ISSN: 1692-7257 - Volumen 1 – Número 47 - 2026

ORIGINAL

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

Codificador-decodificador con autoatención para la segmentación de materia gris en resonancia magnética cerebral

Self-attention encoder–decoder for gray-matter segmentation in brain MRI

Camilo Andres Laiton Bonadiez[1]

German Sanchez Torres[2]

Carlos Nelson Henríquez Miranda[2]

[1] Universidad Nacional de Colombia, Grupo de Investigación y desarrollo en Inteligencia Artificial - GIDIA, Medellín, Antioquia, Colombia.

[2] Universidad del Magdalena, Grupo de Investigación y Desarrollo en Sistemas y Computación, Santa Marta, Magdalena, Colombia.

Recibido: 24 julio 2025. Revisado: 24 septiembre 2025. Aceptado: 20 diciembre 2025. Publicado: 01 enero 2026.

Autor de correspondencia: Carlos Nelson Henríquez Miranda

Cómo citar: C. A. Laiton Bonadiez, G. Sanchez Torres, and C. N. Henríquez Miranda, "Codificador-decodificador con autoatención para la segmentación de materia gris en resonancia magnética cerebral", *Revista Colombiana de Tecnologías de Avanzada (RCTA)*, vol. 1, no. 47, ene. 2026, art. no. 4282, doi: 10.24054/rcta.v1i47.4282

Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial 4.0.

Resumen:

La resonancia magnética cerebral es clave para el diagnóstico asistido por computador, pero la segmentación manual es costosa, lenta y dependiente del operador. Se presenta un pipeline para segmentar materia gris (SG) que combina un preprocesamiento mínimo (armonización de forma, teselado por parches) con un codificador–decodificador 3D que integra autoatención global y conexiones de salto multiescala. En el conjunto de prueba MRBrainS18, el modelo logra un Dice de 0.650 ± 0.043, un IoU de 0.483 ± 0.046, una precisión de 0.700 ± 0.046 y un recall de 0.610 ± 0.059; las distribuciones por sujeto son compactas, lo que refleja consistencia entre casos. Las métricas de solapamiento muestran ausencia de relación monótona con la distancia de Hausdorff ($r \approx 0.08$), lo que resalta la sensibilidad de los bordes aun con buen solapamiento global. El análisis de Bland–Altman evidencia un sesgo volumétrico negativo (RVE = −12.54 % ± 9.53 %), consistente con una precisión mayor que el recall y con predicciones conservadoras en las interfaces tisulares. Las líneas base clásicas (Otsu multiumbral y crecimiento de regiones), evaluadas en un subconjunto representativo, presentan menor rendimiento; el modelo mejora Dice e IoU en un 50 % frente a la mejor línea base.

Palabras clave: imágenes de resonancia magnética, segmentación de imágenes, visión artificial, aprendizaje profundo.

Abstract:

Brain magnetic resonance imaging is key to computer-aided diagnosis, but manual segmentation is costly, time-consuming, and operator-dependent. We present a pipeline for gray-matter (GM) segmentation that combines minimal preprocessing (shape harmonization and patch-based tiling) with a 3D encoder–decoder integrating global self-attention and multiscale skip connections. On the MRBrainS18 test set, the model achieves a Dice score of 0.650 ± 0.043, an IoU of 0.483 ± 0.046, a precision of 0.700 ± 0.046, and a recall of 0.610 ± 0.059; subject-wise distributions are compact, reflecting consistency across cases. Overlap metrics show no monotonic relationship with the Hausdorff distance ($r \approx 0.08$), highlighting boundary sensitivity even with good global overlap. Bland–Altman analysis reveals a negative volumetric bias (RVE = −12.54 % ± 9.53 %), consistent with precision exceeding recall and conservative predictions at tissue interfaces. Classical baselines (multilevel Otsu thresholding and region growing), evaluated on a representative subset, exhibit lower performance; the proposed model improves Dice and IoU by 50 % over the best baseline.

Keywords: brain imaging, image segmentation, computer vision, deep learning.

INTRODUCCIÓN

La neuroimagen es la herramienta principal para diagnosticar enfermedades relacionadas con el cerebro, abarcando un amplio espectro de trastornos y proporcionando información anatómica y funcional del cerebro [1]. Revisiones recientes reafirman la centralidad de la neuroimagen para flujos de trabajo de diagnóstico y segmentación asistidos por computador, consolidando metodologías y prácticas de evaluación [2]. Comprender la estructura del cerebro y el mecanismo neural subyacente es esencial para el monitoreo y el diagnóstico temprano de enfermedades, a fin de evitar que progresen a un nivel grave [3].

Los diagnósticos asistidos por computador se basan en técnicas de imagen no invasivas. Las modalidades más ampliamente disponibles incluyen la imagen por resonancia magnética (MRI), la tomografía por emisión de positrones (PET) y la tomografía computarizada por rayos X (CT), siendo estas las técnicas más populares para imágenes del cerebro[3]. La imagen por resonancia magnética (MRI) es uno de los tipos de imagen más populares; utiliza un campo magnético fuerte y ondas de radiofrecuencia para visualizar órganos, tejidos blandos y huesos[4].

El objetivo principal del análisis de MRI cerebral estructural incluye la clasificación de tipos específicos de tejido, así como la identificación y descripción de estructuras anatómicas específicas. La literatura muestra que la segmentación basada en aprendizaje profundo se ha convertido en el paradigma predominante para la delimitación de tejidos y estructuras en entornos clínicos y de investigación[5]. Sin embargo, es desafiante extraer información de alta calidad de las imágenes cerebrales debido a la baja relación señal-ruido y a los artefactos que se generan en el proceso de adquisición. En[6]se detallan fuentes persistentes de artefactos interferencia de radiofrecuencia y artefactos de adquisición relacionados y su impacto en la segmentación posterior.

La clasificación de datos de MRI en tipos de tejido puede lograrse usando una variedad de métodos, incluidos la inspección visual manual y técnicas de clasificación de tejido semiautomatizadas [7]. La inspección visual manual de datos de MRI es un proceso que consume mucho tiempo y requiere una pericia considerable [8], lo que la hace impráctica para procesar grandes bases de datos. Por ello, se han desarrollado técnicas de clasificación de tejido semiautomatizadas para muchos conjuntos de datos de MRI de bajo costo. Sin embargo, una debilidad de estos enfoques es que pierden precisión debido a imágenes con alta presencia de ruido o artefactos. De hecho, las imágenes médicas incluyen diferentes tipos de ruido que muestran distorsión y muchos problemas durante el diagnóstico de enfermedades [9]. La comunidad dedicada al denoising de MRI y a la mitigación de artefactos continúa reportando ganancias medibles en la segmentación posterior cuando el ruido se aborda explícitamente [10].

