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

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

Digital Object Identifier: 10.24054/rcta.v1i47.4282

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

MSc. Camilo Andres Laiton Bonadiez[1], PhD. German Sanchez Torres[2], PhD. 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.*

Correspondence: *chenriquezm@unimagdalena.edu.co*

Received: *July 24, 2025.* ***Accepted:*** *December 20, 2025.* ***Published:*** *January 01, 2026.*

How to cite: *C. A. Laiton Bonadiez, G. Sánchez Torres, and C. N. Henríquez Miranda, "Self-attention encoder–decoder for gray-matter segmentation in brain MRI", RCTA, vol. 1, n.°. 47, pp. 1-12, Jan. 2026. Recovered from https://ojs.unipamplona.edu.co/index.php/rcta/article/view/4282*

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.

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.

1. INTRODUCTION

Brain imaging is the main tool for diagnosing brain-related diseases, covering a wide spectrum of disorders and providing anatomical and functional information on the brain [1]. Recent surveys reaffirm the centrality of neuroimaging for computer-aided diagnosis and segmentation workflows, consolidating methodology and evaluation practices [2]. Understanding the structure of the brain and the underlying neural mechanism is essential for the monitoring and early diagnosis of diseases to prevent them from progressing to a serious 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) being the most popular techniques for brain imaging [3]. Magnetic resonance imaging (MRI) is one of the most popular types of imaging, it uses a strong magnetic field and radiofrequency waves to visualize organs, soft tissues, and bones [4].

The primary goal of structural brain MRI analysis includes the classification of specific tissue types, as well as the identification and description of specific anatomical structures. Literature show that deep learning–based segmentation has become the prevailing paradigm for tissue and structure delineation in clinical and research settings [5]. However, it is challenging to extract high-quality information from brain images due to the low signal-to-noise ratio and the artifacts that are generated in the acquisition process. In [6] details persistent artifact sources—e.g., radiofrequency interference and related acquisition artifacts—and their impact on downstream segmentation.

Classification of 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. This is why semi-automated tissue classification techniques have been developed for many low-cost MRI data sets. However, a weakness of these approaches is that they lose accuracy due to images with a high presence of noise or artifacts. In fact, medical images include different types of noise that show distortion and many problems during disease diagnosis [9]. Related community on MRI denoising and artifact mitigation continue to report measurable gains for downstream 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 binary GM segmentation: the manual MRBrainS18 annotations 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 higher signal intensity than GM and CSF [12]. Although, tissue properties such as iron content, cell density, tissue anisotropy are factors that alter image contrast [8]. Several methods have been proposed in the literature to segment brain MRIs. 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 extend 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 objective of image segmentation techniques is to divide the image into a set of non-intersecting regions such that each of the regions is homogeneous in some property or characteristic [17], [18]. The characteristics that the regions share are varied and belong to a wide spectrum that is determined by the nature of the images. The result of the segmentation is an image with labels that identify each homogeneous region or a set of contours that describe the limits of the region [11]. Brain image segmentation simplifies the representation of the image making it easy to analyze [19].

