, al profesor Gabriel Alfaro, por haber aceptado ser el lector de mi tesis. Gracias por su apoyo, su tiempo y su disposición para guiarme durante este proceso. Su acompañamiento fue muy valioso para poder llevar este trabajo a buen término.

Hoja de aprobación del proyecto

Tabla de contenido

Lista de ilustraciones

Lista de tablas

Resumen ejecutivo

El diagnóstico de tumores cerebrales mediante imágenes de resonancia magnética (MRI) representa un desafío crítico para los sistemas de salud pública, particularmente en la Caja Costarricense de Seguro Social (CCSS), donde se procesan cerca de 2.970 estudios mensuales con recursos especializados limitados y sin sistemas automatizados de apoyo. Esta situación genera tiempos de espera prolongados y variabilidad interobservador en la interpretación clínica. Ante esta problemática, el presente estudio tuvo como objetivo desarrollar un sistema de diagnóstico asistido por computadora basado en redes neuronales profundas y una arquitectura Big Data en la nube para la clasificación de tumores cerebrales en imágenes MRI.

La solución fue implementada en Microsoft Azure bajo una arquitectura medallón (Bronze–Silver–Gold), utilizando un dataset público de 4.478 imágenes distribuidas en 44 clases, con una partición estratificada de 60% entrenamiento, 20% validación y 20% prueba. Se evaluaron modelos EfficientNet V2 B0, Vision Transformer ViT-B16 y tres métodos de ensamble. El mejor desempeño se obtuvo mediante un ensamble ponderado (0.6/0.4), alcanzando una exactitud de 93.30%, F1-score de 0.9327 y MCC de 0.9304.

Los resultados evidencian la viabilidad técnica y el valor estratégico de integrar inteligencia artificial en el sistema de salud costarricense, contribuyendo a reducir la variabilidad diagnóstica y optimizar el tiempo de análisis clínico.

Capítulo 1. Introducción

1.1 Generalidades

Los diagnósticos de cáncer tienen un rol crucial en la oncología neurológica. Para los médicos, clasificar tumores cerebrales con precisión es indispensable para poder hacer un pronóstico acertado y escoger el tratamiento adecuado. Existen tipos de cáncer en el cerebro que son muy difíciles de clasificar, incluso para los neurólogos más experimentados. Un error en la clasificación de un tumor cerebral puede ser fatal para el paciente. Por ejemplo, puede resultar en un sobre o infratratamiento, toxicidad innecesaria, retrasos en terapias potencialmente efectivas o exclusión de ensayos clínicos. En resumen, puede haber un impacto directo en la supervivencia y la calidad de vida del paciente.

Existen varios estudios, como el de Bruner et al. (1997), que evidencian que los médicos interpretan las imágenes médicas de forma subjetiva. En este estudio, los autores revisaron las primeras 500 biopsias de cerebro o columna vertebral que fueron enviadas a su servicio de consulta de neuropatología para una segunda opinión. Ellos encontraron una tasa alta de discrepancias en los diagnósticos, incluyendo errores considerados como "clínicamente importantes" que podrían cambiar por completo las decisiones terapéuticas. Según Bruner et al. (1997), "hubo algún grado de desacuerdo entre los diagnósticos originales y de revisión en 214 (42,8%) de los 500 casos" (sección **"**Abstract", párr. 2**,** *traducción propia*).

El avance de la inteligencia artificial (IA) en el área de la salud oncológica ha revolucionado el diagnóstico y pronóstico médico mediante herramientas de imágenes, histopatológicas y genómicas, las cuales han permitido una mayor eficiencia en la detección, categorización, predicción de resultados y planificación del tratamiento. La IA se está empezando a utilizar en todas las fases del manejo de tumores cerebrales malignos: el diagnóstico, pronóstico (también conocido como prognosis) y la terapia. Según Khalighi et al. (2024), "Los modelos de IA superan las evaluaciones humanas en términos de exactitud y especificidad. Su capacidad para discernir aspectos moleculares a partir de imágenes puede reducir la dependencia de diagnósticos invasivos y acelerar el tiempo hasta los diagnósticos moleculares" (sección "Abstract", párr. 2, *traducción propia*).

En este trabajo de investigación, nos enfocaremos en cómo los modelos de IA pueden mejorar el proceso de clasificar tumores cerebrales en imágenes de resonancia magnética (MRI). La resonancia magnética es una de las técnicas de imagen más utilizadas para detectar y categorizar tumores cerebrales debido a su alta resolución de contraste tisular y a la ausencia de

radiación ionizante. No obstante, la lectura manual sigue siendo laboriosa, susceptible a variabilidad interobservadores y limitada para explotar patrones sutiles que no son evidentes a simple vista.

La combinación de arquitecturas Big Data en la nube con redes neuronales profundas permite construir flujos reproducibles de preprocesamiento, normalización y segmentación, extraer biomarcadores radiológicos a gran escala e inferir resultados con mayor rapidez y consistencia. Adicionalmente, aporta trazabilidad (versionado de datos y modelos), monitorización del desempeño y mejores condiciones para la gobernanza y auditoría del proceso diagnóstico.

1.2 Antecedentes del problema

En la actualidad, los sistemas de diagnóstico asistido para imágenes médicas han mostrado avances, pero también presentan limitaciones en cuanto a escalabilidad y estandarización. Muchos hospitales y centros médicos en Costa Rica carecen de la infraestructura tecnológica necesaria para manejar grandes volúmenes de imágenes MRI. Por otra parte, la variabilidad en los equipos de adquisición de las imágenes MRI y protocolos genera inconsistencias que afectan el desempeño de los modelos.

Kilim, O. et. al, en el artículo titulado *“Physical imaging parameter variation drives domain shift*” describe la siguiente situación respecto a la adquisición de las imágenes MRI:

> A diferencia de otros factores de generación de conjuntos de datos, los parámetros de imágenes físicas (PIP), a saber, la configuración mecánica, la calibración, el proveedor y el protocolo de adquisición dentro y entre los sitios de recopilación de datos son variables y se pueden medir directamente, además de registrarse de forma rutinaria, pero rara vez se utilizan más allá de las calibraciones (sección “Introduction”, párr. 2, traducción propia).

Proyectos previos han demostrado que la nube puede mitigar estos problemas, pero aún existe la necesidad de integrar soluciones robustas que combinen pipelines de Big Data con algoritmos de redes neuronales para segmentación y clasificación de tumores de manera precisa.

En Costa Rica, los centros de salud cada vez están más presionados para realizar estudios de resonancia magnética. Vega, Y., en el artículo titulado “*Centro Nacional de Imágenes Médicas*

realiza cada mes 2 970 resonancias magnéticas" publicado el 21 de agosto del 2025 en el sitio web de la Caja Costarricense de Seguro Social (CCSS) dice lo siguiente:

> El Centro Nacional de Imágenes Médicas, creado hace 15 años, realiza cada mes 2970 resonancias magnéticas en pacientes de todas las edades, convirtiéndose en la única unidad del país encargada de estos estudios especializados. Las resonancias magnéticas tienen complejidades distintas de acuerdo con el órgano que esté en estudio; por eso, el tiempo dentro de la máquina puede variar desde los 25 minutos hasta las 2 horas (Vega, 2025, párrafo 1).

Sin embargo, el volumen de estudios no es el único problema. Según un informe de la Auditoría Interna de la CCSS, los pacientes con cáncer esperan en promedio 115 días para realizarse una resonancia magnética, con casos extremos de hasta 648 días. Los pacientes de consulta externa, que representan el 60% de los casos, registran esperas promedio de 189 días. A diciembre de 2025, se acumulaban 11.570 estudios pendientes de interpretación, con retrasos de hasta 570 días en imágenes de cerebro y cuello (Segura, 2026).

Para poder tener un punto de comparación, en el 2017 la misma entidad realizaba solamente 66 estudios diarios. Esto refleja la creciente demanda de diagnósticos especializados y la necesidad de optimizar los procesos. Este aumento en la cantidad de estudios genera un reto importante: los radiólogos deben analizar un número cada vez mayor de imágenes en un tiempo limitado, lo que aumenta el riesgo de retrasos y errores humanos en la interpretación.

Aquí es donde la inteligencia artificial aplicada al análisis de imágenes médicas tiene un valor estratégico. Diversos estudios han demostrado que los modelos de *deep learning* pueden alcanzar niveles de exactitud comparables, e incluso hasta superiores, a los de radiólogos experimentados, especialmente en tareas repetitivas como la detección de anomalías y segmentación de tumores. Al estandarizar el proceso de análisis, la IA no solo agiliza la interpretación de grandes volúmenes de imágenes, sino que también minimiza la variabilidad asociada al criterio individual de cada especialista. Esto se traduce en mayor precisión diagnóstica, menor margen de error humano y diagnósticos más consistentes y reproducibles.

Por ejemplo, en el estudio costarricense titulado "*Aplicación de la inteligencia artificial en la salud pública para el diagnóstico temprano y tamizaje de enfermedades oncológicas*", los autores Aguilar, C. et. al explican lo siguiente:

> Por medio de la IA se han logrado nuevas técnicas, a través de ingeniería informática e imágenes a través de análisis estadísticos que permiten una mejora en el reconocimiento y precisión en la identificación de patologías oncológicas para generar diagnósticos más eficaces. Esto ayuda a tener un mejor pronóstico sobre la patología debido a que por medio de diferentes estudios se logró determinar que la precisión es mayor que con los métodos utilizados de forma tradicional (sección “Aplicación en el diagnóstico oncológico”, párr. 1).

En el sector privado costarricense, hospitales como la Clínica Bíblica ya cuentan con equipos avanzados de diagnóstico capaces de generar imágenes de alta resolución. Sin embargo, en el sector público todavía existen limitaciones significativas en infraestructura y en capacidad de procesamiento. La incorporación de soluciones basadas en IA en este contexto no solo permitiría aliviar la carga de trabajo de los especialistas, sino que también podría contribuir a democratizar el acceso a diagnósticos de alta calidad, reduciendo la brecha entre pacientes atendidos en hospitales privados y aquellos que dependen de la CCSS.

1.3 Definición y descripción del problema

El problema central es la dificultad de analizar de manera rápida y precisa un volumen creciente de imágenes de resonancia magnética cerebral para detectar y clasificar tumores. Este análisis, cuando se hace manualmente, demanda tiempo, especialistas altamente capacitados y es susceptible a errores humanos. La magnitud de este problema queda documentada en el informe de la Auditoría Interna de la CCSS, el cual revela que los retrasos en la lectura de imágenes alcanzan hasta 570 días en estudios de cerebro y cuello, y que la institución no cuenta con un registro consolidado de pacientes en espera, lo que impide dimensionar con precisión la lista de espera y aumenta el riesgo de omisiones diagnósticas (Segura, 2026)

A nivel tecnológico, se requiere una solución que permita estandarizar la ingesta, almacenamiento, preprocesamiento y análisis de imágenes MRI, integrando modelos de aprendizaje profundo capaces de identificar regiones tumorales y clasificarlas de forma automática con métricas auditables. A continuación, se explicará por qué se eligió utilizar IA, una arquitectura en la nube y redes neuronales.

En primer lugar, la IA aporta escalabilidad, consistencia y cuantificación objetiva, en comparación con la lectura manual de las imágenes MRI. Se propone utilizar un sistema de aprendizaje profundo (*deep learning* en inglés) ya que permite procesar grandes volúmenes de MRI en paralelo y en segundos, mientras que la lectura humana requiere más tiempo y es

propensa a la variabilidad interobservadora. Además, los modelos de IA permiten detectar patrones sub visuales y relaciones multicanal (por ejemplo, entre T1, T2 y FLAIR) que no siempre son evidentes de forma constante para un especialista. Esto resulta en una mayor habilidad para la detección de lesiones, clasificación más confiable y disponibilidad de salidas cuantitativas que facilitan la toma de decisiones clínicas y su posterior auditoría como, por ejemplo, volúmenes, probabilidades calibradas y mapas de atención.

Dorfner, F. et. al dicen en su estudio titulado "*A review of deep learning for brain tumor analysis in MRI"* lo siguiente:

> En el análisis de tumores cerebrales mediante resonancia magnética, la DL se aplica en la segmentación, cuantificación y clasificación tumoral. Facilita mediciones objetivas y reproducibles, cruciales para el diagnóstico, la planificación del tratamiento y el seguimiento de la enfermedad. Además, tiene el potencial de sentar las bases para la medicina personalizada mediante la predicción del tipo de tumor, el grado, las mutaciones genéticas y la supervivencia del paciente (Dorfner, F. et. al, sección **"**Abstract", párr. 1, *traducción propia*).

En segundo lugar, se propone utilizar una plataforma de Big Data en la nube ya que se considera como el entorno más adecuado para soportar esta solución debido a su elasticidad y su gobernanza. El almacenamiento de imágenes en objetos (DICOM) y el cómputo con GPU bajo demanda permiten absorber picos de carga sin tener que sobredimensionar la infraestructura local. Adicionalmente, el desarrollo de un pipeline estandarizado garantiza ingesta, desidentificación, preprocesamiento (registro, normalización, *bias field correction*), inferencia y postprocesado reproducibles. La nube, además, simplifica el cumplimiento y la trazabilidad gracias a funciones como el control de accesos (IAM), cifrado en tránsito y reposo, registro automático de eventos, versión de datos y modelos, y monitorización de *drift*. Operacionalmente, la separación de almacenamiento y cómputo, el autoscaling y los espacios de trabajo compartidos favorecen la colaboración entre radiólogos, patólogos y científicos de datos.

En tercer y último lugar, las redes neuronales profundas son la opción técnica más idónea para MRI porque aprenden representaciones directamente de los vóxeles, evitando la dependencia de *features* diseñadas a mano. Según Sadr, H. et al. en su artículo científico titulado "*Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate diagnosis":*

> Las técnicas tradicionales de aprendizaje automático empleadas para abordar este problema requieren la creación manual de características para la clasificación. Por otro lado, se pueden formular metodologías de aprendizaje profundo para evitar la extracción manual de características y, al mismo tiempo, lograr resultados de clasificación precisos. Por consiguiente, decidimos proponer un modelo basado en aprendizaje profundo para la clasificación automática de tumores cerebrales a partir de imágenes de resonancia magnética (sección **"**Objective", párr. 1, *traducción propia*).

Se pueden aplicar arquitecturas 3D (como 3D U-Net) que capturan el contexto volumétrico completo, o también modelos híbridos CNN/Transformers los cuales incorporan mecanismos de atención útiles para distinguir subregiones tumorales (realce, edema, necrosis).

1.4 Justificación

El proyecto es relevante porque contribuye a tres dimensiones fundamentales. Primero, en el ámbito clínico, apoya a los especialistas al ofrecer resultados precisos y consistentes que pueden complementar su criterio diagnóstico.

Chen, M. en el artículo titulado *"Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation"* dice lo siguiente:

> Los médicos se enfrentan a una carga de trabajo cada vez mayor en la interpretación de imágenes médicas, y la inteligencia artificial (IA) ofrece un alivio potencial. Este metaanálisis evalúa el impacto de la colaboración humano-IA en la carga de trabajo de interpretación de imágenes. Se buscaron cuatro bases de datos para estudios que comparan el tiempo o la cantidad de lectura para la detección de enfermedades basada en imágenes antes y después de la integración de la IA.
>
> La Evaluación de la Calidad de los Estudios de Precisión Diagnóstica se modificó para evaluar el riesgo de sesgo. La reducción de la carga de trabajo y el rendimiento diagnóstico relativo se agruparon utilizando un modelo de efectos aleatorios. Se incluyeron treinta y seis estudios. La asistencia concurrente de IA redujo el tiempo de lectura en un 27,20 % (intervalo de confianza del 95 %, 18,22 %–36,18 %). La cantidad de lectura disminuyó en un 44,47 % (40,68 %–48,26 %) y un 61,72 % (47,92 %–75,52 %) cuando la IA actuó como segundo lector y preselección, respectivamente (sección "Abstract", párr. 1, *traducción propia*).

Por otro lado, desde el punto de vista tecnológico, introduce una arquitectura Big Data escalable y reproducible que puede adaptarse a diferentes contextos hospitalarios y, desde la perspectiva social, mejora el acceso a herramientas de diagnóstico avanzado en entornos donde los recursos médicos y humanos son limitados. Es importante mencionar que este proyecto de investigación no busca reemplazar al médico, sino proveer un sistema que incremente la eficiencia, reduzca tiempos de espera y mejore la precisión diagnóstica.

1.5 Viabilidad técnica, operativa y económica

- Técnica: existen datasets públicos de imágenes MRI de cerebros y herramientas open source (librerías como TensorFlow, PyTorch, MONAI) que permiten entrenar modelos de segmentación y clasificación. La nube (Azure, AWS) ofrece recursos de cómputo y almacenamiento necesarios.
- Operativa: el proyecto se puede ejecutar con un equipo reducido de una sola persona, aprovechando servicios gestionados de la nube para la orquestación, despliegue y monitoreo.
- Económica: dado que se usarán datasets públicos y planes gratuitos de proveedores cloud, los costos serán controlados y sostenibles para un prototipo académico.

1.6 Objetivo general y específicos

Para esta propuesta de proyecto de investigación, se eligió la taxonomía de Benjamin Bloom, ya que está catalogada como una de las mejores a nivel mundial. Su alcance preciso en tiempos y conceptos facilitará la comprensión del tema principal de estudio y, como su metodología lo indica, cada uno de los objetivos específicos cumplirá de forma indirecta a alcanzar el objetivo principal.

1.6.1 Objetivo general

Desarrollar un sistema de diagnóstico asistido por computadora, con el uso de IA basado en una arquitectura Big Data en la nube y redes neuronales, para clasificar tumores cerebrales en imágenes MRI, con el fin de mejorar la precisión diagnóstica y reducir los tiempos de procesamiento en centros médicos costarricenses.

1.6.2 Objetivos específicos

1. Analizar y documentar los fundamentos teóricos relacionados con el procesamiento de imágenes MRI, el cáncer cerebral, las redes neuronales

profundas, el Big Data y las arquitecturas en la nube, con el fin de establecer el marco conceptual necesario para el desarrollo del sistema propuesto.

2. Aplicar principios de arquitectura de datos para construir un data lake/lakehouse bajo el modelo medallón (bronze–silver–gold), con el fin de gestionar la ingesta y normalización de imágenes MRI y sus metadatos.
3. Examinar y comparar el rendimiento de distintas redes neuronales aplicadas a la segmentación de tumores y a la clasificación de tipos tumorales.
4. Desarrollar pipelines reproducibles de entrenamiento, validación e inferencia en la nube.
5. Evaluar el desempeño técnico de los modelos implementados mediante métricas de exactitud, segmentación y clasificación.
6. Explicar un prototipo funcional que integre los componentes principales del modelo propuesto.
7. Construir un dashboard de monitoreo que visualice métricas técnicas y operativas de los modelos implementados.

1.7 Alcances y limitaciones

1.7.1 Alcances

- Implementación de un prototipo funcional que permita ingresar, procesar y analizar imágenes MRI en la nube.
- Entrenamiento de modelos de segmentación y clasificación de tumores cerebrales.
- Validación técnica con datasets públicos y métricas estándar (accuracy, IoU, F1-score).
- Entrega de un pipeline reproducible y documentado que pueda servir como base para futuros proyectos de investigación aplicada.

1.7.2 Limitaciones

- Dependencia de la disponibilidad y calidad de datasets públicos de imágenes MRI.
- Variabilidad en los equipos de adquisición de imágenes, lo que puede afectar la generalización de los modelos.
- El prototipo no sustituye el criterio clínico de un médico. Se limita a la evaluación técnica en un entorno de investigación.

- Restricciones de cómputo debido a límites de créditos o costos asociados a la infraestructura en la nube.

1.8 Marcos de referencia organizacional y socioeconómico

1.8.1 Marco de referencia organizacional

El proyecto se enfoca en los centros médicos de Costa Rica, principalmente dentro de la Caja Costarricense de Seguro Social (CCSS), cuya misión es garantizar el acceso universal a la salud. El objetivo de la propuesta es fortalecer la capacidad diagnóstica mediante la implementación de soluciones de inteligencia artificial aplicadas al análisis de imágenes de resonancia magnética (MRI), con la finalidad de reducir los tiempos de diagnóstico y aumentar la precisión en la detección de patologías oncológicas.

Dentro de la estructura organizacional, las decisiones clave corresponden a las autoridades médicas y administrativas de la CCSS, mientras que la operación recae en los radiólogos, técnicos en imágenes médicas y personal de informática en salud. También participan actores externos como universidades y proveedores tecnológicos, que pueden aportar conocimiento y soporte especializado en inteligencia artificial.

Los procesos más relevantes abarcan desde la adquisición y almacenamiento de imágenes hasta su análisis e interpretación, incluyendo la generación de reportes para el diagnóstico y seguimiento de pacientes. En este entorno, las capacidades tecnológicas varían considerablemente ya que los hospitales cuentan con equipos MRI de distintas generaciones y con infraestructura digital heterogénea, que incluye sistemas PACS y expedientes digitales, pero con una baja integración de soluciones en la nube y con limitaciones en la adopción de Big Data. Según Vega (2025), en el Centro Nacional de Imágenes Médicas de la CCSS, “las imágenes de las resonancias magnéticas se encuentran disponibles en el servidor de la institución minutos después de que el paciente finaliza su estudio (párr. 5).” Esto quiere decir que, en el centro más innovador de imágenes médicas en Costa Rica, no se están utilizando tecnologías en la nube para almacenar las imágenes MRI.

Es importante mencionar también que existen restricciones importantes relacionadas con los presupuestos limitados del sistema público, las regulaciones estrictas en materia de protección de datos (Ley 8968) y la alta presión sobre los tiempos de atención. Esto se traduce en varios retos actuales, tales como los largos tiempos de espera para resonancias y diagnósticos, las inconsistencias técnicas derivadas de la diversidad de equipos y protocolos, y

la sobrecarga de trabajo que enfrentan los radiólogos ante la escasez de especialistas. Frente a este panorama, los criterios de éxito del proyecto se centran en lograr una reducción significativa en el tiempo de análisis, mejorar la precisión en la clasificación de tumores, obtener la aceptación de médicos y técnicos, y garantizar que la solución cumpla con las normativas vigentes al tiempo que mejore la eficiencia operativa.

1.8.2 Marco de referencia socioeconómico

El contexto externo que enmarca este proyecto está determinado por el sistema de salud costarricense, el cual atiende a más del 90 % de la población bajo un modelo universal y solidario, aunque enfrenta importantes desafíos en materia de sostenibilidad financiera y modernización tecnológica. Tal como señala la OPS, “La Caja Costarricense de Seguro Social (CCSS) cubre a más del 90% de la población costarricense, ofreciendo un esquema de seguridad social con carácter universal y solidario” (Organización Panamericana de la Salud [OPS], s. f., p. 3). En este escenario, el sector salud atraviesa un proceso de transformación en el que la incorporación de nuevas tecnologías representa una oportunidad para aumentar la eficiencia, pero también plantea retos significativos debido a las limitaciones estructurales vigentes.

Algunos indicadores reflejan la magnitud del problema. El Centro Nacional de Imágenes Médicas realiza en promedio 2.970 resonancias magnéticas al mes en pacientes de todas las edades, lo que genera una alta carga de trabajo que no siempre puede ser procesada con la rapidez necesaria. Además, la incidencia creciente de cáncer en Costa Rica, reconocida como una de las principales causas de morbilidad y mortalidad por el Ministerio de Salud, incrementa la demanda de estudios complejos. Para ser más específicos, podemos citar el siguiente estudio que dice que “La incidencia de cáncer en Costa Rica ha ido en aumento desde la década de 1990… pasando de 135,1 a 188,7 por 100.000 habitantes, lo que representa un incremento del 72 % en 30 años” (Calderón, Guzmán, & Murphy, 2023, sec. 3.1). Todo esto se combina con la escasez de radiólogos y especialistas en imágenes, lo cual alarga las listas de espera y limita la capacidad de diagnóstico temprano.

A esta alta demanda se suman problemas estructurales documentados por la Auditoría Interna de la CCSS: la ausencia de un registro consolidado de pacientes en espera, la falta de sistemas automatizados de priorización y retrasos en la lectura de estudios que afectan de forma especialmente grave los casos oncológicos (Segura, 2026). Estos hallazgos confirman que el problema no es solo de volumen, sino de gestión y capacidad de respuesta institucional.

En cuanto a infraestructura, existe una marcada diferencia entre centros urbanos y rurales: mientras algunos hospitales cuentan con conectividad aceptable y sistemas de archivo digital (PACS), otros enfrentan limitaciones importantes que dificultan la integración de soluciones avanzadas. A nivel normativo, rigen la Ley de Protección de la Persona frente al Tratamiento de sus Datos Personales (Ley 8968) y las políticas internas de confidencialidad de la CCSS, además de estándares internacionales como DICOM e ISO para la gestión de imágenes médicas.

Las tendencias tecnológicas en la región, como el uso de inteligencia artificial en diagnóstico, la adopción del expediente digital único en salud (EDUS) y el interés por soluciones en la nube, muestran un camino hacia la modernización, aunque todavía con avances desiguales. En este contexto, los riesgos incluyen posibles vulnerabilidades en ciberseguridad, resistencia al cambio cultural y los costos iniciales de implementación, mientras que las oportunidades se relacionan con la reducción de listas de espera, el acceso equitativo a diagnósticos más precisos y la posibilidad de democratizar la tecnología en hospitales regionales.

1.9 Estado de la cuestión

1.9.1 Planificación de la revisión

La planificación de la revisión se basó en identificar, evaluar y resumir la literatura científica relacionada con el procesamiento de imágenes médicas MRI mediante arquitecturas Big Data en la nube y técnicas de aprendizaje profundo para la clasificación de tumores cerebrales. El proceso se desarrolló siguiendo principios metodológicos para asegurar que las decisiones tomadas fueran sistemáticas y justificadas. La revisión se estructuró para cubrir el panorama actual de las tecnologías de almacenamiento, procesamiento distribuido, redes neuronales convolucionales (CNN), arquitecturas cloud y sus aplicaciones en salud. A través de esta planificación fue posible delimitar los dominios temáticos, establecer las fuentes relevantes y definir los criterios metodológicos que guiaron la selección y evaluación de los estudios.

1.9.2 Formulación de la pregunta

La formulación de la pregunta de investigación surge de la necesidad de comprender cómo las tecnologías emergentes, particularmente las arquitecturas Big Data y los modelos de redes neuronales profundas, pueden responder a los desafíos actuales presentes en los centros médicos públicos de Costa Rica. El procesamiento de imágenes MRI en instituciones de la Caja

Costarricense de Seguro Social (CCSS) enfrenta limitaciones en capacidad computacional, tiempos de espera prolongados para análisis radiológicos, falta de automatización en la clasificación de tumores cerebrales y carencias en infraestructura digital para soportar cargas de trabajo intensivas.

El proceso de formulación se llevó a cabo analizando las condiciones operativas de los hospitales nacionales y regionales, la disponibilidad de equipos de resonancia magnética, el volumen creciente de estudios neurológicos, y las brechas tecnológicas que dificultan la adopción de soluciones escalables basadas en inteligencia artificial. Además, se revisó literatura internacional sobre procesamiento distribuido de imágenes médicas y modelos de deep learning para diagnóstico asistido, identificando una oportunidad clara para adaptar dichas soluciones al contexto costarricense.

1.9.3 Foco de la pregunta

El foco de la pregunta se centró en tres dimensiones principales:

- Tecnológica: arquitecturas Big Data, servicios en la nube, almacenamiento distribuido, cómputo paralelo, frameworks como Hadoop/Spark, y plataformas como AWS, Azure o GCP.
- Algorítmica: modelos de aprendizaje profundo, especialmente CNN, segmentación y clasificación.
- Aplicativa: casos reales de uso en el diagnóstico asistido por computadora, flujos de preprocesamiento MRI y manejo de datasets médicos.

1.9.4 Amplitud y calidad de la pregunta

La pregunta se formuló con la amplitud suficiente para incluir diferentes enfoques (arquitectónicos, algorítmicos y clínicos), pero también manteniendo una precisión metodológica que permitiera excluir estudios que no son relevantes. Se priorizaron investigaciones con autoridad científica, métricas claras de evaluación, metodologías reproducibles y datos derivados de repositorios médicos validados, específicamente datasets de imágenes MRI en Kaggle.

1.9.5 Problema

El sistema de salud público de Costa Rica, administrado por la Caja Costarricense de Seguro Social (CCSS), enfrenta importantes desafíos en el procesamiento oportuno y preciso de

imágenes de resonancia magnética (MRI) utilizadas para el diagnóstico de tumores cerebrales. En los centros médicos nacionales y regionales, la demanda creciente de estudios de neuroimagen supera con frecuencia la capacidad instalada de equipos, personal radiológico especializado y recursos computacionales disponibles, generando tiempos de espera prolongados tanto para la adquisición como para la interpretación de los estudios. Esta situación se ve agravada por la falta de soluciones tecnológicas escalables que permitan automatizar y acelerar el análisis de imágenes médicas de alta complejidad.

En la actualidad, la mayoría de los procesos diagnósticos basados en MRI en los hospitales públicos costarricenses dependen de flujos manuales y de la disponibilidad de especialistas en radiología y neuroimagen. Esta dependencia provoca cuellos de botella, variabilidad en la interpretación de hallazgos y limitaciones en la capacidad del sistema para manejar grandes volúmenes de datos provenientes de estudios clínicos. Las consecuencias de esta dependencia manual ya son medibles. Las consecuencias de esta dependencia manual ya están documentadas. La Auditoría Interna de la CCSS identificó retrasos críticos en la interpretación de estudios y una ausencia de sistemas organizados de gestión, situación que representa riesgos significativos para el diagnóstico oportuno de pacientes oncológicos (Segura, 2026).

A pesar de los avances internacionales en el uso de arquitecturas Big Data y redes neuronales para la clasificación automatizada de tumores cerebrales en imágenes MRI, el sistema de salud público costarricense todavía no ha incorporado de manera sistemática este tipo de tecnologías en sus flujos diagnósticos. Esta brecha tecnológica limita el aprovechamiento de herramientas capaces de reducir significativamente los tiempos de análisis, mejorar la precisión diagnóstica y estandarizar la interpretación de los estudios, especialmente en centros con alta demanda y recursos limitados.

En resumen, existe un problema estructural derivado de la falta de infraestructura escalable, herramientas automatizadas y procesos optimizados para el análisis de imágenes MRI en los centros médicos públicos de Costa Rica. Esta situación afecta directamente la eficiencia operativa, la capacidad de respuesta clínica y la calidad del diagnóstico, evidenciando la necesidad de explorar soluciones basadas en arquitecturas Big Data y redes neuronales profundas que permitan mejorar el desempeño del sistema y brindar soporte al personal sanitario en la detección de tumores cerebrales.

1.9.6 Pregunta

¿De qué manera la implementación de tecnologías Big Data y modelos de redes neuronales profundas puede optimizar el procesamiento de imágenes MRI en los centros médicos públicos de Costa Rica, con el objetivo de reducir los tiempos de análisis radiológico y aumentar la precisión diagnóstica en la clasificación de tumores cerebrales?

1.9.7 Palabras clave y sinónimos

Con el fin de recuperar la literatura científica relacionada con la utilización de tecnologías Big Data y redes neuronales en el procesamiento de imágenes MRI para el diagnóstico de tumores cerebrales en los centros médicos públicos de Costa Rica, se definió un conjunto de palabras clave y sus sinónimos asociados. Estas palabras fueron seleccionadas considerando los términos más utilizados en la literatura especializada, así como las variantes terminológicas empleadas en diferentes bases de datos académicas y repositorios médicos. Las palabras clave se agrupan en seis categorías principales: imagenología médica, tecnologías Big Data, inteligencia artificial, entornos cloud, contexto clínico y procesos diagnósticos.

Tabla 1. *Palabras clave sobre la imagenología médica*

Palabra clave	Sinónimo	Traducción en inglés
Resonancia magnética	MRI	Magnetic Resonance Imaging
Imagen cerebral	Neuroimagen	Brain imaging
Tumor cerebral	Glioma	Brain tumor
Estudio MRI	Escaneo de resonancia	MRI scan
Imagen médica	Imagenología	Medical imaging

Fuente: elaboración propia.

Tabla 2. *Palabras clave sobre la inteligencia artificial y deep learning*

Palabra clave	Sinónimo	Traducción en inglés
Aprendizaje profundo	Deep learning	Deep learning
Red neuronal convolucional	CNN	Convolutional Neural Network
Clasificación automática	Clasificación por IA	Automated classification
Modelo de predicción	Modelo neuronal	Predictive model

Transferencia de aprendizaje	Transfer learning	Transfer learning

Fuente: elaboración propia.

Tabla 3. *Palabras clave sobre las tecnologías Big Data*

Palabra clave	Sinónimo	Traducción en inglés
Big Data	Datos masivos	Big Data
Procesamiento distribuido	Cómputo paralelo	Distributed processing
Sistema distribuido	Arquitectura distribuida	Distributed system
Data Lake	Lago de datos	Data lake
Procesamiento masivo	Análisis masivo de datos	Large-scale data processing

Fuente: elaboración propia.

Tabla 4. *Palabras clave sobre las plataformas y servicios en la nube*

Palabra clave	Sinónimo	Traducción en inglés
Computación en la nube	Cloud computing	Cloud computing
Infraestructura en la nube	IaaS	Cloud infrastructure
Plataforma cloud	Servicio cloud	Cloud platform
Aceleración por GPU	Procesamiento acelerado	GPU acceleration
Almacenamiento distribuido	Storage distribuido	Distributed storage

Fuente: elaboración propia.

Tabla 5. *Palabras clave sobre el diagnóstico médico y procesos clínicos*

Palabra clave	Sinónimo	Traducción en inglés
Diagnóstico asistido	CAD	Computer-Aided Diagnosis
Clasificación de tumores	Detección de tumores	Tumor classification
Análisis radiológico	Interpretación radiológica	Radiological analysis
Flujo diagnóstico	Proceso diagnóstico	Diagnostic workflow

Toma de decisiones clínicas	Soporte clínico	Clinical decision making

Fuente: elaboración propia.

Tabla 6. *Palabras clave sobre el contexto clínico costarricense*

Palabra clave	Sinónimo	Traducción en inglés
Caja Costarricense de Seguro Social	CCSS	Costa Rican Social Security Fund
Hospital público	Centro médico estatal	Public hospital
Sistema de salud costarricense	Red hospitalaria pública	Costa Rican public health system
Radiología médica	Servicio de imagenología	Medical radiology
Lista de espera	Tiempo de espera clínico	Waiting list

Fuente: elaboración propia.

1.9.8 Listado de palabras

Para garantizar una recuperación exhaustiva de la literatura se usó el siguiente listado consolidado: “brain tumor classification”, “MRI deep learning”, “cloud-based medical imaging”, “big data healthcare”, “distributed image processing”, “CNN MRI brain”, “Hadoop medical imaging”, “Spark MRI”, “cloud neural networks”, “medical AI cloud architectures”.

1.9.9 Intervención

La intervención considerada en los estudios es la aplicación de modelos de aprendizaje profundo, principalmente CNN, sobre imágenes MRI preprocesadas, utilizando flujos de trabajo que incluyen normalización, reducción de ruido, segmentación y aumento de datos. La intervención tecnológica analizada incluye la adopción de plataformas cloud para acelerar el entrenamiento y la inferencia mediante GPUs y clústeres distribuidos.

1.9.10 Control

Los estudios control incluyen modelos tradicionales de machine learning, arquitecturas locales no distribuidas o soluciones basadas en hardware limitado. También se consideran controles metodológicos como validación cruzada, partición training/validation/test e indicadores como precisión, sensibilidad y especificidad.

1.9.11 Resultado

Los resultados medidos en la literatura se relacionan con el rendimiento de los modelos, la escalabilidad del procesamiento, la reducción del tiempo total de análisis, la precisión diagnóstica, la capacidad de generalización del modelo y el uso eficiente de los recursos computacionales en entornos distribuidos.

1.9.12 Medida de salida

Las medidas de salida utilizadas son las siguientes:

- Exactitud (Accuracy)
- Sensibilidad y especificidad
- F1-Score
- AUC-ROC
- Tiempo de entrenamiento
- Tiempo promedio de inferencia por estudio
- Consumo de recursos y escalabilidad

1.9.13 Población

La población objetivo son imágenes MRI de cerebro obtenidas de pacientes con sospecha de tumores. Los datasets más utilizados en la literatura incluyen BraTS (2015–2023), TCIA y estudios multidisciplinarios provenientes de instituciones neurológicas.

1.9.14 Aplicación

Las aplicaciones identificadas van desde diagnósticos asistidos por computadora, soporte a radiólogos, análisis longitudinal de tumores, segmentación automatizada, hasta sistemas de priorización de estudios.

1.9.15 Diseño experimental

Los estudios analizados emplean diseños experimentales basados en simulación computacional, evaluación de modelos de IA, experimentos reproducibles en plataformas cloud y comparaciones entre diferentes arquitecturas. La mayoría utiliza métodos cuantitativos.

1.9.16 Selección de fuentes

La selección de fuentes se realizó siguiendo un procedimiento sistemático, con el fin de garantizar la pertinencia y calidad metodológica de los estudios incluidos. El proceso contempló una búsqueda exhaustiva en bases de datos científicas (PubMed, Scopus, SpringerLink, IEEE Xplore, Web of Science, Nature), repositorios regionales (SciELO) y documentos institucionales de la CCSS. A partir de los criterios de inclusión y exclusión previamente definidos, se identificaron inicialmente 184 artículos potencialmente relevantes. Tras una revisión por título y resumen, 54 estudios fueron considerados elegibles. Luego, mediante lectura completa del texto y aplicación de los criterios metodológicos de calidad, se seleccionaron 13 fuentes principales, consideradas las más relevantes para el análisis del uso de tecnologías Big Data, deep learning y arquitecturas cloud aplicadas al procesamiento de imágenes MRI para tumores cerebrales.

1.9.17 Definición del criterio de selección de fuentes

La definición del criterio de selección de fuentes se estableció con el propósito de garantizar que los estudios incluidos en la revisión sistemática fueran pertinentes, metodológicamente sólidos y directamente relacionados con el objetivo de esta investigación. Para ello, se diseñó un conjunto de criterios explícitos que permitieran filtrar la literatura científica de manera consistente y transparente, asegurando que solo los estudios relevantes fueran considerados para el análisis final.

Los criterios de selección se basaron en cuatro dimensiones principales:

a. Pertinencia temática

- Se incluyeron únicamente fuentes que abordaran al menos uno de los siguientes temas:
- Procesamiento de imágenes MRI.
- Clasificación o segmentación de tumores cerebrales.
- Aplicaciones de deep learning o modelos de redes neuronales.

- Arquitecturas Big Data, procesamiento distribuido o cloud computing en salud.
- Diagnóstico asistido por computadora en entornos clínicos.
- Estudios o reportes contextualizados en Costa Rica o Latinoamérica.

b. Calidad metodológica

Los estudios debían presentar:

- Metodologías claras y reproducibles.
- Experimentación documentada (datasets, métricas, modelos, validaciones).
- Resultados cuantitativos verificables.
- Revisión por pares o respaldo institucional reconocido (CCSS, OPS, NIST).

c. Actualidad y relevancia temporal

Se priorizó literatura publicada entre 2015 y 2025, con excepción de artículos seminales o de referencia histórica que fundamentan conceptos esenciales (como Bruner et al., 1997).

d. Accesibilidad y confiabilidad

Solo se consideraron fuentes con acceso a texto completo, repositorios institucionales confiables, evitando la literatura informal o no académica. Estos criterios permitieron estructurar un proceso de filtrado riguroso que redujo el conjunto inicial de 184 artículos a las 13 fuentes más relevantes y representativas del estado actual del conocimiento.

1.9.18 Lenguaje de estudio

El lenguaje de estudio corresponde a los idiomas en los cuales se incluyeron las fuentes científicas analizadas en la revisión sistemática. Dado que la mayor parte de la producción académica relacionada con inteligencia artificial, Big Data y análisis de imágenes médicas se publica en inglés, este fue el idioma predominante. Sin embargo, también se utilizaron fuentes en español.

1.9.19 Identificación de las fuentes

La identificación de las fuentes se realizó utilizando combinaciones de palabras clave en inglés relacionadas con: brain tumor classification, MRI deep learning, cloud computing for medical imaging, Big Data healthcare, distributed image processing, neural networks MRI. Se consultaron bases reconocidas internacionalmente y también documentos institucionales relevantes para el contexto costarricense. En esta etapa se recopilaron todos los estudios que pudieran estar relacionados con la temática.

1.9.20 Método de selección de fuentes

Se aplicó un proceso de filtrado en tres niveles:

1. Filtro inicial por título y resumen: Se excluyeron artículos no relacionados con MRI, IA médica o arquitecturas Big Data.
2. Lectura completa: Se verificó que el estudio cumpliera con los criterios metodológicos y de contenido.
3. Evaluación de calidad: Se valoró la claridad metodológica, el tipo de datos, la validez clínica o técnica y la relevancia para Costa Rica.

Solo las fuentes que cumplieron los tres filtros fueron incorporadas.

1.9.20 Cadena de búsqueda

Ejemplo de cadenas de búsqueda utilizadas:

- *(“MRI brain tumor” AND “deep learning” AND “classification”)*
- *(“cloud computing” AND “medical imaging” AND “Big Data”)*
- *(“distributed processing” AND MRI AND healthcare)*
- *(“neural networks” AND “brain tumor segmentation”)*
- *(“Costa Rica” AND “medical imaging” AND “diagnostics”)*

Las búsquedas se realizaron entre agosto y noviembre de 2025.

1.9.21 Lista de fuentes seleccionadas

Las 13 fuentes consideradas como base del análisis fueron:

1. Bruner et al. (1997)
2. Khalighi et al. (2024)
3. Dorfner et al. (2025)
4. Kilim et al. (2022)
5. Sadr et al. (2024)
6. Hashem et al. (2015)
7. Kitchin (2021)
8. Mell & Grance (2011)
9. Aguilar et al. (2024)
10. Vega (2025) – CCSS
11. OPS (s. f.)
12. Calderón, Guzmán & Murphy (2023)
13. Chen (2024)

Estas fuentes proporcionan evidencia clínica, técnica, contextual y operativa para la investigación aplicada.

1.9.22 Selección de fuentes después de la evaluación

Tras aplicar los criterios de calidad metodológica (relevancia, claridad experimental, aporte técnico o clínico, aplicabilidad al contexto costarricense), se descartaron 41 de los 54 estudios analizados en texto completo. Los 13 artículos finales demostraron tener:

- Relevancia directa con el diagnóstico asistido mediante MRI.
- Rigor científico y validez metodológica.
- Enfoque técnico o clínico aplicable al objetivo del proyecto.

1.9.23 Comprobación de las fuentes

Cada una de las 13 fuentes fue verificada en:

- Autenticidad de la publicación (revista arbitrada o institución reconocida).
- Disponibilidad de texto completo.
- Fecha de publicación dentro del rango definido.
- Pertinencia explícita con deep learning, Big Data o diagnóstico MRI.