Un volumen de MRI puede representarse como una cuadrícula 3D de vóxeles. En MRI cerebral, los vóxeles se categorizan comúnmente en tres tipos de tejido: sustancia blanca (WM), sustancia gris (GM) y líquido cefalorraquídeo (CSF)[11]. En este trabajo, nos enfocamos en la segmentación binaria de GM: las anotaciones manuales de MRBrainS18 se fusionan (etiquetas 1 y 2) para formar una única máscara de GM; la Figura 1 ilustra el conjunto de datos y este mapeo de etiquetas. La WM se caracteriza por una mayor intensidad de señal que la GM y el CSF[12]. Aunque propiedades del tejido como el contenido de hierro, la densidad celular y la anisotropía del tejido son factores que alteran el contraste de la imagen [8]. Se han propuesto varios métodos en la literatura para segmentar MRI cerebrales. Estos pueden agruparse en dos categorías: métodos estadísticos y métodos basados en modelos deformables[7]. Trabajos recientes enfatizan que arquitecturas modernas de aprendizaje profundo (p. ej., variantes de U-Net) dominan la práctica actual, mientras que modelos híbridos CNN/Transformer con self-attention amplían los campos receptivos para dependencias de largo alcance[13]. Los priors basados en atlas han sustentado durante mucho tiempo el análisis de MRI estructural al ofrecer referencias espaciales estandarizadas y mapas de probabilidad de tejido, que reducen la variabilidad entre sujetos y proporcionan fuertes restricciones anatómicas[14]–[16].

En general, el objetivo de las técnicas de segmentación de imágenes es dividir la imagen en un conjunto de regiones no superpuestas tal que cada una de las regiones sea homogénea en alguna propiedad o característica [17], [18]. Las características que las regiones comparten son variadas y pertenecen a un amplio espectro que está determinado por la naturaleza de las imágenes. El resultado de la segmentación es una imagen con etiquetas que identifican cada región homogénea o un conjunto de contornos que describen los límites de la región[11]. La segmentación de imágenes cerebrales simplifica la representación de la imagen, lo que facilita su análisis[19].

Los estudios que incorporan técnicas de aprendizaje automático han mostrado un desempeño sobresaliente dentro del espectro de enfoques para abordar el problema del diagnóstico asistido por computador (CAD), la detección de enfermedades y el pronóstico [20]–[22]. Esto permite mitigar la dependencia del operador y aumentar la precisión de los diagnósticos. Entre sus usos se encuentran la detección y clasificación de tumores mamarios, el desarrollo y crecimiento fetales, la función cerebral, las lesiones cutáneas y las enfermedades pulmonares [23]. Revisiones recientes centradas específicamente en la segmentación de tejidos cerebrales en MRI (GM/WM/CSF) consolidan avances a lo largo del ciclo de vida y destacan desafíos pendientes (ruido, movimiento, difuminado de bordes)[24]. Dentro de las técnicas de aprendizaje automático, el campo del Deep Learning (DL) ha mostrado recientemente su robustez y alto nivel de precisión. DL comprende un enfoque derivado de modelos conexionistas bioinspirados llamados ANN (Artificial Neural Networks). Estas permiten aproximar, en teoría, cualquier función

matemática, lo que habilita un amplio dominio de aplicación en múltiples áreas como simulación, modelado y predicción. La complejidad del problema, representada en reproducir funciones matemáticas no lineales, multidimensionales y complejas, ha requerido el desarrollo de técnicas más robustas que permiten la construcción de mapeos de relaciones matemáticas entre los datos de entrada y la salida esperada. En paralelo, trabajos sintetizan el progreso en backbones de CNN y segmentación basada en Transformers, motivando el uso de mecanismos de autoatención para MRI cerebral.

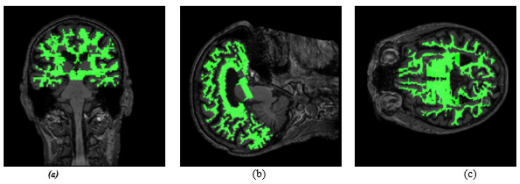

Fig. 1. *Conjunto de datos MRBrainS18 y mapeo binario de etiquetas de sustancia gris: vistas ortogonales—(a) coronal, (b) sagital, (c) axial—que muestran la máscara de GM fusionada (etiquetas 1+2) superpuesta en verde sobre MRI ponderada en T1.*
Fuente: elaboración propia.

En este artículo, se exploran algoritmos básicos derivados de la visión por computador, como la segmentación multiumbral, técnicas de crecimiento de regiones y algoritmos de aprendizaje profundo. Analizamos su desempeño y características en el problema de segmentar la sustancia gris en imágenes de resonancia magnética. Nuestro diseño experimental sigue las directrices de buenas prácticas para la evaluación a fin de asegurar la comparabilidad con la literatura reciente[2]. Describimos algoritmos clásicos (Otsu multiumbral, crecimiento de regiones) y un enfoque moderno basado en aprendizaje. Se reportan resultados cuantitativos para el modelo basado en aprendizaje, mientras que los métodos clásicos se mantienen como referencias metodológicas.

METODOLOGÍA

Los métodos de segmentación de imágenes pueden clasificarse a lo largo de tres ejes ortogonales. Por grado de intervención humana pueden ser manuales, semiautomáticos o automáticos; por modalidad pueden ser monomodales cuando se usa una sola característica o imagen, o multimodales cuando se combinan múltiples características complementarias; y por criterio pueden basarse en homogeneidad (regiones definidas por similitud dentro de la región) o basarse en discontinuidades (fronteras definidas por discontinuidades de características o bordes)[25]–[27].

La segmentación de GM a partir de imágenes cerebrales es un problema de segmentación multietiqueta debido a la presencia de píxeles que pertenecen a diferentes tipos de tejido. Consideramos tres tipos de enfoques de segmentación: segmentación basada en umbrales, segmentación basada en regiones y segmentación basada en aprendizaje profundo.

