ISSN: 1692-7257 - Volume 1 – Issue 47 - 2026

ORIGINAL

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

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

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

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.**

Received: 24 July 2025. Revised: 24 September 2025. Accepted: 20 December 2025. Published: 01 January 2026.
Corresponding Author: Carlos Nelson Henríquez Miranda

How to cite: 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

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

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 ≈ 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.

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 ≈ 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.

INTRODUCTION

Neuroimaging is the main tool for diagnosing brain-related diseases, spanning a broad spectrum of disorders and providing anatomical and functional information about the brain [1]. Recent reviews reaffirm the centrality of neuroimaging for computer-assisted diagnostic and segmentation workflows, consolidating methodologies and evaluation practices [2]. Understanding brain structure and the underlying neural mechanism is essential for monitoring and early diagnosis of diseases, in order to prevent them from progressing to a severe level [3].

Computer-assisted diagnoses are based on non-invasive imaging techniques. The most widely available modalities include magnetic resonance imaging (MRI), positron emission tomography (PET) and X-ray computed tomography (CT), which are the most popular techniques for brain imaging[3]. Magnetic resonance imaging (MRI) is one of the most popular imaging types; it uses a strong magnetic field and radiofrequency waves to visualize organs, soft tissue and bones[4].

The main goal of structural brain MRI analysis includes the classification of specific tissue types, as well as the identification and description of specific anatomical structures. The literature shows that deep learning-based segmentation has become the predominant paradigm for delineating tissue and structures in clinical and research settings[5]. However, extracting high-quality information from brain images is challenging due to the low signal-to-noise ratio and the artifacts generated during the acquisition process.[6]details persistent sources of artifacts, radiofrequency interference and related acquisition artifacts, and their impact on subsequent segmentation.

Classifying MRI data into tissue types can be achieved using a variety of methods, including manual visual inspection and semi-automated tissue classification techniques [7]. Manual visual inspection of MRI data is a time-consuming process that requires considerable expertise [8], which makes it impractical for processing large databases. Therefore, semi-automated tissue classification techniques have been developed for many low-cost MRI datasets. However, a weakness of these approaches is that they lose precision due to images with a high presence of noise or artifacts. Indeed, medical images include different types of noise that show distortion and many problems during disease diagnosis [9]. The community dedicated to MRI denoising and artifact mitigation continues to report measurable gains in subsequent segmentation when noise is explicitly addressed [10].

An MRI volume can be represented as a 3D grid of voxels. In brain MRI, voxels are commonly categorized into three tissue types: white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF)[11]. In this work, we focus on the binary segmentation of GM: the manual annotations of MRBrainS18 are merged (labels 1 and 2) to form a single GM mask; Figure 1 illustrates the dataset and this label mapping. WM is characterized by a higher signal intensity than GM and CSF[12]. However, tissue properties such as iron content, cell density and tissue anisotropy are factors that alter image contrast [8]. Several methods have been proposed in the literature to segment brain MRI. These can be grouped into two categories: statistical methods and methods based on deformable models[7]. Recent works emphasize that modern deep learning architectures (e.g., U-Net variants) dominate current practice, while hybrid CNN/Transformer models with self-attention broaden the receptive fields for long-range dependencies[13]. Atlas-based priors have long underpinned structural MRI analysis by offering standardized spatial references and tissue probability maps, which reduce inter-subject variability and provide strong anatomical constraints[14]–[16].

In general, the goal of image segmentation techniques is to divide the image into a set of non-overlapping regions such that each of the regions is homogeneous in some property or feature [17], [18]. The features that the regions share are varied and belong to a broad spectrum that is determined by the nature of the images. The result of the segmentation is a labeled image that identifies each homogeneous region or a set of contours that describe the region boundaries[11]. Brain image segmentation simplifies the representation of the image, which facilitates its analysis[19].