1.9.24 Selección de los estudios

Los estudios finales fueron elegidos por su aporte específico:

- Clínico (Bruner, Khalighi, Aguilar, OPS).
- Técnico (Kilim, Sadr, Hashem, Kitchin, Mell & Grance).
- Contextual costarricense (Vega, Calderón).
- Operativo (Chen).

1.9.25 Criterios de inclusión y exclusión

Criterios de inclusión:

- Artículos científicos arbitrados.
- Aplicación de IA o deep learning en MRI.
- Uso de Big Data, cloud computing o procesamiento distribuido.
- Estudios clínicos sobre interpretación de imágenes cerebrales.
- Documentos institucionales relevantes al sistema de salud de Costa Rica.

Criterios de exclusion:

- Estudios sin aplicación a MRI.
- Artículos no arbitrados o sin evidencia metodológica verificable.
- Publicaciones sin acceso a texto completo.
- Modelos puramente teóricos sin aplicación clínica o técnica.
- Modalidades de imagen que no incluyeran MRI.

1.9.26 Definición de tipos de estudio

Los estudios seleccionados se categorizan en:

- Revisiones sistemáticas (Khalighi, Dorfner).
- Estudios experimentales computacionales (Sadr, Kilim).
- Análisis clínicos observacionales (Bruner).
- Estudios epidemiológicos (Calderón).
- Documentos institucionales (CCSS, OPS).
- Estudios metodológicos y conceptuales (Hashem, Mell, Kitchin).

1.9.27 Tipos de preguntas para definir artículos

Se clasificaron en tres categorías:

1. Preguntas clínicas:
   - ¿Es la IA más precisa que el radiólogo?
   - ¿Cuánto varía la interpretación humana?
2. Preguntas técnicas:
   - ¿Cómo afecta el domain shift a los modelos de MRI?
   - ¿Qué arquitecturas permiten escalar el procesamiento de imágenes?
3. Preguntas contextuales:
   - ¿Cuál es la carga de trabajo en Costa Rica?
   - ¿Qué limitaciones tiene la CCSS en infraestructura tecnológica?

1.9.28 Ejecución de la revisión

A continuación, se presenta la Tabla 7 que resume las 13 fuentes seleccionadas, con título, autor(es), año y URL (cuando aplica):

Tabla 7. *Fuentes seleccionadas*

Título	Autores	Año	URL / DOI
Physical imaging parameter variation drives domain shift	Kilim, O. et al.	2024	https://doi.org/10.1038/s41467-024-48259-4
Impact of human and artificial intelligence collaboration on workload reduction in medical	Chen, M.	2024	https://doi.org/10.1016/j.acra.2024.05.012

image interpretation			
Application of artificial intelligence in public health for early diagnosis and screening of oncological diseases	Aguilar, C. et al.	2025	https://www.ccss.sa.cr/noticias/aplicacion-inteligencia-artificial-salud-publica
Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate diagnosis	Sadr, H. et al.	2023	https://doi.org/10.1016/j.compbiomed.2023.106785
A review of deep learning for brain tumor analysis in MRI	Dorfner, F. et al.	2021	https://doi.org/10.1016/j.media.2021.102123
Magnetic Resonance Imaging (MRI): How does it work?	National Institute of Biomedical Imaging and Bioengineering (NIBIB)	2024	https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri

Brain tumor symptoms	Mayo Clinic	2024	https://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/syc-20350084
American Cancer Society – Brain and Spinal Cord Tumors in Adults	American Cancer Society	2023	https://www.cancer.org/cancer/types/brain-spinal-cord-tumors-adults.html
Data Mining: Concepts and Techniques	Han, J., Pei, J. & Kamber, M.	2022	https://doi.org/10.1016/C2018-0-02010-2
Big Data and Cloud Computing: Trends and Future Directions	Hashem, I. A. et al.	2015	https://doi.org/10.1016/j.future.2015.12.026
The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences	Kitchin, R.	2021	https://doi.org/10.4135/9781526436317

Centro Nacional de Imágenes Médicas realiza cada mes 2 970 resonancias magnéticas	Vega, Y.	2025	https://www.ccss.sa.cr/noticias/centro-nacional-imagenes-medicas-realiza-resonancias-magneticas
Incidencia del cáncer en Costa Rica desde 1990 a 2020	Calderón, T., Guzmán, M., & Murphy, J.	2023	https://doi.org/10.1590/1980-549720230006

Fuente: elaboración propia.

1.9.29 Revisión de la selección

Se revisó nuevamente el conjunto de fuentes seleccionadas para garantizar su alineación con los objetivos de la investigación. La revisión final confirmó que las 13 fuentes elegidas:

- Cubren todos los dominios temáticos: clínico, técnico, operativo y contextual.
- Permiten sustentar la justificación del uso de Big Data + IA en la CCSS.
- Representan evidencia sólida y reciente (especialmente 2022–2025).
- Incluyen datos relevantes para la realidad costarricense.
- Permiten construir un marco metodológico sólido para la propuesta tecnológica.

Esta revisión final certifica la coherencia entre la evidencia disponible y los objetivos del proyecto, asegurando que la investigación se desarrolla sobre una base teórica robusta y actualizada.

1.9.30 Extracción de la información

La extracción de la información consistió en un proceso sistemático y estructurado para recopilar los datos relevantes de cada uno de los 13 estudios incluidos en la revisión. Este proceso permitió obtener una visión clara y comparable de los métodos, resultados, limitaciones y conclusiones reportadas en la literatura científica sobre el uso de arquitecturas Big Data,

modelos de deep learning y tecnologías en la nube aplicadas al procesamiento de imágenes MRI para la clasificación de tumores cerebrales. Con ello se garantizó la homogeneidad del análisis y la trazabilidad metodológica.

Tabla 8. *Fuente Physical imaging parameter variation drives domain shift. Kilim et al. (2024).*

Título	***Physical imaging parameter variation drives domain shift***
Autor(es)	Kilim, O. et al.
Año	2024
Resumen corto	El estudio analiza cómo pequeñas variaciones en parámetros de adquisición MRI pueden causar *domain shift*, reduciendo drásticamente el rendimiento de modelos de deep learning entrenados en condiciones controladas. El artículo demuestra que las inconsistencias entre máquinas, protocolos y configuraciones pueden causar fallas significativas en la generalización de los modelos.
Aspectos a considerar	Identifica el *domain shift* como una de las mayores limitaciones en IA médica; destaca la necesidad de normalizar o armonizar imágenes MRI; altamente relevante para Costa Rica debido a la variedad de máquinas MRI en el sistema público.

Fuente: elaboración propia.

Tabla 9. *Fuente Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. Chen, M. (2024).*

Título	***Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation***
Autor(es)	Chen, M.
Año	2024

Resumen corto	El estudio evalúa cómo la colaboración entre radiólogos y modelos de inteligencia artificial reduce la carga de trabajo y aumenta la precisión en interpretación de imágenes médicas. Se reportan mejoras significativas en eficiencia diagnóstica y reducción de tiempos de lectura.
Aspectos a considerar	Aporta evidencia directa del beneficio clínico de integrar IA en procesos diagnósticos; destaca el impacto en tiempos de respuesta, factor crítico para hospitales públicos costarricenses.

Fuente: elaboración propia.

Tabla 10*. Fuente Aplicación de inteligencia artificial en salud pública para diagnóstico temprano y tamizaje de enfermedades oncológicas. Aguilar, C. et al. (2025).*

Título	***Aplicación de inteligencia artificial en salud pública para diagnóstico temprano y tamizaje de enfermedades oncológicas***
Autor(es)	Aguilar, C. et al.
Año	2025
Resumen corto	Documento institucional que describe iniciativas de IA en la CCSS enfocadas en diagnóstico temprano de cáncer. Presenta casos de uso, retos de integración y resultados preliminares en entornos clínicos públicos del país.
Aspectos a considerar	Relevancia directa para Costa Rica; evidencia oficial del interés institucional en IA; justifica la viabilidad de implementar soluciones basadas en Big Data y deep learning.

Fuente: Elaboración propia.

Tabla 11. *Fuente Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate diagnosis. Sadr, H. et al. (2023)*

Título	***Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate diagnosis***
Autor(es)	Sadr, H. et al.
Año	2023
Resumen corto	Propone un modelo de deep learning para clasificación de tumores cerebrales utilizando MRI. Reporta métricas competitivas (accuracy, sensibilidad, AUC) y demuestra la efectividad de CNN avanzadas con preprocesamiento optimizado.
Aspectos a considerar	Estudio experimental robusto; ofrece métricas base para comparación; útil para validar el desempeño esperado de tu modelo.

Fuente: elaboración propia.

Tabla 12. *Fuente A review of deep learning for brain tumor analysis in MRI. Dofner, F. et al. (2021)*

Título	***A review of deep learning for brain tumor analysis in MRI***
Autor(es)	Dorfner, F. et al.
Año	2021
Resumen corto	Revisión sistemática que resume los principales avances en deep learning aplicado a MRI para análisis y clasificación de tumores. Discute arquitecturas, desafíos, datasets y brechas tecnológicas.
Aspectos a considerar	Excelente síntesis del estado del arte; útil para contextualizar tu solución; destaca la necesidad de infraestructura computacional escalable.

Fuente: elaboración propia.

Tabla 13. *Fuente Magnetic Resonance Imaging (MRI): How does it work? National Institute of Biomedical Imaging and Bioengineering. (2024)*

Título	***Magnetic Resonance Imaging (MRI): How does it work?***
Autor(es)	National Institute of Biomedical Imaging and Bioengineering
Año	2024
Resumen corto	Explica los fundamentos físicos y técnicos de la resonancia magnética, tipos de señales, secuencias y aplicaciones clínicas. Brinda una base conceptual esencial para comprender el proceso de adquisición de imágenes MRI.
Aspectos a considerar	Fundamental para explicar cómo funciona MRI en el marco conceptual; ayuda a justificar la complejidad del procesamiento y la necesidad de IA.

Fuente: elaboración propia.

Tabla 14. *Fuente Brain tumor symptoms. Mayo Clinic. (2024)*

Título	***Brain tumor symptoms***
Autor(es)	Mayo Clinic
Año	2024
Resumen corto	Describe los síntomas clínicos relacionados con tumores cerebrales y su relevancia diagnóstica. Proporciona contexto médico sobre la importancia de estudios MRI para la detección temprana.
Aspectos a considerar	Útil para la justificación clínica; brinda información sobre la importancia diagnóstica del proyecto.

Fuente: elaboración propia.

Tabla 15. *Fuente Brain and Spinal Cord Tumors in Adults. American Cancer Society. (2023)*

Título	***Brain and Spinal Cord Tumors in Adults***

Autor(es)	American Cancer Society
Año	2023
Resumen corto	Informe clínico que detalla los tipos, características, prevalencia y evolución de tumores cerebrales en adultos. Incluye recomendaciones sobre diagnóstico y uso de MRI.
Aspectos a considerar	Ofrece datos epidemiológicos y clínicos valiosos; fortalece el planteamiento del problema.

Fuente: elaboración propia.

Tabla 16. *Fuente Data Mining: Concepts and Techniques. Han, J., Pei, J. & Kamber, M. (2022)*

Título	***Data Mining: Concepts and Techniques***
Autor(es)	Han, J., Pei, J. & Kamber, M.
Año	2022
Resumen corto	Obra fundamental que describe técnicas de minería de datos, clasificación, reducción de dimensionalidad y fundamentos de machine learning.
Aspectos a considerar	Aporta sustento teórico a técnicas de procesamiento masivo y modelos predictivos; crucial para el marco conceptual y metodológico.

Fuente: elaboración propia.

Tabla 17. *Fuente Big Data and Cloud Computing: Trends and Future Directions. Hashem, I. et al. (2015)*

Título	***Big Data and Cloud Computing: Trends and Future Directions***
Autor(es)	Hashem, I. et al.
Año	2015
Resumen corto	Revisión clave que analiza el surgimiento del Big Data y las plataformas cloud, enfatizando retos, oportunidades y arquitecturas tecnológicas.

Aspectos a considerar	Brinda base conceptual para justificar el enfoque cloud del proyecto; muy usado como referencia en arquitectura Big Data.

Fuente: elaboración propia.

Tabla 18. *Fuente The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. Kitchin, R. (2021)*

Título	***The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences***
Autor(es)	Kitchin, R.
Año	2021
Resumen corto	Analiza las implicaciones sociales, técnicas y estratégicas del Big Data. Presenta modelos de gobernanza, infraestructura y ética en el uso de datos.
Aspectos a considerar	Útil para contextualizar implicaciones éticas y de gobernanza en IA aplicada al sector salud.

Fuente: elaboración propia.

Tabla 19. *Fuente Centro Nacional de Imágenes Médicas realiza cada mes 2 970 resonancias magnéticas. Vega, Y. (2025)*

Título	***Centro Nacional de Imágenes Médicas realiza cada mes 2 970 resonancias magnéticas***
Autor(es)	Vega, Y.
Año	2025
Resumen corto	Artículo institucional que reporta la carga mensual de resonancias MRI en Costa Rica, destacando saturación operativa del sistema público.
Aspectos a considerar	Soporta directamente el problema de congestión diagnóstica; evidencia real del contexto costarricense.

Fuente: elaboración propia.

Tabla 20. *Fuente Incidencia del cáncer en Costa Rica desde 1990 a 2020. Calderón, T., Guzmán, M., & Murphy, J. (2023)*

Título	***Incidencia del cáncer en Costa Rica desde 1990 a 2020***
Autor(es)	Calderón, T., Guzmán, M., & Murphy, J.
Año	2023
Resumen corto	Estudio epidemiológico que analiza patrones de incidencia del cáncer en Costa Rica durante 30 años, destacando el comportamiento de tumores cerebrales y la carga para el sistema público.
Aspectos a considerar	Permite establecer el marco epidemiológico nacional y justificar la necesidad de herramientas de diagnóstico más eficientes.

Fuente: elaboración propia.

2. Marco conceptual

2.1 Imágenes de resonancia magnética (MRI)

La resonancia magnética o MRI (Magnetic Resonance Imaging) es una técnica de imagen médica no invasiva que produce imágenes detalladas del interior del cuerpo en tres dimensiones, especialmente de tejidos blandos (como el cerebro, la médula espinal, músculos, etc.). Los médicos la utilizan comúnmente para detectar enfermedades, realizar diagnósticos y monitorear tratamientos.

De acuerdo con el National Institute of Biomedical Imaging and Bioengineering (2024):

> Las imágenes por resonancia magnética emplean potentes imanes que producen un campo magnético fuerte que obliga a los protones del cuerpo a alinearse con dicho campo. Cuando una corriente de radiofrecuencia se aplica al paciente, los protones se estimulan y giran fuera de su equilibrio, resistiendo la atracción del campo magnético. Al cesar el campo de radiofrecuencia, los sensores de la MRI detectan la energía liberada a medida que los protones se realinean con el campo magnético. El tiempo que tardan en hacerlo, así como la cantidad de energía liberada, varían según el entorno y la naturaleza química de las moléculas. Los médicos pueden distinguir entre varios tipos de tejidos basándose en estas propiedades magnéticas (sección “How does MRI work?”, párr. 1, *traducción propia*).

Una de las ventajas de la MRI es que el cerebro, espina dorsal, nervios, músculos, ligamentos y tendones se aprecian muchísimo mejor que con los rayos X normales o las tomografías computarizadas. Particularmente en el cerebro, la MRI es capaz de diferenciar entre la materia gris (cuerpos celulares de las neuronas) y la materia blanca (los axones mielinizados, es decir, el tejido que se encarga de la transmisión de impulsos nerviosos). Otra ventaja es que al no utilizar rayos X ni otros tipos de radiación, los pacientes pueden hacerse MRI con frecuencia, sin tener efectos negativos. Sin embargo, una desventaja de la MRI es que es más costosa que los rayos X o las tomografías computarizadas.

Es importante mencionar que existe otro tipo de MRI especializado en el cerebro, llamado MRI funcional. Este es utilizado para observar las estructuras del cerebro y determinar cuáles áreas se activan cuando se realizan ciertas actividades cognitivas. En este trabajo de investigación, nos estaremos enfocando en la MRI normal y no en la funcional.

2.2 Cáncer cerebral

El cáncer cerebral es una enfermedad caracterizada por el crecimiento anormal y descontrolado de células en el tejido cerebral, que forman una masa o tumor capaz de invadir estructuras cercanas, alterar las funciones del sistema nervioso y, en algunos casos, propagarse a otras partes del cuerpo. A diferencia de los tumores benignos, que suelen tener un crecimiento lento y no se diseminan, el cáncer cerebral hace referencia específicamente a los tumores malignos del cerebro, que pueden destruir tejido sano, interferir con la comunicación neuronal y aumentar la presión intracraneal. Es importante aclarar que todos los cánceres en el cerebro son tumores, pero no todos los tumores son cáncer.

2.2.1 Tumores cerebrales

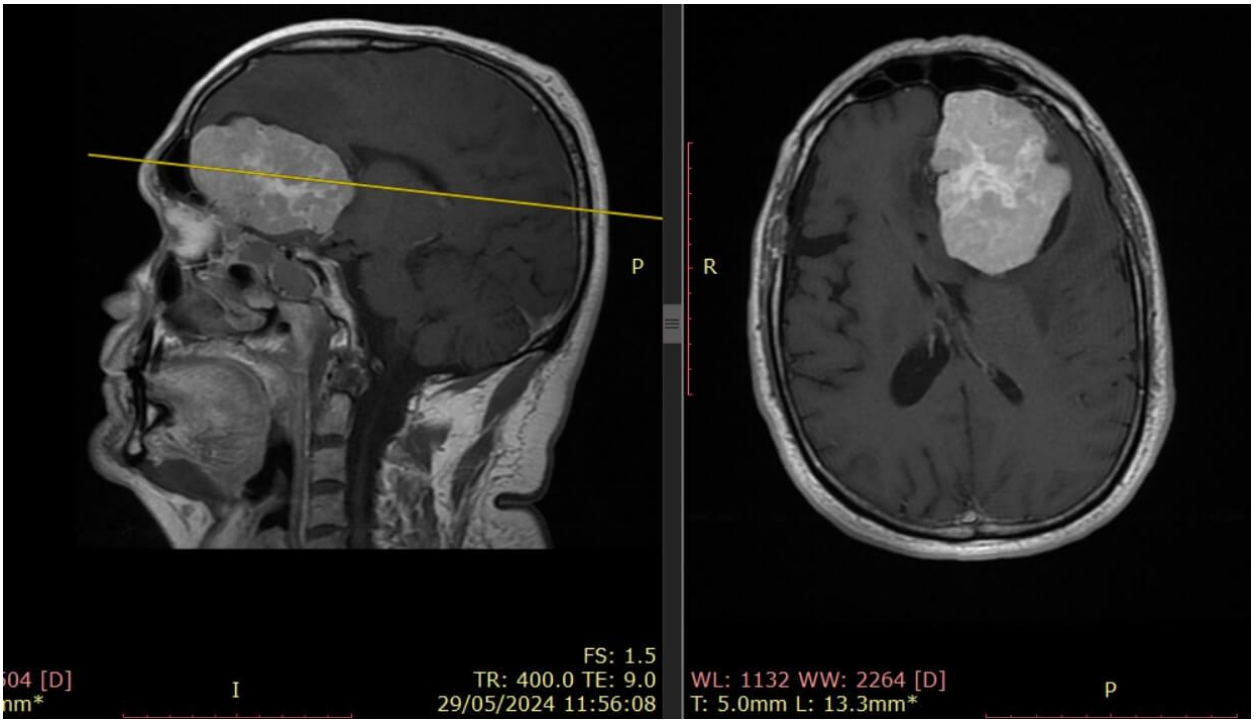
El cáncer cerebral comprende un grupo heterogéneo de neoplasias (crecimientos anormales y descontrolados de células en un tejido) que pueden clasificarse, de forma general, en dos grandes categorías según su origen: tumores primarios y tumores secundarios o metastásicos. Esta distinción es fundamental porque cada grupo difiere en sus mecanismos biológicos, pronóstico, tratamiento y frecuencia de aparición.

Los tumores primarios se originan directamente en los tejidos del sistema nervioso central (SNC), es decir, en el cerebro, el tronco encefálico, el cerebelo o la médula espinal. También pueden desarrollarse a partir de estructuras que lo rodean o lo sustentan, como las meninges, los nervios craneales y la glándula pituitaria (hipófisis).

Los tumores secundarios o metastásicos son aquellos que provienen de cánceres en otras partes del cuerpo y que se trasladan al cerebro mediante metástasis. A diferencia de estos, los tumores primarios no provienen de otro órgano, sino que surgen por mutaciones genéticas o alteraciones celulares en el propio tejido nervioso. Aunque algunos pueden ser benignos (de crecimiento lento y bien delimitado), otros presentan un comportamiento altamente maligno y agresivo.

La Figura 1 ilustra la presencia de un tumor cerebral por medio de una resonancia magnética.

Figura 1. *Ejemplo de tumor cerebral*



Nota. Tomado de *Tumores cerebrales, uno de los mayores desafíos de la neurología moderna*, por Quirónsalud, 2025, https://www.quironsalud.com/es/comunicacion/contenidos-salud/tumores-cerebrales-mayores-desafios-neurologia-moderna

2.2.2 Causas

Las causas de los tumores cerebrales son muy variadas y, en la mayoría de los casos, no pueden determinarse con exactitud. Su origen suele estar relacionado a múltiples factores, generalmente genéticos, ambientales y biológicos, que alteran los mecanismos normales de crecimiento y división celular en el sistema nervioso. En primer lugar, las alteraciones genéticas y moleculares tienen un papel clave. Mutaciones en genes como TP53, PTEN o EGFR pueden provocar una proliferación descontrolada de las células, mientras que cambios epigenéticos, como la metilación del gen MGMT, afectan la reparación del ADN y la respuesta a los tratamientos. Estas alteraciones pueden ser hereditarias o adquiridas durante la vida.

Un pequeño porcentaje de casos se relaciona con síndromes genéticos hereditarios, como la neurofibromatosis tipo 1 y 2, el síndrome de Li-Fraumeni, el síndrome de Turcot y el de von Hippel–Lindau, que aumentan la predisposición a desarrollar tumores del sistema nervioso central. Otro factor reconocido es la exposición a radiación ionizante, especialmente en la

cabeza. Esta puede causar daños en el ADN y favorecer el desarrollo de tumores, aunque la radiación utilizada en estudios médicos se considera segura. También se han estudiado factores ambientales, como la exposición a sustancias químicas o a ondas electromagnéticas, pero la evidencia científica disponible sigue siendo inconclusa.

Según la American Cancer Society (2023):

> Los investigadores ahora comprenden algunos de los cambios genéticos que ocurren en diferentes tipos de tumores cerebrales, pero aún no está claro qué causa la mayoría de estos cambios. Algunos de ellos podrían ser heredados, aunque la mayoría de los tumores cerebrales y de la médula espinal no son el resultado de síndromes hereditarios conocidos. A excepción de la radiación, no existen factores ambientales o relacionados con el estilo de vida que se asocien claramente con los tumores cerebrales. Es probable que la mayoría de los cambios genéticos sean simplemente eventos aleatorios que ocurren dentro de una célula, sin una causa externa identificable (sección “Gene changes acquired during a person's lifetime”, párr. 3, *traducción propia*).

Finalmente, los expertos están de acuerdo con que la causa exacta sigue siendo desconocida en la mayoría de los casos. Se considera que los tumores cerebrales son el resultado de múltiples factores que interactúan entre sí, y las investigaciones actuales se centran en identificar biomarcadores genéticos y moleculares que permitan comprender mejor su origen y desarrollar terapias más precisas.

2.2.3 Síntomas

Los síntomas de los tumores cerebrales dependen principalmente de su tamaño, ubicación, velocidad de crecimiento y grado de malignidad. Debido a que el cerebro controla funciones motoras, sensoriales, cognitivas y emocionales, un tumor puede generar signos muy variados al presionar estructuras vecinas o interferir con la transmisión normal de los impulsos nerviosos.

Según Mayo Clinic (2024):

> Los tumores cerebrales no cancerosos tienden a causar síntomas que se desarrollan lentamente. Estos tumores, también llamados tumores

> cerebrales benignos, pueden provocar síntomas sutiles que al principio pasan desapercibidos. Los síntomas pueden empeorar con el paso de los meses o los años.
>
> Por otro lado, los tumores cerebrales cancerosos provocan síntomas que empeoran rápidamente. Estos también se conocen como cánceres cerebrales o tumores cerebrales malignos. Causan síntomas que aparecen de manera repentina y se agravan en cuestión de días o semanas (sección “Symptoms”, párr. 3, *traducción propia*).

Uno de los síntomas más comunes es la cefalea (dolor de cabeza fuerte) persistente, que suele ser más intensa por la mañana o empeorar al toser, estornudar o realizar esfuerzos físicos, debido al aumento de la presión intracraneal. También pueden presentarse náuseas y vómitos inexplicables, relacionados con ese incremento de presión dentro del cráneo. Las convulsiones o crisis epilépticas son otro signo frecuente, especialmente en tumores que afectan la corteza cerebral. Estas se producen por descargas eléctricas anómalas en el cerebro y pueden manifestarse como movimientos involuntarios, pérdida de conciencia o episodios de confusión súbita.

Además, los pacientes pueden experimentar alteraciones neurológicas focales, es decir, síntomas localizados según el área afectada. Los más comunes son:

- Lóbulos frontales: cambios de personalidad, dificultades para planificar o pérdida del control motor.
- Lóbulos temporales: problemas de memoria, comprensión o lenguaje.
- Lóbulos parietales: alteraciones de la sensibilidad o desorientación espacial.
- Lóbulos occipitales: visión borrosa o pérdida parcial del campo visual.
- Cerebelo: problemas de equilibrio y coordinación.

Asimismo, pueden observarse cambios cognitivos o de comportamiento, debilidad muscular, pérdida de sensibilidad, dificultades para hablar o comprender, somnolencia excesiva, confusión, e incluso alteraciones hormonales cuando el tumor se localiza en la glándula pituitaria. En los niños, es común el aumento del tamaño de la cabeza o retrasos en el desarrollo motor y cognitivo. En general, los síntomas de un tumor cerebral no son específicos y pueden confundirse con otras patologías neurológicas. Por esta razón, la evaluación clínica y los estudios de imagen,

como la resonancia magnética, son esenciales para confirmar el diagnóstico y determinar la causa de los síntomas

2.2.4 Tipos de tumores cerebrales

A continuación, se explican los principales tipos de tumores primarios:

1. Gliomas: Representan alrededor del 80 % de los tumores cerebrales malignos primarios. Se originan en las células gliales, que son las células de soporte y nutrición de las neuronas. Los gliomas se clasifican según el tipo de célula glial de la que derivan:
2. Astrocitomas: surgen de los astrocitos. Pueden variar desde grados bajos (I–II) hasta altamente malignos (III–IV). El glioblastoma multiforme (GBM) es el más agresivo, ya que tiene rápida proliferación y pobre pronóstico.
3. Oligodendrogliomas: provienen de los oligodendrocitos, responsables de la producción de mielina. Generalmente crecen más lentamente y responden mejor a tratamientos combinados.
4. Ependimomas: se desarrollan en las células ependimarias que recubren los ventrículos cerebrales y el canal central de la médula espinal. Son más frecuentes en niños y pueden obstruir la circulación del líquido cefalorraquídeo.
5. Meningiomas: Son tumores originados en las meninges, las membranas que recubren el cerebro y la médula espinal. A pesar de que la mayoría son benignos (grado I), un pequeño porcentaje puede presentar comportamiento maligno (grados II y III). Debido a su localización, pueden comprimir estructuras cerebrales y causar síntomas neurológicos sin invadir directamente el tejido cerebral.
6. Adenomas hipofisarios (pituitarios): Se desarrollan en la glándula pituitaria, ubicada en la base del cráneo. La mayoría son benignos, pero pueden alterar significativamente la producción hormonal, generando trastornos endocrinos como el síndrome de Cushing o la acromegalia.
7. Meduloblastomas: Son tumores embrionarios de alta malignidad, más comunes en la infancia. Se originan en el cerebelo y tienen una fuerte tendencia a diseminarse por el líquido cefalorraquídeo hacia la médula espinal.
8. Otros tumores primarios raros: Incluyen craneofaringiomas (cerca de la hipófisis), schwannomas (de las células de Schwann de los nervios craneales, como el nervio vestibular) y tumores del plexo coroideo, entre otros.

2.2.5 Criterios médicos

El diagnóstico y la evaluación de los tumores cerebrales se basan en una serie de criterios médicos que combinan información clínica, radiológica, histopatológica y molecular. Estos parámetros son esenciales para definir la naturaleza del tumor, su comportamiento biológico y el pronóstico del paciente. En primer lugar, la clasificación histológica y molecular establecida por la Organización Mundial de la Salud (OMS) es uno de los criterios más importantes. Esta clasificación considera tanto la apariencia microscópica de las células tumorales como sus alteraciones genéticas y epigenéticas. Los tumores se agrupan en grados del I al IV, donde los grados bajos indican crecimiento lento y menor agresividad, mientras que los grados altos presentan rápida proliferación, invasión del tejido circundante y peor pronóstico. Por ejemplo, el glioblastoma multiforme se clasifica como grado IV, el más maligno dentro de los gliomas.

En segundo lugar, los hallazgos en las imágenes de resonancia magnética (MRI) son esenciales para el diagnóstico y seguimiento. La MRI permite identificar la localización, tamaño, bordes, edema peritumoral, necrosis, hemorragia o captación de contraste. Las secuencias avanzadas (como difusión, perfusión o espectroscopía) aportan información sobre la celularidad, vascularización y metabolismo tumoral, ayudando a diferenciar tumores de otras lesiones no neoplásicas.

Otro criterio relevante son los síntomas clínicos y el examen neurológico del paciente. Dependiendo de la región afectada, se evalúan la fuerza muscular, la sensibilidad, la coordinación, el lenguaje, la visión y el estado mental. Estos datos permiten correlacionar las manifestaciones clínicas con la localización anatómica del tumor. Asimismo, los estudios moleculares y genéticos han adquirido gran importancia en la práctica médica moderna.

Finalmente, los criterios médicos también consideran el estado funcional y pronóstico global del paciente, utilizando escalas como la Karnofsky Performance Status (KPS), que mide la capacidad para realizar actividades cotidianas. Este tipo de evaluación complementa la información clínica, ayudando a planificar un tratamiento integral que combine cirugía, radioterapia y quimioterapia según las condiciones del paciente. En conjunto, estos criterios permiten un abordaje diagnóstico y terapéutico preciso, asegurando que cada caso sea evaluado desde una perspectiva anatómica, funcional y molecular, conforme a los estándares internacionales más recientes.

2.3 Minería de datos

La minería de datos (Data Mining) es un proceso analítico que consiste en explorar grandes volúmenes de información para descubrir patrones, relaciones, tendencias o conocimientos útiles que no son evidentes a simple vista. Este proceso combina técnicas de estadística, inteligencia artificial, aprendizaje automático (machine learning) y bases de datos, con el fin de transformar los datos en conocimiento accionable. En otras palabras, la minería de datos permite extraer información significativa a partir de conjuntos de datos masivos y heterogéneos, lo que facilita la toma de decisiones informadas en diversos campos como la salud, la banca, el comercio electrónico, la educación o la investigación científica.

Según Han, Pei y Kamber (2022):

> La minería de datos es el proceso de descubrir patrones y conocimientos interesantes a partir de grandes volúmenes de datos. Las fuentes de datos pueden incluir bases de datos, almacenes de datos, la Web y otros repositorios de información o flujos de datos que se incorporan dinámicamente al sistema (sección “What Is Data Mining?”, párr. 1, *traducción propia*).

2.3.1 Técnicas principales de minería de datos

Las principales técnicas de minería de datos incluyen:

- Clasificación: asignar elementos a categorías predefinidas (por ejemplo, diagnosticar una enfermedad según los síntomas).
- Reglas de asociación: identificar relaciones entre variables (por ejemplo, clientes que compran “X” también compran “Y”).
- Agrupamiento o clustering: agrupar datos similares sin categorías previas.
- Detección de anomalías: descubrir comportamientos atípicos (por ejemplo, fraudes o errores).
- Predicción: estimar valores futuros a partir de datos históricos.

2.3.2 Proceso de minería de datos

El proceso forma parte del enfoque más amplio conocido como KDD (Knowledge Discovery in Databases o descubrimiento de conocimiento en bases de datos), el cual incluye varias etapas:

1. Selección y limpieza de datos: se eliminan errores, duplicados o valores inconsistentes.
2. Integración y transformación: los datos se combinan y preparan para el análisis.
3. Minería propiamente dicha: se aplican algoritmos de clasificación, agrupamiento, asociación o predicción.
4. Evaluación e interpretación de resultados: se analizan los patrones detectados y se extrae conocimiento útil.

2.4 Machine learning

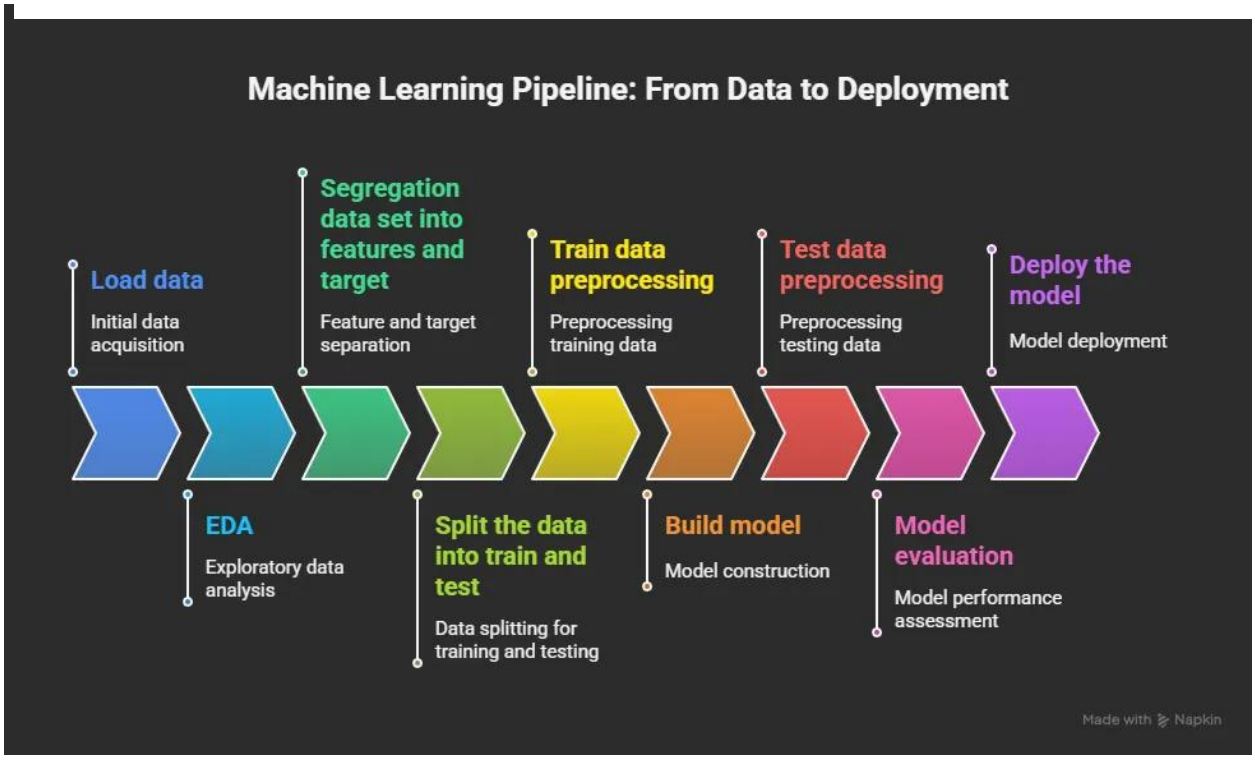
El aprendizaje automático (Machine Learning, en inglés) es una rama de la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos y modelos matemáticos capaces de aprender a partir de datos, identificar patrones y tomar decisiones o hacer predicciones sin ser programados explícitamente. En lugar de seguir instrucciones fijas, los sistemas de machine learning “aprenden” de la experiencia. Estos sistemas analizan grandes conjuntos de datos, ajustan sus parámetros internos y mejoran su rendimiento a medida que se exponen a nueva información. Este enfoque se utiliza en múltiples áreas como la medicina, la economía, la educación, la seguridad informática y el análisis de imágenes o texto.

El principio básico del aprendizaje automático es que una máquina puede generalizar conocimientos a partir de ejemplos. Por ejemplo, un modelo puede aprender a reconocer tumores cerebrales analizando miles de imágenes de resonancia magnética previamente clasificadas por expertos, y luego aplicar lo aprendido para detectar nuevos casos con alta precisión.

Un modelo de aprendizaje automático sigue una serie de etapas organizadas para analizar los datos con el fin de procesarlos para su desarrollo, la figura 2 muestra el flujo de este proceso.

Figura 2

Flujo del proceso de Aprendizaje Automático



Nota. Tomado de *End-to-End Machine Learning Pipeline: From Raw Data to Deployment*, por M. Suvvani, 2025, https://medium.com/@muralisuvvani/end-to-end-machine-learning-pipeline-from-raw-data-to-deployment-624e700f6992

2.4.1 Tipos de machine learning

1. Aprendizaje supervisado: El modelo se entrena con un conjunto de datos etiquetados, donde se conoce la respuesta correcta. Se usa para tareas de clasificación (por ejemplo, diagnosticar enfermedades) y regresión (predecir valores numéricos). Ejemplo: predecir el riesgo de cáncer cerebral según variables clínicas.
2. Aprendizaje no supervisado: Se utilizan datos sin etiquetas para descubrir estructuras ocultas o relaciones entre variables. Las técnicas más comunes son

el agrupamiento (clustering) y la reducción de dimensionalidad. Ejemplo: agrupar pacientes con características clínicas similares.

3. Aprendizaje semi-supervisado: Combina un pequeño conjunto de datos etiquetados con una gran cantidad de datos sin etiquetar, optimizando el rendimiento cuando etiquetar datos es costoso o complejo.
4. Aprendizaje por refuerzo: El sistema aprende mediante ensayo y error, recibiendo recompensas o castigos por sus acciones, similar al aprendizaje humano. Es ampliamente usado en robótica, videojuegos o vehículos autónomos.

2.4.2 Deep learning

El aprendizaje profundo (Deep Learning) es una subrama del aprendizaje automático (Machine Learning) que utiliza redes neuronales artificiales con múltiples capas para modelar y aprender representaciones complejas de los datos. Estas redes imitan, de manera abstracta, el funcionamiento del cerebro humano, procesando la información a través de niveles jerárquicos que permiten reconocer patrones cada vez más sofisticados.

En términos simples, el Deep Learning permite que una computadora aprenda directamente de los datos sin requerir una programación explícita de reglas. A diferencia de los métodos tradicionales, que dependen de características diseñadas manualmente, las redes profundas extraen automáticamente rasgos relevantes a partir de grandes volúmenes de información. Por eso, esta técnica ha revolucionado campos como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la visión por computadora.

2.4.3 Redes neuronales

Las redes neuronales artificiales (RNA) son modelos computacionales inspirados en el funcionamiento del cerebro humano, diseñados para procesar información, reconocer patrones y aprender de los datos. Están compuestas por unidades llamadas neuronas artificiales, que se conectan entre sí formando una estructura en capas capaz de transformar datos de entrada en salidas o predicciones útiles.

El concepto se originó en 1943, cuando Warren McCulloch y Walter Pitts propusieron el primer modelo matemático de una neurona artificial. Posteriormente, Frank Rosenblatt desarrolló el perceptrón en 1958, una de las primeras redes con capacidad de aprendizaje. Según Rosenblatt (1958), “el perceptrón es un modelo probabilístico para el almacenamiento y la

organización de la información en el cerebro, capaz de aprender mediante la modificación de las conexiones sinápticas en respuesta a los estímulos" (p. 386, traducción propia). Desde entonces, el campo ha evolucionado hacia arquitecturas más complejas que constituyen la base del aprendizaje profundo (Deep Learning).

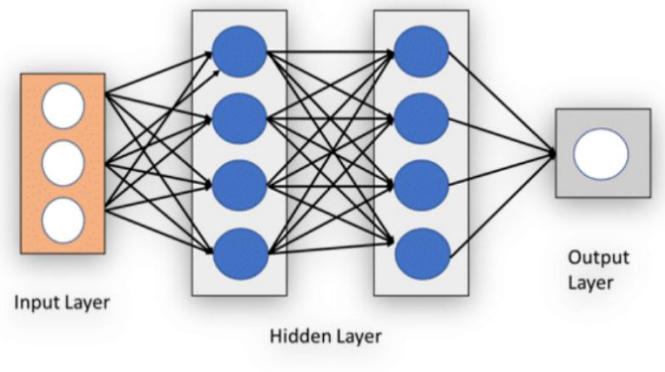
Una red neuronal se organiza en capas interconectadas de neuronas artificiales:

- Capa de entrada: recibe los datos (por ejemplo, los píxeles de una imagen o valores numéricos).
- Capas ocultas: procesan la información aplicando transformaciones matemáticas y funciones de activación (como sigmoid, ReLU o tanh), permitiendo que la red aprenda relaciones no lineales.
- Capa de salida: entrega el resultado final, como una clasificación o predicción.

La figura 3 ilustra el proceso descrito anteriormente.

Figura 3

Estructura de una red neuronal artificial



Nota. Tomado de *Understanding neural networks and their components*, por G. Hansa, 2025,https://www.codecademy.com/article/understanding-neural-networks-and-their-components

Cada conexión entre neuronas tiene un peso que indica la importancia de la señal transmitida. Durante el entrenamiento, la red ajusta estos pesos mediante algoritmos como la retropropagación del error (backpropagation), buscando minimizar la diferencia entre sus predicciones y los valores reales.

Tipos de redes neuronales

Existen múltiples arquitecturas, cada una adaptada a distintos tipos de datos y problemas:

- Perceptrón multicapa (MLP): red clásica de capas totalmente conectadas, usada en tareas generales de predicción o clasificación.
- Redes convolucionales (CNN): ideales para el procesamiento de imágenes y visión computacional, al detectar bordes, formas y texturas.
- Redes recurrentes (RNN): procesan datos secuenciales como texto, audio o series temporales.
- Redes autoencoder: utilizadas para reducción de dimensionalidad y detección de anomalías.
- Redes generativas (GAN): capaces de crear datos nuevos, como imágenes o sonidos sintéticos.