Otsu multiumbral

Una ventaja del umbralado es que es relativamente simple de implementar y puede usarse eficientemente en aplicaciones con pocas clases. La principal desventaja de este enfoque es que la calidad de la segmentación es sensible al ruido y a la presencia de artefactos en la imagen. Seleccionamos el método de segmentación Otsu multiumbral (MTO) debido a que maximiza la varianza entre clases como forma de optimización de umbrales[28]. MTO asume el histograma de una imagen de$L$niveles de gris como una distribución de probabilidad. Si una imagen se divide en N clases(C_1, C_2, …, C_n), MTO estimaN−1umbrales(t_1, t_2, …, t_{n-1}). La varianza entre clases se estima mediante:

$\sum_{k=1}^{N} k^2 k^2 T^2 \sum_{k=1}^{N} k^2 k^2 T^2$	(1)

dondew_(k)=∑_(i□C_(k))^()p_(i),μ(k)=∑_(i□C_(k))^()ip_(i),μ_(k)=∑_(i□C_(k))^()μ(k)/w_(k)son llamados momentos acumulados de orden cero y primero de la clasek_{th},C_k y$p_i$es el número de píxeles en el nivel de gris$i$.

Los umbrales óptimos(t_(1),t_(2), …, t_(N−1))son calculados maximizando la varianza entre clases, así:

Desde la perspectiva de Otsu, la estimación de múltiples umbrales puede requerir un procedimiento iterativo de optimización que suele ser computacionalmente costoso; existen múltiples propuestas dirigidas a reducir los costos computacionales, haciéndolo más eficiente y menos costoso[29]–[31].

Crecimiento de regiones

$\{1, 2, \ldots, N-1\}^{\ldots}$ $1 \le t_1 < \ldots < t_{N-1} < L$ B^2 $1, 2, \ldots, N-1$	(2)

La idea principal de los métodos de segmentación por crecimiento de regiones es que la región crece por agregación de píxeles utilizando medidas de similitud y discontinuidad[32],[33].

Este método parte de un conjunto inicial de puntos semilla, que se utilizan para el cálculo de agrupamiento de similitudes tanto

basadas en distancia como en intensidad entre todos los píxeles contenidos dentro del clúster. A partir de un punto dado, se calcula una medida de similitud para determinar si dos píxeles pertenecen al mismo objeto o clase. La inclusión o exclusión de un píxel depende de las estadísticas de los valores de intensidad circundantes[34].

El punto inicial se selecciona de forma arbitraria y se denomina punto semilla. Esto constituye su principal desventaja, ya que no existe un procedimiento determinista para esta selección y el resultado de la segmentación es altamente sensible al punto semilla seleccionado. En este trabajo, utilizamos el segundo umbral obtenido de MTO para la selección del punto semilla. Una vez estimado este valor, se buscan los índices de los píxeles que tienen un nivel de intensidad específico. De este conjunto, se selecciona aleatoriamente un subconjunto de no más de 5 puntos semilla para iniciar el crecimiento de regiones.

thrds = threshold_multiotsu(inputImage)

seeds = indexPoints(inputImage, thrds [1])

imgGM = regionGrowing(inputImage, seeds)

Métodos de segmentación por Deep Learning

El aprendizaje automático ha ganado mucho interés en amplios dominios de aplicación del procesamiento de imágenes, incluido el campo del procesamiento de imágenes médicas[35]–[37]. Este interés se debe a su desempeño sobresaliente, mostrando mejores resultados en comparación con otros métodos tradicionales existentes en la literatura[35],[38].

El aprendizaje profundo es un campo derivado de las redes neuronales artificiales y se caracteriza por parámetros e hiperparámetros. Este gran número de elementos parece ser la base de la robustez de las técnicas basadas en este enfoque. Sin embargo, también constituye su principal desventaja al incrementar los costos computacionales para su entrenamiento y despliegue.

Las redes neuronales convolucionales (CNN) se han convertido en el enfoque de referencia para el procesamiento de imágenes. La parte central de las arquitecturas CNN son las capas de convolución[38]. Estas se basan en la operación de convolución tradicional, aplicando un conjunto de filtros a la imagen y extrayendo características relevantes de la imagen utilizadas en la parte final con fines de clasificación.

Es bien sabido que, dado que las convoluciones son locales y espaciales, apilar capas produce una extracción jerárquica de características. Esto introduce un sesgo espacial inductivo que asume patrones estructurados en la entrada. El sesgo espacial añade algunos beneficios a las CNN para tareas de clasificación de objetos o segmentación de estructuras. Esto también permite mitigar el requisito de grandes conjuntos de entrenamiento.

Además, reducir las interacciones de las neuronas a un vecindario espacial local puede no ser beneficioso en otras tareas de visión por computador, como la comprensión o la descripción de imágenes. En esta tarea, las características que definen la escena deben estimarse considerando interacciones espacialmente más amplias. Recientemente, el uso de modelos basados en atención ha incorporado estas características, permitiendo la exploración de un dominio espacial más amplio que el permitido por las convoluciones clásicas. Su derivación proviene del campo del procesamiento de lenguaje natural, donde la relación de una palabra no necesariamente corresponde a las otras palabras cercanas[39].

Conjunto de datos

Se utilizó el conjunto de datos MRBrainS18. Proporciona sujetos de entrenamiento con un estándar de referencia manual y casos de prueba adicionales. Para cada sujeto hay tres secuencias de MRI: T1 ponderada (3D, corregida por campo de sesgo), T1 con inversión-recuperación (multicorte, corregida por campo de sesgo) y T2-FLAIR (multicorte, corregida por campo de sesgo); se suministran variantes registradas. Las etiquetas manuales comprenden 11 clases: 0 fondo; 1 sustancia gris cortical (GM); 2 ganglios basales (GM profunda); 3 sustancia blanca (WM); 4 lesiones de WM; 5 CSF (extracerebral); 6 ventrículos; 7 cerebelo; 8 tronco encefálico; 9 infarto; 10 otros.

El conjunto de datos está publicado en DataVerseNL[40], que proporciona los datos de entrenamiento, los datos de prueba y las segmentaciones de referencia manuales, con información adicional y enlaces a código alojados por los organizadores del desafío.

Mapeo de etiquetas

MRBrainS18 proporciona un estándar de referencia manual. Para nuestros experimentos de segmentación binaria de sustancia gris (GM), fusionamos las etiquetas 1 (GM cortical) y 2 (ganglios basales) en una sola clase GM, siguiendo la guía del desafío para protocolos de tres etiquetas que fusionan subetiquetas de GM y WM cuando corresponde. Todas las demás etiquetas se mapearon a fondo para el ajuste binario (etiquetas 3–8, 9–10 ignoradas para la evaluación si están presentes). Esta política de etiquetas asegura consistencia con las recomendaciones de MRBrainS18 para evaluaciones con etiquetas fusionadas.