Studies incorporating machine learning techniques have shown outstanding performance within the spectrum of approaches addressing the problem of computer-aided diagnosis (CAD), disease detection and prognosis [20]–[22]. This makes it possible to mitigate operator dependence and increase diagnostic precision. Among their uses are the detection and classification of breast tumors, fetal development and growth, brain function, skin lesions and lung diseases [23]. Recent reviews focused specifically on brain tissue segmentation in MRI (GM/WM/CSF) consolidate advances across the lifespan and highlight pending challenges (noise, motion, edge blurring)[24]. Within machine learning techniques, the field of deep learning (DL) has recently shown its robustness and high level of precision. DL comprises an approach derived from bioinspired connectionist models called ANN (Artificial Neural Networks). These allow, in theory, approximating any mathematical function, which enables a broad application domain in multiple areas such as simulation, modeling and prediction. The complexity of the problem, represented in reproducing nonlinear, multidimensional and complex mathematical functions, has required the development of more robust techniques that allow the construction of mappings of mathematical relationships between input data and the expected output. In parallel, works synthesize

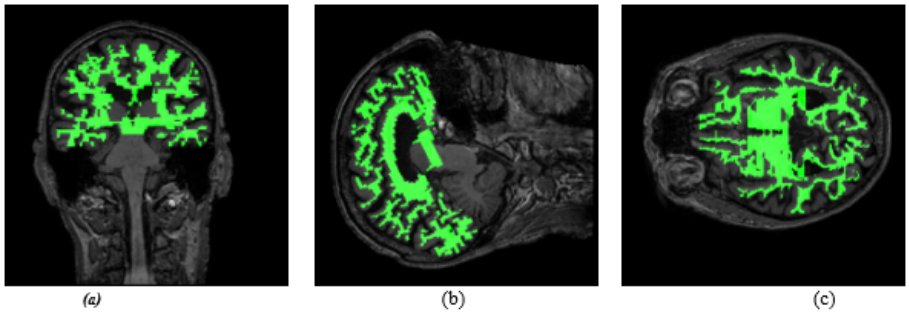

Fig. 1. *MRBrainS18 dataset and binary mapping of gray matter labels: orthogonal views—(a) coronal, (b) sagittal, (c) axial—showing the fused GM mask (labels 1+2) overlaid in green on T1-weighted MRI.*

Source: own elaboration.

This article explores basic algorithms derived from computer vision, such as multilevel Otsu thresholding, region growing techniques and deep learning algorithms. We analyze their performance and characteristics in the problem of segmenting gray matter in magnetic resonance images. Our experimental design follows good-practice guidelines for evaluation in order to ensure comparability with recent literature[2]. We describe classical algorithms (multilevel Otsu thresholding, region growing) and a modern learning-based approach. Quantitative results are reported for the learning-based model, while the classical methods are kept as methodological references.

METHODOLOGY

Image segmentation methods can be classified along three orthogonal axes. By degree of human intervention they can be manual, semiautomatic or automatic; by modality they can be unimodal when a single feature or image is used, or multimodal when multiple complementary features are combined; and by criterion they can be based on homogeneity (regions defined by similarity within the region) or on discontinuities (boundaries defined by feature discontinuities or edges)[25]–[27].

GM segmentation from brain images is a multi-label segmentation problem due to the presence of pixels belonging to different tissue types. We consider three types of segmentation approaches: threshold-based segmentation, region-based segmentation, and deep learning-based segmentation.

Multilevel Otsu thresholding

One advantage of thresholding is that it is relatively simple to implement and can be used efficiently in applications with few classes. The main disadvantage of this approach is that segmentation quality is sensitive to noise and to the presence of artifacts in the image. We selected the multilevel Otsu thresholding (MTO) segmentation method because it maximizes between-class variance as a way of optimizing thresholds[28]. MTO assumes the histogram of an image with L gray levels as a probability distribution. If an image is divided into N classes $(C_1, C_2, \ldots, C_n)$, MTO estimates $N-1$ thresholds $(t_1, t_2, \ldots, t_{n-1})$. The between-class variance is estimated by:

$$\sigma_B^2 = \sum_{k=1}^{N} w_k(\mu_k - \mu_T)^2 = \sum_{k=1}^{N} w_k\mu_k^2 - \mu_T^2 \quad (1)$$

where $w_k = \sum_{i\in C_k} p_i$, $\mu(k) = \sum_{i\in C_k} ip_i$, $\mu_k = \sum_{i\in C_k} \mu(k)/w_k$ are called zeroth- and first-order cumulative moments of class k_{th}, C_k y p_i is the number of pixels at gray level i.