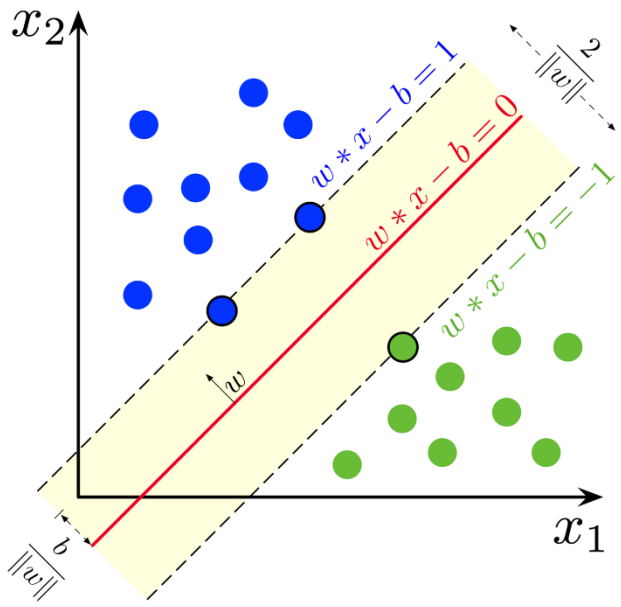
2.4.4 Support Vector Machines (SVM)

Las máquinas de vectores de soporte (Support Vector Machines, SVM) son algoritmos de aprendizaje supervisado utilizados principalmente para clasificación, detección de patrones y regresión dentro del campo del machine learning. Su objetivo fundamental es encontrar una frontera óptima que separe los datos en distintas clases con la mayor precisión posible. En términos simples, las SVM intentan trazar una línea o superficie (llamada hiperplano) que divida los datos en categorías diferentes (por ejemplo, imágenes con tumores y sin tumores). Lo innovador del método es que no solo busca separar las clases, sino maximizar el margen entre ellas, es decir, la distancia más corta entre el hiperplano y los datos más cercanos de cada clase. Estos puntos críticos que definen el margen se denominan vectores de soporte, y son los elementos más importantes del modelo, ya que determinan la posición exacta de la frontera de decisión.

La Figura 4 muestra la representación gráfica de una máquina de vectores de soporte, en la que se ilustran el hiperplano de separación, el margen máximo y los vectores de soporte que determinan la frontera de decisión.

Figura 4

Hiperplano óptimo y margen en una máquina de vectores de soporte



Nota. Tomado de *Maximum-margin hyperplane and margin for an SVM trained on two classes*, por Larhmam, 2018, https://commons.wikimedia.org/wiki/File:SVM_margin.png

2.5 Procesamiento de imágenes

El procesamiento de imágenes es una rama de la computación y la ingeniería que se encarga de adquirir, analizar, transformar y mejorar imágenes digitales con el fin de extraer información útil o facilitar su interpretación. En otras palabras, consiste en aplicar métodos matemáticos y algoritmos computacionales para modificar o analizar imágenes de manera automática o semiautomática.

En el ámbito médico, este campo permite mejorar la calidad de las imágenes obtenidas por resonancia magnética (MRI), tomografía (CT), rayos X o ultrasonido, y extraer rasgos relevantes como el tamaño, forma, textura o color de una lesión, lo cual es esencial para el diagnóstico asistido por computadora.

El procesamiento de imágenes digitales suele dividirse en varias etapas secuenciales:

1. Adquisición de la imagen: Es el proceso mediante el cual se captura la imagen desde un dispositivo (como una cámara o escáner médico). En esta etapa se obtiene la imagen digital en forma de matriz de píxeles.
2. Preprocesamiento: Tiene como objetivo mejorar la calidad visual de la imagen y preparar los datos para su análisis. Incluye técnicas como la reducción de ruido,

aumento del contraste, corrección de brillo o filtrado espacial. Ejemplo: en imágenes MRI, se eliminan artefactos o distorsiones generadas por el movimiento del paciente.

3. Segmentación: Es una de las etapas más críticas. Consiste en dividir la imagen en regiones homogéneas o extraer objetos de interés (por ejemplo, un tumor cerebral). Se utilizan métodos como thresholding, region growing, watershed o algoritmos basados en aprendizaje automático.
4. Extracción de características: A partir de las regiones segmentadas, se calculan atributos numéricos que describen las propiedades del objeto, como textura, forma, color, bordes o intensidad. Estas características luego se usan como entrada para modelos de clasificación, como SVM o redes neuronales.
5. Clasificación o reconocimiento: En esta fase, los algoritmos de machine learning o deep learning clasifican las imágenes según los patrones detectados. Por ejemplo, pueden distinguir entre tejido sano y tumoral o entre diferentes tipos de tumores cerebrales.
6. Postprocesamiento y visualización: Finalmente, los resultados se refinan y presentan al usuario mediante gráficos, máscaras superpuestas o reconstrucciones tridimensionales que facilitan la interpretación médica.

2.5.1 Imágenes digitales

Una imagen digital es una representación numérica de una imagen visual, almacenada y procesada por un computador en forma de una matriz de valores que describe la intensidad o el color de cada punto o elemento de la imagen, llamado píxel (del inglés picture element). Cada píxel contiene información sobre el brillo, color o nivel de gris, dependiendo del tipo de imagen, y su conjunto completo conforma la imagen digital.

En esencia, una imagen digital convierte una escena del mundo real, como una fotografía, una radiografía o una resonancia magnética, en datos discretos que pueden ser almacenados, procesados y analizados por un sistema computacional. Esto permite aplicar técnicas de procesamiento digital, compresión, mejora o análisis automatizado de patrones.

Características fundamentales

- Resolución: cantidad de píxeles que componen la imagen. A mayor resolución, más detalle visual.

- Profundidad de bits: número de bits usados para representar el valor de cada píxel; determina el rango de colores o niveles de gris posibles.
- Tamaño de imagen: producto del número de píxeles en ancho y alto.
- Frecuencia espacial: describe cómo cambian los valores de intensidad o color entre píxeles adyacentes, importante para análisis de textura y bordes.

2.5.2 Preprocesamiento de imágenes MRI

El preprocesamiento de imágenes de resonancia magnética (MRI) es una etapa crítica del flujo de trabajo en análisis y minería de imágenes médicas. Consiste en aplicar transformaciones, correcciones y normalizaciones sobre las imágenes originales con el objetivo de mejorar su calidad, reducir ruido y estandarizar la información antes de aplicar algoritmos de segmentación, extracción de características o clasificación. En otras palabras, el preprocesamiento busca que todas las imágenes, aunque provengan de diferentes pacientes, escáneres o configuraciones, estén en un formato homogéneo y limpio, garantizando que los modelos de machine learning o deep learning trabajen con datos comparables y libres de distorsiones.

2.5.2.1 Conversión y normalización de formato

El primer paso en el preprocesamiento consiste en la conversión y normalización del formato de las imágenes. Las imágenes MRI suelen almacenarse en formato DICOM (Digital Imaging and Communications in Medicine), que es el estándar médico utilizado por los equipos de adquisición. Sin embargo, para fines de análisis computacional, este formato se convierte frecuentemente a NIfTI (.nii) o a estructuras matriciales compatibles con software como Python, MATLAB o 3D Slicer. Una vez convertidas, se realiza la normalización de intensidades, ajustando los niveles de brillo y contraste entre estudios para minimizar las variaciones producidas por diferentes escáneres o configuraciones de adquisición. Este proceso asegura que los valores de los píxeles sean comparables entre distintos pacientes o sesiones de estudio.

2.5.2.2 Filtrado y reducción de ruido

Las imágenes obtenidas mediante resonancia magnética suelen contener ruido aleatorio causado por interferencias electromagnéticas, movimiento del paciente o limitaciones técnicas del equipo. Para mejorar la calidad de la imagen y preservar los detalles anatómicos, se aplican filtros digitales como el Gaussiano, que suaviza pequeñas variaciones de intensidad, el de

mediana, que elimina picos de ruido sin distorsionar los bordes, o el anisotrópico, que reduce el ruido manteniendo las estructuras relevantes del cerebro. Este paso incrementa la relación señal-ruido (SNR) y optimiza la precisión de las etapas posteriores de segmentación y clasificación.

2.5.2.3 Registro o alineación espacial (image registration)

El registro de imágenes consiste en alinear espacialmente distintas imágenes del mismo paciente o de pacientes diferentes en un sistema de coordenadas común. Este procedimiento es indispensable cuando se combinan múltiples secuencias de MRI (como T1, T2 o FLAIR) o cuando se comparan estudios longitudinales. Se utilizan métodos de registro rígido, afín o no lineal, que aplican rotaciones, traslaciones o deformaciones para lograr una coincidencia anatómica precisa. Asimismo, la normalización espacial permite ajustar las imágenes a un atlas cerebral estándar (como MNI o Talairach), garantizando que las regiones equivalentes del cerebro se ubiquen en posiciones similares para análisis comparativos confiables.

2.5.2.4 Eliminación de artefactos y corrección de inhomogeneidad (bias field correction)

Durante la adquisición de imágenes MRI pueden generarse artefactos visuales debido a movimientos involuntarios del paciente, irregularidades del campo magnético o variaciones en la intensidad de señal. Para corregir estos problemas se aplican algoritmos como N4ITK, que compensa las inhomogeneidades de intensidad producidas por campos magnéticos no uniformes. Esta etapa, conocida como bias field correction, garantiza que los tejidos con propiedades similares presenten intensidades homogéneas en toda la imagen, evitando errores en la segmentación o clasificación del tejido cerebral.

2.5.2.5 Extracción del cerebro (skull stripping)

Una etapa fundamental en el análisis de imágenes cerebrales es la eliminación de estructuras no cerebrales, como el cráneo, la piel o los ojos, mediante un proceso conocido como skull stripping. Este procedimiento permite centrar el análisis exclusivamente en el tejido cerebral y reducir el riesgo de falsos positivos en la detección de tumores u otras lesiones. Existen diversas herramientas automatizadas para este fin, como BET (Brain Extraction Tool) del software FSL, 3dSkullStrip de AFNI o ROBEX (Robust Brain Extraction). Al eliminar los elementos externos, se obtiene una representación más precisa del cerebro y se facilita la posterior segmentación del tumor.

2.5.2.6 Estandarización y normalización de intensidad

Dado que los valores de intensidad en las imágenes MRI no son absolutos, sino relativos a las características del escáner y los parámetros de adquisición, es necesario aplicar una normalización de intensidad. Este procedimiento ajusta los valores de los píxeles a un rango estándar, generalmente utilizando métodos estadísticos como la normalización Z-score, que transforma los datos para tener media cero y desviación estándar uno. De esta forma, se asegura que los algoritmos de aprendizaje automático no se vean afectados por diferencias de escala o contraste entre pacientes, mejorando la generalización del modelo.

2.5.2.7 Segmentación preliminar o enmascaramiento

Finalmente, algunas metodologías incluyen una segmentación preliminar o creación de máscaras que delimitan regiones de interés, como posibles zonas tumorales. Este proceso puede realizarse mediante técnicas clásicas como thresholding o region growing, o mediante modelos de machine learning como SVM o k-means clustering. La segmentación inicial reduce el espacio de búsqueda y optimiza el rendimiento computacional de las fases posteriores, permitiendo que los algoritmos se concentren únicamente en las áreas relevantes del cerebro.

2.5.3 Extracción de información

La extracción de información en imágenes de resonancia magnética (MRI) es una etapa fundamental del análisis computacional y del diagnóstico asistido por inteligencia artificial. Consiste en transformar los datos visuales contenidos en la imagen médica en representaciones numéricas o cuantitativas que describen sus propiedades más relevantes, como la textura, la forma, el tamaño o la intensidad del tejido cerebral. El objetivo de este proceso es convertir la información visual en datos matemáticos interpretables por algoritmos de clasificación o predicción, como Support Vector Machines (SVM) o redes neuronales (Parekh & Jacobs, 2019).

Las imágenes MRI son esencialmente matrices tridimensionales de valores de intensidad que representan las propiedades físicas del tejido, como la densidad de protones o los tiempos de relajación T1 y T2. Sin embargo, para los algoritmos de machine learning, esos valores por sí solos carecen de significado contextual.

Por ello, se realiza la extracción de características (features), un proceso que resume la información más representativa de cada región de interés (ROI, Region of Interest), permitiendo a los modelos aprender patrones asociados a la presencia, tipo o gravedad de un tumor.

Existen diversos tipos de características que pueden obtenerse de una imagen MRI, agrupadas principalmente en cuatro categorías:

a. Características de intensidad: Son las más básicas y reflejan la distribución de los valores de los píxeles o vóxeles (volúmenes tridimensionales). Se calculan métricas estadísticas como la media, desviación estándar, curtosis o asimetría de la intensidad dentro del área del tumor. Estas medidas permiten cuantificar diferencias en el brillo o densidad del tejido, las cuales pueden asociarse con tipos específicos de tumores.

b. Características de textura: Analizan la variación espacial de las intensidades y describen cómo se distribuyen los tonos dentro de una región. Se utilizan métodos como:

- Matriz de co-ocurrencia de niveles de gris (GLCM)
- Transformada de wavelet
- Análisis de patrones locales binarios (LBP)

Estas características son muy útiles para identificar heterogeneidad dentro del tumor, un indicador clínico de malignidad o necrosis.

c. Características de forma y morfología: Describen la geometría y estructura del tumor. Incluyen medidas como el área, perímetro, circularidad, compacidad, volumen o relación entre ejes mayores y menores. Estas características permiten distinguir tumores invasivos o irregulares de otros más localizados o esféricos.

d. Características de frecuencia y dominio espectral: Se obtienen mediante transformaciones matemáticas (Fourier, Gabor o Wavelet) que convierten la imagen del dominio espacial al dominio de la frecuencia. Este análisis permite capturar patrones repetitivos o texturas complejas no perceptibles visualmente.

2.5.4 Procesos complementarios: selección y reducción de características

Una vez extraídas, las imágenes pueden contener miles de características, muchas de las cuales son redundantes o irrelevantes. Por eso, se aplica una fase adicional de selección o reducción de dimensionalidad, mediante técnicas como:

- Análisis de Componentes Principales (PCA)
- Linear Discriminant Analysis (LDA)

- Recursive Feature Elimination (RFE)

Estas técnicas eliminan ruido y conservan solo las variables más representativas, reduciendo la complejidad computacional y mejorando la precisión del modelo de clasificación.

En relación con las herramientas y software utilizados para realizar la extracción de características, podemos mencionar las siguientes bibliotecas y plataformas especializadas. Estas herramientas automatizan el proceso de cálculo y aseguran la reproducibilidad en entornos clínicos y de investigación.

- PyRadiomics: biblioteca en Python ampliamente usada para obtener descriptores radiológicos cuantitativos.
- MATLAB Image Processing Toolbox: para cálculos morfológicos y texturales.
- ITK-SNAP o 3D Slicer: herramientas gráficas para segmentación y extracción de ROI.

La extracción de información permite vincular características radiológicas con parámetros clínicos y genéticos, dando origen al campo de la radiómica, que busca correlacionar patrones de imagen con la biología tumoral.

En el caso de los tumores cerebrales, las características extraídas de imágenes MRI han demostrado utilidad en:

- Distinguir entre tumores benignos y malignos.
- Determinar el grado de agresividad del glioma.
- Evaluar la respuesta al tratamiento.
- Estimar la supervivencia del paciente.

De esta forma, la extracción de características no solo mejora la precisión de los modelos predictivos, sino que también aporta información clínica complementaria al diagnóstico humano.

2.6 Tecnologías de Big Data

Las tecnologías de Big Data comprenden el conjunto de herramientas, arquitecturas, lenguajes y plataformas informáticas diseñadas para almacenar, procesar y analizar grandes volúmenes de información que superan la capacidad de los sistemas tradicionales de bases de datos. Estas tecnologías surgieron para enfrentar los desafíos derivados del crecimiento

exponencial de los datos generados por sensores, redes sociales, dispositivos médicos y sistemas empresariales. Según Hashem et al. (2015), el Big Data se caracteriza por cinco dimensiones principales conocidas como las “5 V”: volumen, velocidad, variedad, veracidad y valor. El volumen hace referencia a la enorme cantidad de datos generados continuamente; la velocidad, a la rapidez con la que deben procesarse; la variedad, a los múltiples formatos en los que se presentan (estructurados, semiestructurados y no estructurados); la veracidad, a la fiabilidad de la información; y el valor, al conocimiento útil que puede extraerse a partir de su análisis.

El ecosistema de Big Data se compone de varios elementos que actúan de forma integrada. En primer lugar, la infraestructura y el almacenamiento son fundamentales, ya que permiten gestionar datos distribuidos en diferentes servidores. Tecnologías como el Hadoop Distributed File System (HDFS), Apache HBase o Cassandra posibilitan el almacenamiento y la recuperación eficiente de grandes volúmenes de información con alta disponibilidad y tolerancia a fallos. En entornos modernos, el almacenamiento también se realiza mediante soluciones en la nube, como Amazon S3, Google Cloud Storage o Azure Data Lake, que ofrecen escalabilidad y acceso remoto a los datos.

En segundo lugar, el procesamiento y análisis de datos se lleva a cabo mediante frameworks que permiten ejecutar operaciones en paralelo, aprovechando la potencia de múltiples nodos conectados. Entre las herramientas más destacadas se encuentran MapReduce, un modelo que distribuye las tareas de análisis en distintos fragmentos de datos, y Apache Spark, que permite un procesamiento en memoria mucho más rápido. Para el manejo de flujos de datos en tiempo real, se utilizan tecnologías como Apache Flink y Apache Storm, que son especialmente útiles en aplicaciones médicas donde los datos deben procesarse instantáneamente, como en monitoreo de pacientes o análisis de imágenes en vivo.

Una vez procesados, los datos se analizan utilizando herramientas avanzadas de aprendizaje automático (machine learning) y análisis visual. Librerías como TensorFlow, PyTorch y Scikit-learn permiten desarrollar modelos predictivos capaces de descubrir patrones complejos, mientras que plataformas como Power BI, Tableau o Apache Superset facilitan la visualización interactiva de los resultados. Lenguajes como Python, R y SQL son esenciales en esta fase por su flexibilidad y capacidad para integrar distintas fuentes de datos.

En el ámbito de la salud, las tecnologías de Big Data han adquirido un papel fundamental, ya que permiten integrar, almacenar y analizar enormes volúmenes de datos médicos provenientes de diferentes fuentes, como historiales clínicos electrónicos, dispositivos de monitoreo, laboratorios o sistemas de imágenes médicas (MRI, CT, PET). En el caso específico de la resonancia magnética, estas tecnologías facilitan el procesamiento de grandes colecciones de imágenes de alta resolución, posibilitando la aplicación de modelos de machine learning y deep learning para la detección automática de tumores cerebrales y otras patologías. Asimismo, los sistemas de Big Data permiten crear data lakes médicos, en los que se combinan datos clínicos, genéticos e imagenológicos, impulsando la investigación biomédica y la medicina personalizada.

En este contexto, las tecnologías de Big Data representan la base tecnológica de la inteligencia artificial aplicada a la medicina moderna, ya que permiten transformar grandes volúmenes de datos heterogéneos en conocimiento clínico útil. Como señala Kitchin (2021), el Big Data no solo implica el manejo de grandes cantidades de información, sino la capacidad de extraer valor de esos datos mediante técnicas de análisis avanzado, lo cual resulta esencial para el desarrollo de sistemas predictivos, diagnósticos automatizados y políticas de salud basadas en evidencia.

2.6.1 Apache Hadoop

Apache Hadoop es un marco de software de código abierto diseñado para el almacenamiento y procesamiento distribuido de grandes volúmenes de datos en clústeres de servidores. Su principal objetivo es permitir que organizaciones y sistemas manejen información masiva de manera eficiente, escalable y tolerante a fallos, incluso utilizando hardware común o de bajo costo. Hadoop surgió como una respuesta a las limitaciones de los sistemas tradicionales de bases de datos relacionales, que no podían manejar el crecimiento exponencial de datos generado en la era del Big Data.

La arquitectura de Hadoop se basa en dos componentes principales: el Hadoop Distributed File System (HDFS) y el MapReduce. El HDFS se encarga del almacenamiento distribuido de los datos, dividiendo cada archivo en bloques que se replican y distribuyen entre los nodos del clúster. Esta estructura permite garantizar la disponibilidad y resistencia a fallos, ya que, si un nodo falla, los datos pueden recuperarse desde las réplicas almacenadas en otros nodos. Por su parte, MapReduce es el modelo de procesamiento paralelo que permite dividir una

tarea compleja en subprocesos más pequeños, los cuales se ejecutan simultáneamente en distintos nodos del sistema. Este enfoque distribuye la carga de trabajo y acelera el procesamiento masivo de información.

Además de sus componentes básicos, el ecosistema de Hadoop se ha ampliado para incluir diversas herramientas que complementan su funcionamiento. Apache Hive permite realizar consultas SQL sobre datos almacenados en HDFS; Apache Pig facilita la creación de scripts de análisis de datos; Apache HBase funciona como una base de datos NoSQL distribuida; y YARN (Yet Another Resource Negotiator) actúa como el gestor de recursos del clúster, coordinando la ejecución de las tareas. En conjunto, estas herramientas hacen que Hadoop sea una plataforma integral para el manejo de datos estructurados, semiestructurados y no estructurados.

Una de las principales ventajas de Hadoop es su escalabilidad horizontal, lo que significa que se pueden agregar nuevos nodos al sistema sin interrumpir su funcionamiento. Además, su tolerancia a fallos permite continuar el procesamiento incluso si algunos componentes fallan. Estas características han convertido a Hadoop en una tecnología clave para organizaciones que manejan grandes volúmenes de información, como empresas financieras, de telecomunicaciones, científicas y de salud. En el ámbito médico, Hadoop se utiliza para almacenar y procesar imágenes de resonancia magnética (MRI), historiales clínicos electrónicos y datos genómicos, permitiendo la aplicación de modelos de machine learning y análisis predictivo a gran escala.

En síntesis, Apache Hadoop constituye una de las tecnologías fundamentales del ecosistema Big Data, al ofrecer un entorno robusto, distribuido y de código abierto para el procesamiento de datos masivos. Su adopción ha permitido a las instituciones transformar grandes volúmenes de datos heterogéneos en conocimiento útil, impulsando el desarrollo de aplicaciones analíticas y sistemas de inteligencia artificial en múltiples sectores.

2.6.2 MapReduce

MapReduce es un modelo de programación y un sistema de procesamiento distribuido diseñado para manejar grandes volúmenes de datos en entornos de cómputo paralelo. Fue desarrollado originalmente por Google a principios de la década de 2000 y posteriormente implementado en el marco Apache Hadoop, donde se convirtió en uno de sus componentes fundamentales. El objetivo principal de MapReduce es dividir tareas complejas de análisis o

transformación de datos en partes más pequeñas y ejecutarlas simultáneamente en múltiples nodos del clúster, logrando así un procesamiento masivo eficiente y escalable.

El funcionamiento de MapReduce se basa en dos fases principales: Map y Reduce.

En la fase Map, los datos de entrada se dividen en fragmentos o bloques que se procesan en paralelo por múltiples nodos. Cada nodo aplica una función de mapeo que transforma los datos en pares clave-valor (key-value pairs). Por ejemplo, al analizar un conjunto de documentos, la función Map puede generar una lista con cada palabra (clave) y su frecuencia (valor).

Luego, en la fase Reduce, los pares generados por los distintos nodos se agrupan por clave y se combinan mediante una función reductora que consolida los resultados parciales. Continuando con el ejemplo anterior, la función Reduce sumaría todas las ocurrencias de una misma palabra para obtener un conteo total. Este enfoque permite procesar de manera eficiente terabytes o petabytes de datos distribuidos en miles de máquinas.

Una de las mayores ventajas de MapReduce es su capacidad de tolerar fallos y su escalabilidad horizontal. Si un nodo falla durante el procesamiento, el sistema reasigna automáticamente las tareas a otros nodos disponibles sin interrumpir la ejecución. Además, su diseño se basa en el principio de “mover el cómputo hacia los datos” en lugar de transferir grandes volúmenes de información a través de la red, lo que mejora considerablemente el rendimiento y reduce los tiempos de ejecución.

En el ecosistema de Apache Hadoop, MapReduce trabaja en conjunto con el Hadoop Distributed File System (HDFS): el HDFS gestiona el almacenamiento distribuido de los datos, mientras que MapReduce se encarga del procesamiento paralelo. Esta combinación permite realizar tareas de minería de datos, indexación de grandes colecciones de documentos, análisis de registros (log analysis), procesamiento de imágenes médicas y entrenamientos de modelos de machine learning a gran escala.

En el ámbito de la salud, MapReduce ha demostrado ser una herramienta eficaz para procesar grandes volúmenes de información biomédica, como imágenes de resonancia magnética (MRI), datos genómicos o registros clínicos. Su capacidad para dividir y ejecutar operaciones en paralelo permite analizar miles de estudios de manera simultánea, reduciendo significativamente los tiempos de cómputo y facilitando la detección de patrones o anomalías relevantes en el diagnóstico médico asistido por inteligencia artificial.

2.6.3 Apache Spark

Apache Spark es un marco de procesamiento distribuido de código abierto diseñado para analizar grandes volúmenes de datos de manera rápida, eficiente y en memoria. Fue desarrollado originalmente en la Universidad de California, Berkeley, en el laboratorio AMPLab, y se publicó como proyecto de código abierto en 2010. Posteriormente, se integró dentro del ecosistema de Apache Software Foundation, convirtiéndose en una de las tecnologías más utilizadas en el ámbito del Big Data.

Spark nació como una evolución de MapReduce, el modelo tradicional de Hadoop, superando sus limitaciones de rendimiento y flexibilidad. Mientras que MapReduce almacena los resultados intermedios en disco después de cada operación, Apache Spark permite el procesamiento en memoria (in-memory computing), lo que reduce significativamente los tiempos de ejecución. Gracias a este enfoque, Spark puede ejecutar tareas hasta 100 veces más rápido que MapReduce en cargas de trabajo iterativas o interactivas, como el entrenamiento de modelos de machine learning o el análisis de grandes conjuntos de imágenes.

En el ámbito de la salud, Apache Spark se utiliza para procesar grandes volúmenes de datos médicos, como imágenes de resonancia magnética (MRI), registros electrónicos de pacientes o datos genómicos. Gracias a su motor de procesamiento distribuido, Spark permite realizar tareas de preprocesamiento, segmentación, extracción de características y clasificación de imágenes médicas en menos tiempo que los sistemas tradicionales.

Entre las principales ventajas de Apache Spark destacan su alta velocidad de procesamiento, su capacidad para manejar distintos tipos de datos (estructurados, semiestructurados y no estructurados), su flexibilidad de integración con otras herramientas de Big Data y su tolerancia a fallos. También ofrece una API unificada para múltiples lenguajes de programación (Python, Java, Scala y R), lo que facilita su adopción en entornos multidisciplinarios.

Estas características hacen de Spark una plataforma ideal para análisis de datos complejos en entornos distribuidos, especialmente en proyectos de inteligencia artificial aplicada a la medicina y en sistemas de diagnóstico automatizado por imágenes.

2.7 Tecnologías cloud

Las tecnologías Cloud, también conocidas como tecnologías de computación en la nube (Cloud Computing), hacen referencia a un modelo de prestación de servicios informáticos que permite acceder, almacenar y procesar datos y aplicaciones a través de Internet, en lugar de depender exclusivamente de servidores o infraestructuras locales. Según Mell y Grance (2011), la computación en la nube es un modelo que permite el acceso bajo demanda y de manera ubicua a un conjunto compartido de recursos informáticos configurables, como servidores, almacenamiento, redes, bases de datos y software, que pueden ser rápidamente aprovisionados y liberados con un mínimo esfuerzo de gestión o interacción con el proveedor del servicio.

En otras palabras, las tecnologías Cloud proporcionan una infraestructura flexible y escalable que permite a empresas e instituciones reducir costos, aumentar la eficiencia operativa y adaptarse dinámicamente a las necesidades de procesamiento. En lugar de comprar y mantener hardware físico, los usuarios pueden alquilar recursos informáticos según el modelo “pago por uso” (pay-as-you-go), accediendo a capacidades prácticamente ilimitadas de almacenamiento y cálculo a través de Internet.

En el ámbito médico, las tecnologías Cloud han transformado la forma en que se almacenan y procesan los datos clínicos e imagenológicos. Los hospitales y centros de investigación pueden subir grandes volúmenes de imágenes de resonancia magnética (MRI), tomografías o estudios genómicos a la nube, donde se procesan utilizando algoritmos de machine learning o deep learning distribuidos. Esto elimina las limitaciones del hardware local y permite analizar miles de imágenes en paralelo.

Por ejemplo, los servicios de computación en la nube como Amazon Web Services (AWS HealthLake), Microsoft Azure Health Data Services y Google Cloud Healthcare API ofrecen plataformas especializadas para el manejo seguro de datos médicos, cumpliendo estándares internacionales de protección como HIPAA y GDPR. Estas soluciones posibilitan el desarrollo de sistemas de diagnóstico automatizado, almacenamiento de imágenes DICOM, y análisis predictivo para la detección temprana de enfermedades como el cáncer cerebral.

Además, la nube permite la colaboración remota entre profesionales de la salud y centros de investigación, ya que los datos pueden ser compartidos y analizados en tiempo real desde distintos lugares del mundo. Este enfoque ha impulsado la llamada “medicina digital”, donde la

inteligencia artificial, la analítica avanzada y la computación en la nube trabajan conjuntamente para mejorar la precisión diagnóstica y la eficiencia clínica.

2.7.1 Microsoft Azure

Microsoft Azure, también conocida como la nube de Azure, es una plataforma integral de computación en la nube desarrollada por Microsoft que proporciona una amplia gama de servicios de infraestructura, plataformas y software a través de Internet. Su propósito principal es ofrecer a las organizaciones los recursos necesarios para almacenar, procesar y analizar datos, así como desarrollar, implementar y administrar aplicaciones sin depender de hardware local. Azure es una de las tres principales plataformas de nube pública a nivel mundial, junto con Amazon Web Services (AWS) y Google Cloud Platform (GCP), y ha sido ampliamente adoptada en sectores empresariales, educativos, financieros y de salud por su flexibilidad, escalabilidad y seguridad.

Entre las principales ventajas de Microsoft Azure destacan su alta escalabilidad, que permite aumentar o disminuir los recursos según la demanda; su modelo de pago por uso, que optimiza los costos operativos; su seguridad avanzada, con cifrado en tránsito y en reposo; y su compatibilidad con múltiples lenguajes y frameworks, como Python, R, Java, .NET y TensorFlow.

Adicionalmente, su capacidad para integrarse con herramientas analíticas como Power BI y con entornos de Big Data como Apache Hadoop y Spark la convierte en una plataforma ideal para proyectos de análisis de imágenes médicas, minería de datos clínicos e inteligencia artificial aplicada a la salud. En el campo de la salud y la investigación biomédica, Microsoft Azure ha demostrado ser una herramienta poderosa para el manejo y análisis de grandes volúmenes de datos clínicos e imagenológicos.

Los servicios especializados como Azure Health Data Services, Azure AI Health Bot y Azure Cognitive Services for Vision permiten almacenar imágenes médicas DICOM, aplicar modelos de machine learning para la detección de tumores cerebrales y ejecutar algoritmos de deep learning para el análisis de resonancias magnéticas (MRI). Asimismo, el servicio Azure Data Lake posibilita la construcción de repositorios centralizados de datos médicos que integran información estructurada (historiales clínicos, registros de laboratorio) y no estructurada (imágenes, notas médicas, señales biomédicas). Estos repositorios pueden analizarse mediante Azure Synapse Analytics, combinando análisis masivo de datos con inteligencia artificial para apoyar la toma de decisiones clínicas.

Es importante mencionar también que Azure cumple con los principales estándares internacionales de privacidad y seguridad médica, como HIPAA, ISO/IEC 27001 y GDPR, garantizando la confidencialidad de la información del paciente. Esto ha permitido que hospitales, universidades y centros de investigación utilicen Azure para implementar sistemas de diagnóstico automatizado, plataformas de telemedicina y entornos de análisis colaborativo entre instituciones médicas.

2.8 Arquitectura de medallón

La arquitectura de medallón, también conocida como Medallion Data Architecture, es un modelo de diseño propuesto por Databricks que organiza los datos dentro de un data lake o lakehouse en capas o niveles secuenciales denominados Bronze (Bronce), Silver (Plata) y Gold (Oro). Su objetivo principal es mejorar la calidad, organización y trazabilidad de los datos a medida que avanzan desde su forma bruta hasta convertirse en información lista para el análisis o la inteligencia artificial. En otras palabras, esta arquitectura estructura el flujo de datos en un proceso gradual y controlado, donde cada capa tiene un propósito específico y los datos se enriquecen, limpian y transforman progresivamente. Este enfoque facilita la gobernanza, reproducibilidad y escalabilidad de las soluciones analíticas en entornos de Big Data y nube.

La arquitectura se compone de tres niveles principales:

1. Capa Bronze (Bronce): Datos crudos o de origen): En esta primera capa se almacenan los datos tal como provienen de las fuentes originales, sin transformaciones ni filtrados. Pueden incluir archivos CSV, JSON, imágenes, datos de sensores, logs de sistemas, o incluso información médica en formato DICOM en el caso de imágenes MRI. Su propósito es conservar la integridad y trazabilidad de los datos originales, funcionando como una zona de aterrizaje o staging area. Ejemplo: un conjunto de imágenes MRI en bruto provenientes de diferentes hospitales, junto con sus metadatos clínicos, se almacenan directamente en esta capa para su posterior procesamiento.

2. Capa Silver (Plata): Datos refinados y estandarizados): En esta capa los datos pasan por procesos de limpieza, validación y normalización. Aquí se eliminan duplicados, se corrigen errores, se homogenizan formatos y se aplican las primeras reglas de negocio. El objetivo es producir un conjunto de datos estructurado, confiable y coherente, que sirva como base para análisis posteriores

o integración con otros sistemas. Por ejemplo, en un proyecto de imágenes médicas, en la capa Silver se podrían estandarizar las dimensiones de las imágenes MRI, eliminar registros corruptos y asociar correctamente los metadatos de pacientes y estudios.

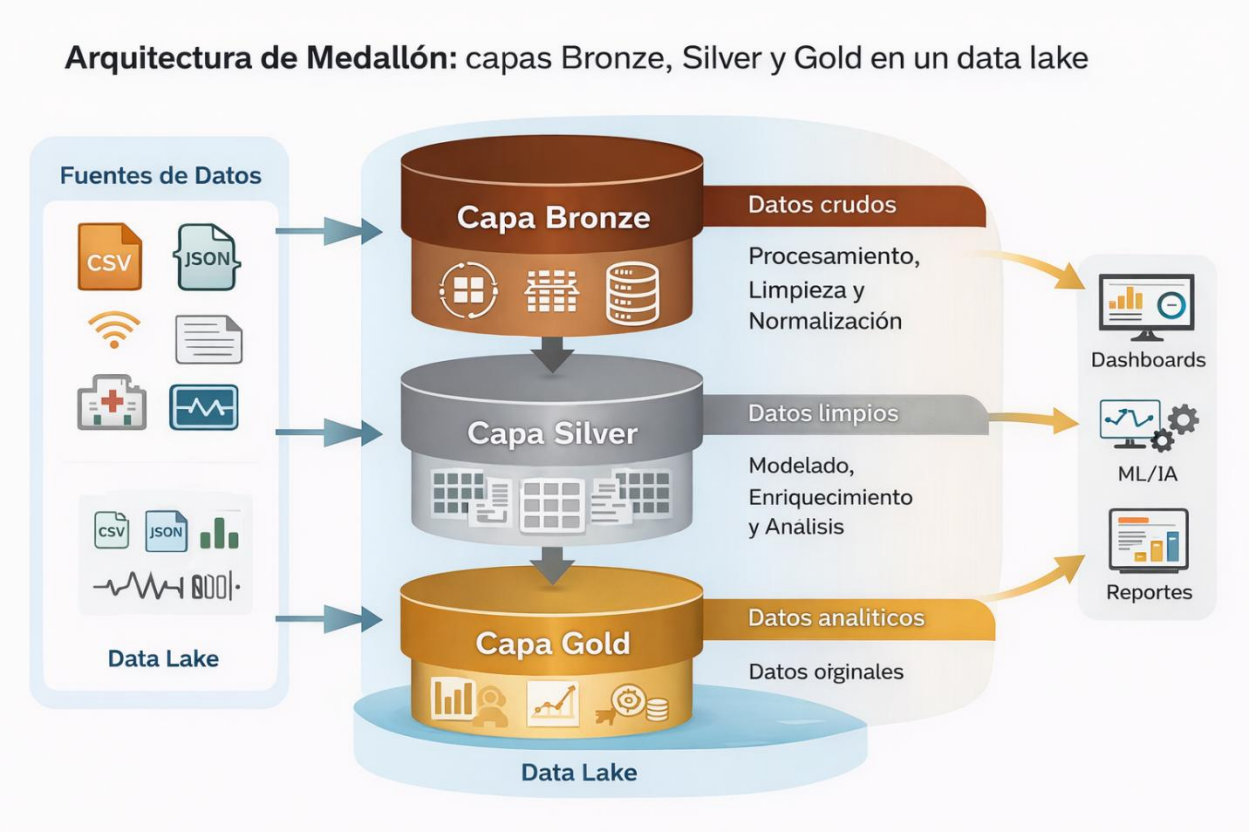
3. Capa Gold (Oro): Datos listos para análisis y machine learning): La capa Gold contiene datos totalmente depurados, enriquecidos y modelados. En esta etapa, los datos se transforman en información analítica lista para alimentar dashboards, modelos predictivos o reportes ejecutivos. Se aplican métricas, agregaciones y cálculos de valor agregado que permiten extraer conocimiento útil para la toma de decisiones.

Siguiendo el ejemplo médico, en la capa Gold se podrían almacenar indicadores derivados de imágenes MRI procesadas (por ejemplo, tamaño de tumor, textura o probabilidad de malignidad), listos para su análisis mediante algoritmos de machine learning o visualización en Power BI.

En la Figura 5 se muestran los tres niveles principales de la arquitectura de medallón y el flujo de los datos a través de esta.

Figura 5

Arquitectura de Medallón



Nota. Elaboración propia.

3. Marco metodológico

El marco metodológico constituye la estructura operativa de la investigación, ya que define el enfoque, los métodos, el diseño, la población de estudio y las técnicas de análisis que permitirán alcanzar los objetivos planteados inicialmente. En el caso de este proyecto, la metodología combina los principios de la investigación aplicada con prácticas experimentales en entornos computacionales, integrando la ingeniería de datos, la inteligencia artificial y la salud digital.

3.1 Tipo de investigación

El tipo de investigación utilizado para este proyecto es aplicado, ya que está orientada la solución de un problema real mediante el uso de fundamentos teóricos, tecnológicos y metodológicos provenientes de la inteligencia artificial, la ingeniería de datos y la computación en la nube. Este tipo de investigación tiene como propósito transformar el conocimiento científico ya existente en una herramienta práctica y verificable, capaz de mejorar los procesos diagnósticos en el área de la salud, particularmente en el análisis automatizado de imágenes médicas de resonancia magnética (MRI) en el contexto de los centros de salud públicos en Costa Rica.

La investigación aplicada se diferencia de la investigación teórica en que no se limita a generar nuevo conocimiento abstracto, sino que utiliza los principios y descubrimientos ya existentes para diseñar, implementar y evaluar una solución concreta a una necesidad identificada. En este caso, el problema central es la dificultad que enfrentan los centros médicos costarricenses, especialmente dentro de la Caja Costarricense de Seguro Social (CCSS), para procesar grandes volúmenes de estudios de resonancia magnética de forma rápida, precisa y estandarizada. Esta limitación es producto de la sobrecarga de trabajo y la falta de infraestructura tecnológica, y afecta directamente la calidad y oportunidad del diagnóstico oncológico para las personas que utilizan los servicios de salud públicos en Costa Rica.

Este proyecto busca aplicar técnicas de inteligencia artificial y Big Data en un entorno cloud con el fin de crear un sistema de diagnóstico asistido que apoye al personal médico en la clasificación automática de tumores cerebrales, incrementando la eficiencia y la confiabilidad de los resultados.

Los aspectos que se mencionan a continuación funcionan como evidencia de que el tipo de investigación de este proyecto es aplicada:

1. Finalidad práctica: se propone una solución funcional (prototipo) que integre algoritmos de aprendizaje profundo con arquitecturas de datos en la nube, permitiendo una posible adaptación futura a contextos clínicos reales.
2. Integración de teoría y práctica: desde el punto de vista teórico se utilizan fundamentos científicos de machine learning, procesamiento de imágenes y arquitecturas Big Data, y desde el punto de vista práctico, su aplicación está orientada a resolver un problema operativo y medible que es la automatización del análisis de imágenes MRI.
3. Evaluación empírica de resultados: el estudio no está limitado a ser solamente un análisis conceptual, sino que incluye fases experimentales que validan la eficacia técnica del modelo, midiendo métricas de desempeño como accuracy, IoU, precision y F1-score.
4. Contribución social y tecnológica: el resultado esperado no es únicamente un producto tecnológico, sino también una propuesta para contribuir al fortalecimiento del sistema de salud público de Costa Rica, mediante el uso de tecnologías emergentes.

Metodológicamente, esta investigación aplicada combina el enfoque cuantitativo con el diseño experimental técnico. Esto permite medir objetivamente los resultados del sistema desarrollado. A través del uso de datasets públicos de resonancias magnéticas anonimizadas, se implementan y comparan modelos de redes neuronales profundas en un entorno de cómputo cloud, y se evalúa su rendimiento bajo condiciones controladas.

Es importante mencionar también que el carácter aplicado de esta investigación se ve reflejado en la intención de que los resultados del estudio sirvan de base para futuras investigaciones o desarrollos tecnológicos en el ámbito médico nacional. En otras palabras, se busca fomentar la integración de soluciones de inteligencia artificial dentro de los flujos de trabajo hospitalarios actuales. Por esta razón, la investigación no solo busca validar una hipótesis técnica, sino también demostrar la viabilidad de aplicar arquitecturas Big Data en la nube para fines diagnósticos, contribuyendo al avance de la transformación digital en el sector salud costarricense.

En síntesis, este proyecto es una investigación aplicada de tipo tecnológico, sustentada en el diseño, implementación y validación de un prototipo funcional, que utiliza herramientas de inteligencia artificial, procesamiento de imágenes médicas y analítica de datos en la nube para abordar una problemática real de impacto clínico, operativo y social.

3.2 Alcance investigativo

Este proyecto posee un alcance descriptivo, correlacional y experimental de tipo técnico, todos propios de una investigación aplicada.

Desde una perspectiva descriptiva, el estudio busca caracterizar el comportamiento de los modelos de inteligencia artificial al procesar y clasificar imágenes médicas de resonancia magnética (MRI). Para lograr esto, se documentan las etapas de ingesta, preprocesamiento, segmentación y clasificación dentro de un entorno de datos en la nube. Se pretende detallar cómo interactúan los diferentes componentes tecnológicos, como el Big Data, redes neuronales y arquitectura cloud, y cómo contribuyen al mejoramiento de la precisión diagnóstica y la eficiencia operativa.