Preprocesamiento

Adoptamos un pipeline para (i) armonizar las dimensiones de volumen entre sujetos, (ii) habilitar entrenamiento e inferencia basados en parches, eficientes en memoria, sin alterar la geometría voxel nativa, y (iii) mantener la ruta de datos reproducible y comparable entre métodos. Los pasos de preprocesamiento son:

- Armonización de forma: Cada volumen 3D se mapea a un campo de visión fijo de 256×256×256 mediante recorte central cuando una dimensión excede 256 vóxeles y relleno con ceros cuando es menor. Los rellenos exactos por eje se registran para restaurar la extensión original después de la inferencia. Esta estrategia preserva la geometría voxel nativa y evita cualquier remuestreo de intensidades.
- Extracción de parches: Del cubo 256^3 extraemos parches cúbicos de 128×128×128 con un paso de 64 (~50% de solapamiento). Durante el entrenamiento y la inferencia, omitimos los parches vacíos/solo de fondo (es decir, parches sin tejido cerebral), lo que reduce el cómputo y evita lotes degenerados, dejando sin cambios el patrón de teselado.
- Etiquetas a imagen: Leemos los metadatos de cabecera NIfTI de los mapas de etiquetas y aplicamos la afinidad almacenada para llevar las anotaciones al espacio de imagen nativo del volumen T1 correspondiente. Las etiquetas se re-mallaron a la red de la imagen usando interpolación del vecino más cercano para preservar los índices discretos. Todas las transformaciones se registraron para que el entrenamiento, la inferencia y la evaluación operen de manera consistente en el espacio original.
- Restauración pos-inferencia: Las predicciones obtenidas en el cubo fijo se mapean de vuelta al campo de visión nativo eliminando los rellenos registrados a lo largo de cada eje, invirtiendo exactamente el Paso 1.

Esta canalización no introduce normalización de intensidad, corrección de sesgo ni remuestreo geométrico; preserva intencionalmente las características de adquisición del conjunto de datos y se apoya en los datos alineados y anotados manualmente del desafío para minimizar la varianza inducida por el preprocesamiento.

Empleamos una red de segmentación codificador–decodificador 3D[41], adaptada a la delimitación de sustancia gris (GM). El codificador combina procesamiento convolucional local con autoatención global sobre parches volumétricos no superpuestos, mientras que el decodificador realiza remuestreo multiescala con conexiones de salto para recuperar detalle anatómico fino. El diseño preserva el contexto espacial en múltiples resoluciones y agrega dependencias de largo alcance que son críticas para los límites de GM (ver Figura 2).

- Codificador: su objetivo es la extracción de características global–local. Los volúmenes se procesan como parches 3D de 128×128×128. Las características se proyectan primero a una representación de tokens, luego se procesan mediante una pila de bloques de autoatención multi-cabeza (4 cabezas) intercalados con mapeos feed-forward livianos (tamaño oculto ≈ 192, ancho de la MLP ≈ 96). Capas residuales y de normalización estabilizan el entrenamiento. Características multiescala de distintas profundidades del codificador se exponen al decodificador mediante conexiones de salto por etapas, habilitando que el decodificador fusione contexto global con detalles de alta resolución.
- Decodificador: su objetivo es la fusión multiescala y la reconstrucción. El decodificador[42]–[44]sigue una vía estándar de arriba hacia abajo con upsampling por convolución transpuesta (stride 2) y concatenación con las características correspondientes del codificador en cada escala. Cada etapa refina la representación fusionada con bloques convolucionales/residuales compactos, restaurando progresivamente la resolución espacial y afinando las interfaces de tejido. El predictor final es una convolución1×1×1que produce dos logits (GM vs. fondo).
- entrenamiento y optimización: optimizamos una pérdida compuesta que suma términos Dice y Focal con pesos iguales. La optimización usa Adam con una tasa de aprendizaje base de 10e−3 y gradient clipping para mejorar la estabilidad. La pérdida es la suma de Dice[45]y Focal Loss[46]calculadas sobre dos canales (GM vs. fondo) con pesos iguales. Esto enfatiza la fidelidad de solapamiento mientras aborda el desbalance de clases[45],[47].
- Inferencia: En tiempo de inferencia, cada volumen se particiona en parches de128×128×128con 50% de solapamiento (stride = 64). Las predicciones de probabilidad de GM por parche se fusionan en las regiones de solapamiento mediante promedio simple (sin windowing). Luego, el mapa de probabilidad fusionado se umbraliza en 0.5 para producir la máscara binaria de GM. Finalmente, se elimina cualquier padding introducido durante la armonización de forma para recuperar la extensión nativa. El solapamiento del 50% es intencionalmente conservador, pero simple y robusto en la práctica.

RESULTADOS

Los experimentos se realizaron en Ubuntu 22.04.5 (kernel 6.8) con 250 GiB de RAM y dos GPU NVIDIA RTX A6000 (48 GB de VRAM cada una).

La arquitectura de aprendizaje profundo se implementó utilizando un stack moderno de desarrollo, basado en Python 3.10 y PyTorch 2.9.0.

Métodos

Se evaluaron tres algoritmos: (i) el método tradicional de Otsu con múltiples umbrales, (ii) el crecimiento de regiones y (iii) una arquitectura de aprendizaje profundo propuesta.

La principal limitación del umbralado múltiple radica en que cada imagen debe procesarse de forma independiente para determinar los valores óptimos de segmentación. Al tratarse de un enfoque basado exclusivamente en intensidades, no es posible definir un conjunto único de umbrales que generalice adecuadamente a diferentes