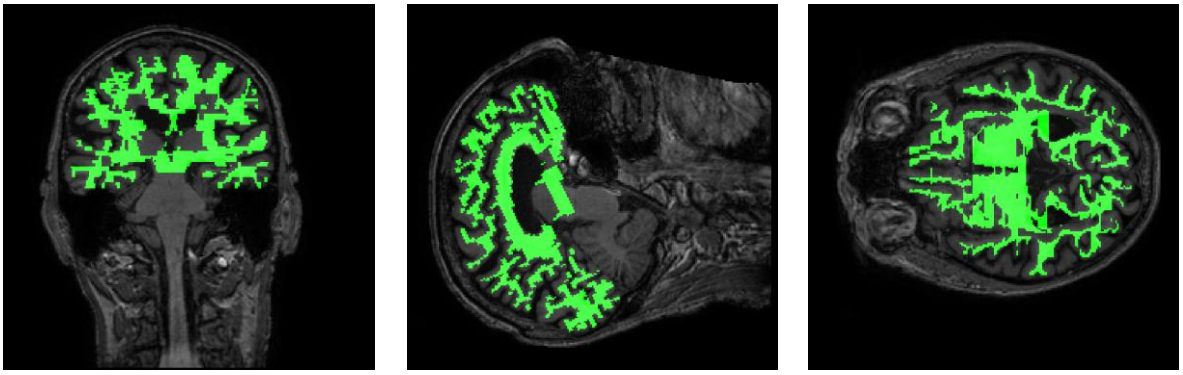
Studies incorporating machine learning techniques have shown an outstanding performance within the spectrum of approaches to address the problem of computer-aided diagnosis - CAD, disease detection, and prognosis [20]–[22]. This allows to mitigate the dependency of the operator and increase the accuracy of the diagnoses. Among its uses are the detection and classification of breast tumors, fetal development and growth, brain function, skin lesions, and lung diseases [23]. Recent surveys focused specifically on MRI brain tissue segmentation (GM/WM/CSF) consolidate advances across the lifespan and highlight remaining 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 accuracy. DL comprises an approach derived from bio-inspired connectionist models called ANN Artificial Neural Networks. These allow approximating, in theory, any mathematical function allowing a wide domain of application 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 maps of mathematical relationships between input data and expected output. In parallel, works synthesize progress in CNN backbones and Transformer-based segmentation, motivating the use of self-attention mechanisms for brain MRI.

In this article, we explore basic algorithms derived from computer vision such as multi-threshold segmentation, region growing techniques, and deep learning algorithms. We analyze its performance and characteristics in the problem of segmenting gray matter in magnetic resonance images. Our experimental design follows best-practice guidance for evaluation to ensure comparability with rencet literature [2]. We describe classical algorithms (multi-threshold Otsu, region growing) and a modern learning-based approach. Quantitative results are reported for the learning-based model, while the classical methods are retained as methodological references.

2. METHODOLOGY

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

GM segmentation from brain images is a multilabel segmentation problem due to the presence of pixels



(a) ***(b)*** ***(c)***

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

belonging to different tissue types. We considered three types of segmentation approaches: threshold-based segmentation, region-based segmentation a deep learning-based segmentation.

2.1. Otsu multi-thresholding

One advantage of thresholding is that it is relatively simple to implement and can be used efficiently in applications with a few classes. The principal drawback of this approach is that the segmentation quality is sensitive to noise and presence of artifacts in the image. We selected the Multi-threshold Otsu -MTO segmentation method due to it maximize between-class variance as a way of threshold optimization [28]. MTO assumes the histogram of L gray levels image as a probability distribution. If an image is divided into N classes $(C_1, C_2, \dots, C_N)$, MTO estimate N-1 thresholds $(t_1, t_2, \dots, t_{N-1})$. 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 $w_k = \sum_{i \in C_k} p_i$, $\mu(k) = \sum_{i \in C_k} i p_i$, $\mu_k = \sum_{i \in C_k} \mu(k)/w_k$ are called the zeroth and first-order cumulative moments of the k_{th} class C_k, and p_i is the number of pixels in the gray level i.
The optimal thresholds $(t_1, t_2, \dots, t_{N-1})$ are calculated maximizing the between-class variance as:

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

From Otsu's perspective, multi thresholds estimation can require an iterative optimization procedure which is usually computationally expensive, there exits multiple proposals addressed to reduce the computational costs making it more efficient and less expensive [29]–[31].

2.2. Region growing

The main idea of the region growing segmentation methods is that the region is grown by pixel aggregation using similarity and discontinuity measures [32], [33].

This method starts from an initial set of seed points, which are used for clustering computing of both distance and intensity-based similarities between all the pixels contained within the cluster. Starting from a set 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 arbitrarily selected and is named seed point. This constitutes its main disadvantage since there is not a determinist procedure for this selection and the segmentation result is highly sensitive to the selected seed point. In this work, we used the second threshold obtained from MTO for the selection of the seed point. Once this value is estimated, the pixels indices that have a specific level of intensity are searched. From this set, a subset of no more than 5 seed points is randomly selected to start growing regions.

```
thrds = threshold_multiotsu(inputImage)
seeds = indexPoints(inputImage, thrds [1])
imgGM = regionGrowing(inputImage, seeds)
```

2.3. Deep Learning segmentation methods

Machine learning has gained a lot of interest in broad image processing application domains including the medical image processing field [35]–[37]. This interest is caused by its outstanding performance showing better results compared to other traditional methods existing in the literature [35], [38].