Por otra parte, el proyecto también tiene un alcance correlacional, ya que examina las relaciones entre las variables técnicas que influyen en el rendimiento del sistema, tales como la calidad del preprocesamiento, los hiperparámetros del modelo, el tamaño del dataset y las métricas de desempeño obtenidas (accuracy, recall, IoU, F1-score). A través de este análisis, se busca identificar los factores que más impactan en la exactitud y generalización de la clasificación automática de tumores cerebrales.

Finalmente, el trabajo también presenta un componente experimental, en el sentido de que se manipulan y evalúan diferentes configuraciones de modelos de aprendizaje profundo, comparando resultados y determinando cuál arquitectura es más eficiente para este tipo de imágenes médicas. Los experimentos se realizan en un entorno controlado de cómputo en la nube, garantizando la reproducibilidad y la trazabilidad de los resultados.

3.3 Enfoque

El presente estudio adopta un enfoque metodológico mixto, que combina los componentes cuantitativo, tecnológico y experimental, fundamentándose en la evaluación objetiva de variables técnicas y en el análisis estadístico de resultados reproducibles.

El enfoque cuantitativo se refleja en el uso de métricas estandarizadas para evaluar la efectividad de los modelos de aprendizaje profundo aplicados al diagnóstico asistido por imágenes. Cada modelo desarrollado se analizará mediante indicadores numéricos que permitirán cuantificar su precisión, sensibilidad y capacidad de generalización. Este enfoque busca garantizar la objetividad y la comparabilidad de los resultados obtenidos.

El componente tecnológico del enfoque deriva de la integración de herramientas de inteligencia artificial, procesamiento distribuido y almacenamiento masivo de datos. El proyecto hace uso de tecnologías modernas como TensorFlow, PyTorch, Apache Spark, Databricks y Azure Machine Learning, que forman parte del ecosistema de Big Data y computación en la nube. Estas herramientas permiten gestionar flujos de datos complejos y ejecutar modelos de redes neuronales en entornos escalables, reproducibles y auditables.

Además, el enfoque experimental dentro de la investigación aplicada se evidencia en el diseño de pruebas y validaciones empíricas, donde se manipulan múltiples variables técnicas (por ejemplo, número de capas, función de activación, tasa de aprendizaje) para observar su impacto en el desempeño del sistema. Este enfoque permite comprobar, a través de evidencia medible, si la solución propuesta cumple con los objetivos de eficiencia, precisión y aplicabilidad clínica.

3.4 Diseño

El diseño metodológico adoptado de tipo experimental y de carácter aplicado, implementado en un entorno controlado de simulación y despliegue en la nube. Este diseño permite validar la hipótesis central del proyecto: que la combinación de arquitecturas Big Data y redes neuronales profundas puede mejorar la precisión y eficiencia del diagnóstico de tumores cerebrales a partir de imágenes MRI.

El desarrollo metodológico se estructura en cinco fases principales, que están interrelacionadas entre sí y alineadas con el enfoque de investigación aplicada:

- Fase 1. Adquisición y preparación de los datos: En esta fase se seleccionan datasets públicos y anonimizados de resonancias magnéticas cerebrales. Las imágenes se almacenan en un entorno cloud y se someten a procesos de preprocesamiento estandarizado, que incluyen:
    - Normalización de intensidades y formatos.
    - Eliminación de artefactos y reducción de ruido.
    - Registro espacial y alineación anatómica.
    - Extracción del cerebro (skull stripping).
    - Segmentación preliminar de regiones tumorales.

  Este conjunto de pasos garantiza la calidad y homogeneidad de los datos que alimentarán el modelo de inteligencia artificial.

- Fase 2. Diseño e implementación del modelo: Se desarrollarán distintas arquitecturas de redes neuronales profundas, priorizando modelos convolucionales tridimensionales (3D CNN) y U-Net para la segmentación y clasificación de tumores. También se evaluarán modelos híbridos que incorporen mecanismos de atención o Transformers visuales. El entrenamiento se llevará a cabo en entornos de cómputo cloud y se utilizarán librerías especializadas en procesamiento de imágenes médicas como MONAI (Medical Open Network for AI). Los parámetros del modelo (número de capas, tasa de aprendizaje, optimizadores) serán ajustados empíricamente para maximizar la precisión diagnóstica.
- Fase 3. Evaluación y validación: En esta etapa se realiza la validación técnica y estadística del modelo. Se utilizarán conjuntos de datos independientes para pruebas (train–validation–test split) y técnicas de validación cruzada k-fold para garantizar la robustez de los resultados. Las métricas de desempeño incluirán:
    - Exactitud (Accuracy)
    - Sensibilidad (Recall)
    - Precisión (Precision)
    - Puntaje F1 (F1-score)
    - Intersección sobre Unión (IoU)

  Además, se analizarán los tiempos de procesamiento y el consumo de recursos en la nube para evaluar la eficiencia operativa del sistema.

- Fase 4. Integración en la arquitectura Big Data: La solución propuesta se implementará en una arquitectura medallón (Bronze–Silver–Gold) que garantice la trazabilidad, el versionamiento y la calidad de los datos:
    - Capa Bronze: almacenamiento de datos crudos (imágenes originales en DICOM).
    - Capa Silver: datos transformados, normalizados y enriquecidos con metadatos.
    - Capa Gold: resultados analíticos y salidas del modelo (clasificaciones y métricas).

  El procesamiento distribuido se gestionará mediante Apache Spark y Databricks, aprovechando los servicios de almacenamiento y cómputo escalable de plataformas como Azure Data Lake o AWS S3.

- Fase 5. Visualización y monitoreo: Finalmente, se desarrollará un dashboard de monitoreo técnico y operativo, que mostrará el rendimiento de los modelos, los tiempos de inferencia y las métricas de calidad. Esta capa de visualización permitirá realizar seguimiento en tiempo real, analizar tendencias y facilitar la interpretación de resultados para usuarios técnicos y clínicos.

En conjunto, este diseño metodológico aplicado garantiza un flujo de trabajo reproducible, escalable y alineado con las buenas prácticas internacionales de desarrollo de soluciones basadas en inteligencia artificial médica.

3.5 Población y muestreo

La población de estudio está conformada por imágenes de resonancia magnética cerebral correspondientes a pacientes diagnosticados con tumores cerebrales primarios o metastásicos, disponibles en Kaggle, un repositorio de libre acceso y uso académico. Dado que el proyecto no involucra la recolección directa de datos de pacientes ni intervenciones clínicas, el uso de bases públicas garantiza el cumplimiento de los principios éticos y la confidencialidad de la información.

El muestreo será no probabilístico por conveniencia. Este tipo de muestreo se basa en la selección intencionada de imágenes MRI que están disponibles o son más accesibles para el investigador, en lugar de utilizar el azar. Se emplea este método ya que no se cuenta con una lista completa de la población y también porque se busca obtener casos representativos y útiles para el estudio. En este contexto, se eligen datasets públicos de imágenes MRI en Kaggle que cumplan con los criterios de disponibilidad, relevancia y calidad técnica necesarios para los objetivos de la investigación, los cuales se describen a continuación:

- Disponibilidad de imágenes MRI en formato DICOM o NIfTI.
- Inclusión de etiquetas o máscaras de segmentación validadas por expertos.
- Representación equilibrada de diferentes tipos y grados de tumores cerebrales.
- Calidad visual suficiente (sin artefactos ni distorsiones excesivas).

Se proyecta trabajar con un conjunto aproximado de 1.000 a 2.500 estudios MRI, distribuidos entre distintas secuencias (T1, T2, FLAIR, T1ce). Estos datos se utilizarán exclusivamente con fines de validación técnica y científica, respetando las políticas de licenciamiento de los repositorios de origen.

Los resultados derivados de este muestreo permitirán evaluar la capacidad del modelo de IA para generalizar sus predicciones y mantener un nivel alto de desempeño frente a

variaciones en la calidad y el origen de los datos. De esta forma, la población y la muestra seleccionadas son representativas del problema técnico que se aborda, y se garantiza la pertinencia y aplicabilidad de los hallazgos.

4. Análisis de la situación

El análisis de la situación tiene como propósito contextualizar el problema desde una perspectiva organizacional, técnica y social, identificando las limitaciones actuales en la gestión de imágenes médicas en Costa Rica, las oportunidades tecnológicas emergentes y las implicaciones del proyecto en el fortalecimiento del sistema de salud.

4.1 Diagnóstico general

El sistema de salud costarricense, liderado por la Caja Costarricense de Seguro Social (CCSS), enfrenta una demanda creciente de estudios de resonancia magnética. Según Vega (2025), el Centro Nacional de Imágenes Médicas realiza 2.970 resonancias mensuales, lo que representa un incremento del 350% con respecto a 2017. Este crecimiento no ha venido acompañado de un aumento proporcional en infraestructura tecnológica ni en la cantidad de radiólogos especializados, lo que genera cuellos de botella en el diagnóstico. Los hallazgos de la Auditoría Interna de la CCSS confirman las consecuencias: retrasos en la interpretación de imágenes y ausencia de sistemas de priorización automatizada que afectan especialmente a los pacientes oncológicos (Segura, 2026).

La ausencia de automatización en los procesos de análisis de imágenes limita la capacidad de respuesta del sistema público, mientras que los hospitales privados han avanzado con equipos más modernos y software especializado. Esta brecha tecnológica crea una desigualdad diagnóstica entre sectores, afectando la equidad en el acceso a servicios de salud.

4.2 Problemática tecnológica y operativa

La principal problemática tecnológica y operativa identificada es la ausencia de plataformas integradas que posibiliten el procesamiento, y análisis automatizado y reproducible de grandes volúmenes de imágenes de resonancia magnética (MRI). Actualmente, los flujos de trabajo dependen en gran medida de la interpretación manual y del almacenamiento en servidores locales, sin un aprovechamiento significativo de las tecnologías Big Data, de los servicios en la nube ni de la inteligencia artificial en general.

Por otra parte, los sistemas PACS no se encuentran interconectados con entornos cloud ni con motores de análisis basados en inteligencia artificial. En la CCSS, estos datos suelen almacenarse en formatos propietarios o heterogéneos, lo que dificulta su estandarización. Otro problema importante es la variabilidad de los equipos MRI y de sus protocolos de adquisición ya que esto genera inconsistencias entre estudios. Además, las limitaciones presupuestarias de la CCSS restringen la adquisición de servidores locales con GPU de alto rendimiento. En conjunto,

estas condiciones generan una alta dependencia del trabajo humano, una escasa trazabilidad y una falta de estandarización que afectan la calidad y eficiencia del diagnóstico médico.

4.3 Factores socioeconómicos

Desde la perspectiva socioeconómica, el sistema público enfrenta presiones de sostenibilidad financiera, un envejecimiento poblacional creciente y un aumento en la incidencia de cáncer. Como indican Calderón, Guzmán y Murphy (2023), la tasa de cáncer en Costa Rica ha aumentado un 72 % en los últimos 30 años, lo que incrementa la demanda de estudios de diagnóstico por imagen.

Otro factor importante es que el acceso a estos servicios sigue concentrado en el Gran Área Metropolitana, generando inequidades regionales y listas de espera prolongadas. La implementación de soluciones basadas en inteligencia artificial en la nube permitiría descentralizar la capacidad diagnóstica, ofreciendo resultados automáticos y reproducibles a centros periféricos, sin requerir infraestructura costosa.

4.4 Análisis de oportunidades

El entorno actual presenta múltiples oportunidades de innovación:

1. Disponibilidad de datos públicos y software open source (como PyTorch o MONAI), que facilitan la experimentación académica.
2. Acceso a servicios cloud gratuitos o de bajo costo para investigación (Azure for Students, AWS Educate, Databricks Community Edition).
3. Avances en redes neuronales profundas tridimensionales, que mejoran la segmentación y clasificación de tumores.
4. Normativas de protección de datos claras que permiten desarrollar soluciones éticas y seguras, particularmente las que están incluidas en la Ley N.º 8968 de Costa Rica, que es la “Ley de Protección de la Persona frente al Tratamiento de sus Datos Personales”.
5. Interés institucional en modernizar la infraestructura de la CCSS mediante estrategias de transformación digital.

Estas condiciones hacen viable el desarrollo de un prototipo reproducible, escalable y alineado con las necesidades del país, que pueda posteriormente adaptarse a entornos clínicos reales.

4.5 Conclusión del análisis

El análisis evidencia que Costa Rica cuenta con el talento humano y las condiciones tecnológicas necesarias para implementar soluciones innovadoras de diagnóstico asistido por IA, pero, al mismo tiempo, enfrenta desafíos en infraestructura y estandarización.

La adopción de una arquitectura Big Data en la nube y de redes neuronales profundas ofrece una alternativa viable para:

- Reducir el tiempo de interpretación de imágenes MRI.
- Aumentar la precisión diagnóstica.
- Garantizar trazabilidad, reproducibilidad y gobernanza de datos.
- Democratizar el acceso a tecnologías médicas avanzadas.

El proyecto propuesto no solo responde a una necesidad técnica, sino también a una prioridad de salud pública y de equidad tecnológica, que podría posicionar a Costa Rica en la ruta hacia una medicina más digital, eficiente y sostenible.

5. Propuesta de solución

5.1 Descripción general de la solución propuesta

La solución propuesta en esta investigación consiste en el desarrollo e implementación de un sistema de diagnóstico asistido por computadora basado en técnicas de inteligencia artificial, en particular mediante el uso de redes neuronales profundas para el análisis de imágenes médicas de resonancia magnética (MRI). Esta propuesta surge como respuesta a una problemática presente en el sistema de salud pública costarricense, donde el aumento constante en la cantidad de estudios de MRI, que ya alcanza miles de resonancias magnéticas cada mes, genera una alta presión sobre los recursos humanos disponibles en los centros de salud pública, especialmente en el área de radiología.

Existen desafíos significativos, como los prolongados tiempos de espera, que según Segura (2026) pueden alcanzar hasta 648 días para pacientes oncológicos en la CCSS, así como limitaciones en la capacidad de análisis de los estudios y variabilidad en la interpretación de las imágenes entre especialistas. En este contexto, la solución desarrollada busca apoyar el criterio de los médicos mediante el uso de modelos de aprendizaje profundo capaces de identificar patrones complejos en imágenes MRI. El objetivo final es mejorar la precisión de los diagnósticos, hacer más consistentes los procesos de análisis y contribuir a optimizar el flujo de trabajo en los entornos médicos.

Desde la perspectiva tecnológica, la solución propuesta integra diferentes componentes de ingeniería de datos, computación en la nube y aprendizaje automático en un proceso que permite gestionar de forma eficiente y unificada todo el ciclo de vida de los datos, desde su recolección hasta la generación de los resultados predictivos. En este caso, se construye una arquitectura basada en los servicios en la nube de Microsoft Azure ya que facilita el almacenamiento, procesamiento y análisis de grandes volúmenes de datos médicos de manera escalable y segura. Se implementan los modelos de deep learning utilizando técnicas de transfer learning sobre esta infraestructura, aprovechando arquitecturas preentrenadas que han demostrado un buen desempeño en tareas de clasificación de imágenes.

Este enfoque no solo reduce los recursos computacionales y la cantidad de datos necesarios para entrenar modelos desde cero, sino que también mejora la capacidad del modelo para generalizar cuando se le presentan nuevos datos. Además, la solución incluye un flujo estructurado de procesamiento que abarca etapas de preprocesamiento, aumento de datos, entrenamiento, validación y evaluación. Así se garantiza que haya un proceso ordenado,

consistente y reproducible, alineado con las buenas prácticas en proyectos de inteligencia artificial aplicada.

Desde el punto de vista metodológico, la solución se desarrolla dentro de un enfoque de investigación aplicada, en el cual se combinan los fundamentos teóricos de la inteligencia artificial y el procesamiento de imágenes con una implementación práctica orientada a resolver un problema real en el contexto de la salud pública. En este sentido, el sistema propuesto no se limita únicamente a construir un modelo predictivo, sino que plantea un flujo completo que incluye la preparación de los datos, el diseño del experimento, la evaluación rigurosa de los resultados y la posibilidad de escalar la solución a entornos productivos.

Finalmente, se realiza un análisis comparativo entre los dos modelos y también tres técnicas de ensamble, con el objetivo de identificar aquellas configuraciones que nos den los mejores resultados. Este enfoque no solo permite validar la viabilidad técnica de la solución, sino también establecer una base sólida para futuras implementaciones en entornos clínicos reales, donde el uso de herramientas de inteligencia artificial puede contribuir significativamente a mejorar la calidad del diagnóstico y la eficiencia operativa del sistema de salud.

5.2 Configuración del entorno en la nube de Azure

5.2.1 Cuenta, subscripción y resource group

La implementación de la solución propuesta se realizó utilizando la plataforma de computación en la nube Microsoft Azure, la cual fue seleccionada porque ofrece recursos escalables, servicios especializados en machine learning y una buena integración con herramientas de almacenamiento y procesamiento de datos. Se utilizó Azure Machine Learning como el entorno principal para el desarrollo, entrenamiento y evaluación de los modelos de inteligencia artificial. Esto permitió gestionar de forma centralizada tanto los recursos de cómputo como los experimentos realizados.

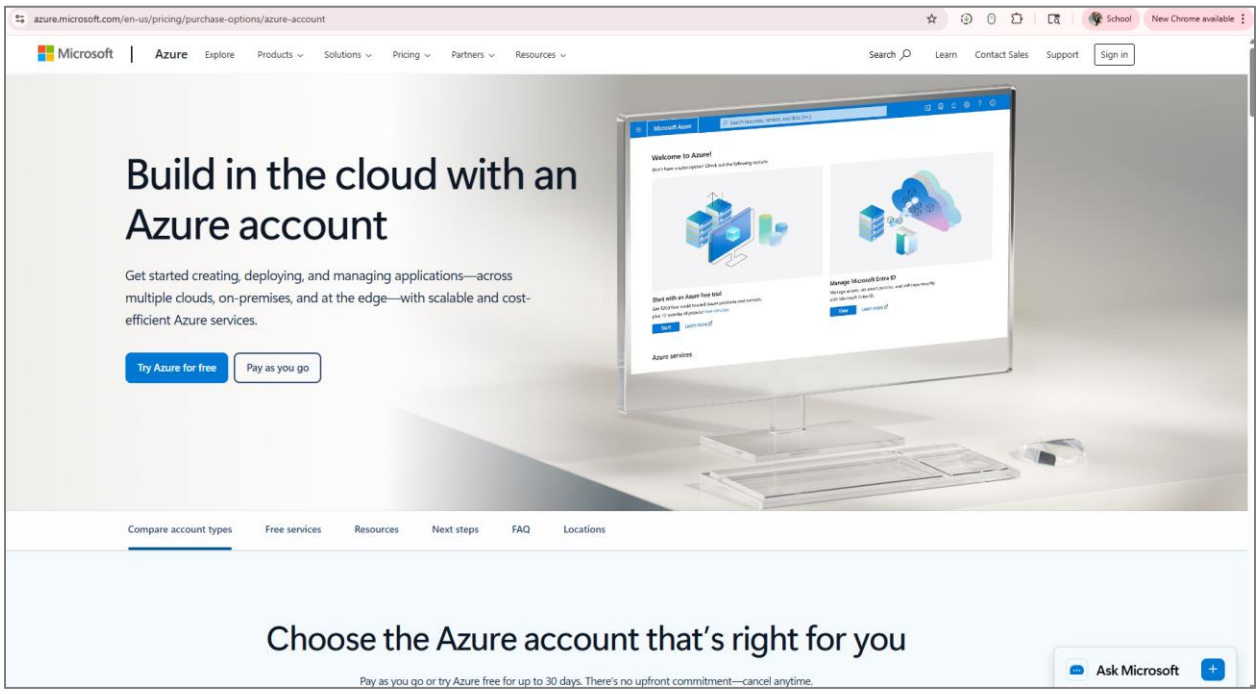
Se eligió esta plataforma debido a que los modelos de deep learning requieren un alto consumo computacional. También, porque es una gran ventaja contar con un entorno que sea reproducible y que permita la ejecución de experimentos bajo distintas configuraciones sin depender de infraestructura local. Azure es un componente clave dentro de la arquitectura de esta solución, ya que proporciona las capacidades necesarias para soportar todo el ciclo de desarrollo de modelos de aprendizaje profundo.

El primer paso del proyecto consistió en la creación de una cuenta en Azure, seguida de la configuración de una suscripción y, posteriormente, de un grupo de recursos (Resource Group).

En primer lugar, la cuenta de Azure representa el punto de acceso principal a la plataforma, permitiendo autenticarse y gestionar todos los servicios disponibles en la nube. Sin una cuenta, no es posible crear ni administrar recursos. La figura 6 representa la creación de la cuenta en la plataforma.

Figura 6

Configuración de suscripción en Microsoft Azure

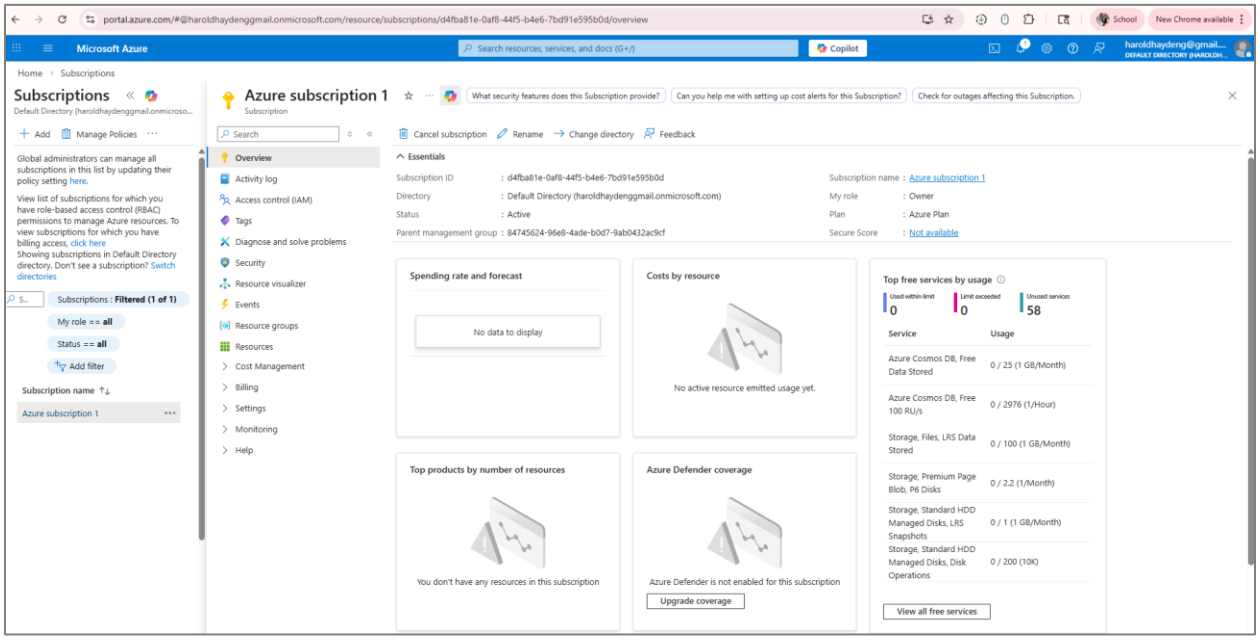


Nota: Captura de pantalla del proceso de configuración en Microsoft Azure. Elaboración propia.

Una vez creada la cuenta, el siguiente paso es la suscripción como se muestra en la figura 7, la cual actúa como un contenedor lógico para los recursos y define aspectos clave como la facturación, los límites de uso y los permisos. En otras palabras, la suscripción permite controlar los costos del proyecto y establecer quién puede acceder o modificar los recursos.

Figura 7

Interfaz de gestión de suscripción en Microsoft Azure

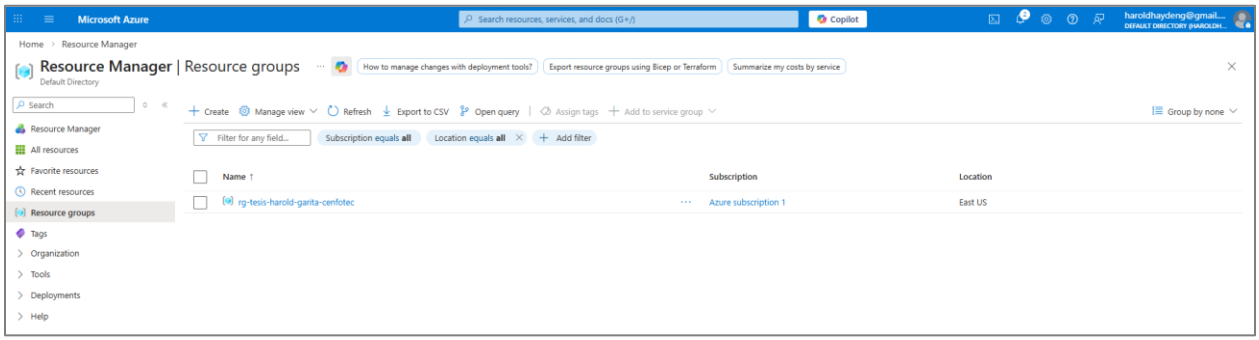


Nota. Captura de pantalla del proceso de configuración de la suscripción en Microsoft Azure. Elaboración propia.

Posteriormente, se crea un Resource Group, que funciona como un contenedor dentro de la suscripción para agrupar todos los recursos relacionados con un proyecto específico, como máquinas virtuales, almacenamiento, servicios de machine learning, entre otros, como se muestra en la figura 8. Esto facilita la organización, el monitoreo y la administración de los recursos, ya que permite gestionarlos de forma conjunta.

Figura 8

Creación y administración de Resource Groups en Microsoft Azure



Nota. Captura de pantalla del proceso de gestión de recursos en Microsoft Azure. Elaboración propia.

Además, el uso de Resource Groups permite aplicar configuraciones de seguridad, políticas y control de acceso de manera centralizada, así como eliminar todos los recursos asociados a un proyecto de forma eficiente cuando ya no sean necesarios.

5.2.2 Azure Machine Learning workspace

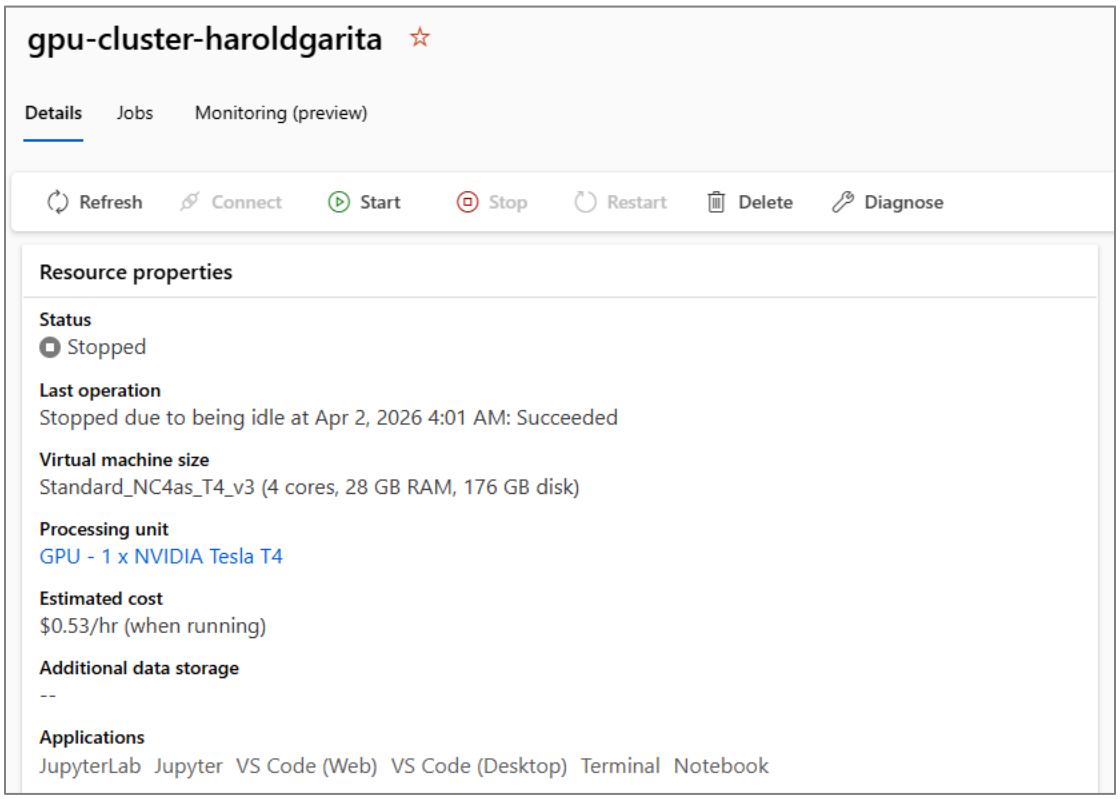
Dentro de Azure Machine Learning, se creó un workspace que funciona como el espacio central donde se gestionan todos los recursos del proyecto, incluyendo los experimentos, conjuntos de datos, modelos y configuraciones de cómputo. Para poder entrenar los modelos de deep learning en Azure, es necesario configurar recursos de cómputo especializados, preferiblemente máquinas virtuales con capacidad de procesamiento gráfico (GPU), ya que estas permiten acelerar de forma significativa las operaciones necesarias en las redes neuronales convolucionales.

5.2.3 Compute instance

En este proyecto se utilizó una máquina virtual Standard_NC4as_T4_v3 con 4 cores, 28 GB RAM y un disco de 176 GB. La unidad de procesamiento seleccionada fue 1 GPU NVIDIA Tesla T4. Es importante mencionar que el costo de utilizar esta máquina virtual es de $0.53/hora mientras está corriendo. Esta decisión es clave debido a la complejidad de las arquitecturas utilizadas y al tamaño del conjunto de datos. También, se seleccionó el kernel de Python 3.10 - SDK v2 que es uno de los ambientes curados (curated environments) disponibles en Azure Machine Learning. Esto quiere decir que el ambiente ya tenía un conjunto de librerías instaladas y no fue necesario crearlo desde cero, como se ilustra en la Figura 9.

Figura 9

Configuración de la máquina virtual utilizada en Azure Machine Learning



Nota. Captura de pantalla de las especificaciones de la instancia de cómputo utilizada en Azure Machine Learning. Elaboración propia.

Para proporcionar contexto sobre la decisión de utilizar una GPU en lugar de una CPU, inicialmente se intentó ejecutar una versión más liviana del modelo en una computadora local con las siguientes características (Figura 10):

- Modelo: HP EliteBook 84
- Procesador: Gen Intel® Core™ i7-1355U, 1700 Mhz, 10 Cores y 12 Procesadores Lógicos
- Sistema Operativo: Microsoft Windows 11 Pro
- Memoria RAM: 16.0 GB

Sin embargo, el tiempo de entrenamiento resultó ser considerablemente alto, alcanzando aproximadamente 1,5 horas por época. Adicionalmente, la computadora presentaba fallos durante la ejecución, lo que impedía completar el proceso de entrenamiento de manera

consistente. Particularmente, la computadora no lograba manejar las necesidades de memoria RAM para entrenar el modelo.

Figura 10

Características del equipo utilizado para el entrenamiento local del modelo

Item	Value
OS Name	Microsoft Windows 11 Pro
Version	10.0.26200 Build 26200
Other OS Description	Not Available
OS Manufacturer	Microsoft Corporation
System Name	HAROLDGARITA
System Manufacturer	HP
System Model	HP EliteBook 840 14 inch G10 Notebook PC
System Type	x64-based PC
System SKU	846V7LT#ABM
Processor	13th Gen Intel(R) Core(TM) i7-1355U, 1700 Mhz, 10 Core(s), 12 Logical Process...
BIOS Version/Date	HP V70 Ver. 01.11.00, 11/6/2025
SMBIOS Version	3.4
Embedded Controller Version	81.69
BIOS Mode	UEFI
BaseBoard Manufacturer	HP
BaseBoard Product	8B41
BaseBoard Version	KBC Version 51.45.00
Platform Role	Mobile
Secure Boot State	On
PCR7 Configuration	Elevation Required to View
Windows Directory	C:\WINDOWS
System Directory	C:\WINDOWS\system32
Boot Device	\Device\HarddiskVolume1
Locale	United States
Hardware Abstraction Layer	Version = "10.0.26100.1"
User Name	HaroldGarita\harol
Time Zone	Central America Standard Time
Installed Physical Memory (RAM)	16.0 GB

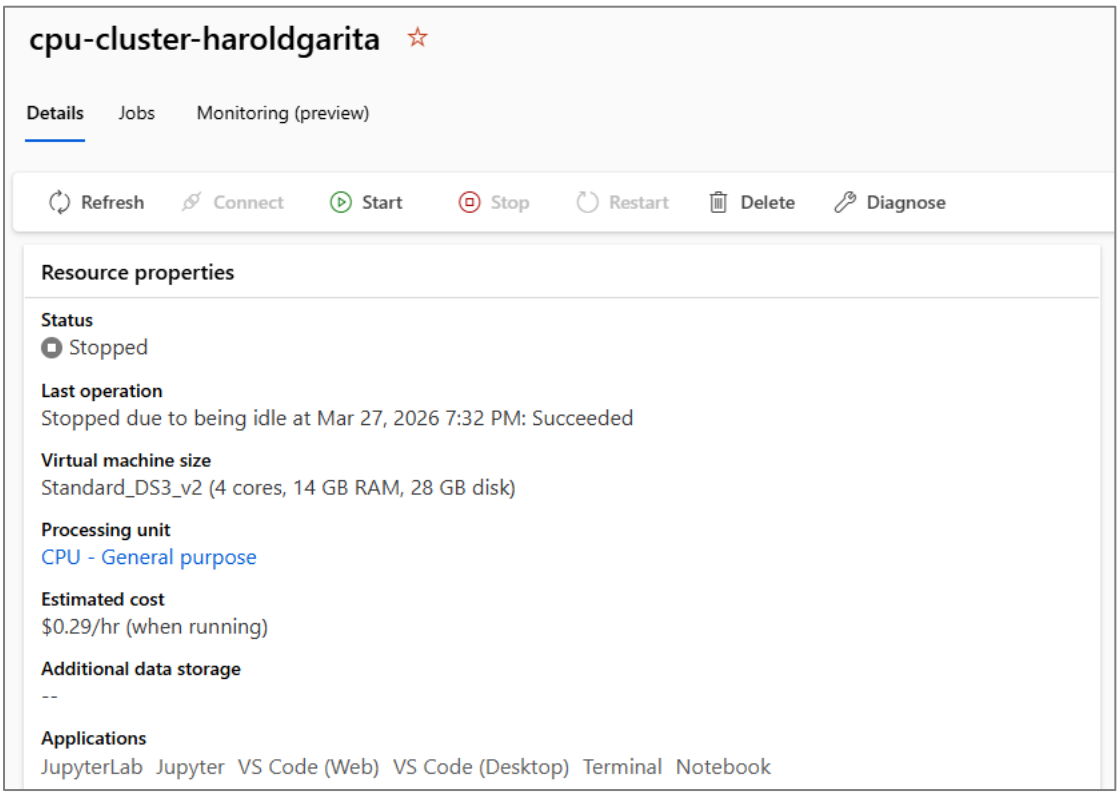
*Nota***.** Información del sistema del equipo utilizado para el entrenamiento local del modelo, incluyendo sistema operativo, procesador y memoria RAM. Elaboración propia.

Después de no haber tenido éxito desde la computadora local, se intentó utilizar una máquina virtual Standard_DS3_v2 en Azure con 4 cores, 14 GB de RAM y un disco de 28GB, según la Figura 11. Esta tiene un costo de $0.29/hora mientras está corriendo. Si bien esta

alternativa permitió completar el entrenamiento, el tiempo por época se mantenía en aproximadamente 15 minutos. Aunque era funcional, este rendimiento implicaba varias horas de espera para evaluar resultados, realizar ajustes y ejecutar nuevas iteraciones del modelo.

Figura 11

Configuración de la máquina virtual Standard_DS3_v2 en Microsoft Azure



*Nota***.** Configuración de la máquina virtual Standard_DS3_v2 en Microsoft Azure, donde se detallan sus características principales como CPU, memoria RAM y costo estimado por hora. Elaboración propia.

El GPU NVIDIA Tesla T4 fue seleccionado principalmente porque ofrece un buen balance entre rendimiento, costo y compatibilidad con tareas de aprendizaje profundo en entornos en la nube como Azure. Debido principalmente a restricciones de tiempo, se decidió finalmente utilizar este GPU, ya que permitió reducir significativamente los tiempos de entrenamiento, alcanzando aproximadamente 20 segundos por época en el modelo EfficientNet V2 y cerca de 1 minuto por época en el modelo ViT-B16.

Esto se debe a que el GPU Tesla T4 está diseñado para ejecutar tareas de entrenamiento e inferencia en deep learning, ya que incluye tecnologías como CUDA y Tensor Cores. Estas herramientas permiten acelerar de forma importante las operaciones matemáticas que utilizan modelos como las redes neuronales convolucionales (CNN) y los modelos basados en transformadores, como EfficientNet V2 y ViT-B16. Esto explica la gran reducción en los tiempos de entrenamiento, pasando de varios minutos por época en CPU a solo segundos en GPU.

Por otro lado, desde el punto de vista de costo-beneficio, la T4 es un GPU de gama media muy eficiente dentro de plataformas en la nube. En comparación con opciones más avanzadas, como la V100 o la A100, ofrece un buen nivel de rendimiento a un costo por hora más bajo, lo que lo hace adecuado para proyectos académicos y procesos donde se requiere probar el modelo varias veces.

Finalmente, dado que el desarrollo de modelos es un proceso iterativo, que implica entrenar, ajustar parámetros y evaluar resultados varias veces, el GPU T4 permite realizar estos ciclos mucho más rápido, lo cual fue especialmente importante considerando las limitaciones de tiempo del proyecto.

5.2.4 Storage account

Dentro de la solución implementada en Microsoft Azure, fue necesario crear una cuenta de almacenamiento (Storage Account) para gestionar de forma eficiente los datos del proyecto. En particular, se utilizó Azure Data Lake Storage como el principal componente de almacenamiento, lo cual permitió manejar grandes volúmenes de datos no estructurados, como las imágenes de resonancia magnética empleadas en esta investigación.

Este servicio facilitó la implementación de la arquitectura de medallón (Bronze, Silver y Gold), permitiendo organizar los datos en diferentes niveles de procesamiento y calidad. Asimismo, su integración con Azure Machine Learning permitió acceder a los datos de forma directa durante los procesos de entrenamiento y evaluación del modelo, mejorando la eficiencia del flujo de trabajo. Algunas de las ventajas de usar una cuenta de almacenamiento son que proporciona escalabilidad, alta disponibilidad y una separación clara entre los datos y los recursos de cómputo, lo cual constituye una buena práctica en arquitecturas en la nube.

5.3 Arquitectura de la solución

La arquitectura de la solución propuesta se basa en un enfoque moderno de ingeniería de datos que combina principios de data lakehouse con técnicas de procesamiento orientadas a

machine learning, utilizando específicamente el modelo de arquitectura de medallón. Esta arquitectura permite organizar el flujo de datos en capas progresivas que aseguran la calidad, trazabilidad y disponibilidad de la información a lo largo de todo el proceso analítico.

En el contexto de esta investigación, la arquitectura fue diseñada para soportar el procesamiento de imágenes médicas en un entorno en la nube, garantizando que los datos pasen de forma controlada desde su estado original hasta convertirse en insumos adecuados para el entrenamiento de modelos de deep learning. La implementación de esta arquitectura no solo responde a la necesidad de manejar grandes volúmenes de datos no estructurados, como imágenes MRI, sino que también facilita la reproducibilidad del proceso, permitiendo que cada etapa del flujo de datos pueda ser revisada, replicada y optimizada de manera independiente.

5.3.1 Capa Bronze

En la capa Bronze, los datos se almacenan en su forma original, sin aplicar transformaciones, lo que permite conservar la integridad y fidelidad de las imágenes médicas tal como fueron obtenidas del dataset fuente. Esta capa funciona como un repositorio de datos crudos dentro del entorno de almacenamiento en la nube, específicamente en Azure Data Lake Storage, donde se organizan las imágenes de resonancia magnética junto con sus etiquetas correspondientes. La importancia de esta capa se basa en que constituye la base del sistema, ya que permite mantener un registro completo de los datos originales, facilitando su reutilización en futuros experimentos o procesos de validación. Además, al no modificar los datos en esta etapa, se reduce el riesgo de introducir sesgos o errores durante el procesamiento inicial, lo cual es especialmente importante en aplicaciones médicas donde la precisión y la trazabilidad son aspectos fundamentales.

5.3.2 Capa Silver

En la capa Silver se realiza el procesamiento y la preparación de los datos, etapa en la cual las imágenes son transformadas para que puedan ser utilizadas en el entrenamiento de modelos de aprendizaje profundo. En esta fase se aplican diferentes técnicas de preprocesamiento, como el redimensionamiento de las imágenes a una resolución uniforme compatible con las arquitecturas de redes neuronales utilizadas, así como la normalización de los valores de los píxeles para mejorar la estabilidad durante el entrenamiento. Además, se implementan estrategias de data augmentation, como rotaciones, volteos horizontales, ajustes de brillo y zoom, con el objetivo de aumentar la variabilidad del conjunto de datos sin necesidad de recolectar nuevas imágenes. Este proceso es fundamental para reducir el sobreajuste y

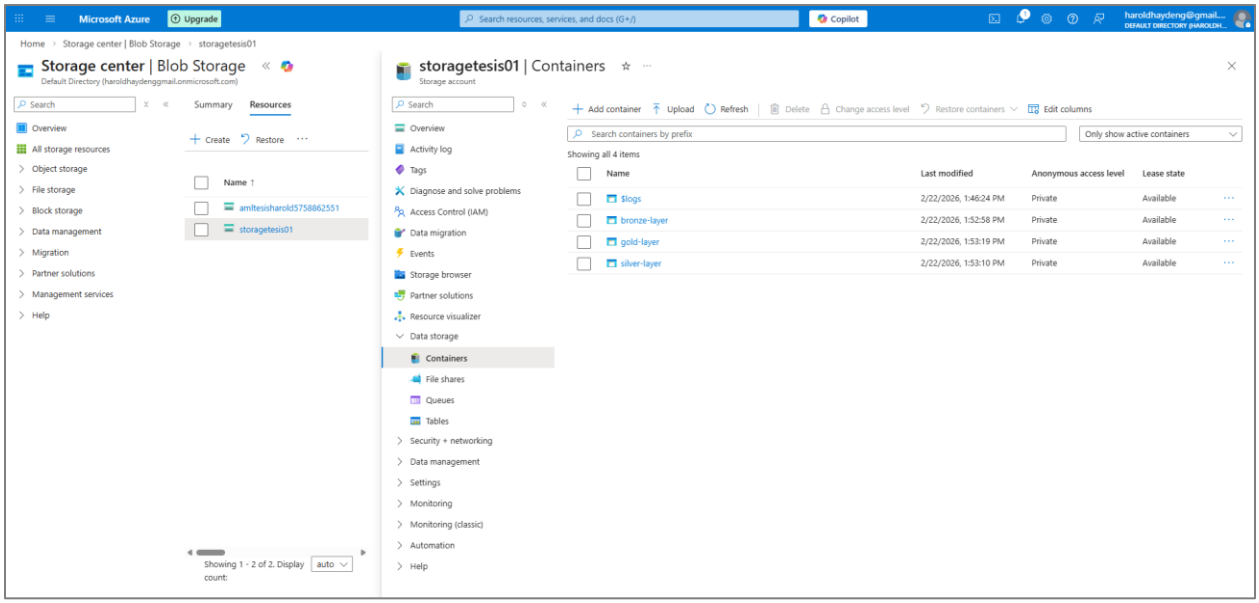
mejorar la capacidad del modelo para generalizar, especialmente en escenarios donde la cantidad de datos es limitada o existe desbalance entre clases. De esta manera, la capa Silver funciona como un punto clave de transformación, donde los datos pasan de ser información cruda a convertirse en insumos organizados y enriquecidos, listos para ser utilizados en el proceso de modelado.