The optimal thresholds $(t_1, t_2, \ldots, t_{N-1})$ are computed by maximizing the between-class variance, as follows:

$$\{t_1, t_2, \ldots, t_{N-1}\} = arg\max_{1\le t_1<\ldots<t_{N-1}<L}\{\sigma_B^2(t_1, t_2, \ldots, t_{N-1})\} \quad (2)$$

From the Otsu perspective, estimating multiple thresholds may require an iterative optimization procedure that is usually computationally expensive; several proposals aim to reduce computational costs, making it more efficient and less costly[29]–[31].

Region growing

The main idea of region growing segmentation methods is that the region grows by aggregating pixels using similarity and discontinuity measures[32],[33].

This method starts from an initial set of seed points, which are used to compute the clustering of similarities based on both distance and intensity among all the pixels contained within the cluster. From a given point, a similarity measure is computed to determine whether two pixels belong to the same object or class. The inclusion or exclusion of a pixel depends on the

statistics of the surrounding intensity values[34].

The initial point is selected arbitrarily and is called the seed point. This constitutes its main disadvantage, since there is no deterministic procedure for this selection and the segmentation result is highly sensitive to the selected seed point. In this work, we use the second threshold obtained from MTO for seed point selection. Once this value is estimated, the indices of the pixels that have a specific intensity level are searched. From this set, a subset of no more than 5 seed points is randomly selected to start region growing.

thrds = threshold_multiotsu(inputImage)

seeds = indexPoints(inputImage, thrds [1])

imgGM = regionGrowing(inputImage, seeds)

Deep Learning segmentation methods

Machine learning has gained considerable interest across broad application domains of image processing, including the field of medical image processing[35]–[37]. This interest stems from its outstanding performance, showing better results compared with other traditional methods existing in the literature[35],[38].

Deep learning is a field derived from artificial neural networks and is characterized by parameters and hyperparameters. This large number of elements appears to be the basis of the robustness of techniques based on this approach. However, it also constitutes their main drawback by increasing the computational costs for their training and deployment.

Convolutional neural networks (CNNs) have become the reference approach for image processing. The core part of CNN architectures is the convolution layers[38]. These are based on the traditional convolution operation, applying a set of filters to the image and extracting relevant image features used in the final part for classification purposes.

It is well known that, since convolutions are local and spatial, stacking layers produces a hierarchical extraction of features. This introduces an inductive spatial bias that assumes structured patterns in the input. The spatial bias adds some benefits to CNNs for object classification or structure segmentation tasks. This also makes it possible to mitigate the requirement for large training datasets.

Furthermore, reducing the interactions of neurons to a local spatial neighborhood may not be beneficial in other computer vision tasks, such as image understanding or image captioning. In this task, the features that define the scene must be estimated by considering spatially broader interactions. Recently, the use of attention-based models has incorporated these features, enabling the exploration of a broader spatial domain than that allowed by classical convolutions. Their derivation comes from the field of natural language processing, where the relationship of a word does not necessarily correspond to the other nearby words[39].

Dataset

The MRBrainS18 dataset was used. It provides training subjects with a manual reference standard and additional test cases. For each subject there are three MRI sequences: T1-weighted (3D, bias field corrected), T1 with inversion-recovery (multi-slice, bias field corrected) and T2-FLAIR (multi-slice, bias field corrected); registered variants are supplied. The manual labels comprise 11 classes: 0 background; 1 cortical gray matter (GM); 2 basal ganglia (deep GM); 3 white matter (WM); 4 WM lesions; 5 CSF (extracerebral); 6 ventricles; 7 cerebellum; 8 brainstem; 9 infarct; 10 other.

The dataset is published on DataVerseNL[40], which provides the training data, the test data and the manual reference segmentations, with additional information and links to code hosted by the challenge organizers.

Label mapping

MRBrainS18 provides a manual reference standard. For our binary gray matter (GM) segmentation experiments, we merged labels 1 (cortical GM) and 2 (basal ganglia) into a single GM class, following the challenge guidance for three-label protocols that merge GM and WM sublabels where appropriate. All other labels were mapped to background for the binary setting (labels 3–8, 9–10 ignored for evaluation if present). This label policy ensures consistency with the MRBrainS18 recommendations for evaluations with merged labels.