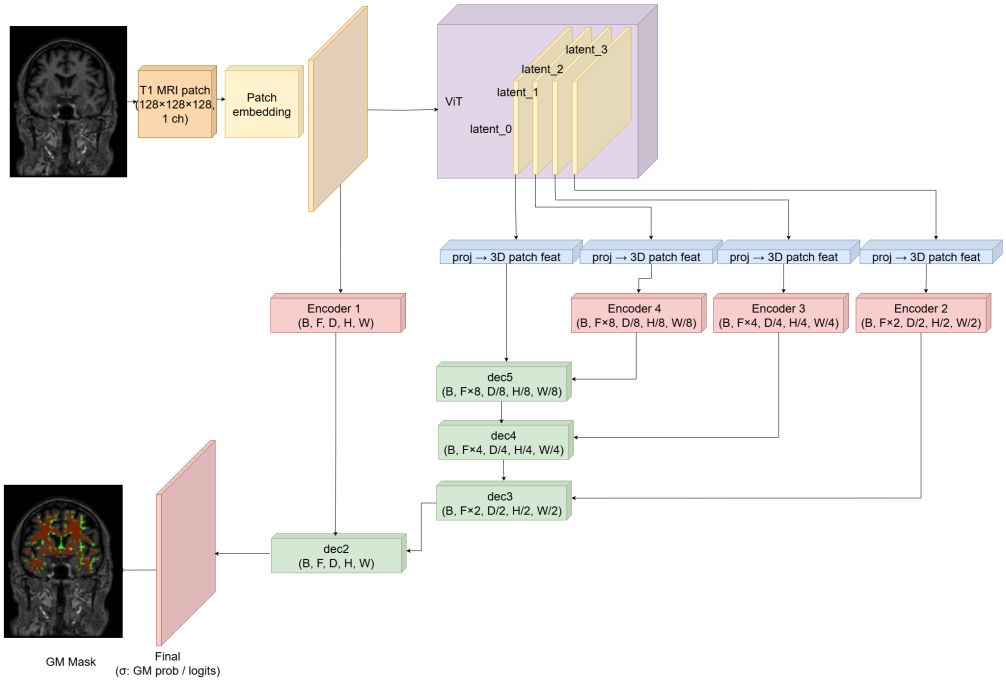

Fig. 2. *Codificador–decodificador 3D para segmentación de GM.*
Fuente: elaboración propia.

Por otro lado, los métodos basados en técnicas de crecimiento de regiones dependen de la selección de las semillas. Este valor suele ser arbitrario y lo establece el usuario. En este enfoque, usamos los umbrales MTO como referencia. Estos enfoques requieren menos parámetros. Sin embargo, es necesario no solo especificar los valores de semilla, sino también los límites mínimo y máximo permitidos en la intensidad para aplicar el algoritmo. Estos valores siguen siendo arbitrarios y el resultado final es altamente sensible a ellos.

Para el último modelo, entrenamos la red neuronal profunda propuesta usando los hiperparámetros descritos en la Tabla 1, utilizando el optimizador Adam[48], con un tamaño de imagen de entrada de128×128×128y 32 imágenes por batch.

Tabla 1: hiperparámetros usados para entrenar el codificador–decodificador 3D.

Hyperparametros	Valor
Tamaño de entrada	128×128×128
Tamaño de lote	4
Función de pérdida	Dice + Binary Focal
Reducción de Focal	mean
Optimizador	Adam, lr=1e-3 [48]
Inicialización de pesos	He initialization [49]
Tasa de aprendizaje de Adam	0.001
Adam β_1, β_2	0.9, 0.999
Adam ε	1e-08
Número de épocas	300
Detención temprana	None
Precisión mixta	off
Umbral de inferencia	0.5

Fuente: elaboración propia.

Deep Learning

El modelo basado en aprendizaje muestra una disminución monótona tanto en las pérdidas de entrenamiento como de validación, con una pequeña y estable brecha de generalización hacia el final del entrenamiento (Figura 3a). La mejor pérdida de validación ocurre en la época 297 (val = 0.1662) y coincide estrechamente con la pérdida de entrenamiento (0.1673), lo que sugiere convergencia sin sobreajuste en el checkpoint seleccionado. Una vista complementaria de la brecha de generalización (validación − entrenamiento) a lo largo de las épocas (Figura 3b) confirma que la brecha tiende a cero. Las últimas 10 épocas (Figura 3c) exhiben baja varianza en la pérdida de validación (0.1739 ± 0.0034), reforzando la estabilidad del régimen de entrenamiento final.

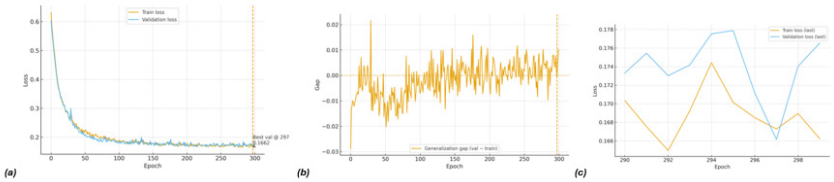

Fig. 3. *a) Pérdida de entrenamiento y validación a lo largo de las épocas, b) brecha de generalización a lo largo de las épocas, y c) curvas de aprendizaje en las últimas 10 épocas.*
Fuente: elaboración propia.

Desempeño de segmentación

Se evaluo el modelo basado en aprendizaje en el conjunto de prueba. El modelo alcanza Dice de 0.650 ± 0.043 (mediana 0.660; min–max 0.548 – 0.718) e IoU 0.483 ± 0.046 (mediana 0.492; min–max 0.377 – 0.561), con Precisión 0.700 ± 0.046 y Recall 0.610 ± 0.059 (Tabla 2). Las distribuciones por sujeto (Fig. 4a–d) muestran una dispersión moderada con rangos intercuartílicos compactos, lo que indica un rendimiento consistente entre casos. Un análisis complementario frente al error de borde (Fig. 5a) no mostró evidencia de una relación monótona entre solapamiento y distancia de Hausdorff (Pearson r=0.076, p=0.69; Spearman ρ=0.081, p=0.671). El histograma de Dice (Fig. 5b) confirma una distribución centrada con una pequeña fracción de outliers de menor desempeño.

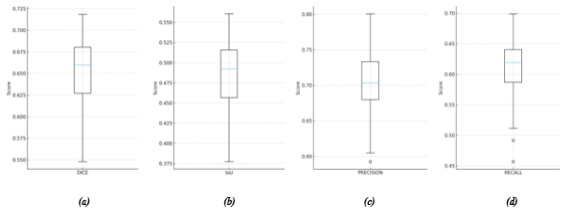

Fig. 4. *Puntajes de segmentación por sujeto en el conjunto de prueba. (a) Dice; (b) Intersección-sobre-Unión (IoU); (c) Precisión; (d) Recall. Las cajas muestran la mediana y el rango intercuartílico; los bigotes se extienden hasta 1,5×IQR; los puntos denotan sujetos individuales.*

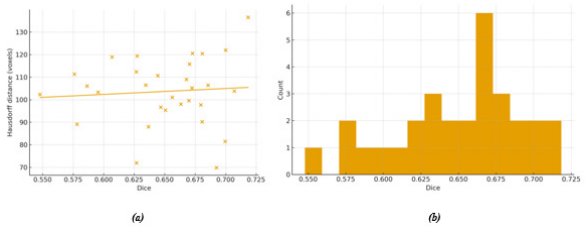

Fig. 5. *Solapamiento vs. error de borde y distribución de Dice. (a) Dice vs. distancia de Hausdorff (vóxeles) para los sujetos de prueba; la línea sólida muestra el ajuste por mínimos cuadrados (menor Hausdorff es mejor). (b) Histograma de los puntajes Dice por sujeto (más alto es mejor).*
Fuente: elaboración propia.