5.3.3 Capa Gold

La capa Gold representa el nivel más refinado dentro de la arquitectura de medallón, y corresponde al conjunto de artefactos analíticos y resultados de alto valor que se generan como producto final del proceso de entrenamiento, evaluación y comparación de los modelos de aprendizaje profundo (Figura 12). A diferencia de las capas anteriores, cuyo propósito es la ingesta y transformación de los datos, la capa Gold está orientada al consumo directo de la información por parte de investigadores, profesionales clínicos o sistemas de apoyo a la decisión, por lo que los artefactos almacenados en esta capa deben ser de la más alta calidad, estar correctamente organizados y ser completamente trazables. Esta capa funciona como el repositorio final de los resultados del experimento.

Figura 12

Organización de contenedores Bronze, Silver y Gold en Azure Blob Storage



Nota. Organización del almacenamiento de datos en Azure Blob Storage de acuerdo con la arquitectura de medallón, utilizada para la gestión y almacenamiento de los resultados del proyecto. Elaboración propia.

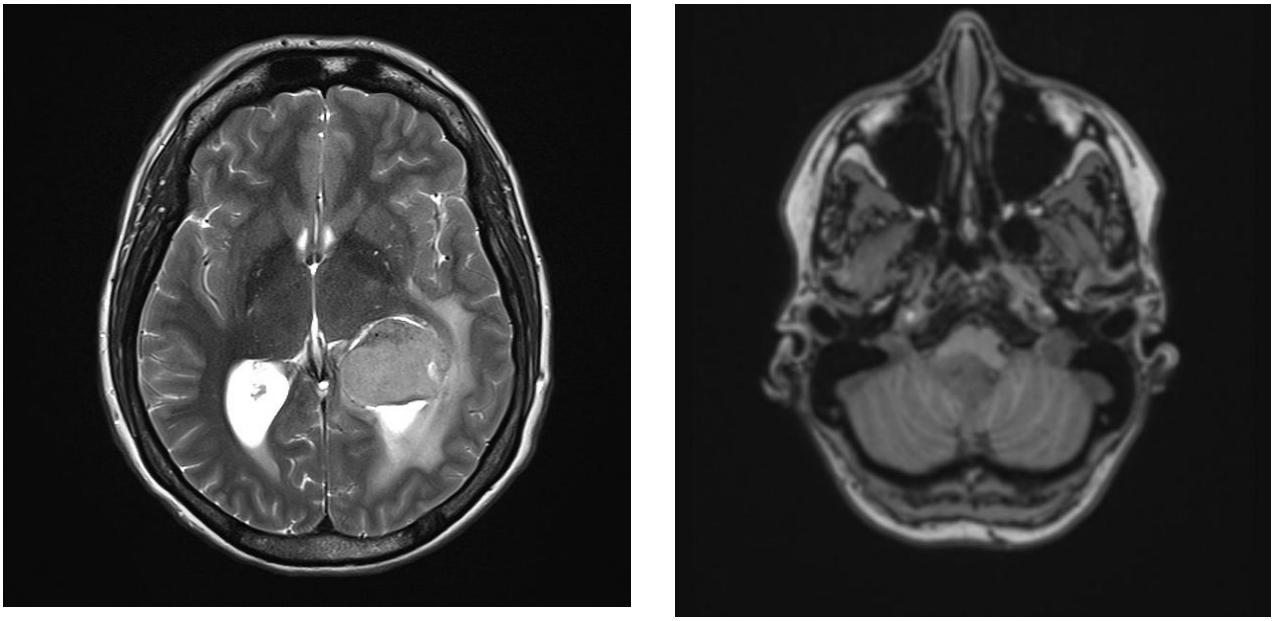
5.4 Fuente de datos y exploración

5.4.1 Fuente de datos

En esta investigación, se trabajó con un conjunto de datos público de imágenes de resonancia magnética cerebral (MRI), disponible en la plataforma Kaggle (Feltrin, 2023), compuesto por 4.478 imágenes organizadas en 44 clases. Estas clases representan 14 tipos de tumores cerebrales y una clase adicional de cerebros normales. Los tumores incluyen imágenes en modalidades MRI T1, T2 y T1 con contraste (T1C+), mientras que los casos normales solo están disponibles en T1 y T2. Este dataset fue desarrollado por Fernando Feltrin, ingeniero en computación y técnico radiólogo en GRS Imagem, Brasil, lo que aporta confiabilidad técnica y respaldo clínico en la selección y organización de los datos. El conjunto fue elegido por su variedad de clases y su utilidad para problemas de clasificación multiclase en el ámbito médico para ilustrarlo, se incluyen ejemplos en la figura 13.

Figura 13

Ejemplos de imágenes de resonancia magnética cerebral (MRI) del conjunto del dataset



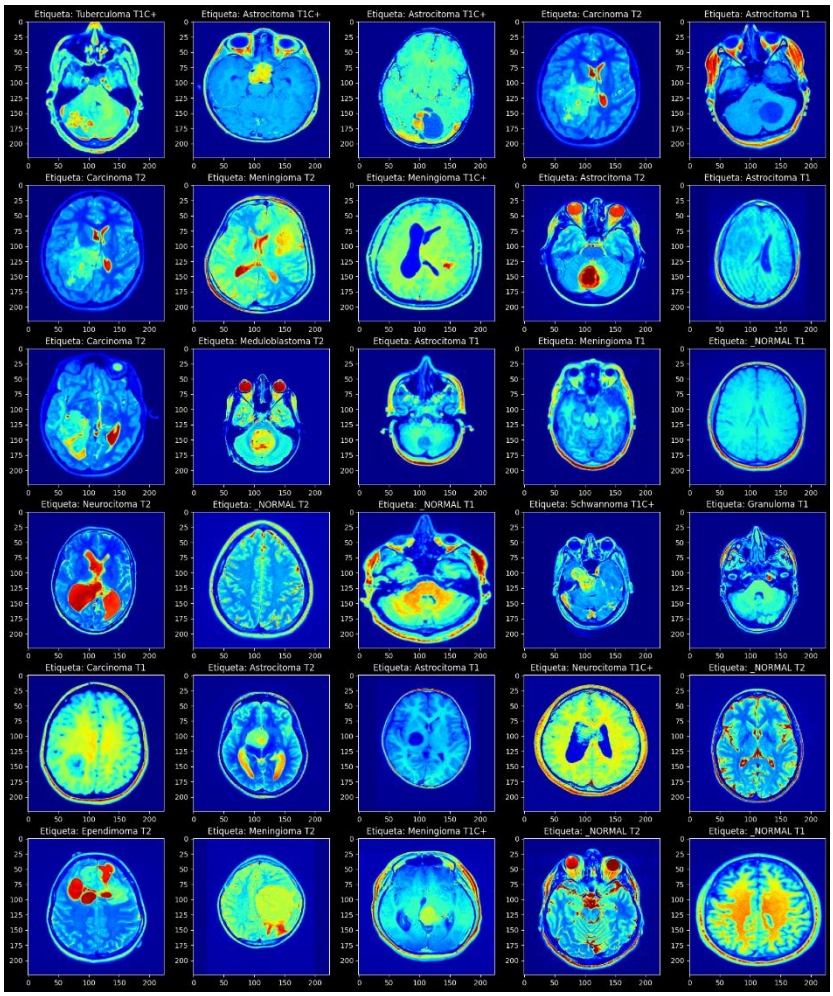
*Nota***.** Tomado de *Brain Tumor MRI Images 44 Classes* [Conjunto de datos], por F. Feltrin, 2023, Kaggle, https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c

5.4.2 Exploración de los datos

La etapa de exploración y preprocesamiento de datos constituye un componente clave dentro de la solución propuesta, dado que la calidad y representatividad del conjunto de datos influyen directamente en el desempeño de los modelos de aprendizaje profundo. En una primera fase, se llevó a cabo un análisis exploratorio con el objetivo de comprender la distribución de las clases, identificar posibles desbalances y evaluar características generales de las imágenes, tales como dimensiones, formatos y calidad visual.

Figura 14

Visualización de muestras de imágenes MRI por clase durante el análisis exploratorio de datos



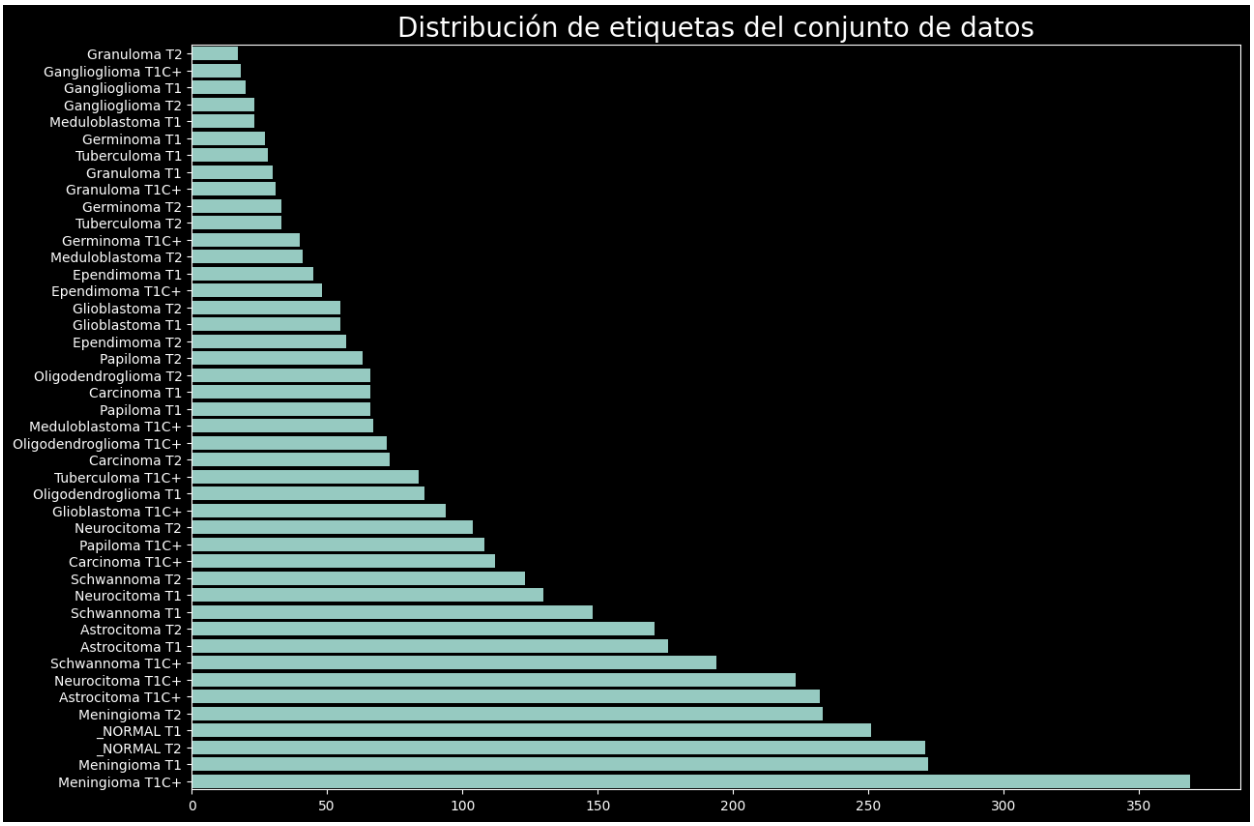
Nota: Elaboración propia a partir del conjunto de datos Brain Tumor MRI Images 44 Classes (Feltrin, 2021).

Como resultado de este análisis, se observó una distribución considerablemente desbalanceada entre las clases (Figura 15), donde algunas categorías, como Meningioma T1C+, Meningioma T1 y NORMAL T2, presentan una alta cantidad de muestras, mientras que otras clases cuentan con una representación significativamente menor. En particular, se identificó que clases como Granuloma T2, Ganglioglioma T1C+, Ganglioglioma T1 y Ganglioglioma T2 corresponden a las categorías con menor número de muestras dentro del dataset.

Es importante destacar que este desbalance no solo representa un desafío técnico, sino que también refleja en cierta medida la distribución real de estas condiciones en la práctica clínica, ya que algunos tipos de tumores son naturalmente menos frecuentes que otros. Sin embargo, desde la perspectiva del aprendizaje automático, esta situación puede introducir sesgos en el modelo, favoreciendo las clases mayoritarias durante el proceso de entrenamiento si no se aplican técnicas adecuadas de balanceo.

Figura 15

Distribución de frecuencias por clase del conjunto de datos de resonancia magnética cerebral



Nota: Elaboración propia a partir del conjunto de datos Brain Tumor MRI Images 44 Classes (Feltrin, 2021).

Asimismo, se identificaron variaciones en el tamaño, resolución y proporciones de las imágenes, lo cual introduce inconsistencias en los datos de entrada. Estas diferencias pueden afectar la capacidad del modelo para aprender patrones relevantes de manera uniforme.

A partir de estos hallazgos, se definieron estrategias de preprocesamiento orientadas a estandarizar las imágenes (por ejemplo, mediante redimensionamiento y normalización) y a mitigar el desbalance de clases, con el fin de mejorar la calidad del entrenamiento y reducir posibles sesgos. De esta manera, se estableció una base de datos más consistente y adecuada para las etapas posteriores de modelado.

5.5 División del conjunto de datos

La división del conjunto de datos es una etapa clave dentro del proceso de desarrollo de modelos de aprendizaje automático, ya que permite evaluar de forma objetiva la capacidad del modelo para generalizar ante datos que no ha visto antes. En esta investigación, el dataset se dividió en tres subconjuntos principales: entrenamiento, validación y prueba, como se muestra en la Figura 16. Esta es una estrategia común en problemas de clasificación supervisada. La separación se realiza para evitar el sobreajuste y asegurar que el modelo no solo aprenda patrones específicos del conjunto de entrenamiento, sino que también pueda desempeñarse correctamente en situaciones nuevas.

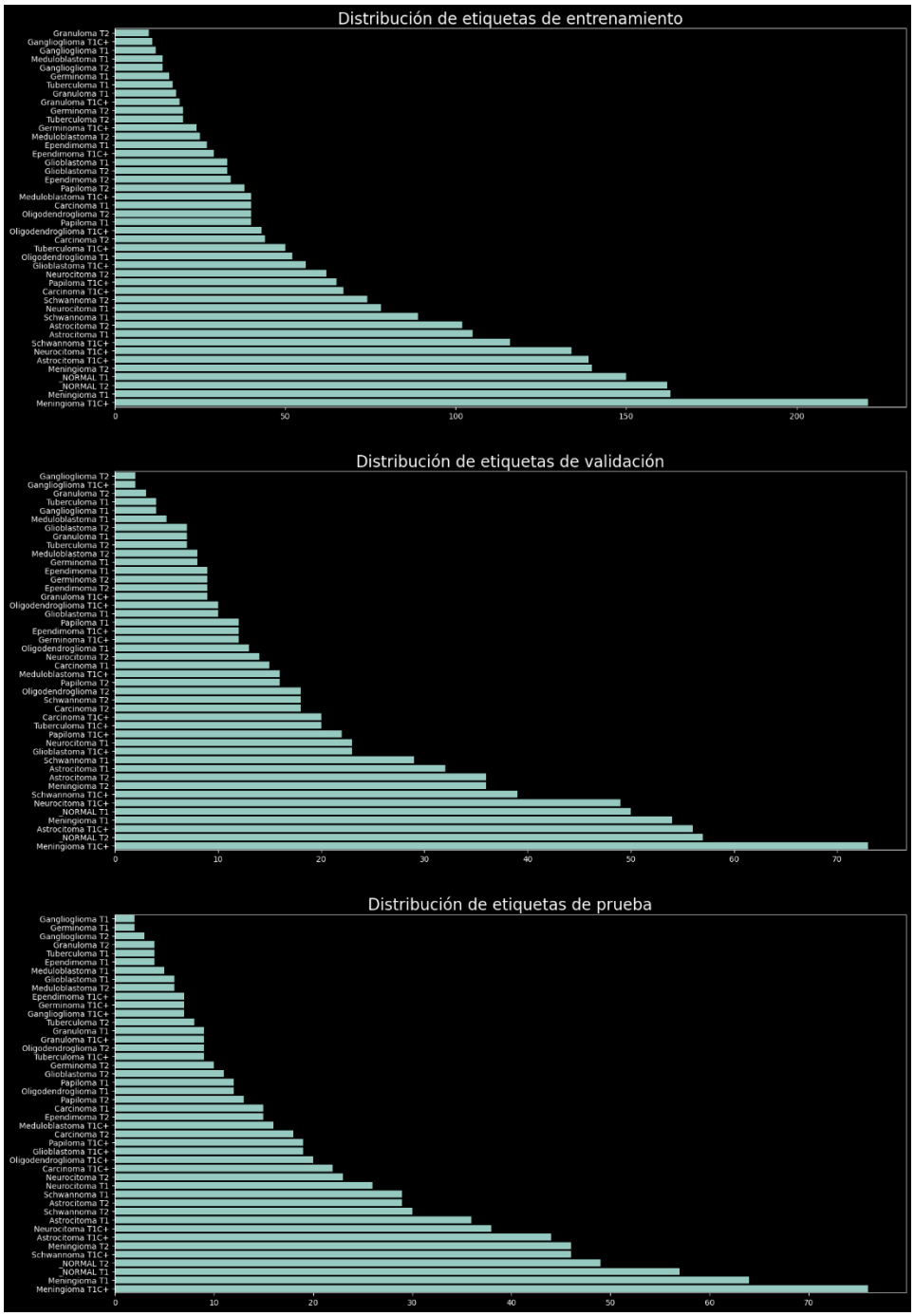
El conjunto de entrenamiento se utilizó para ajustar los parámetros del modelo, mientras que el conjunto de validación permitió monitorear su desempeño durante el aprendizaje y apoyar la selección de hiperparámetros. Por su parte, el conjunto de prueba se reservó exclusivamente para la evaluación final, proporcionando una medida imparcial del rendimiento del modelo. Esta estrategia garantiza una adecuada separación entre las fases de entrenamiento y evaluación, lo cual es fundamental para asegurar la validez de los resultados obtenidos.

Debido a que este dataset se considera mediano, porque tiene entre 1.000 y 10.000 imágenes, se realizó la división del dataset utilizando una distribución aproximada de 60% para entrenamiento, 20% para validación y 20% para prueba. Esto permite mantener un equilibrio adecuado entre la cantidad de datos disponibles para entrenar el modelo y la necesidad de contar con subconjuntos representativos para su evaluación.

Se aplicó una división estratificada, para asegurarse que la proporción de cada tipo de tumor cerebral se mantuviera similar en los tres subconjuntos. Esta decisión es especialmente importante en problemas de clasificación multiclase, donde algunas clases pueden tener menos muestras y generar sesgos si no se distribuyen correctamente. La estratificación ayuda a mantener la representatividad del dataset original en cada subconjunto, lo que permite obtener métricas de evaluación más confiables y comparables. Asimismo, esta estrategia facilita la identificación de posibles problemas de desbalance durante el entrenamiento, permitiendo realizar ajustes adicionales en caso de ser necesario.

Figura 16

Distribución de frecuencias por clase en los conjuntos de entrenamiento, validación y prueba



Nota. Elaboración propia a partir del conjunto de datos Brain Tumor MRI Images 44 Classes (Feltrin, 2021).

Desde el punto de vista de implementación, la división del conjunto de datos se integró directamente en el pipeline de carga y procesamiento de imágenes mediante el uso de generadores de datos en TensorFlow y Keras, lo que permite manejar grandes volúmenes de información sin necesidad de cargar todo el dataset en memoria. Estos generadores se configuraron para apuntar a los directorios correspondientes a cada subconjunto (entrenamiento, validación y prueba), asegurando que cada etapa del proceso utilice únicamente los datos que le corresponden.

5.6 Preparación del conjunto de datos

Como parte del proceso de preparación, se aplicaron técnicas de preprocesamiento orientadas a estandarizar las características de las imágenes y hacerlas compatibles con las arquitecturas de redes neuronales utilizadas. En primer lugar, se redimensionaron todas las imágenes a una resolución uniforme de 224x224 píxeles, como se ilustra en la Figura 16. Esta normalización min-max es requerida por los modelos preentrenados (EfficientNet V2 y ViT-B16) que esperan entradas en ese rango. Esta transformación permite asegurar que todas las imágenes tengan el mismo tamaño, evitando errores durante el entrenamiento y facilitando el procesamiento por lotes.

Posteriormente, se aplicó un proceso de normalización utilizando funciones específicas de la arquitectura empleada, con el objetivo de escalar los valores de los píxeles desde el rango original de 0 a 255 hacia un rango entre 0.0 y 1.0. Este ajuste es más adecuado para el entrenamiento de modelos y contribuye a mejorar la estabilidad del proceso de aprendizaje. La normalización se aplica a todas las imágenes antes de que ingresen al pipeline y sean procesadas por los modelos. Este paso es crítico, ya que ayuda a que el modelo converja de manera más eficiente y mejore su capacidad de generalización. En conjunto, estas transformaciones permiten convertir un conjunto de datos heterogéneo en una estructura homogénea y adecuada para el entrenamiento de modelos de deep learning.

Figura 17

Implementación del preprocesamiento de imágenes mediante redimensionamiento y normalización

Cargar y visualizar una imagen de muestra aleatoria

```python
def _load(image_path):
    # Leer y decodificar un archivo de imagen en un tensor de tipo uint8
    image = tf.io.read_file(image_path)
    image = tf.io.decode_jpeg(image, channels=3)

    # Redimensionar la imagen
    image = tf.image.resize(image, [CFG.HEIGHT, CFG.WIDTH],
                            method=tf.image.ResizeMethod.LANCZOS3)

    # Convertir el tipo de dato de la imagen a float32 y normalizarla
    image = tf.cast(image, tf.float32)/255.

    # Retornar imagen
    return image

def view_sample(image, label, color_map='gray', fig_size=(8, 10)):
    plt.figure(figsize=fig_size)
    plt.imshow(tf.image.rgb_to_grayscale(image), cmap=color_map)
    plt.title(f'Label: {label}', fontsize=16)
    return
```

Nota. Elaboración propia.

Se construyó una capa de aumento de datos (data augmentation) utilizando la API Sequential de Keras (Figura 18), encadenando dos transformaciones aleatorias aplicadas a las imágenes durante el entrenamiento con el objetivo de incrementar artificialmente la diversidad del conjunto de datos y reducir el sobreajuste (overfitting).

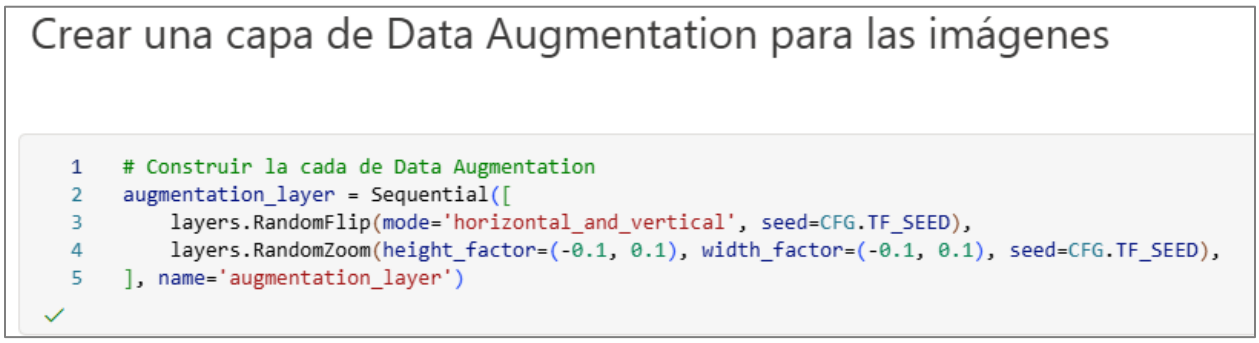
La primera transformación, RandomFlip, aplica volteos aleatorios sobre ambos ejes de la imagen: horizontal (espejo izquierda-derecha) y vertical (espejo arriba-abajo), con una probabilidad del 50% para cada eje de forma independiente. La segunda transformación, RandomZoom, aplica un zoom aleatorio tanto en altura como en anchura dentro de un rango de ±10%, lo que simula variaciones en la distancia de captura de la imagen.

Ambas transformaciones utilizan una semilla fija (CFG.TF_SEED) para garantizar la reproducibilidad de los experimentos. Esta capa de aumento se aplicó exclusivamente al conjunto

de entrenamiento, ya que en imágenes de resonancia magnética cerebral la orientación de un tumor no posee relevancia diagnóstica, lo que hace que los volteos y cambios de escala sean transformaciones válidas para enriquecer el conjunto de datos sin introducir información clínicamente incorrecta.

Figura 18

Implementación de la capa de aumento de datos (data augmentation) en Keras

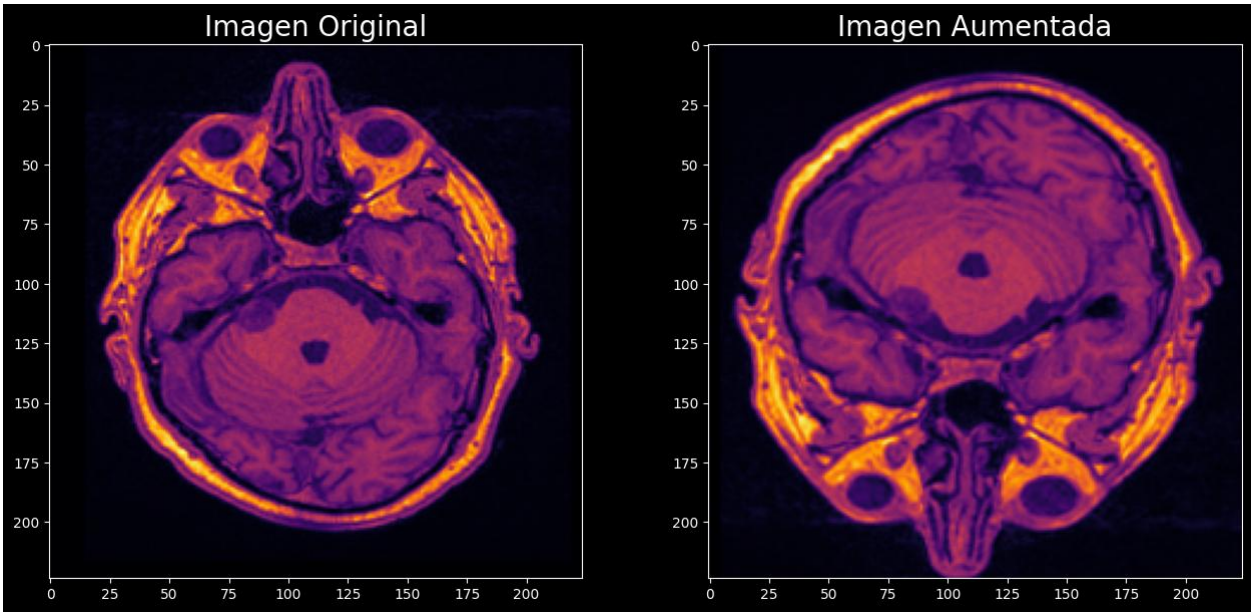


Nota. Elaboración propia

Estas transformaciones se implementaron mediante generadores de datos durante el entrenamiento, lo que permite aplicarlas de forma dinámica en cada época y aumentar la diversidad del conjunto de entrenamiento, como se ilustra en la Figura 19. Como resultado, el modelo es expuesto a una mayor variedad de patrones, lo que mejora su robustez y su capacidad para generalizar ante datos no vistos. Este enfoque no solo incrementa el tamaño efectivo del dataset, sino que también mejora el desempeño del modelo en escenarios reales donde las condiciones de captura pueden variar considerablemente.

Figura 19

Ejemplo de imagen original y su versión aumentada mediante data augmentation



Nota. Elaboración propia a partir del conjunto de datos Brain Tumor MRI Images 44 Classes (Feltrin, 2021), mediante técnicas de aumento de datos implementadas en el pipeline de entrenamiento.

Como resultado de estas etapas, se obtuvo un conjunto de datos organizado y optimizado para su uso en el entrenamiento de los modelos de aprendizaje profundo. En esta fase final de preparación, las imágenes se organizaron en directorios o estructuras compatibles con los generadores de datos utilizados en TensorFlow y Keras, lo que permite cargarlas de manera eficiente en memoria durante el entrenamiento.

Además, se definieron configuraciones específicas para los generadores de datos, como el tamaño de lote y la forma en que se leen las imágenes, lo cual ayuda a optimizar el uso de los recursos computacionales. Esta etapa consolida el pipeline de preparación de datos, garantizando que el conjunto final cumpla con los requisitos necesarios para alimentar de manera eficiente y confiable a los modelos de redes neuronales convolucionales en las etapas posteriores del proceso.

5.7 Implementación de los modelos de redes neuronales convolucionales

La implementación de los modelos de aprendizaje profundo representa el núcleo técnico de la solución propuesta, ya que en esta etapa se lleva a cabo el proceso de modelado orientado a la clasificación multiclase de imágenes de resonancia magnética cerebral en 44 categorías de condiciones neurológicas. Para abordar el problema planteado, se adoptó un enfoque basado en transfer learning, en el que se utilizaron arquitecturas preentrenadas de alto desempeño como punto de partida, adaptando su salida al problema específico mediante la adición de capas densas de clasificación.

Este enfoque permite aprovechar representaciones visuales previamente aprendidas a partir de grandes conjuntos de datos de imágenes generales, como ImageNet e ImageNet21K, lo que reduce significativamente la necesidad de contar con grandes volúmenes de datos etiquetados y disminuye el costo computacional del entrenamiento. En el contexto de esta investigación, esta estrategia resulta especialmente adecuada dada la naturaleza especializada del conjunto de datos médico y la complejidad inherente de un problema de clasificación con 44 clases. Concretamente, se aplicó una estrategia de extracción de características (feature extraction), en la cual los pesos del modelo base se mantuvieron completamente congelados durante todo el proceso de entrenamiento, permitiendo que únicamente las capas de clasificación añadidas se ajustaran a los datos del dominio médico.

En total, se implementaron cinco modelos: dos modelos base individuales y tres variantes de ensamble derivadas de sus predicciones combinadas. A continuación, se describe en detalle cada uno de ellos.

Es importante señalar que el código de entrenamiento implementado en esta investigación tomó como punto de partida el notebook de Jansen (2025) en la plataforma Kaggle, el cual desarrolla un pipeline de referencia basado en transferencia de aprendizaje para la clasificación de tumores cerebrales en imágenes MRI. Sobre esa base, se realizaron dos modificaciones principales:

1. Se ajustaron los hiperparámetros del modelo, tasa de aprendizaje, número de épocas, tamaño de lote y configuración de callbacks, para adaptarlos al conjunto de datos de 44 clases utilizado, el cual es significativamente más amplio y variado que el empleado en el notebook original.
2. Se integró el pipeline de entrenamiento dentro de una arquitectura medallón sobre Azure Machine Learning, con almacenamiento estructurado en capas Bronze–

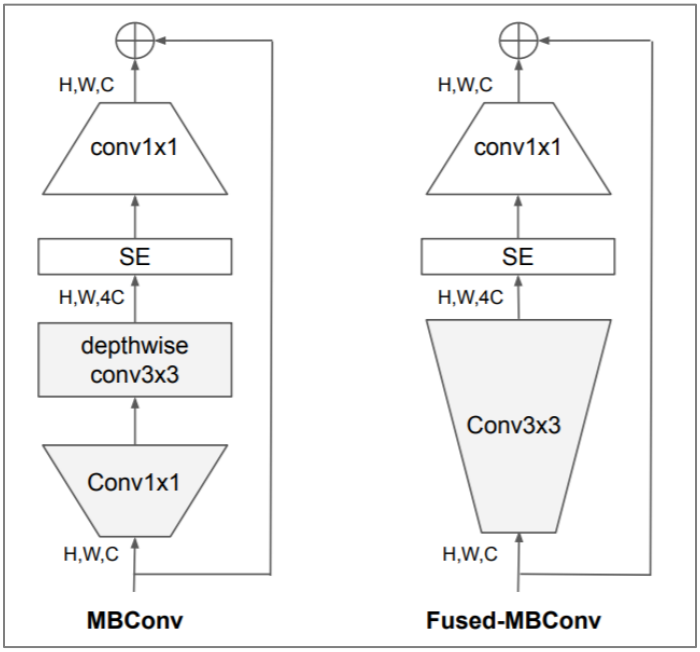
Silver–Gold en Azure Blob Storage, garantizando la trazabilidad y reproducibilidad completa del proceso.

5.7.1 Modelo EfficientNet V2 B0

El primero de los modelos individuales se construyó sobre la arquitectura EfficientNet V2 B0, una red neuronal convolucional desarrollada por Tan & Le (2021) que introduce mejoras significativas respecto a la versión original de EfficientNet en cuanto a velocidad de entrenamiento y eficiencia en el uso de parámetros. EfficientNetV2 emplea una combinación de búsqueda de arquitecturas neurales sensible al entrenamiento (training-aware neural architecture search) y escalado compuesto, lo cual permite optimizar simultáneamente la profundidad, el ancho y la resolución de la red. Entre sus innovaciones más destacadas se encuentra la incorporación de bloques Fused-MBConv, que sustituyen a los bloques MBConv tradicionales en las primeras capas y ofrecen un mejor equilibrio entre precisión y eficiencia computacional (Figura 20).

Figura 20

Comparación entre los bloques MBConv y Fused-MBConv en EfficientNetV2



Nota: Tomado de *EfficientNetV2*, por A. Arora (2021), Weights & Biases, https://wandb.ai/wandb_fc/pytorch-image-models/reports/EfficientNetV2--Vmlldzo2NTkwNTQ

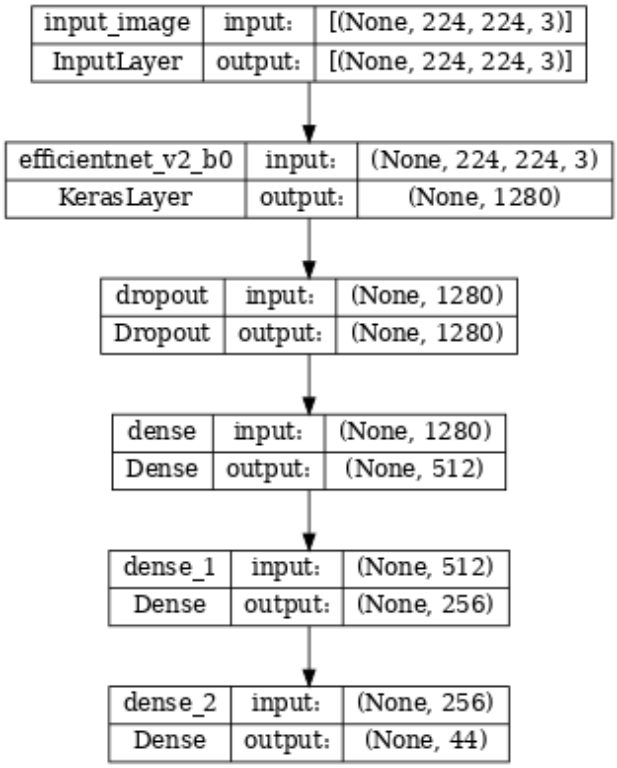
Para esta implementación, el modelo base fue obtenido a través de TensorFlow Hub, cargando los pesos preentrenados en el conjunto de datos ImageNet21K mediante la siguiente URL: https://tfhub.dev/google/imagenet/efficientnet_v2_imagenet21k_b0/feature_vector/2. El modelo fue cargado como una capa de Keras (hub.KerasLayer) configurada en modo no entrenable (trainable=False), lo que garantiza que los 5.919.312 parámetros del modelo base permanezcan fijos durante el entrenamiento, actuando exclusivamente como extractor de características visuales. La capa produce un vector de características de 1.280 dimensiones para cada imagen de entrada.

El modelo completo fue definido utilizando la API Sequential de Keras, encadenando las siguientes capas sobre la base preentrenada como se muestra en la Figura 21:

1. Capa de entrada: dimensiones (224, 224, 3), tipo float32, con nombre input_image.
2. Capa base EfficientNet V2 B0: extractor de características con pesos congelados, salida de (None, 1280). Aquí las imágenes de entrada son transformadas en vectores de alta dimensión (1280 características), los cuales son posteriormente procesados por las capas densas para realizar la clasificación final. Aunque en el diagrama EfficientNetV2B0 aparece como una sola capa (KerasLayer), en realidad corresponde a un modelo profundo compuesto por múltiples capas internas.
3. Dropout (tasa = 0.2): para reducir el sobreajuste durante el entrenamiento de las capas densas.
4. Capa densa (512 unidades, activación ReLU): primera capa de clasificación intermedia.
5. Capa densa (256 unidades, activación ReLU): segunda capa de clasificación intermedia.
6. Capa densa de salida (44 unidades, activación Softmax): produce las probabilidades de pertenencia a cada una de las 44 clases.

Figura 21

Estructura del modelo de clasificación con EfficientNetV2B0 como extractor de características



Nota. Elaboración propia.

Las tres capas densas fueron inicializadas con el esquema GlorotNormal (también conocido como Xavier Normal), con una semilla fija (CFG.SEED = 42) para garantizar la reproducibilidad. El resumen de parámetros del modelo resultante es el siguiente: 6.717.820 parámetros totales (25,63 MB), de los cuales 798.508 (3,05 MB) son entrenables y 5.919.312 (22,58 MB) son no entrenables (correspondientes al modelo base congelado).

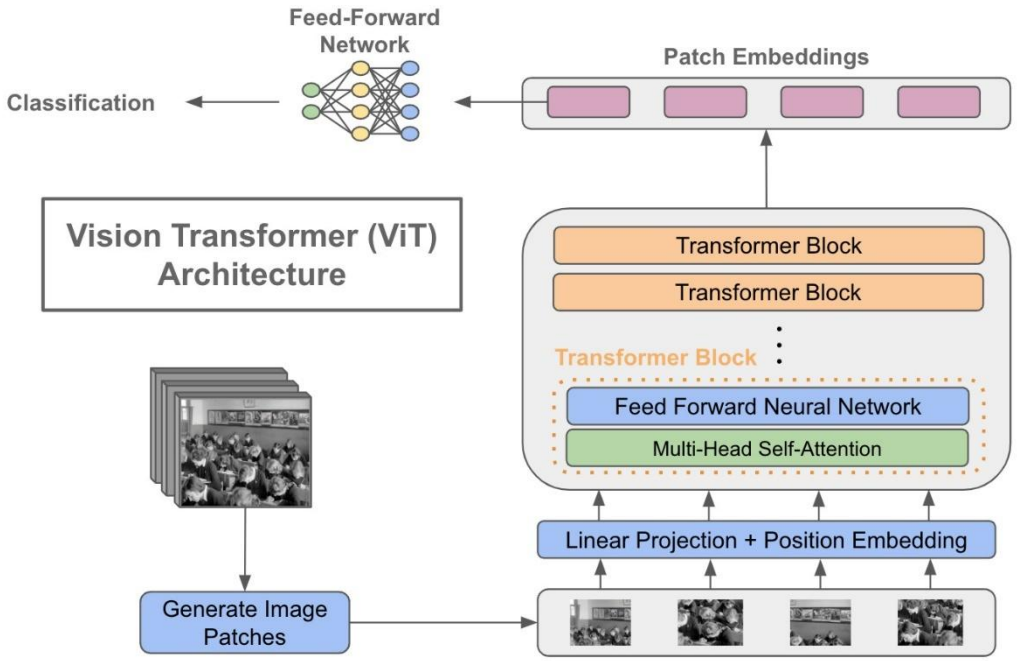
5.7.2 Modelo Vision Transformer ViT-B16

El segundo modelo individual se construyó sobre la arquitectura Vision Transformer ViT-B16 (Vision Transformer with Base configuration and patch size 16×16), introducida por Dosovitskiy et al. en el artículo "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale" (ICLR, 2021). A diferencia de las redes convolucionales tradicionales, los Vision Transformers dividen la imagen de entrada en parches de tamaño fijo (16×16 píxeles en este caso) y los procesan como una secuencia de tokens, aplicando mecanismos de atención multi-cabeza (multi-head self-attention) propios de los transformers originalmente diseñados para

procesamiento de lenguaje natural. La arquitectura del codificador de transformers se compone de tres elementos principales: normalización de capa (Layer Normalization), atención multi-cabeza y redes de retroalimentación (Feed-Forward Networks) como se ilustra en la Figura 22.

Figura 22

Arquitectura del Vision Transformer (ViT-B16)



Nota. Tomado de *Vision Transformers*, por C. R. Wolfe, 2021, Substack, https://cameronrwolfe.substack.com/p/vision-transformers

El modelo fue cargado utilizando la librería vit-keras (versión 0.2.0), la cual proporciona una implementación de los modelos ViT compatibles con TensorFlow/Keras. El modelo fue inicializado con los pesos preentrenados en ImageNet21K + ImageNet2012 mediante la instrucción vit.vit_b16(image_size=224, activation='softmax', pretrained=True, include_top=False, pretrained_top=False, classes=num_classes). La configuración include_top=False permite excluir la cabeza de clasificación original del modelo, sustituyéndola por las capas densas personalizadas definidas para las 44 clases del presente problema. Al igual que en el modelo EfficientNet, todos los pesos del modelo base fueron congelados mediante la instrucción layer.trainable = False aplicada a cada capa del modelo.