Preprocessing

Adoptamos a pipeline to (i) harmonize volume dimensions across subjects, (ii) enable memory-efficient patch-based training and inference without altering the native voxel geometry, and (iii) keep the data path reproducible and comparable across methods. The preprocessing steps are:

- Shape harmonization: Each 3D volume is mapped to a fixed 256×256×256 field of view by center cropping when a

dimension exceeds 256 voxels and zero-padding when it is smaller. The exact per-axis paddings are recorded to restore the original extent after inference. This strategy preserves the native voxel geometry and avoids any intensity resampling.

- Patch extraction: From the 256^3 cube we extract 128×128×128 cubic patches with a stride of 64 (~50% overlap). During training and inference, we skip empty/background-only patches (i.e., patches without brain tissue), which reduces computation and avoids degenerate batches, leaving the tiling pattern unchanged.
- Labels to image: We read the NIfTI header metadata of the label maps and apply the stored affine to bring the annotations into the native image space of the corresponding T1 volume. Labels were resampled to the image grid using nearest-neighbor interpolation to preserve discrete indices. All transformations were recorded so that training, inference and evaluation operate consistently in the original space.
- Post-inference restoration: Predictions obtained in the fixed cube are mapped back to the native field of view by removing the recorded paddings along each axis, exactly inverting Step 1.

This pipeline does not introduce intensity normalization, bias correction or geometric resampling; it intentionally preserves the acquisition characteristics of the dataset and relies on the challenge's aligned, manually annotated data to minimize preprocessing-induced variance.

We employed a 3D encoder–decoder segmentation network[41], adapted to gray matter (GM) delineation. The encoder combines local convolutional processing with global self-attention over non-overlapping volumetric patches, while the decoder performs multiscale upsampling with skip connections to recover fine anatomical detail. The design preserves spatial context at multiple resolutions and aggregates long-range dependencies that are critical for GM boundaries (see Figure 2).

- Encoder: its goal is global–local feature extraction. Volumes are processed as 3D patches of 128×128×128. Features are first projected to a token representation and then processed through a stack of multi-head self-attention blocks (4 heads) interleaved with lightweight feed-forward mappings (hidden size ≈ 192, MLP width ≈ 96). Residual and normalization layers stabilize training. Multiscale features from different encoder depths are exposed to the decoder via staged skip connections, which lets the decoder fuse global context with high-resolution details.
- Decoder: its goal is multiscale fusion and reconstruction. The decoder[42]–[44]follows a standard top-down pathway with transposed-convolution upsampling (stride 2) and concatenation with the corresponding encoder features at each scale. Each stage refines the fused representation with compact convolutional/residual blocks, progressively restoring spatial resolution and refining tissue interfaces. The final predictor is a $1 \times 1 \times 1$ convolution that produces two logits (GM vs. background).
- Training and optimization: we optimize a composite loss that sums Dice and Focal terms with equal weights. Optimization uses Adam with a base learning rate of 10e−3 and gradient clipping to improve stability. The loss is the sum of Dice[45] and Focal Loss[46]computed over two channels (GM vs. background) with equal weights. This emphasizes overlap fidelity while addressing class imbalance[45],[47].
- Inference: At inference time, each volume is partitioned into patches of $128 \times 128 \times 128$ with 50% overlap (stride = 64). The per-patch GM probability predictions are fused in the overlap regions by simple averaging (no windowing). The fused probability map is then thresholded at 0.5 to produce the binary GM mask. Finally, any padding introduced during shape harmonization is removed to recover the native extent. The 50% overlap is intentionally conservative, yet simple and robust in practice.

RESULTS

The experiments were performed on Ubuntu 22.04.5 (kernel 6.8) with 250 GiB of RAM and two NVIDIA RTX A6000 GPUs (48 GB of VRAM each).

The deep learning architecture was implemented using a modern development stack, based on Python 3.10 and PyTorch 2.9.0.

Methods

Three algorithms were evaluated: (i) the traditional multilevel Otsu thresholding method, (ii) region growing, and (iii) a proposed deep learning architecture.