En términos de volumetría, el modelo presenta un sesgo de volumen negativo (RVE medio = −12.54% ± 9.53%, mediana −11.45%), es decir, una tendencia a subsegmentar la sustancia gris en relación con la referencia manual (Tabla 2).

Tabla 2: desempeño de segmentación en el conjunto de prueba: Dice, IoU, Precisión, Recall, distancia de Hausdorff (vóxeles) y error relativo de volumen (%).

Métrica	Media ± DE	Mediana [RIC]	Mín–Máx
Dice	0.650 ± 0.043	0.660 [0.627, 0.680]	0.548–0.718
IoU	0.483 ± 0.046	0.492 [0.457, 0.516]	0.377–0.561
Precisión	0.700 ± 0.046	0.703 [0.680, 0.733]	0.592–0.800
Recall	0.610 ± 0.059	0.619 [0.587, 0.640]	0.457–0.699
Hausdorff (vóxeles)	103.656 ± 14.785	104.494 [96.936, 112.084]	69.871–136.532
Error relativo de volumen (%)	-12.541 ± 9.535	-11.451 [-18.821, -5.975]	-33.266–5.119
RVE absoluto (%)	12.882 ± 9.052	11.451 [5.975, 18.821]	0.277–33.266

Fuente: elaboración propia.

Un análisis de Bland–Altman (Fig. 6) confirma un desplazamiento negativo sistemático con límites de acuerdo que reflejan la variabilidad morfológica entre sujetos, lo que sugiere que el modelo es conservador en las interfaces tisulares—consistente con el patrón Precisión > Recall. El mejor sujeto (por Dice) alcanza 0.718 mientras que el peor logra 0.548, ilustrando la dispersión observada.

Los resultados muestran que el codificador–decodificador con contexto global produce métricas de solapamiento estables entre sujetos, aunque tiende a subestimar el volumen de GM (algunos ejemplos se muestran en la Figura 7). Refinamientos futuros podrían dirigirse al Recall y a la adherencia al borde para reducir el sesgo de volumen observado sin sacrificar la precisión.

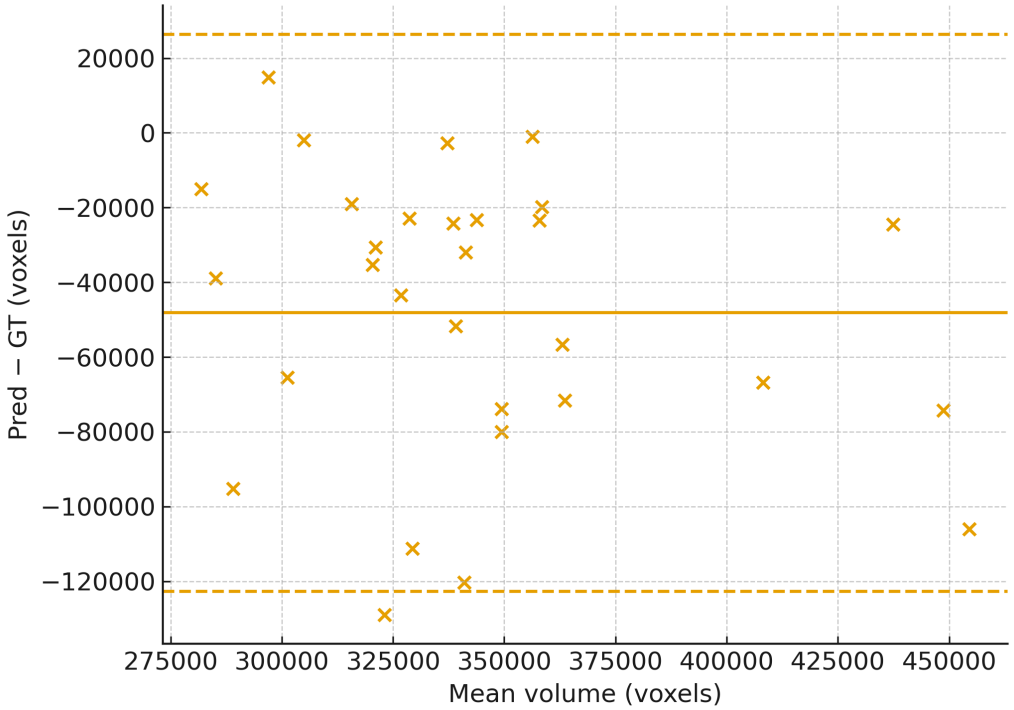

Fig. 6. *Análisis de Bland–Altman del volumen de sustancia gris (Pred − GT) con sesgo medio y límites de acuerdo del 95%.*
Fuente: elaboración propia.

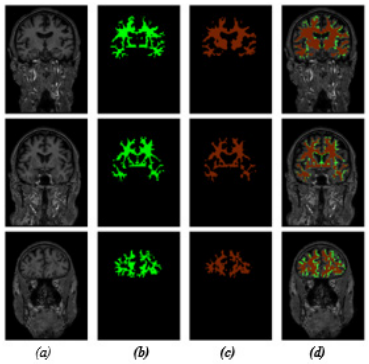

Fig 7. Segmentación cualitativa de sustancia gris en tres sujetos de prueba retenidos (cortes coronales): (a) MRI ponderada en T1; (b) referencia manual (verde); (c) predicción del modelo (naranja); (d) superposición sobre la MRI.
Fuente: elaboración propia.

Para contextualizar el desempeño del modelo de DL, calculamos resultados de referencia para dos enfoques clásicos: Otsu multiumbral y crecimiento de regiones. Ambos se aplicaron a un subconjunto representativo de MRBrainS18. Como se resume en la Tabla 3, estos métodos clásicos arrojan solapamientos limitados debido a su sensibilidad a la inhomogeneidad de intensidad y al ruido. El modelo propuesto de deep learning mejora Dice e IoU en más de un 50% con respecto a la mejor línea base, evidenciando la eficacia del contexto global aprendido y la reconstrucción con skips para la delimitación de sustancia gris.

Tabla 3: desempeño comparativo de segmentación de los métodos de línea base y el método propuesto.

Método	Dice	IoU	Precision	Recall
Otsu multiumbral	0.41 ± 0.07	0.28 ± 0.06	0.49 ± 0.08	0.36 ± 0.05
Crecimiento de regiones	0.46 ± 0.06	0.33 ± 0.05	0.55 ± 0.07	0.39 ± 0.06
Deep Learning (nuestro)	0.65 ± 0.04	0.48 ± 0.05	0.70 ± 0.05	0.61 ± 0.06

Fuente: elaboración propia.