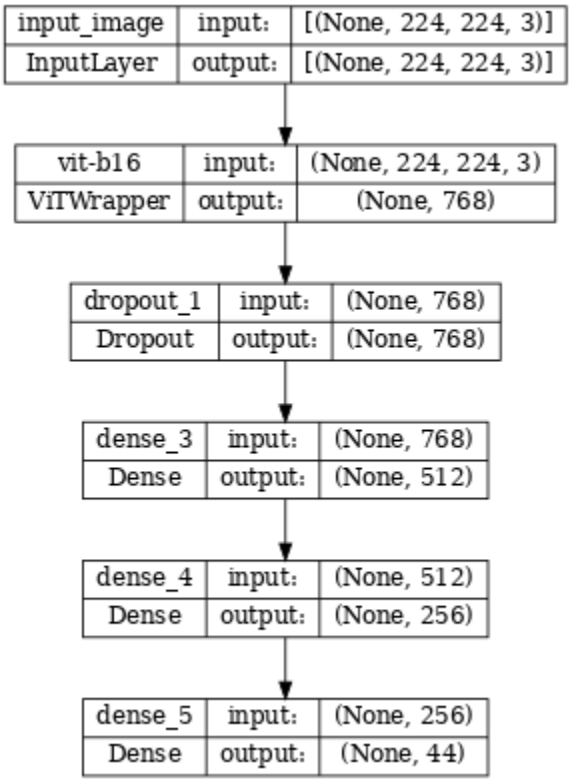
Dado que la librería vit-keras está basada en la API de Keras 3 mientras que el pipeline del notebook utiliza tf_keras (Keras 2), fue necesario implementar una clase envoltorio personalizada denominada ViTWrapper, la cual extiende tf.keras.layers.Layer y encapsula el modelo ViT para hacerlo compatible con la API Sequential de tf_keras. Esta clase delega la llamada al modelo interno durante la inferencia y el entrenamiento, preservando el comportamiento correcto de la capa en ambos modos. El modelo ViT-B16 genera un vector de características de 768 dimensiones para cada imagen de entrada.

El modelo completo fue definido también con la API Sequential de Keras, con la siguiente arquitectura (Figura 23):

1. Capa de entrada: dimensiones (224, 224, 3), tipo float32, con nombre input_image.
2. ViTWrapper (ViT-B16 congelado): extractor de características basado en atención, salida de (None, 768). Aquí las imágenes de entrada son transformadas en vectores de alta dimensión (768 características), los cuales son posteriormente procesados por las capas densas para realizar la clasificación final. Aunque en el diagrama vit_b16 aparece como una sola capa (ViTWrapper), en realidad corresponde a una arquitectura compuesta por múltiples capas internas.
3. Dropout (tasa = 0.2): regularización durante el entrenamiento de las capas densas.
4. Capa densa (512 unidades, activación ReLU): primera capa de clasificación intermedia.
5. Capa densa (256 unidades, activación ReLU): segunda capa de clasificación intermedia.
6. Capa densa de salida (44 unidades, activación Softmax): probabilidades de pertenencia a cada clase.

Figura 23

Estructura del modelo de clasificación con ViT-B16 como extractor de características



Nota. Elaboración propia

Las capas densas fueron inicializadas con GlorotNormal (seed=42). Es importante destacar que, a diferencia del modelo EfficientNet, los parámetros del modelo ViT base no son contabilizados dentro del resumen del modelo Sequential debido a la encapsulación mediante ViTWrapper; por esta razón, el total de parámetros entrenables reportado asciende a 536.364 (2,05 MB), correspondientes únicamente a las capas densas añadidas.

5.7.3 Modelos de ensamble

El aprendizaje por ensamble (ensemble learning) es una estrategia dentro del aprendizaje automático que consiste en combinar las predicciones de múltiples modelos base con el objetivo de obtener una predicción final más precisa, robusta y estable que la que podría producir cualquiera de los modelos individuales por separado. El principio fundamental detrás del ensamble es que distintos modelos cometen errores distintos, y que al combinar sus salidas se

pueden compensar estos errores individuales, reducir la varianza en las predicciones y mejorar la capacidad de generalización del sistema en su conjunto.

En el contexto de esta investigación, el ensamble resulta especialmente significativo desde el punto de vista arquitectónico: los dos modelos base implementados, EfficientNet V2 B0 y ViT-B16, pertenecen a paradigmas fundamentalmente distintos de procesamiento visual. EfficientNet V2 B0 es una red neuronal convolucional que extrae características mediante filtros locales y jerarquías espaciales convolucionales, lo que le confiere una fuerte capacidad para detectar texturas, bordes y patrones locales en las imágenes. ViT-B16, por su parte, es un Vision Transformer que divide la imagen en parches y modela las relaciones globales entre ellos mediante mecanismos de atención multi-cabeza, capturando dependencias de largo alcance (long-range dependencies) que los modelos convolucionales tienden a representar de manera menos efectiva. Esta diversidad arquitectónica constituye una base sólida para el ensamble, ya que cada modelo aporta una perspectiva complementaria sobre las características presentes en las imágenes MRI.

Para esta implementación, se optó por la modalidad de ensamble por promediado de probabilidades (ensemble via probability averaging), que opera directamente sobre los vectores de probabilidades a posteriori generados por cada modelo al evaluar el conjunto de prueba, sin requerir entrenamiento adicional. Una vez obtenidos los vectores combinados, la clase predicha para cada muestra se determina aplicando la función argmax sobre el vector resultante:

$$\hat{y} = \arg\max_{c} \bar{p}(c)$$

donde $\bar{p}(c)$ es la probabilidad combinada asignada a la clase *c* y $\hat{y}$ es la predicción final del ensamble. Se implementaron tres variantes de este enfoque, cada una con diferente método de combinación de probabilidades, las cuales se explican a continuación.

5.7.3.1 Ensamble por promedio simple

En esta variante, la probabilidad combinada para cada clase se obtiene calculando la media aritmética de los vectores de probabilidades de ambos modelos:

$$\bar{p}_{\text{avg}}(c) = \tfrac{1}{2}\left(p_{\text{EfficientNet}}(c) + p_{\text{ViT}}(c)\right)$$

En términos generales, dado un conjunto de n modelos con predicciones x_1, x_2,…,xn, la media aritmética se define como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Este método trata a ambos modelos como igualmente confiables, asignando el mismo peso relativo a sus predicciones. Es la variante más sencilla del ensamble por promediado y resulta efectiva cuando los modelos tienen un desempeño similar o cuando no se dispone de información previa suficiente para justificar una ponderación diferenciada. Su implementación en el notebook fue la siguiente como se muestra en la Figura 24:

Figura 24

Código para el cálculo del promedio de probabilidades y predicción final en el ensamble

```
# Calcular las probabilidades promedio
avg_probabilities = np.mean([
    efficientnet_v2_test_probabilities,
    vit_b16_test_probabilities], axis=0)

# Calcular las predicciones promedio del ensamble
avg_predictions = tf.argmax(avg_probabilities, axis=1)
```

Nota. Elaboración propia

5.7.3.2 Ensamble por promedio ponderado

En esta variante, se asignan pesos diferenciados a cada modelo en función de su desempeño relativo, de modo que el modelo con mayor precisión contribuye proporcionalmente más a la predicción final. La probabilidad combinada se calcula como una combinación lineal convexa de los vectores individuales donde los pesos satisfacen w_1, $w_2 \geq 0$ y $w_1+w_2 = 1$:

$$\bar{p}_{\text{weighted}}(c) = w_1 \cdot p_{\text{EfficientNet}}(c) + w_2 \cdot p_{\text{ViT}}(c)$$

En la forma general con *n* modelos y pesos normalizados que se representa de la siguiente manera:

,

$$w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j}$$

La expresión puede escribirse como:

$$\bar{x} = \sum_{i=1}^{n} w_i' x_i$$

En esta implementación, los pesos se fijaron de manera manual tomando como referencia el desempeño observado en la evaluación individual de los modelos: se asignó un peso de 0.6 al modelo EfficientNet V2 B0 y de 0.4 al modelo ViT-B16, reflejando la diferencia sustancial en precisión entre ambas arquitecturas (una diferencia de aproximadamente 10 puntos porcentuales en accuracy). Esta elección de pesos otorga mayor protagonismo al modelo convolucional sin descartar completamente la contribución del Vision Transformer. Su implementación fue (Figura 25):

Figura 25

Código para el cálculo del promedio ponderado de probabilidades y predicción final del ensamble

```
# Definir los pesos y listar las probabilidades de los modelos
weights = [0.6, 0.4]
model_probabilities = [efficientnet_v2_test_probabilities, vit_b16_test_probabilities]

# Calcular las probabilidades promedio ponderadas
weighted_avg_probabilities = sum([w * p for w, p in zip(weights, model_probabilities)])

# Calcular las predicciones del ensamble mediante promedio ponderado
weighted_avg_predictions = tf.argmax(weighted_avg_probabilities, axis=1)
```

Nota. Elaboración propia

5.7.3.3 Ensamble por media geométrica

La media geométrica es una alternativa al promedio aritmético que calcula el promedio multiplicativo de los valores, tomando la raíz de orden *n* del producto de todos los valores. Para dos modelos, la expresión es:

.
$$\bar{p}_{\text{geom}}(c) = \sqrt{p_{\text{EfficientNet}}(c) \cdot p_{\text{ViT}}(c)}$$

En su forma general para *n* valores:

$$\bar{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln x_i\right)$$

La propiedad más importante de la media geométrica en el contexto del ensamble es su sensibilidad reducida ante valores extremos: si un modelo asigna una probabilidad muy baja a una clase, el producto tiende a atenuar el efecto de las predicciones exageradamente altas del otro modelo, produciendo estimaciones más conservadoras y robustas ante distribuciones de probabilidad muy dispares entre los modelos. Esto la hace especialmente útil cuando uno de los modelos muestra picos de confianza (overconfidence) en clases específicas.

Su implementación fue (Figura 26):

Figura 26

Código para el cálculo de la media geométrica de probabilidades y predicción final del ensamble

```
# Listar las probabilidades de los modelos
model_probabilities = [efficientnet_v2_test_probabilities, vit_b16_test_probabilities]

# Calcular las probabilidades mediante media geométrica
geometric_mean_probabilities = np.power(np.multiply(model_probabilities[0],
                                                    model_probabilities[1]),
                                        1/len(model_probabilities))

# Calcular las predicciones del ensamble mediante promedio ponderado
geometric_mean_predictions = tf.argmax(geometric_mean_probabilities, axis=1)
```

Nota. Elaboración propia

5.8 Proceso de entrenamiento

El proceso de entrenamiento de los modelos individuales fue diseñado para maximizar el desempeño predictivo, asegurar la estabilidad del aprendizaje y hacer un uso eficiente de los recursos computacionales disponibles. Ambos modelos, EfficientNet V2 B0 y ViT-B16, fueron entrenados bajo una configuración idéntica de hiperparámetros e infraestructura, lo que garantiza la comparabilidad directa de sus resultados. En lo que respecta a los modelos de ensamble, estos no requieren un proceso de entrenamiento propio, ya que su funcionamiento se basa exclusivamente en la combinación post-hoc de los vectores de probabilidades generados por los modelos individuales ya entrenados sobre el conjunto de prueba.

5.8.1 Configuración del entrenamiento

5.8.1.1 Función de pérdida

Para ambos modelos se utilizó la entropía cruzada categórica (Categorical Crossentropy) como función de pérdida, la cual es la elección estándar para problemas de clasificación multiclase con

codificación one-hot en la capa de salida. Esta función mide la diferencia entre la distribución de probabilidad predicha por el modelo y la distribución real (etiqueta verdadera), calculando su valor para cada muestra del lote (batch) y promediándolo. Valores de pérdida cercanos a cero indican que el modelo asigna alta probabilidad a la clase correcta.

5.8.1.2 Optimizador

Se empleó el optimizador Adam (Adaptive Moment Estimation) con una tasa de aprendizaje inicial de 0.001. Adam es un optimizador de gradiente descendente estocástico que combina las ventajas de los métodos AdaGrad y RMSProp: adapta dinámicamente la tasa de aprendizaje para cada parámetro del modelo a partir de estimaciones de primer y segundo momento de los gradientes, lo que acelera la convergencia y produce actualizaciones más estables que el descenso de gradiente estocástico clásico.

5.8.1.3 Métricas y épocas

La métrica monitoreada durante el entrenamiento fue la exactitud (accuracy), tanto sobre el conjunto de entrenamiento como sobre el de validación. El número máximo de épocas se estableció en 50 (CFG.EPOCHS = 50), aunque este límite fue alcanzado en la práctica únicamente por el modelo ViT-B16, ya que el mecanismo de parada temprana interrumpió el entrenamiento del modelo EfficientNet V2 B0 antes de completar todas las épocas.

5.8.1.4 Tamaño de lote y distribución de datos

El tamaño de lote fue fijado en 64 muestras (CFG.BATCH_SIZE = 64). El conjunto de datos fue dividido previamente en tres subconjuntos: entrenamiento con 2.686 muestras, validación con 896 muestras y prueba con 896 muestras (esta última no participó en ninguna etapa del entrenamiento ni en la selección de hiperparámetros). En cada época, el modelo procesó los datos de entrenamiento organizados en 42 lotes (steps per epoch), evaluando su desempeño al finalizar cada época con el conjunto de validación. El parámetro shuffling fue configurado como False, ya que el conjunto de datos había sido previamente aleatorizado durante la etapa de construcción de los DataFrames.

5.8.2 Mecanismos de control: callbacks

Con el objetivo de optimizar el proceso de entrenamiento y evitar el sobreajuste o el uso innecesario de recursos computacionales, se implementaron dos callbacks que monitorean el desempeño del modelo en tiempo real y toman decisiones automáticas durante la ejecución:

5.8.2.1 EarlyStopping

El callback de parada temprana fue configurado para monitorear la pérdida de validación (val_loss) con una paciencia de 7 épocas (patience=7). Esto significa que el entrenamiento se detiene automáticamente si la métrica monitoreada no mejora durante 7 épocas consecutivas. Adicionalmente, se activó la opción restore_best_weights=True, la cual garantiza que al finalizar el entrenamiento, ya sea por parada temprana o por alcanzar el máximo de épocas, el modelo recupera los pesos correspondientes a la época en que se obtuvo la mejor pérdida de validación. Esta estrategia evita que el modelo final corresponda a un estado de sobreajuste producido en épocas tardías.

5.8.2.2 ReduceLROnPlateau

Este callback implementa un mecanismo de reducción adaptativa de la tasa de aprendizaje: cuando la pérdida de validación deja de mejorar durante 4 épocas consecutivas (patience=4), la tasa de aprendizaje vigente se reduce multiplicándola por un factor de 0.1 (factor=0.1). De esta forma, la tasa de aprendizaje disminuye de manera progresiva y automática a lo largo del entrenamiento, permitiendo ajustes más finos en los pesos del modelo en las etapas tardías del proceso, cuando la convergencia ya es cercana y son necesarias actualizaciones de menor magnitud para mejorar.

Ambos callbacks monitorean la misma métrica (val_loss) y fueron definidos con anterioridad al inicio del entrenamiento, configurándose en una lista compartida que se aplica a ambos modelos (Figura 27):

Figura 27

Configuración de callbacks EarlyStopping y ReduceLROnPlateau para el entrenamiento del modelo

```python
# Definir el callback de parada temprana
early_stopping_callback = tf.keras.callbacks.EarlyStopping(
    monitor='val_loss',
    patience=7,
    restore_best_weights=True)

# Definir el callback de reducción de la tasa de aprendizaje
reduce_lr_callback = tf.keras.callbacks.ReduceLROnPlateau(
    monitor='val_loss',
    patience=4,
    factor=0.1,
    verbose=1)

# Definir las listas de callbacks y métricas
CALLBACKS = [early_stopping_callback, reduce_lr_callback]
METRICS = ['accuracy']
```

Nota. Elaboración propia

5.8.3 Entrenamiento del modelo EfficientNet V2 B0

El modelo EfficientNet V2 B0 fue compilado y entrenado con la configuración descrita, con la semilla aleatoria de TensorFlow fijada en CFG.SEED = 42 inmediatamente antes del inicio del entrenamiento para garantizar la reproducibilidad del proceso. La evolución del entrenamiento mostró una convergencia progresiva y controlada ,como se muestra en la Figura 28 , con la tasa de aprendizaje siendo ajustada automáticamente en tres momentos distintos a lo largo de las épocas:

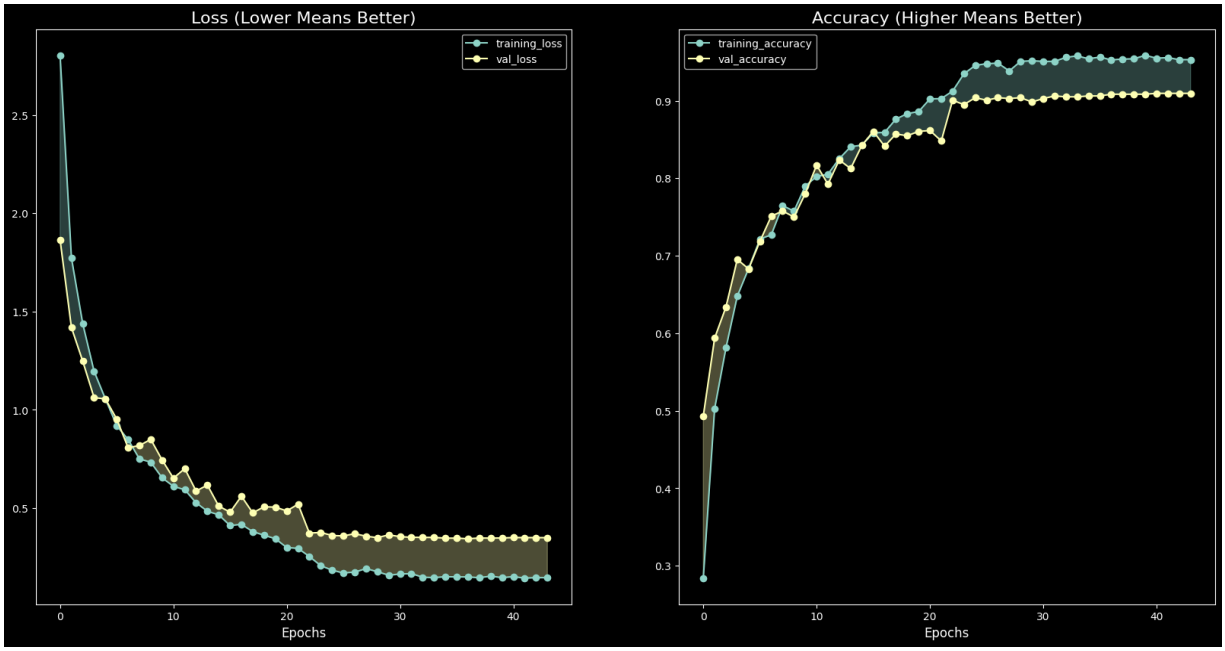
1. **Épocas 1–21 (LR = 0.001):** El modelo inicia con una exactitud de entrenamiento del 28.41% y una exactitud de validación del 49.33% en la primera época, mostrando una mejora sostenida en ambas métricas a lo largo de las épocas iniciales. La pérdida de validación disminuye de forma consistente a medida que el modelo aprende a discriminar entre las 44 clases.
2. **Época 22 (reducción de LR a 0.0001):** El callback ReduceLROnPlateau detecta que la pérdida de validación no ha mejorado durante las 4 épocas previas, reduciendo la tasa de aprendizaje de 0.001 a 0.0001. Esta reducción produce una

mejora notable e inmediata: en la época 23, la exactitud de validación salta al 90.07%, lo que evidencia que el modelo se encontraba en un mínimo local y que la reducción de la tasa de aprendizaje le permitió escapar de él y converger hacia una región de mejor rendimiento.

3. **Época 33 (reducción de LR a 0.00001):** Ante una nueva meseta en la pérdida de validación, la tasa se reduce nuevamente a 0.00001. El modelo continúa mejorando marginalmente, alcanzando una exactitud de validación del 90.85% en la época 37, que representa el mejor valor obtenido durante todo el entrenamiento.
4. **Época 41 (reducción de LR a 0.000001):** El callback realiza una tercera reducción, llevando la tasa de aprendizaje a 0.000001. El modelo entra en una fase de plateau, sin mejoras adicionales en la pérdida de validación.
5. **Época 44 (parada temprana):** El callback EarlyStopping detecta que la pérdida de validación no ha mejorado en 7 épocas consecutivas desde la época 37, deteniendo el entrenamiento en la época 44 y restaurando los pesos del modelo a los valores obtenidos en la época 37. El entrenamiento total duró aproximadamente 15 minutos sobre GPU Tesla T4.

Figura 28

Evolución de la pérdida y exactitud durante el entrenamiento del modelo EfficientNet V2 B0



Nota. Elaboración propia a partir del notebook de Jansen (2025), con modificaciones en los hiperparámetros y el proceso de entrenamiento para adaptarlo a un conjunto de datos desbalanceado y con mayor número de clases.

5.8.4 Entrenamiento del modelo ViT-B16

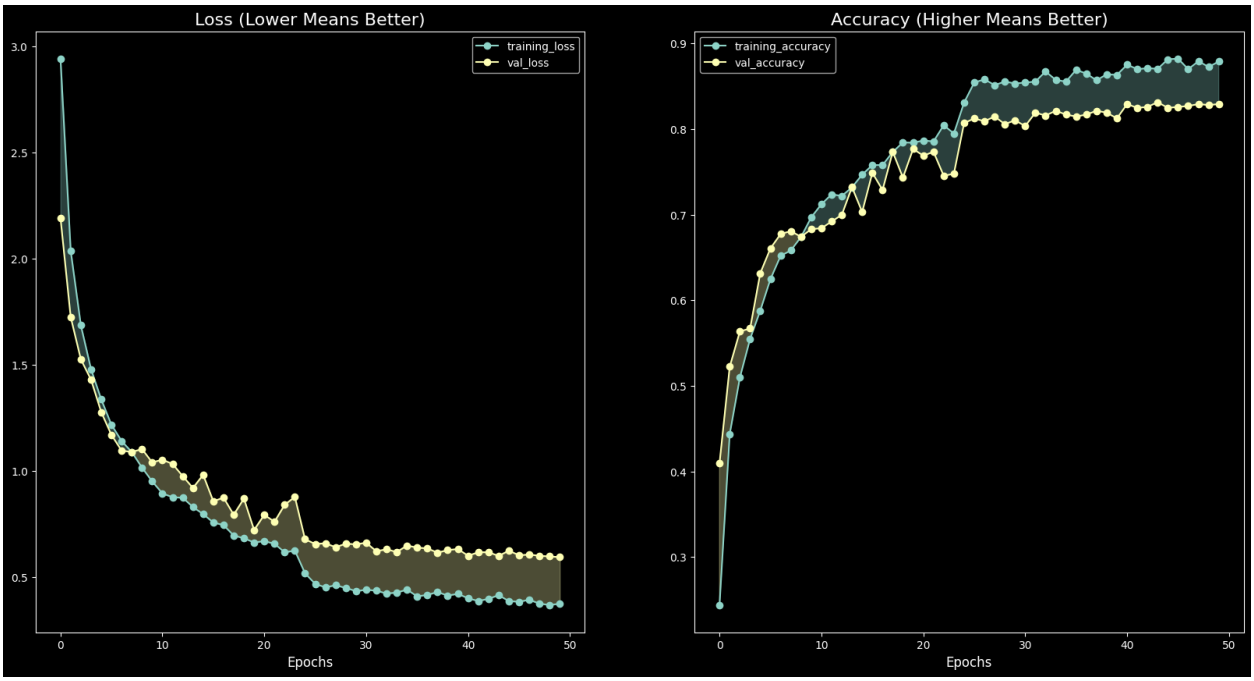
El proceso de entrenamiento del modelo ViT-B16 siguió la misma configuración de compilación que el modelo EfficientNet (Figura 29) : pérdida categórica, optimizador Adam con LR inicial de 0.001 y los mismos callbacks. Sin embargo, el comportamiento de convergencia fue muy diferente, reflejando las características propias de la arquitectura transformer al ser aplicada a un conjunto de datos de tamaño mediano.

1. **Épocas 1–23 (LR = 0.001):** El modelo muestra una convergencia considerablemente más lenta que EfficientNet. En la primera época, la exactitud de entrenamiento es del 24.39% y la de validación del 40.96%, valores inferiores a los del modelo convolucional en la misma etapa. La mejora es gradual y sostenida, pero con mayor variabilidad en las métricas de validación entre épocas consecutivas, lo cual es consistente con el comportamiento conocido de los Vision Transformers en datasets de tamaño limitado, donde su sesgo inductivo más débil dificulta la generalización en las etapas tempranas del aprendizaje.
2. **Época 24 (reducción de LR a 0.0001):** El callback ReduceLROnPlateau reduce la tasa de aprendizaje de 0.001 a 0.0001, produciéndose nuevamente una mejora apreciable: la exactitud de validación sube del 74.78% (época 24) al 80.69% (época 25) y continúa mejorando en épocas posteriores.
3. **Épocas 25–47 (LR = 0.0001):** El modelo continúa entrenando con mejora gradual. La pérdida de validación más baja alcanzada en esta fase se produce alrededor de la época 44 (val_loss ≈ 0.5985), con una exactitud de validación del 83.15%.
4. **Época 48 (reducción de LR a 0.00001):** Se produce una nueva reducción automática de la tasa de aprendizaje a 0.00001, tras lo cual el modelo logra su mejor desempeño al final del entrenamiento, con una pérdida de validación de 0.5925 y una exactitud de validación del 82.92% en la época 50.
5. **Época 50 (límite máximo alcanzado):** A diferencia del modelo EfficientNet, el modelo ViT-B16 no activó el callback de parada temprana, ya que seguía mostrando mejoras marginales en la pérdida de validación al llegar a la época 50. El entrenamiento se detuvo al alcanzar el límite máximo de épocas establecido en CFG.EPOCHS = 50. Los pesos finales del modelo corresponden a la época de

mejor val_loss, restaurados por la opción restore_best_weights=True. El entrenamiento total del modelo ViT-B16 fue considerablemente más costoso en tiempo que el de EfficientNet, con aproximadamente 2 segundos por paso versus los ~0.5 segundos por paso del modelo convolucional, lo que se traduce en un tiempo total de entrenamiento de aproximadamente 55 minutos sobre GPU Tesla T4.

Figura 29

Evolución de la pérdida y la exactitud en entrenamiento y validación del modelo ViT-B16



Nota. Elaboración propia a partir del notebook de Jansen (2025), con modificaciones en los hiperparámetros y el proceso de entrenamiento para adaptarlo a un conjunto de datos desbalanceado y con mayor número de clases.

5.8.5 Consideraciones sobre la ausencia de entrenamiento en los modelos de ensamble

Los tres modelos de ensamble implementados, promedio simple, promedio ponderado y media geométrica, no requieren ningún proceso de entrenamiento adicional. Su construcción se realiza en la fase de inferencia, operando exclusivamente sobre los vectores de probabilidades a posteriori que los modelos individuales ya entrenados generan al procesar el conjunto de prueba. Esta característica los hace computacionalmente muy eficientes y elimina el riesgo de

sobreajuste inherente a los métodos de aprendizaje por ensamble que requieren un conjunto de validación adicional para ajustar los pesos del combinador.

Sin embargo, es importante mencionar que, en el caso del ensamble ponderado, los pesos asignados a cada modelo (0.6 para EfficientNet y 0.4 para ViT) fueron determinados de forma manual con base en el desempeño individual observado, lo cual representa una decisión de diseño que podría optimizarse en trabajos futuros mediante técnicas de búsqueda automática como validación cruzada o búsqueda en cuadrícula (grid search).

5.9 Evaluación y métricas

Se definió un conjunto de seis métricas que permiten analizar el desempeño de los modelos desde múltiples perspectivas, considerando tanto el rendimiento global como el comportamiento por clase. Esta decisión es especialmente importante en el contexto de un problema de clasificación con 44 clases y desbalance entre ellas, donde una única métrica puede resultar insuficiente para caracterizar correctamente la calidad del modelo.

5.9.1Exactitud global (Accuracy Score)

La exactitud mide la proporción de predicciones correctas sobre el total de muestras evaluadas:

$$\text{Accuracy} = \frac{\text{Predicciones correctas}}{\text{Total de Predicciones}}$$

Es la métrica más intuitiva y de más rápida interpretación. Sin embargo, en presencia de desbalance de clases puede ser engañosa, ya que un modelo que clasifique siempre hacia las clases mayoritarias podría obtener una exactitud aparentemente alta sin haber aprendido a reconocer las clases minoritarias. Por esta razón, es obligatorio complementarla con las métricas siguientes.

5.9.2 Exactitud Top-3 (Top-3 Accuracy Score)

La exactitud Top-3 evalúa si la clase verdadera de una muestra se encuentra entre las tres categorías a las que el modelo le asignó mayor probabilidad, en lugar de exigir que la predicción principal sea la correcta. Formalmente, se define como la proporción de muestras del conjunto de prueba que cumplen esta condición sobre el total de muestras evaluadas.

Esta métrica adquiere especial relevancia en el contexto del diagnóstico médico asistido por computadora, ya que refleja de una forma más realista el modo en que un sistema de apoyo

clínico sería utilizado en la práctica. Por ejemplo, en un escenario real, un radiólogo o neurólogo no necesariamente acepta la primera sugerencia del sistema de forma automática. Más bien, consulta una lista de diagnósticos candidatos y aplica su criterio clínico para tomar la decisión final. En este sentido, que la clase correcta aparezca entre las tres primeras opciones indica que el modelo ha identificado información discriminativa relevante sobre la muestra, aun cuando no haya logrado posicionarla como la predicción de mayor confianza.

Esta característica es particularmente importante en un problema de clasificación con 44 clases, donde la proximidad morfológica entre ciertos tipos de tumores y modalidades de imagen puede dificultar la distinción entre categorías similares. Una exactitud Top-3 alta, combinada con una exactitud Top-1 sólida, sugiere que el modelo no solo clasifica bien en términos generales, sino que cuando se equivoca, sus errores tienden a ser "cercanos" a la respuesta correcta dentro del espacio de probabilidades. La métrica fue calculada mediante la función top_k_accuracy_score(y_true, y_probabilities, k=3) de la biblioteca scikit-learn.

5.9.3 Precisión ponderada (Weighted Precision)

La precisión responde a la siguiente pregunta: de todas las veces que el modelo predijo una clase determinada, ¿cuántas veces acertó? Esta metrica se define utilizando la siguiente fórmula:

$$\text{Precision} = \frac{\text{Verdaderos positivos (TP)}}{\text{Verdaderos positivos (TP)} + \text{Falsos positivos (FP)}}$$

Una precisión baja significa que el modelo predice frecuentemente una clase cuando en realidad la muestra pertenece a otra categoría.

En el contexto médico, una precisión baja para una clase concreta implica que el sistema estaría generando diagnósticos incorrectos de esa condición con demasiada frecuencia, lo que podría derivar en tratamientos innecesarios o en la realización de pruebas adicionales que no corresponden al cuadro clínico real del paciente.

Para obtener un indicador global del modelo, se utiliza la versión ponderada: el valor de precisión de cada clase se promedia proporcionalmente al número de muestras reales de esa clase en el conjunto de prueba (support). Esto garantiza que las clases con mayor representación en el dataset influyan más en el resultado final, produciendo un indicador que refleja el desempeño del modelo de manera coherente con la distribución real de los datos.

5.9.4 Recall ponderado (Weighted Recall)

El recall responde a la pregunta inversa a la precisión: de todas las muestras que realmente pertenecen a una clase, ¿cuántas fue capaz de detectar correctamente el modelo?

También se le conoce como sensibilidad o tasa de verdaderos positivos. Esta se define utilizando la siguiente fórmula:

$$Recall = \frac{\text{Verdaderos positivos (TP)}}{\text{Verdaderos positivos (TP)} + \text{Falsos negativos (FN)}}$$

Un recall bajo para una condición específica equivale a que el sistema falla en detectar pacientes que sí padecen esa enfermedad, clasificándolos incorrectamente como otra categoría, lo cual puede resultar en diagnósticos tardíos o erróneos con consecuencias potencialmente graves para el paciente. En un problema multiclase como el presente, donde cada clase representa un tipo de tumor y una modalidad de imagen distintos, el recall por clase permite identificar cuáles condiciones específicas el modelo tiende a pasar por alto con mayor frecuencia, información que es directamente relevante para evaluar la utilidad clínica del sistema en cada escenario diagnóstico concreto.

5.9.5 F1-score ponderado (Weighted F1-score)

El F1-score responde a la siguiente necesidad: ¿cómo evaluar el desempeño de un modelo cuando tanto los falsos positivos como los falsos negativos son relevantes, y se requiere un único indicador que los considere de forma simultánea? Esta métrica se define como la media armónica entre la precisión y el recall. Para calcularla utilizamos la siguiente fórmula:

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

La media armónica tiene una propiedad importante: penaliza de forma más severa los valores extremos que la media aritmética. Esto significa que para que un F1-score sea alto solo es posible cuando tanto la precisión como el recall son altos de manera simultánea. Si uno de

los dos es bajo, el F1-score lo reflejará de inmediato, independientemente de cuán alto sea el otro.

Por ejemplo, un modelo que tiene precisión perfecta (1.0) pero recall muy bajo (0.2) obtendrá un F1-score de apenas 0.33, lo cual captura adecuadamente que el modelo, aunque no genera predicciones incorrectas frecuentemente, está fallando en detectar la mayoría de los casos reales de esa clase.

En el contexto de este problema, donde existen 44 clases con distribución desigual de muestras, el F1-score por clase es especialmente útil para identificar categorías en las que el modelo presenta un comportamiento deficiente de forma integral, sin que dicha deficiencia quede oculta por un valor alto en solo una de las dos métricas componentes.

Para obtener un indicador global, se utiliza la versión ponderada: el F1-score de cada clase se promedia proporcionalmente al support de esa clase en el conjunto de prueba, de la misma forma que la precisión y el recall ponderados descritos anteriormente. Esta versión fue calculada mediante precision_recall_fscore_support(y_true, y_pred, average="weighted") de scikit-learn.

5.9.6 Coeficiente de Correlación de Matthews (Matthews Correlation Coefficient, MCC)

El MCC responde a una necesidad específica que las métricas anteriores no resuelven completamente: ¿cómo obtener un indicador global de desempeño que sea verdaderamente imparcial cuando las clases tienen tamaños muy distintos?

La exactitud global puede ser engañosa cuando hay clases desbalanceadas, el F1-score ponderado depende de cómo se distribuye el peso entre clases, y tanto la precisión como el recall analizan solo una dimensión del error a la vez. El MCC, en cambio, considera de forma simultánea todos los tipos de aciertos y errores del modelo a través de la matriz de confusión completa, lo que lo convierte en uno de los indicadores más robustos y equilibrados disponibles para problemas de clasificación multiclase.

Intuitivamente, el MCC puede entenderse como una medida de correlación entre las predicciones del modelo y las etiquetas reales: cuantifica en qué medida las predicciones del modelo y los valores verdaderos "se mueven juntos". Un valor de +1 indica predicción perfecta en todas las clases; un valor de 0 equivale a un modelo que predice de forma aleatoria, sin ninguna capacidad discriminativa; y un valor de −1 representa el caso opuesto a la perfección, donde el modelo se equivoca sistemáticamente en todas sus predicciones. En la práctica, valores

por encima de 0.8 se consideran indicativos de un desempeño muy sólido en tareas de clasificación complejas.

Su formulación general para el caso multiclase es:

$$\textit{Coeficiente de correlación de Matthews (MCC)} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

TP: Verdaderos Positivos → Correctamente clasificados como positivos.

TN: Verdaderos Negativos → Correctamente clasificados como negativos.

FP: Falsos Positivos → Incorrectamente clasificados como positivos.

FN: Falsos Negativos → Incorrectamente clasificados como negativos.

La explicación de la fórmula es la siguiente: el numerador mide la diferencia entre los aciertos y los errores del modelo considerando todas las combinaciones de clases posibles, mientras que el denominador normaliza ese valor para que el resultado siempre se encuentre en el rango [−1,+1], independientemente del número de clases o de la distribución del dataset.

A diferencia del F1-score ponderado, que asigna mayor peso a las clases con más muestras, el MCC trata los errores de todas las clases de forma más equitativa, lo que lo hace especialmente valioso en este problema donde coexisten clases con decenas de muestras de prueba y clases con apenas dos o tres. Por esta razón, el MCC es utilizado en esta investigación como el indicador de referencia principal para la selección del mejor modelo, complementando al F1-score y a la exactitud global para proporcionar una visión completa y confiable del desempeño del sistema.

5.9.7 Resultados globales

La Tabla 20 presenta el resumen consolidado de las seis métricas de evaluación obtenidas por los cinco modelos sobre el conjunto de prueba.

Tabla 21. *Resultados globales de los modelos evaluados.*

Modelo	Accuracy	Top-3 Accuracy	Precisión	Recall	F1-score	MCC
EfficientNet V2 B0	0.9252	0.9777	0.9311	0.9252	0.925	0.9222
Vision Transformer (ViT-B16)	0.8237	0.9498	0.8482	0.8237	0.8233	0.8172
Ensamble (promedio simple)	0.9196	0.9754	0.9294	0.9196	0.9194	0.9166
Ensamble (promedio ponderado)	0.933	0.9754	0.939	0.933	0.9327	0.9304
Ensamble (media geométrica)	0.9263	0.9732	0.9342	0.9263	0.9261	0.9235

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

Todas las métricas fueron calculadas sobre el conjunto de prueba (896 muestras, 44 clases). Precision, Recall y F1-score corresponden a la versión ponderada por frecuencia de clase.

El modelo con mayor exactitud top-1 es el ensamble por promedio ponderado con un 93.30%, seguido del modelo EfficientNet V2 B0 individual con 92.52%. En cuanto a la exactitud Top-3, el modelo EfficientNet individual obtiene el valor más alto (97.77%), lo que indica que para casi la totalidad de las muestras del conjunto de prueba, la clase correcta se encuentra dentro de las tres predicciones de mayor probabilidad.

En lo que respecta a la precisión ponderada, el ensamble ponderado obtiene nuevamente el mejor resultado con 0.9390, lo que significa que el 93.90% de todas las predicciones realizadas por este modelo corresponden a la clase correcta, ponderando la contribución de cada clase según su frecuencia en el dataset. Le siguen el ensamble por media geométrica (0.9342), EfficientNet individual (0.9311), el ensamble por promedio simple (0.9294) y, finalmente, el modelo ViT-B16 con el valor más bajo (0.8482).

El comportamiento del recall ponderado refleja exactamente los mismos valores que la exactitud global para todos los modelos, lo cual es una consecuencia directa de que ambas métricas comparten el mismo numerador en este contexto multiclase cuando se calcula el promedio ponderado. El ensamble ponderado lidera con 0.9330, confirmando que es también el modelo que detecta correctamente la mayor proporción de muestras reales de cada clase, considerando su peso relativo en el dataset.

El F1-score ponderado, al combinar precisión y recall en un único indicador, consolida la imagen de desempeño de cada modelo: el ensamble por promedio ponderado obtiene (0.9327), seguido de EfficientNet individual (0.9250), el ensamble por media geométrica (0.9261), el ensamble por promedio simple (0.9194) y el modelo ViT-B16 (0.8233). Estos valores son coherentes con los observados en las métricas individuales y confirman que el ensamble ponderado mantiene un equilibrio sólido entre no generar predicciones incorrectas y no omitir casos reales en ninguna de las 44 clases.

Finalmente, el MCC más alto corresponde también al ensamble ponderado (0.9304), lo que confirma que este modelo exhibe el mejor desempeño global considerando todas las clases de forma equilibrada e independientemente del desbalance presente en el dataset, siendo este el indicador de mayor confiabilidad para la comparación entre modelos en este contexto.

5.9.8 Reportes de clasificación por clase

Adicionalmente a las métricas globales, se generaron reportes de clasificación detallados para cada uno de los cinco modelos mediante la función classification_report() de scikit-learn. Estos reportes desglosan las métricas de precisión, recall y F1-score para cada una de las 44 clases del dataset, junto con el support (número de muestras reales por clase), lo que permite identificar aquellas categorías en las que el modelo presenta mayor dificultad para generalizar.

El reporte de clasificación por clase para el modelo EfficientNet V2 B0 muestra un desempeño sólido en la mayoría de las clases. Entre las categorías con F1-score perfecto o cercano a 1.0 se encuentran: Carcinoma T2 (1.00), Ependimoma T1 (1.00), Gangliоglioma T1 (1.00), Germinoma T1 (1.00), Neurocitoma T1C+ (0.97), Schwannoma T1 (0.98), Oligodendroglioma T1 (1.00) y NORMAL T1 (0.97). Por otro lado, las clases con menor desempeño corresponden principalmente a categorías con pocos ejemplos de prueba, como Gangliоglioma T2 (F1=0.40, support=3), Germinoma T2 (F1=0.72) y _NORMAL T2 (F1=0.88), donde el desbalance de clases y la escasez de muestras dificultan la generalización. En términos

de promedios globales, el modelo obtiene: accuracy=0.93, macro avg* F1=0.91 y weighted avg F1=0.92. Los resultados se muestran en la Tabla 22.

Tabla 22. Reporte de clasificación por clase para el Modelo EfficientNet V2 B0

Clase	Precision	Recall	F1-score	Support
Astrocitoma T1	0.97	0.94	0.96	36
Astrocitoma T1C+	0.95	0.91	0.93	44
Astrocitoma T2	0.95	0.72	0.82	29
Carcinoma T1	1.00	0.87	0.93	15
Carcinoma T1C+	1.00	0.86	0.93	22
Carcinoma T2	1.00	1.00	1.00	18
Ependimoma T1	1.00	1.00	1.00	4
Ependimoma T1C+	1.00	0.86	0.92	7
Ependimoma T2	0.87	0.87	0.87	15
Ganglioglioma T1	1.00	1.00	1.00	2
Ganglioglioma T1C+	1.00	1.00	1.00	7
Ganglioglioma T2	0.50	0.33	0.40	3
Germinoma T1	1.00	1.00	1.00	2
Germinoma T1C+	0.78	1.00	0.88	7
Germinoma T2	0.67	1.00	0.80	10
Glioblastoma T1	1.00	0.83	0.91	6
Glioblastoma T1C+	0.94	0.89	0.92	19
Glioblastoma T2	1.00	0.73	0.84	11
Granuloma T1	1.00	1.00	1.00	9
Granuloma T1C+	1.00	0.89	0.94	9
Granuloma T2	1.00	0.50	0.67	4
Meduloblastoma T1	0.83	1.00	0.91	5
Meduloblastoma T1C+	1.00	0.94	0.97	16
Meduloblastoma T2	1.00	0.83	0.91	6
Meningioma T1	0.92	0.95	0.94	64
Meningioma T1C+	0.95	0.92	0.93	76
Meningioma T2	0.91	0.87	0.89	46
Neurocitoma T1	0.93	1.00	0.96	26
Neurocitoma T1C+	0.95	1.00	0.97	38
Neurocitoma T2	0.92	0.96	0.94	23
Oligodendroglioma T1	1.00	1.00	1.00	12

Oligodendroglioma T1C+	0.95	0.95	0.95	20
Oligodendroglioma T2	1.00	0.89	0.94	9
Papiloma T1	1.00	0.92	0.96	12
Papiloma T1C+	1.00	0.95	0.97	19
Papiloma T2	1.00	0.92	0.96	13
Schwannoma T1	1.00	1.00	1.00	29
Schwannoma T1C+	0.92	1.00	0.96	46
Schwannoma T2	0.90	0.90	0.90	30
Tuberculoma T1	1.00	1.00	1.00	4
Tuberculoma T1C+	1.00	0.89	0.94	9
Tuberculoma T2	1.00	1.00	1.00	8
_NORMAL T1	0.97	1.00	0.98	57
_NORMAL T2	0.77	1.00	0.87	49
accuracy			**0.93**	**896**
macro avg	**0.94**	**0.91**	**0.92**	**896**
weighted avg	**0.94**	**0.93**	**0.93**	**896**

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

Por otro lado, el reporte para el modelo ViT-B16 refleja un desempeño inferior al de EfficientNet en la mayoría de las clases, con mayor dificultad especialmente en categorías de bajo support. Las clases más problemáticas incluyen: Oligodendroglioma T2 (precision=0.50, recall=0.11, F1=0.18, support=9), Ganglioglioma T1 (F1=0.40), Germinoma T1 (precision=0.25, recall=1.00, F1=0.40) y Ganglioglioma T2 (F1=0.50). Este comportamiento es consistente con la literatura sobre Vision Transformers, ya que al tener un sesgo inductivo más débil que las CNN, los modelos ViT son más dependientes del volumen de datos de entrenamiento y tienden a presentar mayor variabilidad en la clasificación de categorías con pocas muestras. El reporte global del modelo ViT arroja: accuracy=0.82, macro avg F1=0.77 y weighted avg F1=0.82. Los resultados se muestran en la Tabla 23.