The main limitation of multilevel thresholding is that each image must be processed independently to determine the optimal segmentation values. Because the approach relies exclusively on intensities, no single set of thresholds can be defined that generalizes adequately to different

Methods based on region growing techniques, on the other hand, depend on the selection of seeds. This value is usually arbitrary and is set by the user. In this approach, we use the MTO thresholds as a reference. These approaches require fewer parameters. However, it is necessary not only to specify the seed values, but also the minimum and maximum intensity limits allowed to apply the algorithm. These values remain arbitrary and the final result is highly sensitive to them.

For the last model, we trained the proposed deep neural network using the hyperparameters described in Table 1, with the Adam optimizer[48], an input image size of $128 \times 128 \times 128$, and 32 images per batch.

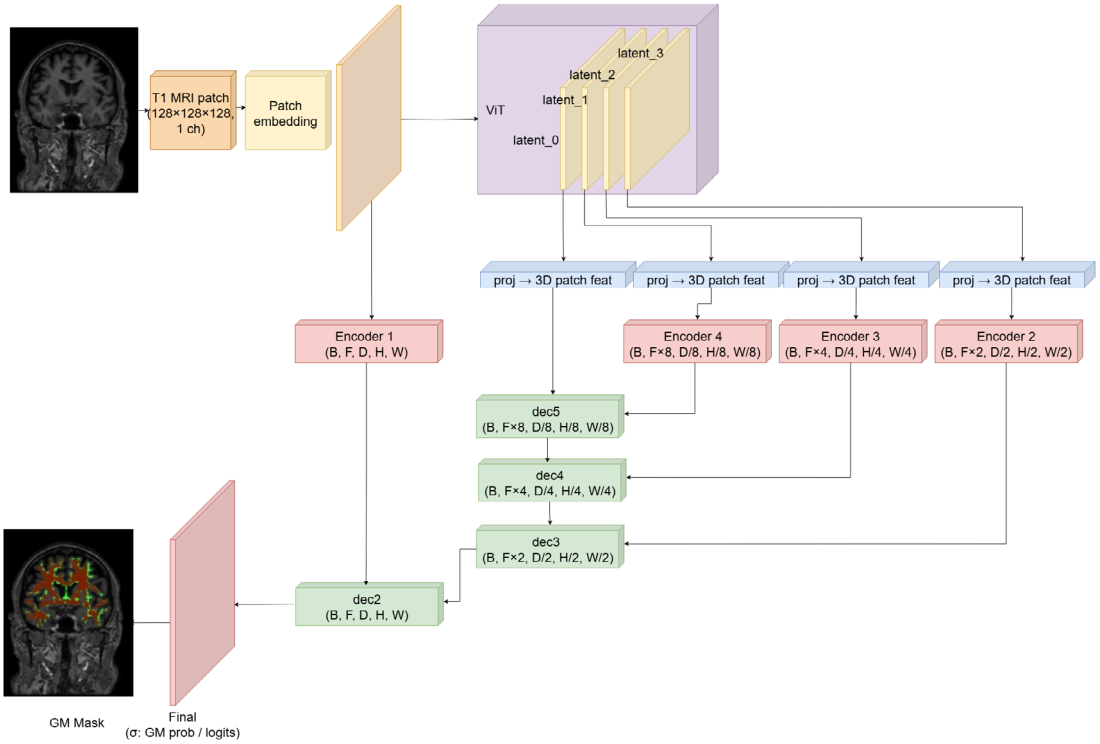

Fig. 2. *3D encoder–decoder for GM segmentation.*
Source: own elaboration.

Table 1: hyperparameters used to train the 3D encoder–decoder.

Hyperparameters	Value
Input size	128×128×128
Batch size	4
Loss function	Dice + Binary Focal
Focal reduction	mean
Optimizer	Adam, lr=1e-3 [48]
Weight initialization	He initialization [49]
Adam learning rate	0.001
Adam β_1, β_2	0.9, 0.999
Adam ε	1e-08
Number of epochs	300
Early stopping	None
Mixed precision	off
Inference threshold	0.5

Source: own elaboration.