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

Convolutional neural networks (CNN) had become the referring approach for image processing. The central part of CNN architectures is the convolutional 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 because convolutions are local and spatial, stacking layers yields hierarchical feature extraction. This introduces an inductive spatial bias that assumes structured patterns in the input. The spatial bias adds some benefits to the CNN for object classification or structure segmentation tasks. This also allows mitigating the requirement of large training sets.

Moreover, reducing the interactions of the neurons to a local spatial neighborhood may not be beneficial in other computer vision tasks such as image comprehension or description. In this task, the features that define the scene should be estimated considering spatially broader interactions. Recently, the use of attention-based models has incorporated these characteristics, allowing the exploration of a broader spatial domain than those allowed by classical convolutions. Its 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].

2.4. Dataset

We use the MRBrainS18 dataset. 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-weighted inversion recovery (multi-slice, bias-field corrected), and T2-FLAIR (multi-slice, bias-field corrected); registered variants are provided. 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 infarction; 10 other.

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

2.5. 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's guidance for three-label protocols that merge GM and WM sublabels when 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 MRBrainS18 recommendations for merged-label evaluations.

2.6. Preprocessing

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

No intensity normalization, bias correction, or geometric resampling is introduced by this pipeline; it intentionally preserves the dataset's acquisition characteristics and relies on the challenge's aligned, manually annotated data to minimize preprocessing-induced variance

We employ a 3D encoder–decoder segmentation network [41], tailored to gray-matter (GM) delineation. The encoder combines local convolutional processing with global self-attention over non-overlapping volumetric patches, while the decoder performs multi-scale 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:* It aims to global-local feature extraction. Volumes are processed as 3D patches of 128×128×128. Features are first projected into a