CONCLUSIONES

Se propone un pipeline para la segmentación de sustancia gris que combina una estrategia mínima de preprocesamiento con una red codificador–decodificador 3D que integra autoatención global y conexiones de salto multiescala.

En el conjunto de prueba retenido, el método alcanzó Dice 0.650 ± 0.043 (mediana 0.660) e IoU 0.483 ± 0.046 (mediana 0.492), con Precisión 0.700 ± 0.046 y Recall 0.610 ± 0.059. Estos resultados indican un solapamiento confiable con la referencia manual entre sujetos. El análisis de Dice vs. distancia de Hausdorff no mostró una asociación monótona, reflejando la sensibilidad de las métricas de borde a errores localizados incluso cuando el solapamiento global es adecuado. El histograma de Dice revela una distribución centrada con una pequeña fracción de casos de menor desempeño, consistente con la variabilidad entre sujetos.

Desde una perspectiva volumétrica, el modelo exhibe un sesgo negativo (RVE medio = −12.54% ± 9.53%), es decir, una tendencia a subsegmentar la GM en relación con la referencia manual. La gráfica de Bland–Altman confirma este desplazamiento sistemático y cuantifica los límites de acuerdo. Esto sugiere predicciones conservadoras cerca de las interfaces tisulares.

Comparado con los métodos clásicos de referencia (Otsu y crecimiento de regiones), el modelo propuesto de deep learning mejora Dice e IoU en aproximadamente 50–60%, evidenciando la ventaja del contexto global aprendido para la delimitación de GM.

Las fortalezas del enfoque incluyen (i) una ruta de preprocesamiento ligera y reproducible que preserva la geometría voxel nativa, (ii) dinámicas de entrenamiento estables con sobreajuste limitado, y (iii) consistencia entre sujetos en las métricas de solapamiento. Las limitaciones incluyen un ajuste binario de etiquetas, un sesgo de subsegmentación medible y el uso de un único conjunto de datos público.

El trabajo futuro se orientará a la adherencia al borde y al recall, la calibración de volumen y una validación más amplia a través de escáneres y cohortes. Extensiones a segmentación multiclase (GM/WM/CSF) y modelado de incertidumbre podrían mejorar aún más la robustez, mientras que priors anatómicos débiles o adaptación de dominio pueden reducir el sesgo residual sin sacrificar la precisión.

REFERENCIAS

[1] L. Wang, T. Chitiboi, H. Meine, M. Günther, and H. K. Hahn, “Principles and methods for automatic and semi-automatic tissue segmentation in MRI data,” *MAGMA*, vol. 29, pp. 95–110, 2016, doi: 10.1007/s10334-015-0520-5.

[2] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, “Medical image segmentation: A comprehensive review of deep learning-based methods,” *Tomography*, vol. 11, Art. no. 52, 2025, doi: 10.3390/tomography11050052.

[3] B. Kim, H. Kim, S. Kim, and Y. Hwang, “A brief review of non-invasive brain imaging technologies and the near-infrared optical bioimaging,” *Applied Microscopy*, vol. 51, Art. no. 9, 2021, doi: 10.1186/s42649-021-00058-7.

[4] C. A. Nelson, “Incidental findings in magnetic resonance imaging (MRI) brain research,” *J. Law Med. Ethics*, vol. 36, pp. 315–213, 2008, doi: 10.1111/j.1748-720X.2008.00275.x.

[5] Md. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, “Deep learning for medical image segmentation: State-of-the-art advancements and challenges,” *Informatics in Medicine Unlocked*, vol. 47, Art. no. 101504, 2024, doi: 10.1016/j.imu.2024.101504.

[6] W. E. Kwok, “Radiofrequency interference in magnetic resonance imaging: Identification and rectification,” *J. Clin. Imaging Sci.*, vol. 14, Art. no. 33, 2024, doi: 10.25259/JCIS_74_2024.

[7] E. D. Angelini, T. Song, B. D. Mensh, and A. F. Laine, “Brain MRI segmentation with multiphase minimal partitioning: A comparative study,” *Int. J. Biomed. Imaging*, vol. 2007, Art. no. 10526, 2007, doi: 10.1155/2007/10526.

[8] D. Feng, L. Tierney, and V. Magnotta, “MRI tissue classification using high-resolution Bayesian hidden Markov normal mixture models,” *J. Amer. Statist. Assoc.*, vol. 107, pp. 102–119, 2012, doi: 10.1198/jasa.2011.ap09529.

[9] P. Kaur, G. Singh, and P. Kaur, “A review of denoising medical images using machine learning approaches,” *Curr. Med. Imaging Rev.*, vol. 14, pp. 675–685, 2018, doi: 10.2174/1573405613666170428154156.

[10] A. Pereira Neto and F. J. B. Barros, “Noise reduction in brain magnetic resonance imaging using adaptive wavelet thresholding based on linear prediction factor,” *Front. Neurosci.*, vol. 18, 2025, doi: 10.3389/fnins.2024.1516514.

[11] I. Despotović, B. Goossens, and W. Philips, “MRI segmentation of the human brain: Challenges, methods, and applications,” *Computational and Mathematical Methods in Medicine*, vol. 2015, Art. no. 450341, 2015, doi: 10.1155/2015/450341.

[12] A. L. Alexander, S. A. Hurley, A. A. Samsonov, N. Adluru, A. P. Hosseinbor, P. Mossahebi, D. P. M. Tromp, E. Zakszewski, and A. S. Field, “Characterization of cerebral white matter properties using quantitative magnetic resonance imaging stains,” *Brain Connect.*, vol. 1, pp. 423–446, 2011, doi: 10.1089/brain.2011.0071.

[13] Md. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, “Deep learning for medical image segmentation: State-of-the-art advancements and challenges,” *Informatics in Medicine Unlocked*, vol. 47, Art. no. 101504, 2024, doi: 10.1016/j.imu.2024.101504.

[14] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, and the Brain Development Cooperative Group, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, pp. 313–327, 2011, doi: 10.1016/j.neuroimage.2010.07.033.

[15] V. Fonov, A. Evans, R. McKinstry, C. Almli, and D. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage*, vol. 47, p. S102, 2009, doi: 10.1016/S1053-8119(09)70884-5.

[16] D. L. Collins, A. P. Zijdenbos, W. F. C. Baaré, and A. C. Evans, "ANIMAL+INSECT: Improved cortical structure segmentation," in Proc. 16th Int. Conf. Information Processing in Medical Imaging (IPMI), Berlin, Heidelberg, Germany: Springer, Jun. 1999, pp. 210–223.

[17] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, pp. 1277–1294, 1993, doi: 10.1016/0031-3203(93)90135-J.