Deep Learning

The learning-based model shows a monotonic decrease in both training and validation losses, with a small, stable generalization gap toward the end of training (Figure 3a). The best validation loss occurs at epoch 297 (val = 0.1662) and closely matches the training loss (0.1673), suggesting convergence without overfitting in the selected checkpoint. A complementary view of the generalization gap (validation − training) across epochs (Figure 3b) confirms that the gap tends to zero. The last 10 epochs (Figure 3c) show low variance in validation loss (0.1739 ± 0.0034), reinforcing the stability of the final training regime.

Segmentation performance

The learning-based model was evaluated on the test set. The model achieved a Dice of 0.650 ± 0.043 (median 0.660; min–max 0.548 – 0.718) and an IoU of 0.483 ± 0.046 (median 0.492; min–max 0.377 – 0.561), with Precision 0.700 ± 0.046 and Recall 0.610 ± 0.059 (Table 2). The per-subject distributions (Fig. 4a–d) show moderate dispersion with compact interquartile ranges, indicating consistent performance across cases. A complementary analysis of boundary error (Fig. 5a) showed no evidence of a monotonic relationship between overlap and Hausdorff distance (Pearson r=0.076, p=0.69; Spearman ρ=0.081, p=0.671). The Dice histogram (Fig. 5b) confirms a centered distribution with a small fraction of lower-performing outliers.

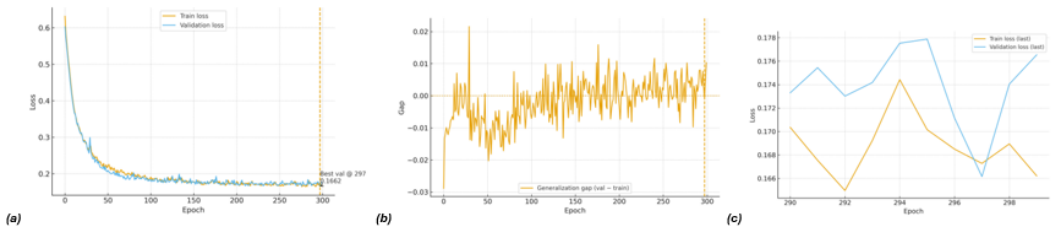

Fig. 3. *a) Training and validation loss across epochs, b) generalization gap across epochs, and c) learning curves over the last 10 epochs.*
Source: own elaboration.

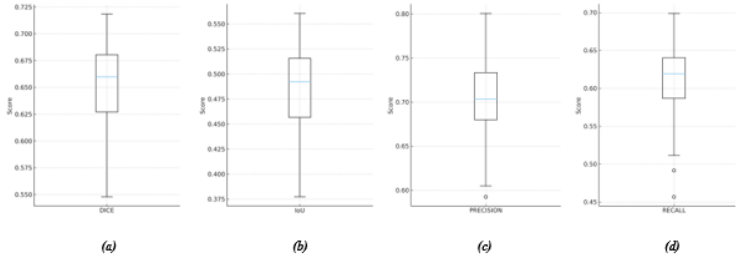

Fig. 4. *Per-subject segmentation scores on the test set. (a) Dice; (b) Intersection-over-Union (IoU); (c) Precision; (d) Recall. Boxes show the median and interquartile range; whiskers extend to 1.5×IQR; dots denote individual subjects.*
Source: own elaboration.

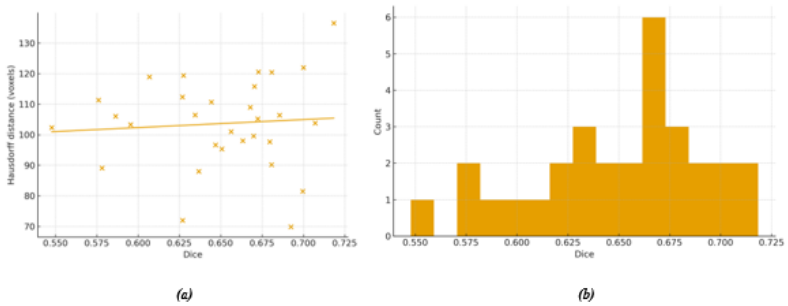

Fig. 5. *Overlap vs. boundary error and Dice distribution. (a) Dice vs. Hausdorff distance (voxels) for the test subjects; the solid line shows the least-squares fit (lower Hausdorff is better). (b) Histogram of per-subject Dice scores (higher is better).*
Source: own elaboration.