Tabla 23. Reporte de clasificación por clase para el Modelo Vision Transformer ViT-B16

Clase	Precision	Recall	F1-score	Support
Astrocitoma T1	0.94	0.86	0.90	36
Astrocitoma T1C+	0.90	0.80	0.84	44
Astrocitoma T2	0.86	0.62	0.72	29

Carcinoma T1	0.87	0.87	0.87	15
Carcinoma T1C+	1.00	0.77	0.87	22
Carcinoma T2	0.94	0.94	0.94	18
Ependimoma T1	0.67	1.00	0.80	4
Ependimoma T1C+	0.67	0.86	0.75	7
Ependimoma T2	1.00	0.60	0.75	15
Ganglioglioma T1	0.33	0.50	0.40	2
Ganglioglioma T1C+	1.00	0.71	0.83	7
Ganglioglioma T2	1.00	0.33	0.50	3
Germinoma T1	0.25	1.00	0.40	2
Germinoma T1C+	0.50	0.86	0.63	7
Germinoma T2	0.60	0.90	0.72	10
Glioblastoma T1	0.83	0.83	0.83	6
Glioblastoma T1C+	0.94	0.84	0.89	19
Glioblastoma T2	1.00	0.55	0.71	11
Granuloma T1	0.64	1.00	0.78	9
Granuloma T1C+	1.00	1.00	1.00	9
Granuloma T2	0.67	0.50	0.57	4
Meduloblastoma T1	1.00	0.80	0.89	5
Meduloblastoma T1C+	0.82	0.88	0.85	16
Meduloblastoma T2	0.56	0.83	0.67	6
Meningioma T1	0.85	0.81	0.83	64
Meningioma T1C+	0.91	0.89	0.90	76
Meningioma T2	0.72	0.63	0.67	46
Neurocitoma T1	0.92	0.88	0.90	26
Neurocitoma T1C+	0.85	0.92	0.89	38
Neurocitoma T2	0.81	0.96	0.88	23
Oligodendroglioma T1	0.92	1.00	0.96	12
Oligodendroglioma T1C+	0.90	0.95	0.93	20
Oligodendroglioma T2	0.50	0.11	0.18	9
Papiloma T1	0.90	0.75	0.82	12
Papiloma T1C+	0.89	0.89	0.89	19
Papiloma T2	0.78	0.54	0.64	13
Schwannoma T1	0.84	0.93	0.89	29

Schwannoma T1C+	0.98	0.87	0.92	46
Schwannoma T2	0.90	0.63	0.75	30
Tuberculoma T1	1.00	0.75	0.86	4
Tuberculoma T1C+	0.75	0.67	0.71	9
Tuberculoma T2	0.80	0.50	0.62	8
_NORMAL T1	0.87	0.93	0.90	57
_NORMAL T2	0.56	0.98	0.71	49
accuracy			**0.82**	**896**
macro avg	**0.81**	**0.78**	**0.77**	**896**
weighted avg	**0.85**	**0.82**	**0.82**	**896**

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

En el caso del ensamble por promedio simple, el reporte muestra un desempeño similar al de EfficientNet individual en la mayoría de las clases, aunque con algunas mejoras y retrocesos puntuales respecto a ambos modelos base. Entre las clases con F1-score perfecto (1.00) se encuentran: Ganglioglioma T1, Ganglioglioma T1C+, Germinoma T1, Granuloma T1, Granuloma T1C+, Meduloblastoma T1, Oligodendroglioma T1, Schwannoma T1, Schwannoma T1C+ y Tuberculoma T1. Las clases con mayor dificultad continúan siendo aquellas con bajo support: Ganglioglioma T2 (F1=0.50, support=3), Granuloma T2 (F1=0.67) y Ependimoma T2 (F1=0.74). Este último resultado es notable, ya que EfficientNet obtuvo F1=0.80 para Ependimoma T2, lo que evidencia que la contribución sin ponderación del modelo ViT-B16, cuyo desempeño en esa clase es inferior, arrastra hacia abajo el resultado combinado. En términos globales, el ensamble por promedio simple obtiene: accuracy=0.92, macro avg F1=0.90 y weighted avg F1=0.92. Los resultados se muestran en la Tabla 24.

Tabla 24. Reporte de clasificación por clase para el Modelo Ensamble por Promedio Simple

Clase	Precision	Recall	F1-score	Support
Astrocitoma T1	0.97	0.89	0.93	36
Astrocitoma T1C+	0.95	0.91	0.93	44
Astrocitoma T2	0.95	0.72	0.82	29
Carcinoma T1	1.00	0.87	0.93	15
Carcinoma T1C+	1.00	0.86	0.93	22
Carcinoma T2	1.00	0.94	0.97	18
Ependimoma T1	0.80	1.00	0.89	4

Ependimoma T1C+	1.00	0.86	0.92	7
Ependimoma T2	0.83	0.67	0.74	15
Ganglioglioma T1	1.00	1.00	1.00	2
Ganglioglioma T1C+	1.00	1.00	1.00	7
Ganglioglioma T2	1.00	0.33	0.50	3
Germinoma T1	1.00	1.00	1.00	2
Germinoma T1C+	0.70	1.00	0.82	7
Germinoma T2	0.62	1.00	0.77	10
Glioblastoma T1	1.00	0.83	0.91	6
Glioblastoma T1C+	0.94	0.89	0.92	19
Glioblastoma T2	1.00	0.73	0.84	11
Granuloma T1	1.00	1.00	1.00	9
Granuloma T1C+	1.00	1.00	1.00	9
Granuloma T2	1.00	0.50	0.67	4
Meduloblastoma T1	1.00	1.00	1.00	5
Meduloblastoma T1C+	0.94	0.94	0.94	16
Meduloblastoma T2	0.71	0.83	0.77	6
Meningioma T1	0.91	0.95	0.93	64
Meningioma T1C+	0.95	0.92	0.93	76
Meningioma T2	0.86	0.83	0.84	46
Neurocitoma T1	0.93	1.00	0.96	26
Neurocitoma T1C+	0.95	1.00	0.97	38
Neurocitoma T2	0.88	0.96	0.92	23
Oligodendroglioma T1	1.00	1.00	1.00	12
Oligodendroglioma T1C+	0.95	0.95	0.95	20
Oligodendroglioma T2	1.00	0.67	0.80	9
Papiloma T1	1.00	0.92	0.96	12
Papiloma T1C+	1.00	0.95	0.97	19
Papiloma T2	0.92	0.92	0.92	13
Schwannoma T1	1.00	1.00	1.00	29
Schwannoma T1C+	1.00	1.00	1.00	46
Schwannoma T2	0.93	0.87	0.90	30
Tuberculoma T1	1.00	1.00	1.00	4
Tuberculoma T1C+	1.00	0.89	0.94	9

Tuberculoma T2	0.86	0.75	0.80	8
_NORMAL T1	0.93	1.00	0.97	57
_NORMAL T2	0.71	1.00	0.83	49
accuracy			**0.92**	**896**
macro avg	**0.94**	**0.89**	**0.90**	**896**
weighted avg	**0.93**	**0.92**	**0.92**	**896**

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

En el caso del ensamble por promedio ponderado, el reporte refleja el mejor desempeño por clase de los cinco modelos evaluados, con el mayor número de categorías en F1=1.00: Carcinoma T2, Ependimoma T1, Ganglioglioma T1, Ganglioglioma T1C+, Germinoma T1, Granuloma T1, Granuloma T1C+, Oligodendroglioma T1, Schwannoma T1, Tuberculoma T1 y Tuberculoma T2. Se destaca especialmente la mejora respecto al modelo EfficientNet individual en las clases que este tenía dificultades: Oligodendroglioma T2 sube de F1=0.80 a F1=0.94, Ependimoma T2 de 0.80 a 0.87, y Meduloblastoma T2 de 0.80 a 0.91. Estas mejoras se atribuyen a que el modelo ViT-B16, aunque es globalmente inferior, logra clasificar correctamente algunas de estas muestras específicas que EfficientNet clasifica mal, y la ponderación 0.6/0.4 permite que esa información complementaria influya en la predicción final sin comprometer el desempeño global. La clase con mayor dificultad sigue siendo Ganglioglioma T2 (F1=0.40, support=3), lo que refleja la limitación inherente al bajo número de muestras disponibles para esa categoría. En términos globales, el ensamble ponderado obtiene: accuracy=0.93, macro avg F1=0.92 y weighted avg F1=0.93, siendo este último el valor más alto entre todos los modelos. Los resultados se presentan en la Tabla 25.

Tabla 25. Reporte de clasificación por clase para el Modelo Ensamble por Promedio Ponderado

Clase	Precision	Recall	F1-score	Support
Astrocitoma T1	0.97	0.94	0.96	36
Astrocitoma T1C+	0.95	0.91	0.93	44
Astrocitoma T2	0.95	0.72	0.82	29
Carcinoma T1	1.00	0.87	0.93	15
Carcinoma T1C+	1.00	0.86	0.93	22
Carcinoma T2	1.00	1.00	1.00	18
Ependimoma T1	1.00	1.00	1.00	4
Ependimoma T1C+	1.00	0.86	0.92	7
Ependimoma T2	0.87	0.87	0.87	15

Ganglioglioma T1	1.00	1.00	1.00	2
Ganglioglioma T1C+	1.00	1.00	1.00	7
Ganglioglioma T2	0.50	0.33	0.40	3
Germinoma T1	1.00	1.00	1.00	2
Germinoma T1C+	0.78	1.00	0.88	7
Germinoma T2	0.67	1.00	0.80	10
Glioblastoma T1	1.00	0.83	0.91	6
Glioblastoma T1C+	0.94	0.89	0.92	19
Glioblastoma T2	1.00	0.73	0.84	11
Granuloma T1	1.00	1.00	1.00	9
Granuloma T1C+	1.00	0.89	0.94	9
Granuloma T2	1.00	0.50	0.67	4
Meduloblastoma T1	0.83	1.00	0.91	5
Meduloblastoma T1C+	1.00	0.94	0.97	16
Meduloblastoma T2	1.00	0.83	0.91	6
Meningioma T1	0.92	0.95	0.94	64
Meningioma T1C+	0.95	0.92	0.93	76
Meningioma T2	0.91	0.87	0.89	46
Neurocitoma T1	0.93	1.00	0.96	26
Neurocitoma T1C+	0.95	1.00	0.97	38
Neurocitoma T2	0.92	0.96	0.94	23
Oligodendroglioma T1	1.00	1.00	1.00	12
Oligodendroglioma T1C+	0.95	0.95	0.95	20
Oligodendroglioma T2	1.00	0.89	0.94	9
Papiloma T1	1.00	0.92	0.96	12
Papiloma T1C+	1.00	0.95	0.97	19
Papiloma T2	1.00	0.92	0.96	13
Schwannoma T1	1.00	1.00	1.00	29
Schwannoma T1C+	0.92	1.00	0.96	46
Schwannoma T2	0.90	0.90	0.90	30
Tuberculoma T1	1.00	1.00	1.00	4
Tuberculoma T1C+	1.00	0.89	0.94	9
Tuberculoma T2	1.00	1.00	1.00	8
_NORMAL T1	0.97	1.00	0.98	57

_NORMAL T2	0.77	1.00	0.87	49
accuracy			**0.93**	**896**
macro avg	**0.94**	**0.91**	**0.92**	**896**
weighted avg	**0.94**	**0.93**	**0.93**	**896**

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

Finalmente, en el caso del ensamble por media geométrica, el reporte muestra un desempeño muy cercano al del ensamble por promedio ponderado, aunque hay algunas diferencias en clases específicas. Entre las categorías con F1-score perfecto o cercano a 1.0 se encuentran: Ependimoma T1, Ganglioglioma T1C+, Germinoma T1, Granuloma T1, Granuloma T1C+, Schwannoma T1C+ (F1=0.99) y Tuberculoma T1. Una diferencia notable respecto al ensamble ponderado es el comportamiento en Ganglioglioma T1 (support=2), donde la media geométrica obtiene F1=0.67 frente al F1=1.00 del ensamble por promedio ponderado. Esto sugiere que la atenuación de valores extremos propia de la media geométrica penaliza en este caso particular, donde uno de los modelos base asigna probabilidades muy bajas a la clase correcta. Por otra parte, Granuloma T2 mejora a F1=0.86 respecto al F1=0.67 del promedio simple, lo que evidencia que la media geométrica produce combinaciones más estables en ciertas clases con alta variabilidad entre modelos. En términos globales, el ensamble por media geométrica obtiene: accuracy=0.93, macro avg F1=0.91 y weighted avg F1=0.93. Los resultados se muestran en la Tabla 26.

Tabla 26. Reporte de clasificación por clase para el Modelo Ensamble por Media Geométrica

Clase	Precision	Recall	F1-score	Support
Astrocitoma T1	0.97	0.92	0.94	36
Astrocitoma T1C+	0.95	0.93	0.94	44
Astrocitoma T2	1.00	0.76	0.86	29
Carcinoma T1	1.00	0.87	0.93	15
Carcinoma T1C+	1.00	0.86	0.93	22
Carcinoma T2	1.00	0.94	0.97	18
Ependimoma T1	1.00	1.00	1.00	4
Ependimoma T1C+	1.00	0.86	0.92	7
Ependimoma T2	0.86	0.80	0.83	15
Ganglioglioma T1	1.00	0.50	0.67	2
Ganglioglioma T1C+	1.00	1.00	1.00	7
Ganglioglioma T2	1.00	0.33	0.50	3

Germinoma T1	1.00	1.00	1.00	2
Germinoma T1C+	0.70	1.00	0.82	7
Germinoma T2	0.71	1.00	0.83	10
Glioblastoma T1	1.00	0.83	0.91	6
Glioblastoma T1C+	0.94	0.89	0.92	19
Glioblastoma T2	1.00	0.73	0.84	11
Granuloma T1	1.00	1.00	1.00	9
Granuloma T1C+	1.00	1.00	1.00	9
Granuloma T2	1.00	0.75	0.86	4
Meduloblastoma T1	0.83	1.00	0.91	5
Meduloblastoma T1C+	0.88	0.94	0.91	16
Meduloblastoma T2	0.71	0.83	0.77	6
Meningioma T1	0.90	0.95	0.92	64
Meningioma T1C+	0.96	0.92	0.94	76
Meningioma T2	0.89	0.89	0.89	46
Neurocitoma T1	0.93	1.00	0.96	26
Neurocitoma T1C+	0.95	1.00	0.97	38
Neurocitoma T2	0.88	0.96	0.92	23
Oligodendroglioma T1	0.92	1.00	0.96	12
Oligodendroglioma T1C+	0.95	0.95	0.95	20
Oligodendroglioma T2	1.00	0.78	0.88	9
Papiloma T1	1.00	0.92	0.96	12
Papiloma T1C+	1.00	0.95	0.97	19
Papiloma T2	0.86	0.92	0.89	13
Schwannoma T1	0.93	0.97	0.95	29
Schwannoma T1C+	1.00	0.98	0.99	46
Schwannoma T2	0.92	0.80	0.86	30
Tuberculoma T1	1.00	1.00	1.00	4
Tuberculoma T1C+	1.00	0.89	0.94	9
Tuberculoma T2	1.00	0.88	0.93	8
_NORMAL T1	0.97	1.00	0.98	57
_NORMAL T2	0.75	1.00	0.86	49
accuracy			**0.93**	**896**
macro avg	**0.94**	**0.90**	**0.91**	**896**

weighted avg	**0.93**	**0.93**	**0.93**	**896**

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

5.9.9 Matrices de confusión

Para cada uno de los cinco modelos se generó una matriz de confusión de dimensiones 44×44, que permite visualizar de forma gráfica las predicciones realizadas por el modelo comparadas con las etiquetas reales. Cada celda (x, y) de la matriz indica el número de muestras de la clase real x que el modelo clasificó como clase y. X serían las filas mientras que Y serían las columnas en la matriz. La diagonal principal corresponde a las predicciones correctas, mientras que los valores fuera de la diagonal representan errores de clasificación.

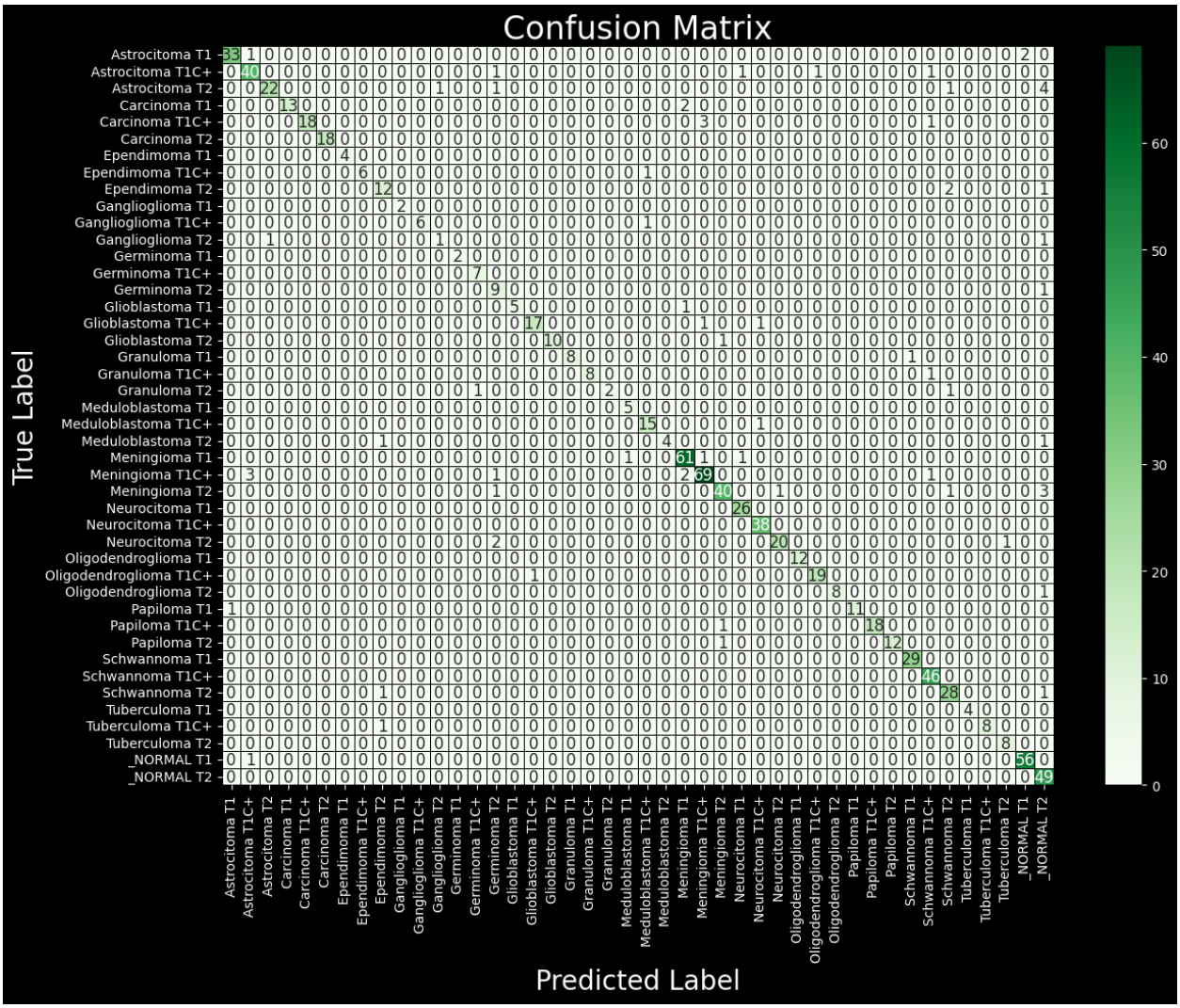
Las matrices de confusión se generaron utilizando la función personalizada plot_confusion_matrix() del notebook, que internamente utiliza confusion_matrix() de scikit-learn y visualiza el resultado como un mapa de calor. Estas visualizaciones permiten identificar patrones de confusión sistemáticos entre clases específicas como, por ejemplo, clases morfológicamente similares o que comparten el mismo tipo de secuencia MRI (T1, T1C+ o T2), lo cual es información valiosa para el análisis clínico y para guiar posibles mejoras futuras en el modelo.

La matriz del modelo EfficientNet V2 B0 muestra un desempeño sólido en la mayoría de las clases. Las clases con mayor número de predicciones correctas en la diagonal incluyen Meningioma T1 (61), Meningioma T1C+ (69), Meningioma T2 (40), Schwannoma T1C+ (46), NORMAL T1 (56) y _NORMAL T2 (49), lo que sugiere que el modelo aprendió representaciones robustas para estas categorías, que suelen ser las de mayor frecuencia en el conjunto de datos.

Se observan errores dispersos en clases con menor representación, como Ganglioglioma, Ependimoma, Granuloma T2 y Tuberculoma. En el caso particular de Granuloma T2, únicamente 2 muestras fueron clasificadas correctamente de forma consistente en todos los modelos, lo que refleja tanto la dificultad intrínseca de esta clase como el reducido número de muestras disponibles. Astrocitoma T2 registra cuatro errores de predicción hacia la clase NORMAL T2, lo que sugiere que el modelo encuentra similitud visual entre ambas categorías en la secuencia T2.

Figura 30

Matriz de confusión del Modelo EfficientNet V2 B0



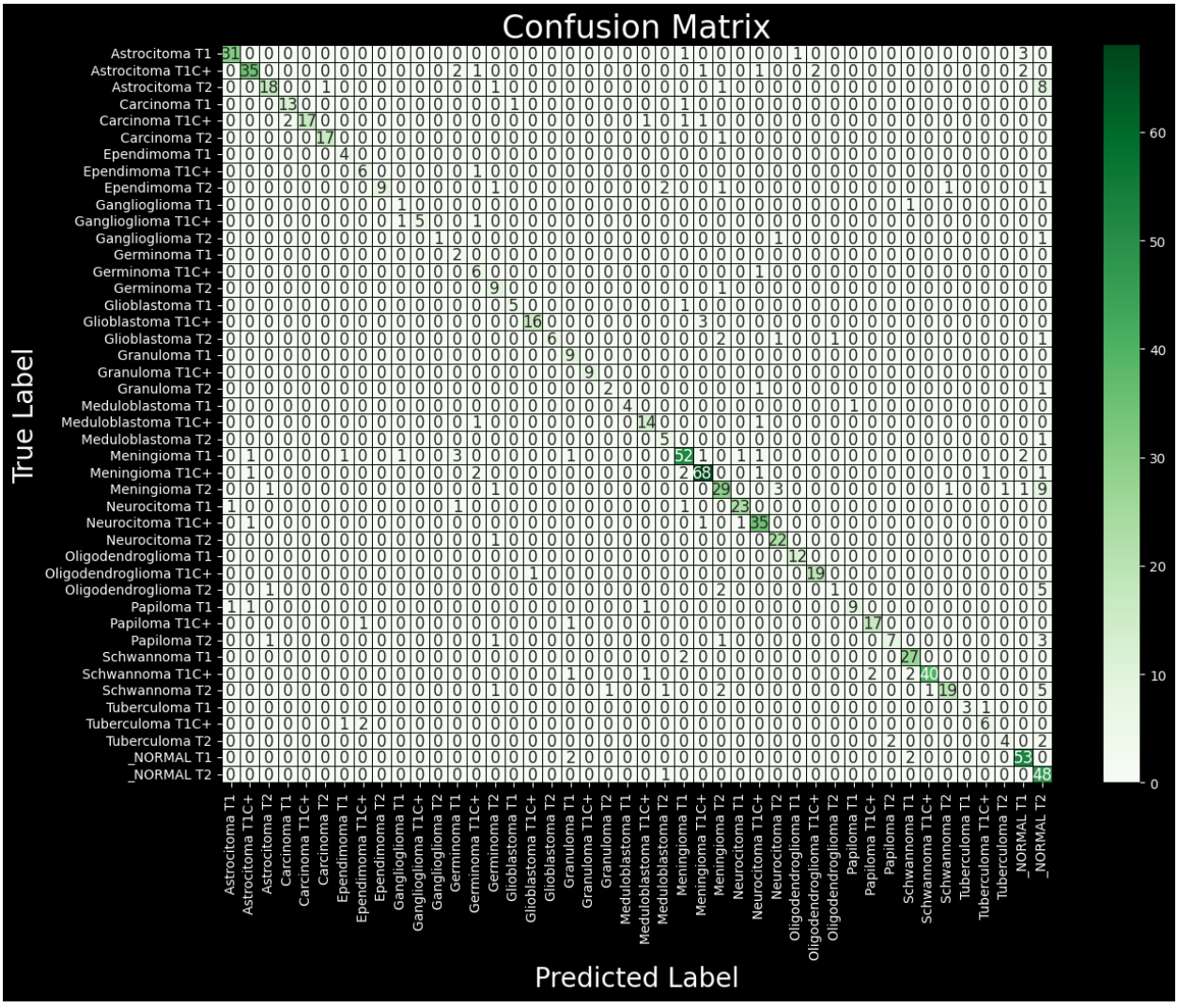
Nota. Elaboración propia a partir de los resultados obtenidos en el experimento.

La matriz del modelo ViT-B16 presenta una distribución diagonal similar a la de EfficientNet, pero con una mayor dispersión de errores en varias clases clave. Comparado con EfficientNet, ViT-B16 clasifica correctamente menos instancias en clases como Astrocitoma T1C+ (35 vs. 40), Meningioma T1 (52 vs. 61), Meningioma T2 (29 vs. 40), Schwannoma T1C+ (40 vs. 46) y Schwannoma T2 (19 vs. 28). Estas diferencias son consistentes con el menor desempeño global de ViT-B16 reportado en las métricas agregadas.

A pesar de esto, ViT-B16 mantiene un nivel de precisión aceptable en clases bien representadas como Meningioma T1C+ (68) y NORMAL T2 (48). La mayor dispersión de errores en ViT-B16 se concentra especialmente en las variantes T2 de múltiples tumores, lo que podría indicar que la arquitectura basada en transformadores requeriría más ejemplos de entrenamiento para capturar patrones visuales propios de esa secuencia de imagen.

Figura 31

Matriz de confusión del Modelo Vision Transformer ViT-B16



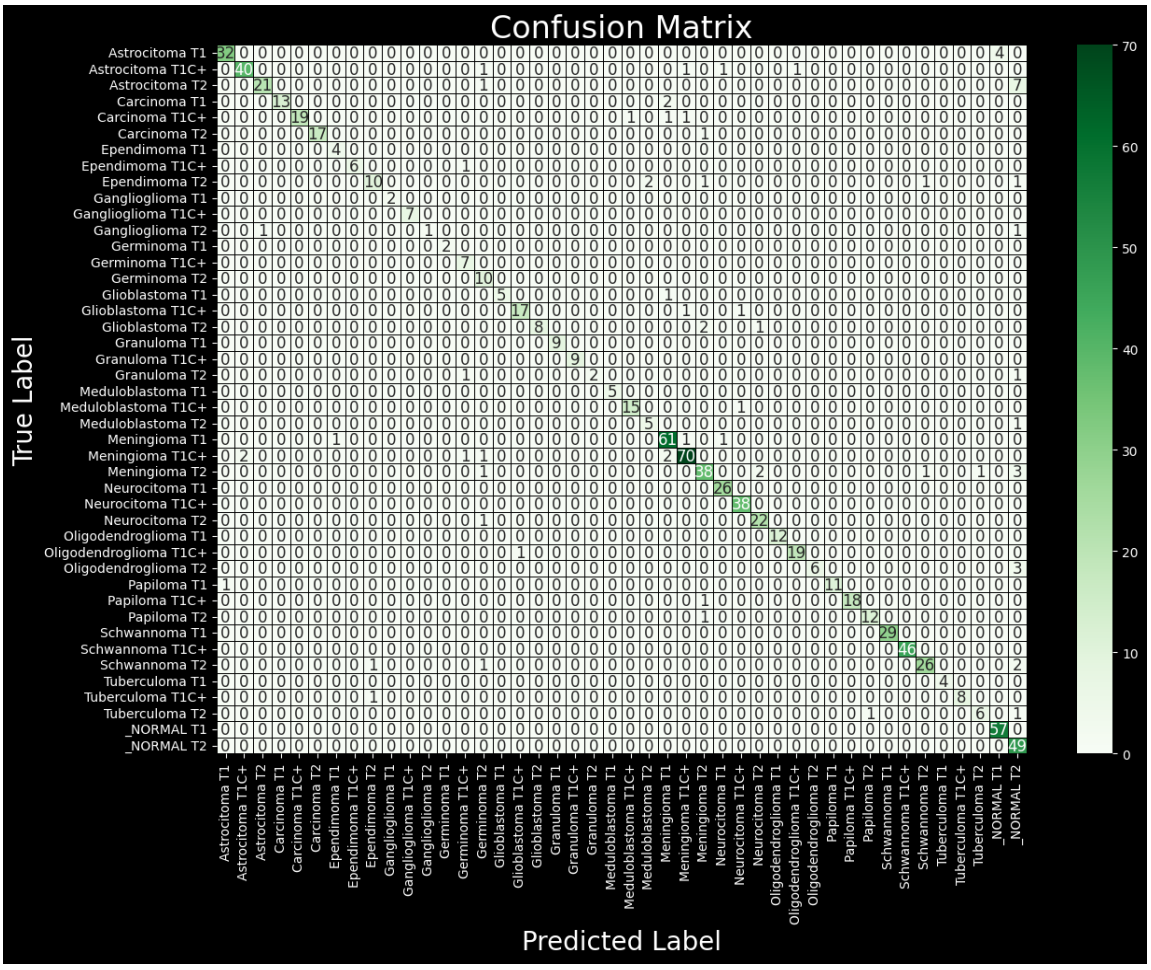
Nota. Elaboración propia a partir de los resultados obtenidos en el experimento.

La matriz del ensamble por promedio simple muestra un patrón muy similar al de EfficientNet V2 B0, pero con ligeras mejoras en algunas clases. Por ejemplo, Meningioma T1C+ alcanza 70 predicciones correctas (frente a 69 de EfficientNet) y NORMAL T1 sube a 57 (frente a 56). Estas mejoras, aunque son marginales, evidencian que la combinación de probabilidades de ambos modelos logra compensar algunos errores individuales.

Las clases con mayor dificultad (Ganglioglioma, Ependimoma T1, Granuloma T2, Tuberculoma) continúan presentando errores similares, lo que indica que el promedio simple no es suficiente para resolver las confusiones estructurales en clases con poca representación o alta variabilidad visual.

Figura 32

Matriz de confusión del Modelo de Ensamble por Promedio Simple



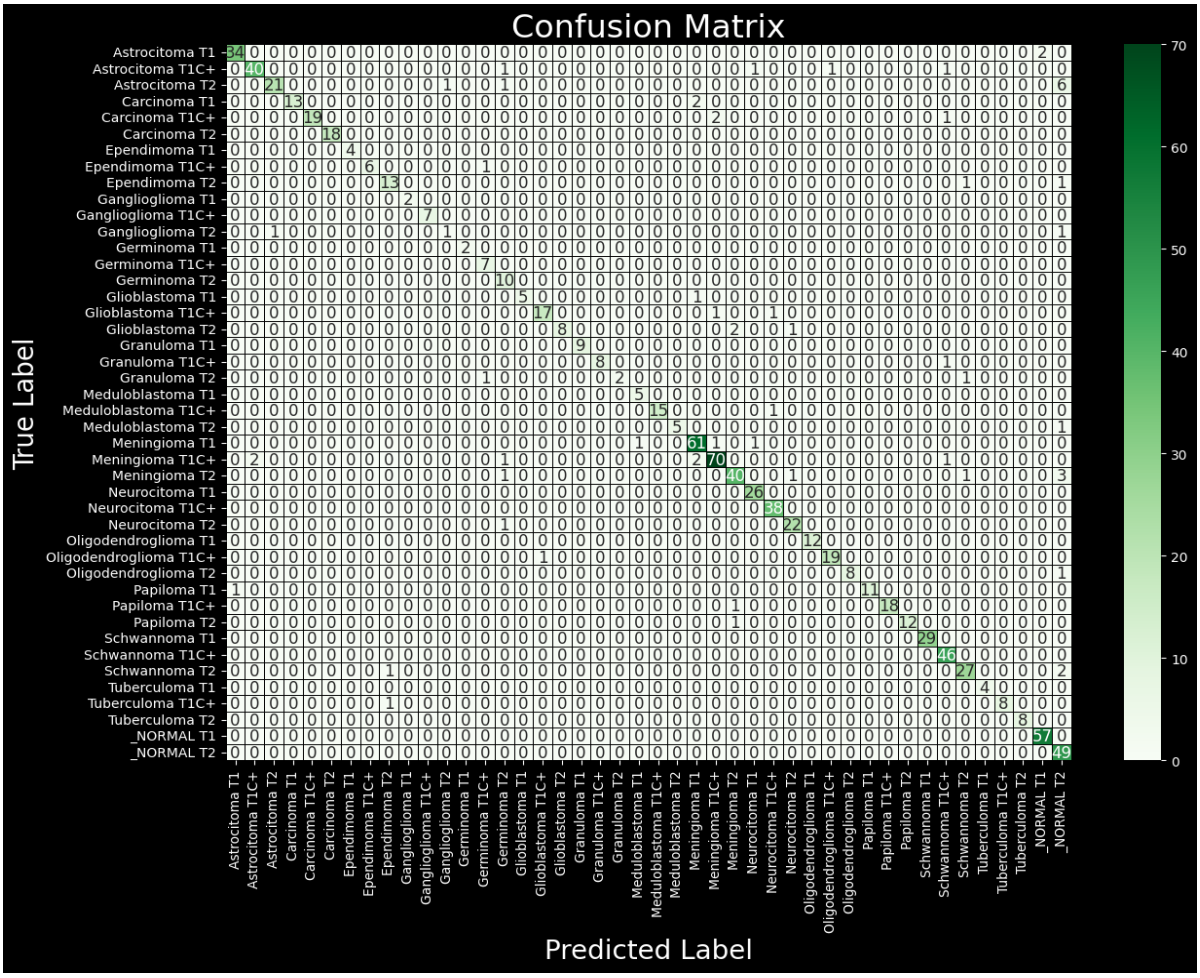
Nota. Elaboración propia a partir de los resultados obtenidos en el experimento.

La matriz del ensamble con pesos (0.6 para EfficientNet, 0.4 para ViT-B16) es la que más se asemeja a EfficientNet V2 B0 de manera individual, lo cual es esperable dado que EfficientNet tiene mayor peso en la combinación. Se observan mejoras puntuales en Astrocitoma T1 (54 vs. 53 de EfficientNet), Meningioma T1C+ (70) y NORMAL T1 (57).

La ponderación mayor hacia EfficientNet le permite al ensamble preservar las fortalezas de ese modelo mientras que la contribución del ViT reduce ligeramente algunos errores específicos. El resultado es una matriz prácticamente idéntica al promedio simple, lo que sugiere que la diferencia entre ambos métodos de combinación es mínima en este conjunto de datos.

Figura 33

Matriz de confusión del Modelo de Ensamble por Promedio Ponderado



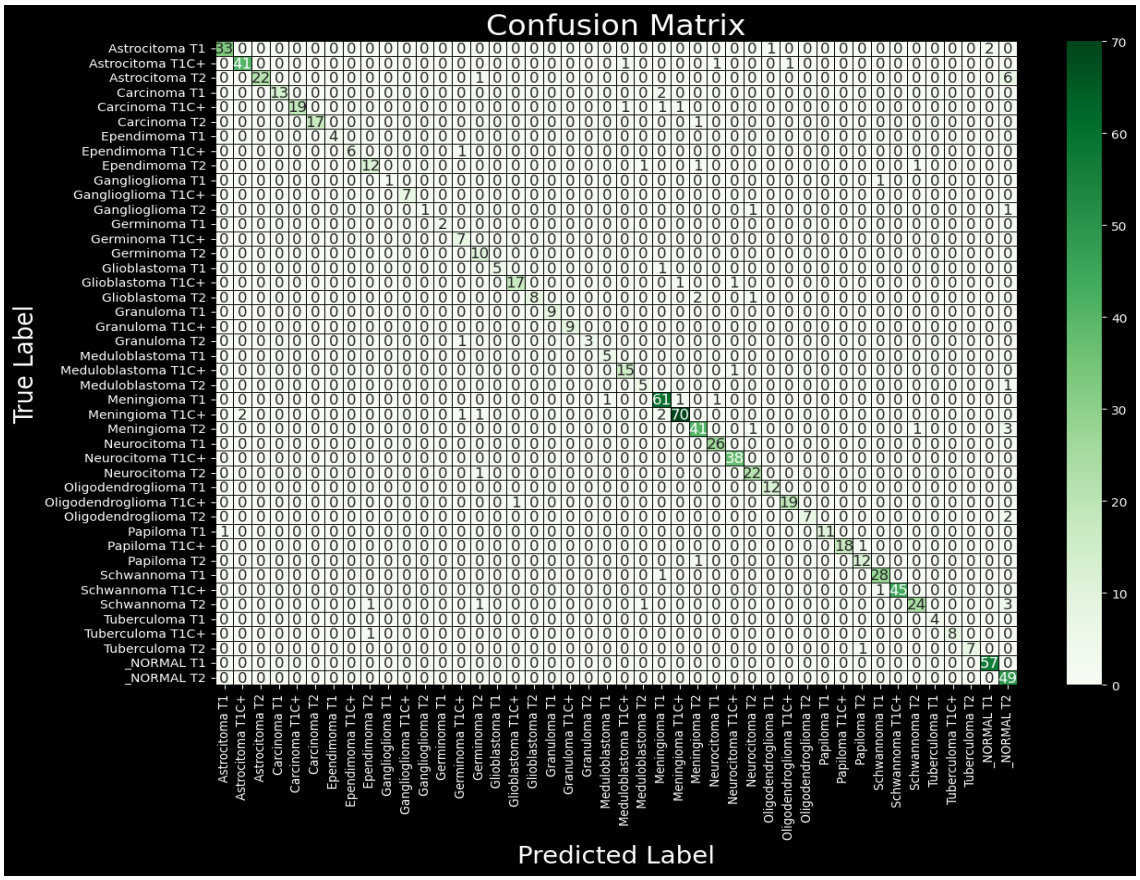
Nota. Elaboración propia a partir de los resultados obtenidos en el experimento.

La matriz del ensamble por media geométrica presenta la diagonal más consistente entre los modelos evaluados en clases de alta frecuencia. Se destacan Astrocitoma T1C+ (41, el más alto entre todos los modelos), Meningioma T1C+ (70), Meningioma T2 (41) y NORMAL T1 (57). En contraste, algunas clases de menor representación como Schwannoma T2 (24) muestran una ligera disminución respecto a EfficientNet (28), lo que podría explicarse por la sensibilidad de la media geométrica ante valores de probabilidad muy bajos, que pueden suprimir la contribución del modelo más débil.

De forma general, se puede decir que la media geométrica produce predicciones más calibradas al enfatizar la consistencia entre los dos modelos base: cuando ambos coinciden en alta probabilidad para una clase, la media geométrica refuerza esa predicción; cuando uno de los modelos es poco confiante, la media geométrica tiende a penalizar esa clase, lo que puede reducir errores en clases bien aprendidas pero incrementar la incertidumbre en clases difíciles.

Figura 34

Matriz de confusión del Modelo de Ensamble por Media Geométrica



Nota. Elaboración propia a partir de los resultados obtenidos en el experimento.

A partir del análisis conjunto de las cinco matrices, se pueden extraer cuatro hallazgos relevantes:

1. Clases con desempeño consistentemente alto: Meningioma (en sus tres variantes de secuencia), Schwannoma T1 y T1C+, Neurocitoma T1C+, y las clases NORMAL T1 y _NORMAL T2 son clasificadas correctamente por todos los modelos con alta frecuencia, lo que refleja su representación suficiente en el conjunto de entrenamiento y la diferenciabilidad visual de sus características.
2. Clases con dificultad estructural: Ganglioglioma, Ependimoma T1, Granuloma T2 y Tuberculoma presentan errores persistentes en todos los modelos. Estas clases tienen en común un menor número de muestras en el conjunto de datos, lo que limita la generalización de los modelos durante el entrenamiento, independientemente de la arquitectura empleada.
3. Confusión entre secuencias del mismo tipo: En varios tumores se observa confusión entre variantes de imagen del mismo tipo (por ejemplo, T1 clasificado como T1C+). Aunque el tipo de tumor es identificado correctamente, la distinción entre secuencias tiene relevancia clínica, dado que T1 con contraste y T1 sin contraste revelan información fisiológica diferente sobre el comportamiento tumoral. Este patrón está presente en los cinco modelos, sugiriendo que se trata de un reto inherente a la granularidad del problema de clasificación.
4. Confusión entre Astrocitoma T2 y tejido normal: En las cinco matrices se observa de forma consistente que algunas imágenes cuya etiqueta real es Astrocitoma T2 fueron clasificadas erróneamente como _NORMAL T2. Este patrón aparece en todos los modelos, incluyendo los ensambles, lo que indica que no se trata de un error aislado de una arquitectura particular.

   Una posible explicación es que los astrocitomas en secuencia T2, especialmente en etapas tempranas o de bajo grado, pueden presentar características visuales muy similares a las del tejido cerebral sano en esa misma secuencia. A diferencia de la secuencia T1C+, donde el uso de contraste permite resaltar el tumor con mayor claridad, la secuencia T2 no siempre permite distinguir con facilidad entre tejido tumoral y tejido normal.

   Este tipo de error es particularmente importante en un contexto clínico, ya que implica que el modelo clasificaría como sano a un paciente que en realidad

presenta un tumor. Este hallazgo subraya la importancia de no depender exclusivamente de una sola secuencia de resonancia magnética en sistemas de apoyo diagnóstico.