[18] M. K. Kar, M. K. Nath, and D. R. Neog, "A review on progress in semantic image segmentation and its application to medical images," *SN Comput. Sci.*, vol. 2, Art. no. 397, 2021, doi: 10.1007/s42979-021-00784-5.

[19] R. Kadri, M. Tmar, and B. Bouaziz, "Brain image processing using deep learning: An overview," in Digital Health in Focus of Predictive, Preventive and Personalised Medicine, L. Chaari, Ed. Cham, Switzerland: Springer, 2020, pp. 77–86.

[20] G. Zhu, B. Jiang, L. Tong, Y. Xie, G. Zaharchuk, and M. Wintermark, "Applications of deep learning to neuro-imaging techniques," *Front. Neurol.*, vol. 10, 2019.

[21] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, pp. 387–399, 2011, doi: 10.1016/j.neuroimage.2010.11.004.

[22] A.-I. Barranco-Gutiérrez, "Machine learning for brain images classification of two language speakers," *Comput. Intell. Neurosci.*, vol. 2020, Art. no. 9045456, 2020, doi: 10.1155/2020/9045456.

[23] S. Wang and R. M. Summers, "Machine learning and radiology," *Med. Image Anal.*, vol. 16, pp. 933–951, 2012, doi: 10.1016/j.media.2012.02.005.

[24] L. Wu, S. Wang, J. Liu, L. Hou, N. Li, F. Su, X. Yang, W. Lu, J. Qiu, M. Zhang, et al., "A survey of MRI-based brain tissue segmentation using deep learning," *Complex Intell. Syst.*, vol. 11, Art. no. 64, 2024, doi: 10.1007/s40747-024-01639-1.

[25] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep learning-based image segmentation on multimodal medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, pp. 162–169, 2019, doi: 10.1109/TRPMS.2018.2890359.

[26] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–337, 2000, doi: 10.1146/annurev.bioeng.2.1.315.

[27] Md. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep learning for medical image segmentation: State-of-the-art advancements and challenges," *Informatics in Medicine Unlocked*, vol. 47, Art. no. 101504, 2024, doi: 10.1016/j.imu.2024.101504.

[28] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst., Man, Cybern., vol. 9, pp. 62–66, 1979, doi: 10.1109/TSMC.1979.4310076.

[29] W. A. Hussein, S. Sahran, and S. N. H. S. Abdullah, "A fast scheme for multilevel thresholding based on a modified bees algorithm," *Knowl.-Based Syst.*, vol. 101, pp. 114–134, 2016, doi: 10.1016/j.knosys.2016.03.010.

[30] D.-Y. Huang and C.-H. Wang, "Optimal multi-level thresholding using a two-stage Otsu optimization approach," *Pattern Recognit. Lett.*, vol. 30, pp. 275–284, 2009, doi: 10.1016/j.patrec.2008.10.003.

[31] R. Panda, L. Samantaray, A. Das, S. Agrawal, and A. Abraham, "A novel evolutionary row class entropy based optimal multi-level thresholding technique for brain MR images," *Expert Syst. Appl.*, vol. 168, Art. no. 114426, 2021, doi: 10.1016/j.eswa.2020.114426.

[32] S. Tan, L. Li, W. Choi, M. K. Kang, W. D. D'Souza, and W. Lu, "Adaptive region-growing with maximum curvature strategy for tumor segmentation in ^{18}F-FDG PET," *Phys. Med. Biol.*, vol. 62, no. 13, pp. 5383–5402, 2017, doi: 10.1088/1361-6560/aa6e20.

[33] M. Tamal, "A hybrid region growing tumour segmentation method for low contrast and high noise nuclear medicine images by combining a novel nonlinear diffusion filter and global gradient measure," *Heliyon*, vol. 5, no. 11, Art. no. e02993, 2019, doi: 10.1016/j.heliyon.2019.e02993.

[34] S. Bama, R. Velumani, N. B. Prakash, G. R. Hemalakshmi, and A. Mohanarathinam, "Automatic segmentation of melanoma using superpixel region growing technique," *Mater. Today Proc.*, vol. 45, pp. 1726–1732, 2021, doi: 10.1016/j.matpr.2020.08.618.

[35] X. Zhao and X.-M. Zhao, "Deep learning of brain magnetic resonance images: A brief review," *Methods*, vol. 192, pp. 131–140, 2021, doi: 10.1016/j.ymeth.2020.09.007.

[36] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "A review of deep learning based methods for medical image multi-organ segmentation," *Physica Medica*, vol. 85, pp. 107–122, 2021, doi: 10.1016/j.ejmp.2021.05.003.

[37] R. Balakrishnan, M. del C. Valdés Hernández, and A. J. Farrall, "Automatic segmentation of white matter hyperintensities from brain magnetic resonance images in the era of deep learning and big data—A systematic review," *Comput. Med. Imaging Graph.*, vol. 88, Art. no. 101867, 2021, doi: 10.1016/j.compmedimag.2021.101867.

[38] H. Izadkhah, "Medical image processing: An insight to convolutional neural networks," in Deep Learning in Bioinformatics, H. Izadkhah, Ed. Academic Press, 2022, pp. 175–213.

[39] Y. Xu et al., "Transformers in computational visual media: A survey," *Comput. Visual Media*, vol. 8, no. 1, pp. 33–62, 2022, doi: 10.1007/s41095-021-0247-3.

[40] H. J. Kuijf et al., "MR brain segmentation challenge 2018 data," 2024.

[41] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI,

Cham, Switzerland: Springer, 2015, pp. 234–241.

[42] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," arXiv:1603.07285, 2018.

[43] H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel transposed convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1218–1227, 2020, doi: 10.1109/TPAMI.2019.2893965.

[44] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in Proc. 27th Int. Conf. Mach. Learn. (ICML), Madison, WI, USA, 2010, pp. 807–814.

[45] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in Proc. MICCAI, 2017, pp. 240–248, doi: 10.1007/978-3-319-67558-9_28.

[46] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," arXiv:1708.02002, 2018.

[47] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imaging*, vol. 25, no. 11, pp. 1451–1461, 2006, doi: 10.1109/TMI.2006.880587.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980, 2017.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," arXiv:1502.01852, 2015.

FINANCIACIÓN

Ninguna

CONFLICTOS DE INTERÉS

No existen

CONTRIBUCIÓN DE AUTORÍA

Redacción - borrador original: Camilo Andres Laiton Bonadiez, German Sanchez Torres, Carlos Nelson Henríquez Miranda.

Redacción - revisión y edición: Camilo Andres Laiton Bonadiez, German Sanchez Torres, Carlos Nelson Henríquez Miranda.