In terms of volumetry, the model shows a negative volume bias (mean RVE = −12.54% ± 9.53%, median −11.45%), that is, a tendency to undersegment gray matter relative to the manual reference (Table 2).

Table 2: segmentation performance on the test set: Dice, IoU, Precision, Recall, Hausdorff distance (voxels) and relative volume error (%).

Metric	Mean ± SD	Median [IQR]	Min–Max
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
Precision	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 (voxels)	103.656 ± 14.785	104.494 [96.936, 112.084]	69.871–136.532
Relative volume error (%)	-12.541 ± 9.535	-11.451 [-18.821, -5.975]	-33.266–5.119
Absolute RVE (%)	12.882 ± 9.052	11.451 [5.975, 18.821]	0.277–33.266

Source: own elaboration.

A Bland–Altman analysis (Fig. 6) confirms a systematic negative shift, with limits of agreement that reflect morphological variability between subjects. This suggests that the model is conservative at tissue interfaces, consistent with the pattern Precision > Recall. The best subject (by Dice) reaches 0.718, while the worst achieves 0.548, illustrating the observed dispersion.

To contextualize the performance of the DL model, we computed reference results for two classical approaches: multilevel Otsu thresholding and region growing. Both were applied to a representative subset of MRBrainS18. As summarized in Table 3, these classical methods yield limited overlap due to their sensitivity to intensity inhomogeneity and noise. The proposed deep learning model improves Dice and IoU by more than 50% with respect to the best baseline, which shows the effectiveness of the learned global context and of the reconstruction with skips for gray matter delineation.

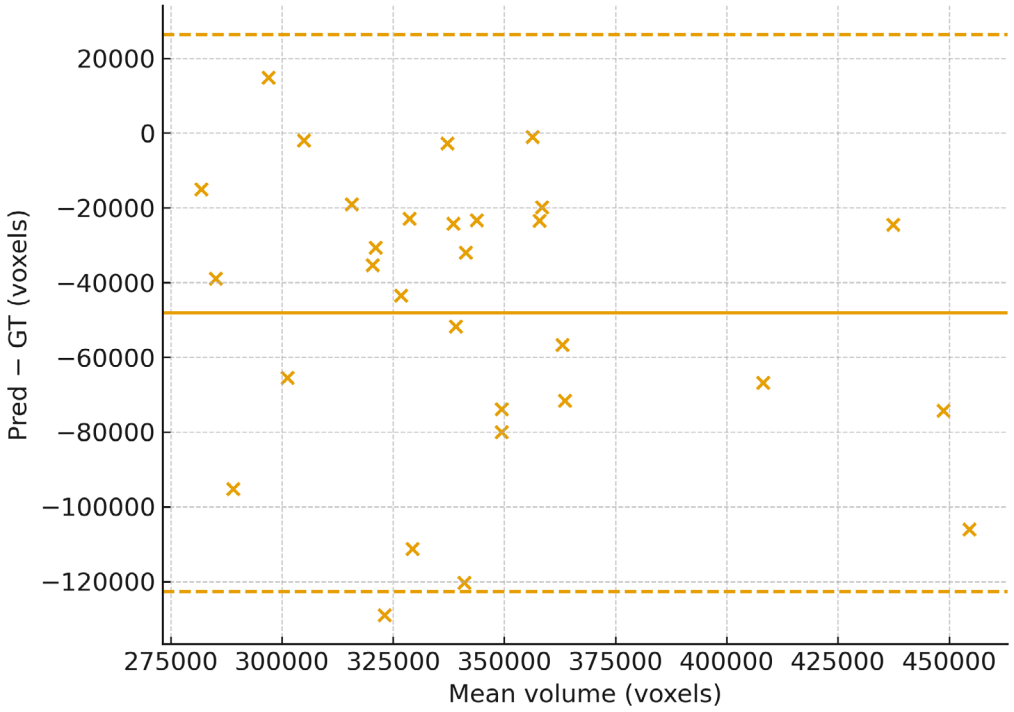

Fig. 6. *Bland–Altman analysis of gray matter volume (Pred − GT) with mean bias and 95% limits of agreement.*
Source: own elaboration.