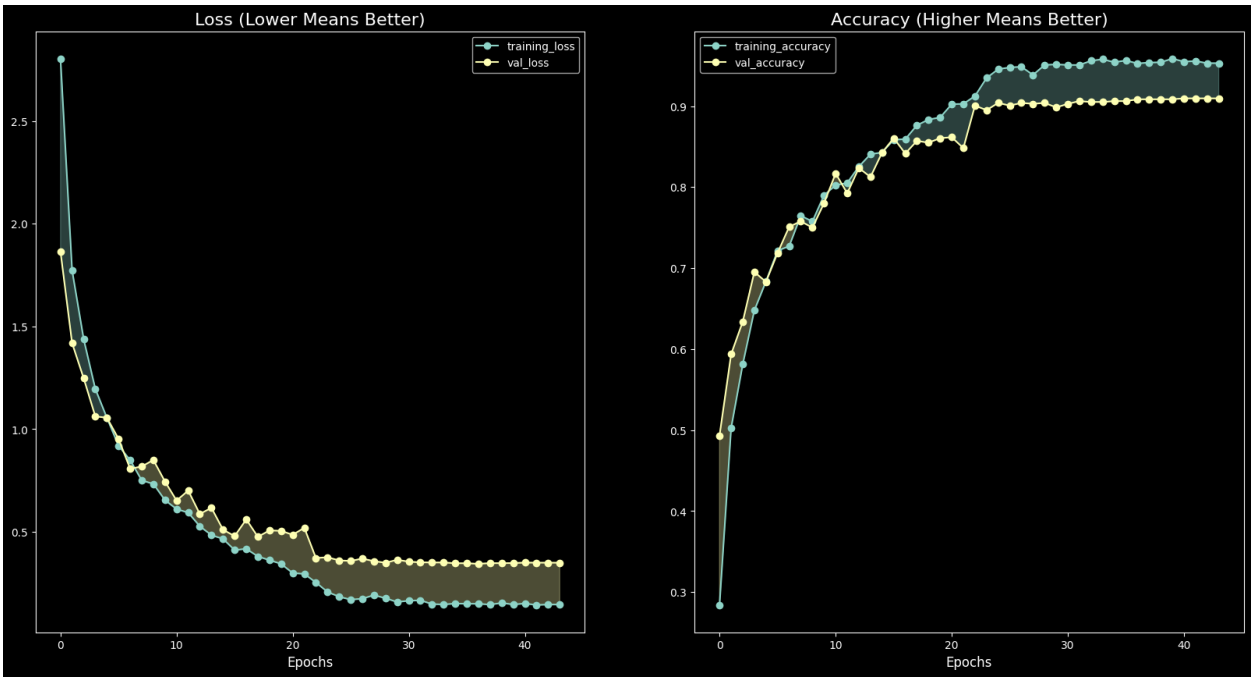
5.9.10 Curvas de historial de entrenamiento

Complementariamente, se graficaron las curvas de historial de entrenamiento para los dos modelos individuales, exactitud y pérdida tanto de entrenamiento como de validación a lo largo de las épocas, con el objetivo de analizar la dinámica del aprendizaje, detectar posibles señales de sobreajuste y verificar la estabilidad de la convergencia.

En el modelo EfficientNet V2 B0, las curvas muestran una separación creciente entre la pérdida de entrenamiento y la de validación en las épocas finales, lo que indica un grado moderado de sobreajuste mitigado por el mecanismo de restauración de mejores pesos.

Figura 35

Evolución de la pérdida y la exactitud en entrenamiento y validación del modelo EfficientNet V2 B0

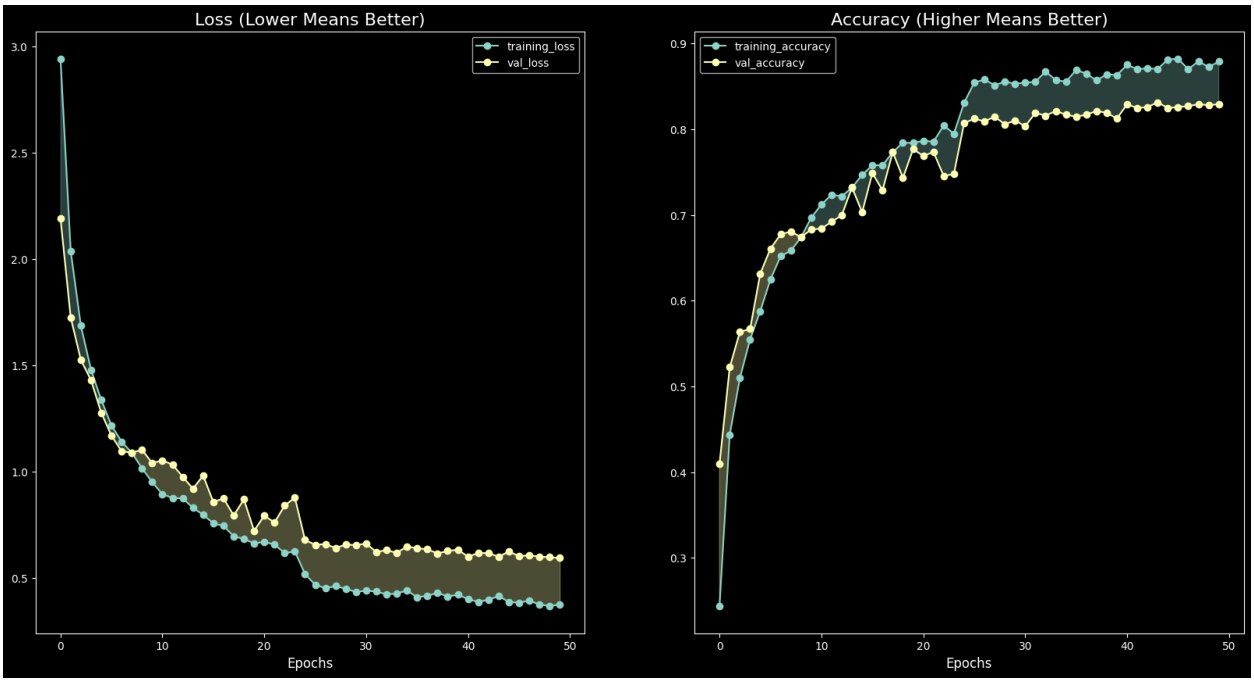


Nota. Elaboración propia a partir del notebook de Jansen (2025), con modificaciones en los hiperparámetros y el proceso de entrenamiento para adaptarlo a un conjunto de datos desbalanceado y con mayor número de clases.

En el modelo ViT-B16, la evolución más lenta y progresiva de ambas curvas refleja la mayor inercia de la arquitectura transformer al aprender sobre el conjunto de datos médico como se muestra en la Figura 36. En ambos casos, las reducciones de la tasa de aprendizaje generadas por el callback ReduceLROnPlateau son claramente visibles como saltos de mejora en las curvas de validación.

Figura 36

Evolución de la pérdida y la exactitud en entrenamiento y validación del modelo ViT-B16



Nota. Elaboración propia a partir del notebook de Jansen (2025), con modificaciones en los hiperparámetros y el proceso de entrenamiento para adaptarlo a un conjunto de datos desbalanceado y con mayor número de clases.

5.10 Comparación de resultados

La comparación sistemática de los resultados obtenidos por los cinco modelos implementados permite identificar cuál configuración ofrece el mejor equilibrio entre precisión, robustez y capacidad de generalización para el problema de clasificación multiclase de tumores cerebrales en imágenes MRI. Este análisis se realiza sobre el conjunto completo de métricas presentadas en el subcapítulo 5.9, considerando tanto el desempeño global agregado como el

comportamiento por clase, con el objetivo de justificar la selección del modelo final sobre la base de evidencia empírica sólida.

5.10.1 Comparación entre los modelos individuales

La primera dimensión de análisis consiste en comparar el desempeño de los dos modelos base, EfficientNet V2 B0 y ViT-B16, para comprender las diferencias arquitectónicas y sus implicaciones prácticas en este dominio de aplicación.

EfficientNet V2 B0 obtuvo una exactitud del 92.52%, un F1-score ponderado de 0.9250 y un MCC de 0.9222. Su exactitud Top-3 fue la más alta de todos los modelos evaluados, con un 97.77%, lo que indica que para prácticamente la totalidad del conjunto de prueba, la clase correcta figura entre las tres predicciones de mayor probabilidad. Estos resultados posicionan a EfficientNet como la arquitectura individual más sólida para este problema, mostrando una excelente capacidad de generalización incluso en clases con bajo support, gracias al fuerte sesgo inductivo convolucional que le permite detectar patrones espaciales locales de forma eficiente.

ViT-B16 obtuvo una exactitud del 82.37%, un F1-score ponderado de 0.8233 y un MCC de 0.8172, resultados que son aproximadamente 10 puntos porcentuales inferiores en todas las métricas respecto a EfficientNet. Esta diferencia de rendimiento, aunque es significativa, consistente con lo reportado en la literatura científica sobre Vision Transformers aplicados a conjuntos de datos de tamaño limitado. A diferencia de las redes convolucionales, los ViT carecen del sesgo inductivo de localidad e invarianza a la traslación que caracteriza a las CNN, lo que los hace más dependientes de grandes volúmenes de datos de preentrenamiento y ajuste fino para alcanzar su potencial máximo. Con un conjunto de entrenamiento de 2.686 imágenes, que es relativamente pequeño para una arquitectura transformer. el modelo ViT-B16 muestra mayor dificultad para generalizar, especialmente en las clases con menor cantidad de muestras, como se evidenció en su reporte de clasificación por clase. No obstante, su exactitud Top-3 del 94.98% sigue siendo alta, lo que sugiere que el modelo sí captura información discriminativa relevante, aunque con menor confianza en la predicción de la clase principal.

La brecha de desempeño entre ambos modelos puede atribuirse también a la naturaleza del conjunto de datos: las imágenes MRI presentan patrones de textura, bordes y estructuras anatómicas locales que son precisamente el tipo de información que las redes convolucionales están diseñadas para capturar. Los mecanismos de atención global de ViT, aunque son poderosos para tareas que requieren modelar dependencias de largo alcance, resultan menos eficientes cuando el discriminador clave está en características locales de la imagen.

5.10.2 Análisis de los modelos de ensamble

Los tres ensambles implementados combinan las predicciones de EfficientNet V2 B0 y ViT-B16 mediante distintas estrategias de promediado de probabilidades. Su análisis revela patrones importantes sobre cómo la diversidad arquitectónica interactúa con cada método de combinación.

Ensamble por promedio simple (accuracy = 0.9196, MCC = 0.9166): Este ensamble obtiene un resultado inferior al del modelo EfficientNet individual en todas las métricas. Este comportamiento, aunque parece ser contraintuitivo a primera vista, se explica por la asimetría de desempeño entre los dos modelos base. Al promediar con igual peso las probabilidades de un modelo con 92.52% de exactitud y uno con 82.37%, el efecto del modelo más débil domina en aquellos casos donde ViT asigna alta probabilidad a una clase incorrecta, arrastrando el vector combinado lejos de la predicción correcta de EfficientNet. El promedio simple es más efectivo cuando los modelos base tienen niveles de desempeño similares. Pero, en este caso, la diferencia de ~10 puntos porcentuales entre ambos hace que la combinación sin ponderación resulte perjudicial.

Ensamble por promedio ponderado (accuracy = 0.9330, MCC = 0.9304): Este ensamble, con pesos de 0.6 para EfficientNet y 0.4 para ViT-B16, obtiene el mejor desempeño de los cinco modelos evaluados en todas las métricas principales. Su exactitud supera al modelo EfficientNet individual en 0.78 puntos porcentuales (93.30% vs. 92.52%), mientras que su MCC mejora en 0.0082 puntos (0.9304 vs. 0.9222) y su F1-score ponderado en 0.0077 puntos (0.9327 vs. 0.9250). Aunque estas mejoras pueden parecer marginales en términos absolutos, son consistentes en todas las métricas, lo que confirma que no se trata de una variación estadística aleatoria sino de una mejora real. Este resultado demuestra que la contribución ponderada del modelo ViT-B16, aunque es inferior en términos de rendimiento individual, introduce información complementaria, derivada de su paradigma de atención global, que permite al ensamble corregir algunos errores que EfficientNet comete de forma sistemática. La asignación de pesos 0.6/0.4 resulta ser un balance efectivo porque da mayor autoridad al modelo más preciso sin descartar completamente la perspectiva del modelo transformer.

Ensamble por media geométrica (accuracy = 0.9263, MCC = 0.9235): Este ensamble supera ligeramente al modelo EfficientNet individual (92.63% vs. 92.52%) pero no alcanza al ensamble por promedio ponderado. La media geométrica, al atenuar el impacto de probabilidades extremas, produce predicciones más conservadoras que benefician la clasificación en escenarios donde alguno de los modelos exhibe exceso de confianza (overconfidence) en clases incorrectas. Su comportamiento se sitúa entre el promedio simple y

el ponderado, lo que sugiere que la ponderación explícita es una estrategia más efectiva que la regulación implícita de valores extremos para este conjunto de datos particular.

5.10.3 Selección del modelo final

A partir del análisis comparativo presentado en la Tabla 27, el ensamble por promedio ponderado (pesos: EfficientNet=0.6, ViT-B16=0.4) es seleccionado como el modelo final de la solución propuesta, al obtener los mejores valores en cinco de las seis métricas evaluadas:

Tabla 27. Análisis comparativo entre el Modelo EfficientNet V2 B0 y el Modelo Ensamble por Promedio Ponderado

Métrica	Ensamble Ponderado	EfficientNet V2 B0	Diferencia
Accuracy	0.9330	0.9252	+0.0078
Top-3 Accuracy	0.9754	0.9777	-0.0023
Precision (pond.)	0.9390	0.9311	+0.0079
Recall (pond.)	0.9330	0.9252	+0.0078
F1-score (pond.)	0.9327	0.9250	+0.0077
MCC	0.9304	0.9222	+0.0082

Nota: Elaboración propia a partir de los resultados obtenidos en el experimento.

La única métrica en que el modelo EfficientNet supera al ensamble ponderado es la exactitud Top-3, con una diferencia de 0.23 puntos porcentuales, lo cual es estadísticamente marginal. Esta diferencia es atribuible a que EfficientNet, como modelo individual más confiado en sus predicciones de clase correcta, tiende a asignar mayor probabilidad a la clase verdadera incluso cuando no la selecciona como predicción principal; la combinación con ViT modera levemente esta confianza en algunos casos.

La elección del ensamble ponderado como modelo final se justifica por las siguientes razones:

1. Mejor desempeño en las métricas más relevantes para el dominio médico: El F1-score ponderado y el MCC son los indicadores más fiables cuando existe desbalance de clases, y el ensamble ponderado lidera en ambos. Un MCC de 0.9304 indica una correlación muy alta entre las predicciones del modelo y las etiquetas reales, considerando todas las clases de forma equilibrada.
2. Mayor robustez por diversidad arquitectónica: La combinación de una red convolucional y un Vision Transformer aprovecha las fortalezas complementarias de ambos paradigmas: la detección de patrones locales de EfficientNet y la

captura de dependencias globales de ViT. Esta diversidad reduce la probabilidad de que ambos modelos cometan el mismo error simultáneamente.

3. Corrección de errores sistemáticos: El análisis de los reportes de clasificación por clase sugiere que ViT-B16, aunque es menos preciso en general, clasifica correctamente algunas muestras que EfficientNet clasifica incorrectamente. La ponderación 0.6/0.4 permite incorporar estas correcciones sin sacrificar el fuerte desempeño base de EfficientNet.
4. Sin sobrecosto de entrenamiento: El ensamble no requiere entrenamiento adicional, lo que lo convierte en una estrategia computacionalmente eficiente que aprovecha al máximo los modelos ya entrenados.
5. Relevancia en contexto de diagnóstico asistido: En el dominio médico, incluso mejoras marginales en precisión y recall tienen potencial impacto clínico. Un sistema con 93.30% de exactitud global sobre 44 clases de condiciones neurológicas representa un nivel de desempeño relevante para su consideración como herramienta de apoyo a la decisión diagnóstica.

5.10.4 Interpretación en el contexto médico

Más allá de los valores numéricos, es importante contextualizar estos resultados en el problema clínico que motiva la investigación. El dataset evaluado comprende 44 clases de condiciones neurológicas, diferenciadas tanto por el tipo de lesión (astrocitoma, meningioma, glioblastoma, schwannoma, entre otras) como por la modalidad de imagen MRI (T1, T1C+ y T2). Esta granularidad es notablemente mayor que la mayoría de los benchmarks reportados en la literatura de clasificación de tumores cerebrales, que típicamente trabajan con 3 a 5 clases.

En ese contexto, una exactitud global del 93.30% sobre 44 clases y 896 muestras de prueba, con un MCC de 0.9304, constituye un resultado técnicamente sólido. La exactitud Top-3 del 97.54% del ensamble ponderado es particularmente significativa: en la práctica clínica, un sistema que ofrece las tres categorías diagnósticas más probables puede asistir al especialista reduciendo el espacio de diagnósticos diferenciales a considerar, aun cuando la predicción principal no sea la correcta.

Sin embargo, es importante reconocer que los resultados obtenidos corresponden a un entorno experimental controlado. Las clases con menor support en el conjunto de prueba, como Ganglioglioma T2 (3 muestras), Tuberculoma T1 (4 muestras) o Germinoma T1 (2 muestras), presentan mayor variabilidad en sus métricas individuales, lo que limita la confiabilidad de las conclusiones sobre el comportamiento del modelo en esas categorías específicas. La

consolidación de estos resultados en un entorno de producción requeriría un dataset más equilibrado y de mayor volumen para estas clases subrepresentadas.

5.11 Reproducibilidad de la solución

La reproducibilidad es un principio fundamental de la investigación científica en inteligencia artificial, ya que permite validar de forma independiente los resultados obtenidos, detectar errores o sesgos no identificados previamente, y establecer una base sólida sobre la cual otros investigadores o equipos técnicos puedan construir mejoras. En el contexto del aprendizaje profundo, la reproducibilidad es especialmente desafiante debido a la naturaleza estocástica de los procesos de optimización, la dependencia del hardware y las versiones de las bibliotecas, y la variabilidad introducida por las operaciones paralelas en GPU. Esta solución fue diseñada desde su concepción para minimizar estas fuentes de variabilidad, documentando de forma exhaustiva cada componente técnico relevante.

5.11.1 Control de aleatoriedad

La principal fuente de no determinismo en el entrenamiento de modelos de aprendizaje profundo es la aleatoriedad introducida por los generadores de números pseudoaleatorios de las distintas bibliotecas utilizadas. Para garantizar resultados consistentes entre ejecuciones, se implementó una función de fijación de semillas (seed_everything) que establece valores deterministas en todos los módulos relevantes de manera simultánea:

Figura 37

Definición de configuración global (CFG) y función de inicialización de semillas en el entorno de entrenamiento

```
[7]
class CFG:
    EPOCHS = 50
    BATCH_SIZE = 64
    SEED = 42
    TF_SEED = 768
    HEIGHT = 224
    WIDTH = 224
    CHANNELS = 3
    IMAGE_SIZE = (224, 224, 3)
```

```
[8]
def create_dir(path):
    if os.path.exists(path):
        print('Path already exists.')
        return
    os.mkdir(path)


def seed_everything(seed=CFG.SEED):
    random.seed(seed)
    np.random.seed(seed)
    tf.random.set_seed(seed)

seed_everything(CFG.SEED)
```

Nota. Elaboración propia

Esta función es invocada al inicio del notebook antes de cualquier operación que dependa de aleatoriedad. Adicionalmente, la semilla de TensorFlow fue restablecida inmediatamente antes de cada llamada a la compilación y entrenamiento de cada modelo (tf.random.set_seed(CFG.SEED)), asegurando que la inicialización de los pesos de las capas densas sea idéntica en cada ejecución. Las semillas utilizadas en los distintos componentes fueron:

- CFG.SEED = 42: semilla principal para Python, NumPy, TensorFlow, inicialización de pesos GlorotNormal, y división estratificada del dataset (random_state=42).
- CFG.TF_SEED = 768: semilla secundaria utilizada exclusivamente en las capas de aumento de datos (RandomFlip y RandomZoom), garantizando que las transformaciones aleatorias aplicadas durante el entrenamiento sean reproducibles entre ejecuciones.

5.11.2 Entorno de ejecución

La solución fue desarrollada y ejecutada sobre un entorno de Azure Machine Learning, cuya configuración exacta se detalla a continuación para facilitar su replicación:

Infraestructura cloud:

- Plataforma: Azure Machine Learning workspace
- Tipo de recurso de cómputo: instancia de cómputo con acelerador GPU
- GPU: NVIDIA Tesla T4 (13.775 MB de memoria de video, capacidad de cómputo 7.5)

Stack de software:

- Sistema operativo: Linux (entorno de contenedor Azure ML)
- Entorno Conda: azureml_py310_sdkv2
- Versión de Python: 3.10
- Variable de entorno requerida: TF_USE_LEGACY_KERAS=1 (necesaria para compatibilidad entre Keras 3 y tf_keras)

- CUDA Driver: 12.2.0 | CUDA Runtime: 12.2.0 | CUDA Toolkit: 12.5.0 | cuDNN: 9.17.1

5.11.3 Dependencias y versiones exactas

La tabla 28 registra las versiones exactas de todas las bibliotecas utilizadas en el notebook, verificadas mediante la salida de pip install y los comandos de impresión de versiones al inicio de la ejecución. Estas versiones son esenciales para garantizar la reproducibilidad, ya que cambios en las APIs o en los comportamientos internos de estas bibliotecas pueden producir resultados distintos.

Tabla 28. Bibliotecas utilizadas en el proyecto.

Biblioteca	Versión	Restricción de instalación
TensorFlow	2.21.0	—
Keras	3.12.1	—
tf-keras	2.21.0	—
TensorFlow Hub	0.16.1	—
vit-keras	0.2.0	—
NumPy	1.26.4	"numpy<2.0"
pandas	1.5.3	—
scikit-learn	1.7.2	—
scikit-plot	0.3.7	—
SciPy	1.11.4	"scipy<1.12"
seaborn	0.13.2	—
opencv-python-headless	4.11.0.86	—

Nota: Elaboración propia.

Las restricciones de versión para NumPy (<2.0) y SciPy (<1.12) son condiciones necesarias para mantener la compatibilidad con las versiones de TensorFlow y scikit-plot utilizadas. El comando de instalación completo es:

```
pip install seaborn "numpy<2.0" "scipy<1.12" scikit-plot tensorflow \
    tensorflow-hub opencv-python-headless vit-keras
```

5.11.4 Arquitectura de datos en la nube

El conjunto de datos de imágenes MRI fue almacenado en Azure Data Lake Storage Gen2 (ADLS Gen2), siguiendo la arquitectura de medallón definida en el subcapítulo 5.3. Para

reproducir la solución es necesario contar con esta estructura de almacenamiento. Los parámetros de conexión son los siguientes:

- Cuenta de almacenamiento: storagetesis01.dfs.core.windows.net
- Sistema de archivos (contenedor): bronze-layer
- Ruta del dataset: tesis-mri-archive/
- Formato de imágenes: JPEG/JPG (extensiones reconocidas: *.jpg, *.jpeg, *.JPG, *.JPEG)
- Estructura de directorios: una carpeta por clase, nombrada con el nombre del tipo de tumor y la modalidad MRI (ejemplo: Meningioma T1C+/, _NORMAL T2/)

La descarga del dataset desde ADLS Gen2 al entorno de cómputo local se realiza mediante la biblioteca azure-storage-file-datalake, que recorre recursivamente el sistema de archivos y descarga todas las imágenes manteniendo la estructura de directorios original. Esta estructura es la que permite la extracción automática de etiquetas por nombre de carpeta durante la construcción del DataFrame principal.

5.11.5 Guía de reproducción paso a paso

A continuación, se presenta el flujo completo de pasos necesarios para reproducir la solución desde cero, en el orden exacto en que deben ejecutarse:

- **Paso 1**: Configuración del entorno cloud: Crear un workspace de Azure Machine Learning con una instancia de cómputo que incluya aceleración por GPU (equivalente a Tesla T4 o superior). Configurar el grupo de recursos, la cuenta de almacenamiento y el espacio de trabajo siguiendo las especificaciones descritas en el subcapítulo 5.2.
- **Paso 2**: Preparación del almacenamiento: Crear la cuenta de Azure Data Lake Storage Gen2, configurar el sistema de archivos bronze-layer y cargar el dataset de imágenes MRI en la ruta tesis-mri-archive/, respetando la estructura de un directorio por clase con el nombre de la condición neurológica y la modalidad de imagen.
- **Paso 3**: Configuración del entorno de ejecución: En la instancia de cómputo de Azure ML, activar el entorno Conda azureml_py310_sdkv2 y establecer la variable de entorno TF_USE_LEGACY_KERAS=1. Ejecutar el comando de instalación de dependencias con las versiones especificadas en el subcapítulo 5.11.3.
- **Paso 4**: Descarga del dataset: Ejecutar la celda de descarga del notebook, que establece la conexión con ADLS Gen2 mediante la clave de cuenta y descarga recursivamente todas las imágenes al directorio local ./tesis-mri-archive.
- **Paso 5**: Fijación de semillas y configuración global: Ejecutar las celdas de configuración del notebook: definición de la clase CFG con los hiperparámetros

globales, e invocación de seed_everything(CFG.SEED) para establecer el estado determinista de todos los generadores de números aleatorios.

- **Paso 6**: Exploración y preparación del dataset: Ejecutar las celdas de exploración de datos (conteo de imágenes, visualización de distribución de clases) y construcción del DataFrame principal con rutas y etiquetas. Aplicar la codificación de etiquetas con LabelEncoder y realizar la división estratificada del dataset en los subconjuntos de entrenamiento (60%), validación (20%) y prueba (20%) con random_state=42.
- **Paso 7**: Preprocesamiento y construcción del pipeline de datos: Ejecutar las celdas de definición de la función _load() (lectura, decodificación, redimensionamiento a 224×224 y normalización a [0,1]) y de la capa de aumento de datos (RandomFlip y RandomZoom con seed=CFG.TF_SEED). Construir los tres pipelines de TensorFlow (train_ds, val_ds, test_ds) con batch_size=64 y aumento de datos activado exclusivamente para el conjunto de entrenamiento.
- **Paso 8**: Entrenamiento del modelo EfficientNet V2 B0: Cargar el modelo base desde TensorFlow Hub con trainable=False, construir el modelo completo con la función efficientnet_v2_model(), compilar con pérdida categórica y Adam (lr=0.001), y ejecutar el entrenamiento con los callbacks EarlyStopping y ReduceLROnPlateau configurados según los parámetros del subcapítulo 5.8.
- **Paso 9**: Entrenamiento del modelo ViT-B16: Descargar el modelo ViT-B16 preentrenado mediante vit-keras, congelar todas sus capas, crear el ViTWrapper, construir el modelo completo con vit_b16_model(), compilar con la misma configuración que EfficientNet y ejecutar el entrenamiento con los mismos callbacks.
- **Paso 10**: Generación de predicciones y ensambles: Generar los vectores de probabilidades de ambos modelos sobre el conjunto de prueba mediante model.predict() y calcular las predicciones de los tres ensambles (promedio simple, promedio ponderado con pesos [0.6, 0.4], y media geométrica).
- **Paso 11**: Evaluación y visualización: Ejecutar la función generate_performance_scores() para cada uno de los cinco modelos, generar los reportes de clasificación por clase, construir el DataFrame comparativo de métricas, y visualizar las matrices de confusión y las curvas de historial de entrenamiento.

5.11.6 Limitaciones de reproducibilidad

A pesar de las medidas implementadas, existen ciertas limitaciones inherentes al entorno de ejecución que deben considerarse:

- No determinismo en operaciones GPU: Las operaciones matemáticas en GPU, especialmente la convolución y la reducción paralela, pueden producir pequeñas diferencias numéricas entre ejecuciones debido al no determinismo de las operaciones de punto flotante en hardware paralelo. Aunque las semillas fijas minimizan esta variabilidad, no la eliminan completamente en todos los entornos de GPU.
- Dependencia de modelos preentrenados externos: El modelo EfficientNet V2 B0 se descarga en tiempo de ejecución desde TensorFlow Hub y el modelo ViT-B16 desde los servidores de vit-keras. Si estas fuentes externas modificaran los pesos preentrenados o dejaran de estar disponibles, sería necesario utilizar copias locales previamente guardadas de dichos modelos.
- Versión del entorno Conda de Azure ML: El entorno azureml_py310_sdkv2 es gestionado por Azure Machine Learning y puede actualizarse con el tiempo. Se recomienda verificar que las versiones de las bibliotecas listadas en el subcapítulo 5.11.3 coincidan con las del entorno disponible en el momento de la ejecución, instalando explícitamente las versiones correctas mediante el comando de instalación proporcionado.

6. Conclusiones

Los resultados de este proyecto de investigación permiten concluir que la integración de tecnologías Big Data, arquitecturas en la nube y modelos de redes neuronales profundas representa una alternativa viable, escalable y técnicamente sólida para enfrentar los desafíos actuales del procesamiento de imágenes MRI en los centros médicos públicos de Costa Rica. Los hallazgos obtenidos se sustentan tanto en la revisión sistemática de la literatura como en la implementación y evaluación experimental de los modelos propuestos.

En primer lugar, se confirma que la variabilidad interobservador en la interpretación manual de imágenes MRI, documentada en la literatura por autores como Bruner et al. (1997) con tasas de desacuerdo de hasta 42.8%, puede mitigarse mediante el uso de sistemas de clasificación automática. El modelo de ensamble por promedio ponderado implementado en este trabajo alcanzó una exactitud global del 93.30% y un MCC de 0.9304 sobre un conjunto de prueba de 896 imágenes distribuidas en 44 clases, lo que demuestra que estos sistemas pueden actuar como una referencia objetiva y consistente que complementa, sin reemplazar, el criterio del especialista. La colaboración entre el humano y la inteligencia artificial documentada por Chen (2024), que muestra reducciones de hasta 44.47% en la cantidad de lecturas requeridas, representa un horizonte alcanzable para la CCSS al adoptar este tipo de soluciones.

En segundo lugar, la comparación de los modelos implementados evidencia que las redes neuronales convolucionales, representadas por EfficientNet V2 B0, mantienen una ventaja significativa sobre los Vision Transformers (ViT-B16) cuando se trabaja con conjuntos de datos de tamaño moderado y alta variabilidad entre clases. EfficientNet V2 B0 obtuvo una exactitud del 92.52% frente al 82.37% del ViT-B16, diferencia que se atribuye a la capacidad superior de las CNN para capturar patrones espaciales locales en imágenes médicas, tal como lo evidencian Dorfner et al. (2025) y Sadr et al. (2025). Este resultado sugiere que los modelos Transformer requieren mayor volumen de datos de entrenamiento para alcanzar su potencial máximo en este dominio.

En tercer lugar, se demuestra que el uso de métodos de ensamble es una estrategia efectiva para mejorar el desempeño sin incurrir en costos adicionales de entrenamiento. El ensamble por promedio ponderado (pesos: EfficientNet=0.6, ViT=0.4) superó a todos los modelos individuales en cinco de seis métricas, logrando el mejor F1-score ponderado (0.9327) y MCC (0.9304). Este resultado confirma que las arquitecturas de los modelos convolucionales y transformers se complementan y pueden producir sistemas más robustos. Por ejemplo, donde EfficientNet presenta errores, el ViT puede contribuir a corregirlos, y viceversa. La única métrica

en que EfficientNet supera al ensamble es la exactitud Top-3, con una diferencia marginal de 0.23 puntos porcentuales.

Un cuarto hallazgo de importancia clínica es la confusión sistemática identificada entre las clases Astrocitoma T2 y NORMAL T2 en todos los modelos evaluados. Esta confusión se explica por las similitudes visuales entre los astrocitomas de bajo grado en secuencia T2 y el tejido cerebral normal en esa misma modalidad, particularmente en estadios tempranos donde la alteración del tejido no es suficientemente pronunciada. Este hallazgo subraya que los sistemas de diagnóstico asistido deben ser diseñados con alertas explícitas para los casos donde la probabilidad de clasificar tejido tumoral como normal es más alta, y refuerza la importancia de no depender exclusivamente del criterio automático en categorías clínicamente críticas.

En quinto lugar, se concluye que la arquitectura medallón (Bronze–Silver–Gold) implementada sobre Azure Machine Learning ofrece el entorno más adecuado para soportar este tipo de soluciones en el contexto del sistema público de salud. Esta arquitectura garantiza que los datos pasen de forma controlada desde su ingesta original hasta la generación de resultados analíticos, con trazabilidad completa, control de versiones y reproducibilidad verificada. El control de semillas (SEED=42, TF_SEED=768) y la documentación detallada del entorno de ejecución permiten replicar los resultados de forma exacta, lo que es un requisito fundamental para cualquier implementación clínica.

Finalmente, desde una perspectiva estratégica, este proyecto demuestra que la adopción de IA en el sector público de salud no solo es técnicamente posible, sino necesaria y urgente. Los hallazgos de la Auditoría Interna de la CCSS documentados por Segura (2026) evidencian que los tiempos de espera actuales representan un riesgo real para el diagnóstico oportuno de pacientes oncológicos. El sistema implementado puede transformar el flujo diagnóstico actual, que es reactivo y secuencial, hacia un modelo de triage asistido donde la clasificación automática prioriza casos críticos, optimiza el tiempo de los especialistas y estandariza el proceso de análisis. En ese sentido, la solución propuesta sienta las bases para futuras implementaciones institucionales orientadas a mejorar la calidad, consistencia y equidad del diagnóstico oncológico en Costa Rica.

7. Referencias bibliográficas

Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Carvalho, S., ... & Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature Communications, 5, 4006.

Aguilar, C., Barrantes, K., Ibañez, S., Ketelhohn, H. W., & Meza, F. (2024). Aplicación de la inteligencia artificial en la salud pública para el diagnóstico temprano y tamizaje de enfermedades oncológicas. *Revista Ciencia y Salud Integrando Conocimientos, 8*(2). https://revistacienciaysalud.ac.cr/ojs/index.php/cienciaysalud/article/view/582/896

American Cancer Society. (2023). Signs and symptoms of brain and spinal cord tumors in adults. https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/detection-diagnosis-staging/signs-symptoms.html

American Cancer Society. (2023). *What causes brain and spinal cord tumors in adults?* https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/causes-risks-prevention/what-causes.html

American College of Radiology (ACR). (2025). ACR Appropriateness Criteria®: Brain Tumors. https://www.acr.org/Clinical-Resources/ACR-Appropriateness-Criteria/Brain-Tumors

Armbrust, M., Das, T., Torres, J., van der Lans, J., & Zaharia, M. (2020). Delta Lake: High-performance ACID table storage over cloud object stores. Proceedings of the VLDB Endowment, 13(12), 3411–3424.

Arora, A. (2021). EfficientNetV2. Weights & Biases. https://wandb.ai/wandb_fc/pytorch-image-models/reports/EfficientNetV2--Vmlldzo2NTkwNTQ

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. NeuroImage, 26(3), 839–851.

Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. BioMed Research International, 2015, 1–16. https://doi.org/10.1155/2015/370194

Benfatto, S., Sill, M., Jones, D. T. W., Pfister, S. M., Sahm, F., von Deimling, A., Capper, D., & Hovestadt, V. (2025). Explainable artificial intelligence of DNA methylation-based brain tumor diagnostics. *Nature Communications, 16*, Artículo 1787. https://doi.org/10.1038/s41467-025-57078-0

Bilsky, M. H. (2024, julio). *Overview of brain tumors*. En *Merck Manual* (Consumer Version). https://www.merckmanuals.com/home/brain-spinal-cord-and-nerve-disorders/tumors-of-the-nervous-system/overview-of-brain-tumors

Bruner, J. M., Inouye, L., Fuller, G. N., & Langford, L. A. (1997). Diagnostic discrepancies and their clinical impact in a neuropathology referral practice. *Cancer, 79*(4), 796–803. https://doi.org/10.1002/(SICI)1097-0142(19970215)79:4<796::AID-CNCR17>3.0.CO;2-V

Calderon, A., Guzman, P., & Murphy, J. D. (2023). Epidemiological patterns of common cancers in Costa Rica: An overview up to 2020. *Open Journal of Social Sciences, 11*(6), 500–517. https://doi.org/10.4236/jss.2023.116033

Chaplot, S., Patnaik, L. M., & Jagannathan, N. R. (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomedical Signal Processing and Control, 1(1), 86–92.

Chen, M., Wang, Y., Wang, Q., Shi, J., Wang, H., Ye, Z., Xue, P., & Qiao, Y. (2024). Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. *npj Digital Medicine, 7*, 349. https://doi.org/10.1038/s41746-024-01328-w

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., & Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Transactions on Medical Imaging, 27(4), 425–441.

Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.

Databricks. (2023). What is the medallion architecture? Databricks Documentation. https://docs.databricks.com/en/delta/medallion-architecture.html

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107–113.

Dorfner, F. J., Patel, J. B., Kalpathy-Cramer, J., Gerstner, E. R., & Bridge, C. P. (2025). A review of deep learning for brain tumor analysis in MRI. *npj Precision Oncology, 9*, Article 2. https://doi.org/10.1038/s41698-024-00789-2

Feltrin, F. (2023). Brain Tumor MRI Images 44 Classes [Conjunto de datos]. Kaggle. https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c

Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing (4th ed.). Pearson.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182.

Han, J., Pei, J., & Kamber, M. (2022). Data Mining: Concepts and Techniques (4th ed.). Morgan Kaufmann.

Hansa, G. (2025). Understanding neural networks and their components. Codecademy. https://www.codecademy.com/article/understanding-neural-networks-and-their-components

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610–621.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, 98–115.

Haykin, S. (2009). Neural Networks and Learning Machines (3rd ed.). Pearson.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. Communications of the ACM, 57(7), 86–94.

Jain, A. K. (1989). Fundamentals of Digital Image Processing. Prentice Hall.

Jansen, M. (2025). Transfer Learning | Brain Tumor Classification. [Notebook].

Kaggle. https://www.kaggle.com/code/matthewjansen/transfer-learning-brain-tumor-classification

Karau, H., & Warren, R. (2017). High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark. O'Reilly Media.

Khalighi, S., Reddy, K., Midya, A., Pandav, K. B., Madabhushi, A., & Abedalthagafi, M. (2024). Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. *npj Precision Oncology, 8*, Article 80. https://doi.org/10.1038/s41698-024-00575-0

Kharrat, A., Benamrane, N., & Abid, M. (2010). Detection of brain tumor in medical images using hybrid approach. Computers in Biology and Medicine, 40(4), 383–392.

Kilim, O., Olar, A., Joó, T., Palicz, T., Pollner, P., Csabai, I., … et. al. (2022). Physical imaging parameter variation drives domain shift. *Scientific Reports, 12*, Article 21302. https://doi.org/10.1038/s41598-022-23990-4

Kitchin, R. (2021). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (2nd ed.). SAGE Publications.

Larhmam. (2018). *Maximum-margin hyperplane and margin for an SVM trained on two classes*. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:SVM_margin.png

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining (2nd ed.). Wiley.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., … Ellison, D. W. (2021). *The 2021 WHO classification of tumors of the central nervous system: A summary. Neuro-Oncology, 23*(8), 1231–1251.

Marques, G., Pitarma, R., Garcia, N. M., & Pombo, N. (2022). Cloud computing for healthcare: Review and future directions. Journal of Medical Systems, 46(6), 1–15. https://doi.org/10.1007/s10916-022-01841-2

Mayo Clinic. (2024). Brain tumor – Symptoms and causes. https://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/syc-20350084

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115–133.

Microsoft. (2024). What is Azure?. Microsoft Learn. https://learn.microsoft.com/en-us/azure/what-is-azure

Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing. National Institute of Standards and Technology (NIST), Special Publication 800-145.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging, 34(10), 1993–2024.

Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.

Mount Sinai Health System. (s. f.). *Brain tumors – primary*. https://www.mountsinai.org/health-library/report/brain-tumors-primary

National Cancer Institute (NCI). (2024). *Brain and spinal cord tumors in adults: Patient version.* U.S. Department of Health & Human Services. https://www.cancer.gov/types/brain

National Institute of Biomedical Imaging and Bioengineering. (2024, enero 25). *Magnetic resonance imaging (MRI).* U.S. Department of Health & Human Services. https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri

Parekh, V. S., & Jacobs, M. A. (2019). Radiomics: A new application from established techniques. Expert Review of Precision Medicine and Drug Development, 4(1), 1–15.

Quirónsalud. (2025). *Tumores cerebrales, uno de los mayores desafíos de la neurología moderna*. https://www.quironsalud.com/es/comunicacion/contenidos-salud/tumores-cerebrales-mayores-desafios-neurologia-moderna

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408.

Russ, J. C. (2016). The Image Processing Handbook (7th ed.). CRC Press.

Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., & Ahuja, C. K. (2016). A hybrid automatic brain tumor classification system using texture features and support vector machine. Journal of Medical Engineering & Technology, 40(2), 80–87.

Sadr, H., Nazari, M., Yousefzadeh-Chabok, S., Emami, H., Rabiei, R., & Ashraf, A. (2025). Enhancing brain tumor classification in MRI images. *Image and Vision Computing*. https://doi.org/10.1016/j.imavis.2025.105555

Segura, A. (2026). *Pacientes con cáncer esperan más de 600 días por una resonancia magnética en la CCSS.* CRHoy. https://crhoy.com/nacionales/pacientes-con-cancer-esperan-mas-de-600-dias-por-una-resonancia-magnetica-en-la-ccss/

Shah, V., Kumar, V., Guo, Y., & Ahuja, C. K. (2018). Brain tumor segmentation and classification using SVM and CNN. International Journal of Computer Applications, 181(9), 8–14.

Suvvani, M. (2025). End-to-End machine learning pipeline: From raw data to deployment. Medium. https://medium.com/@muralisuvvani/end-to-end-machine-learning-pipeline-from-raw-data-to-deployment-624e700f6992

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging, 29(6), 1310–1320.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... & Yu, T. (2014). scikit-image: Image processing in Python. PeerJ, 2, e453.

Van den Bent, M. J. (2010). Interobserver variation of the histopathological diagnosis in clinical trials on glioma: A clinician's perspective. *Acta Neuropathologica, 120*(3), 297–304. https://doi.org/10.1007/s00401-010-0725-7

Vega, Y. (2025, 21 agosto). *Centro Nacional de Imágenes Médicas realiza cada mes 2 970 resonancias magnéticas.* Recuperado de: https://www.ccss.sa.cr/noticias/noticia?v=012136000152

White, T. (2015). Hadoop: The Definitive Guide (4th ed.). O'Reilly Media.

World Health Organization (WHO). (2023). Electromagnetic fields and public health: Mobile phones. https://www.who.int/news-room/questions-and-answers/item/radiation-electromagnetic-fields-and-public-health-mobile-phones

Wolfe, C. R. (2021). *Vision transformers*. Substack. https://cameronrwolfe.substack.com/p/vision-transformers

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. Communications of the ACM, 59(11), 56–65.