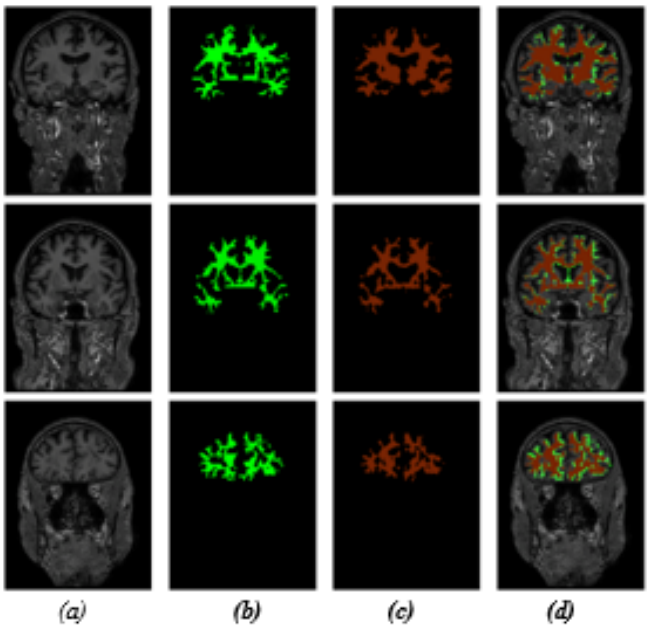

Fig 7. Qualitative segmentation of gray matter in three held-out test subjects (coronal slices): (a) T1-weighted MRI; (b) manual reference (green); (c) model prediction (orange); (d) overlay on the MRI.
Source: own elaboration.

Table 3: comparative segmentation performance of the baseline methods and the proposed method.

Method	Dice	IoU	Precision	Recall
Multilevel Otsu thresholding	0.41 ± 0.07	0.28 ± 0.06	0.49 ± 0.08	0.36 ± 0.05
Region growing	0.46 ± 0.06	0.33 ± 0.05	0.55 ± 0.07	0.39 ± 0.06
Deep Learning (ours)	0.65 ± 0.04	0.48 ± 0.05	0.70 ± 0.05	0.61 ± 0.06

Source: own elaboration.

CONCLUSIONS

We propose a pipeline for gray matter segmentation that combines a minimal preprocessing strategy with a 3D encoder–decoder network that integrates global self-attention and multiscale skip connections.

On the held-out test set, the method achieved Dice 0.650 ± 0.043 (median 0.660) and IoU 0.483 ± 0.046 (median 0.492), with precision 0.700 ± 0.046 and recall 0.610 ± 0.059. These results indicate reliable overlap with the manual reference across subjects. The analysis of Dice vs. Hausdorff distance did not show a monotonic association, reflecting the sensitivity of boundary metrics to localized errors even when the global overlap is adequate. The Dice histogram shows a centered distribution with a small fraction of lower-performance cases, consistent with inter-subject variability.

From a volumetric perspective, the model shows a negative bias (mean RVE = −12.54% ± 9.53%), that is, a tendency to undersegment GM relative to the manual reference. The Bland–Altman plot confirms this systematic shift and quantifies the limits of agreement. This suggests conservative predictions near the tissue interfaces.

Compared with classic reference methods (Otsu and region growing), the proposed deep learning model improves Dice and IoU by approximately 50–60%, which shows the advantage of the learned global context for GM delineation.

The strengths of the approach include (i) a lightweight and reproducible preprocessing pipeline that preserves the native voxel geometry, (ii) stable training dynamics with limited overfitting, and (iii) cross-subject consistency in the overlap metrics. The limitations include a binary label setting, a measurable undersegmentation bias, and the use of a single public dataset.

Future work will focus on boundary adherence and recall, volume calibration, and broader validation across scanners and cohorts. Extensions to multiclass segmentation (GM/WM/CSF) and uncertainty modeling could further improve robustness, while weak anatomical priors or domain adaptation may reduce the residual bias without sacrificing precision.

REFERENCES

[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.

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Camilo Andres Laiton Bonadiez, German Sanchez Torres, Carlos Nelson Henríquez Miranda.

Writing–review and editing: Camilo Andres Laiton Bonadiez, German Sanchez Torres, Carlos Nelson Henríquez Miranda.