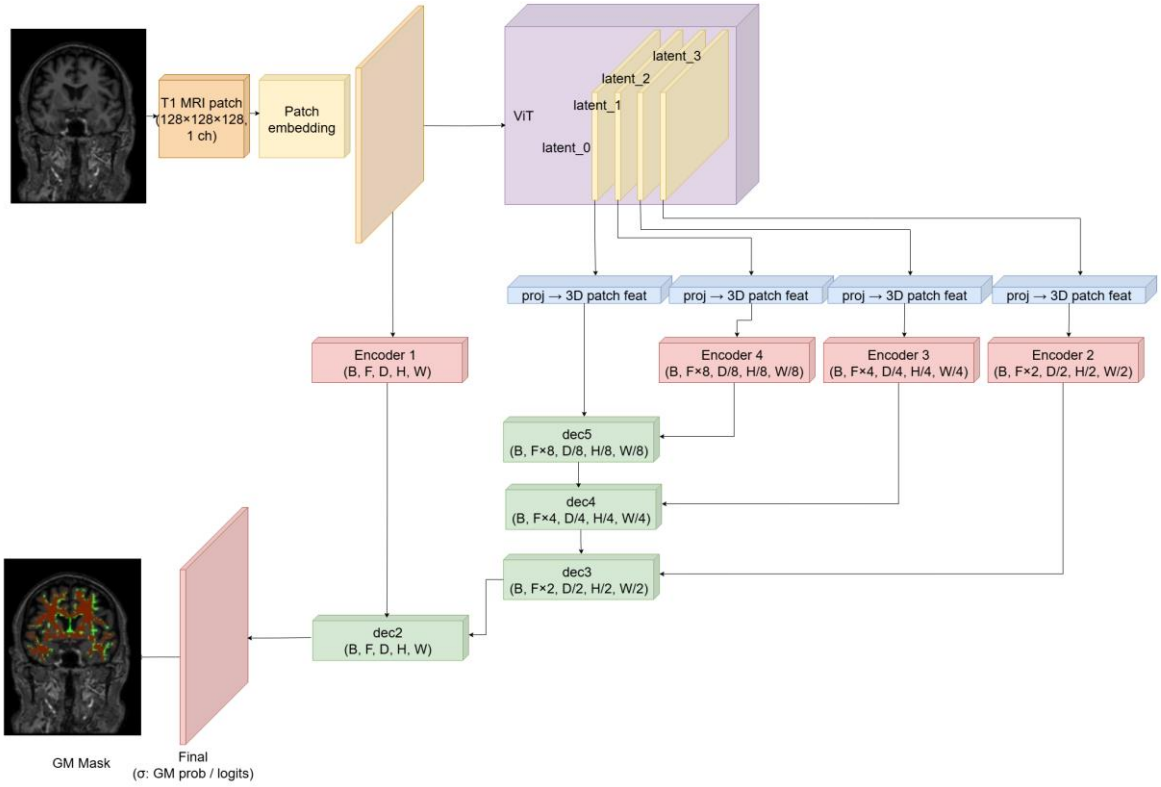


Fig. 2. *3D encoder–decoder for GM segmentation.*
Source: *own elaboration.*

token representation, then processed by 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. Multi-scale features from distinct depths of the encoder are exposed to the decoder through stage-wise skip connections, enabling the decoder to fuse global context with high-resolution details.

- *Decoder:* It aims to multi-scale 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 sharpening tissue interfaces. The final predictor is a 1×1×1 convolution producing two logits (GM vs. background).

- *Training and optimization*: We optimize a compound loss that adds Dice and Focal terms with equal weights. Optimization uses Adam with a base learning rate of 10^{-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 128×128×128 patches with 50% overlap (stride = 64). Patch-wise 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 but simple and robust in practice.

3. RESULTS

Experiments were conducted on Ubuntu 22.04.5 (kernel 6.8) with 250 GiB RAM and two NVIDIA RTX A6000 GPUs (48 GB VRAM each), using a modern deep-learning stack (Python 3.10, PyTorch 2.9.0).

3.1. Methods

We evaluated three algorithms: first, the traditional multi thresholds Otsu; second, the region growing; and finally, our deep learning architecture.

The main disadvantage of the multiple thresholding method is that each image must be processed to find the optimal value. As it is a basic method, it is not possible to have a unique set of intensity values that allow the application of the method to various data sets. The effects of the noise level, the presence of artifacts and the acquisition parameters are weaknesses of this approach.

On the other hand, the methods based on region growing techniques depend on the selection of the seeds. This value is usually arbitrary and 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 limits allowed in the intensity 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 model using the hyperparameters described in Table 1, using Adam optimizer [48], with an input image size of 128 ×128 ×128 and 32 images per batch.

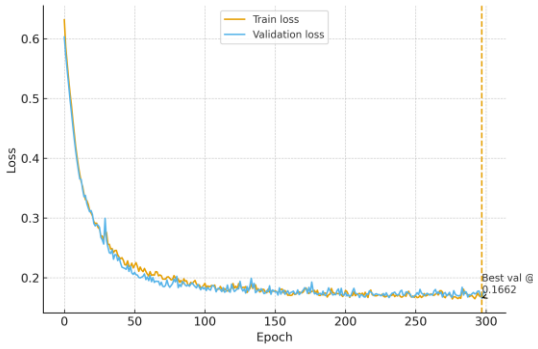
Table 1. *Hyperparameters Used for Training the 3D Encoder–Decoder.*

Hyperparameter	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

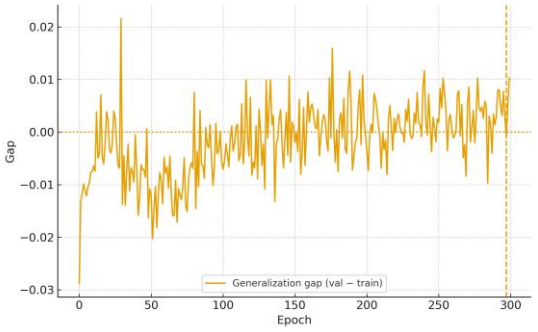
Source: *own elaboration.*

3.2. Deep Learning

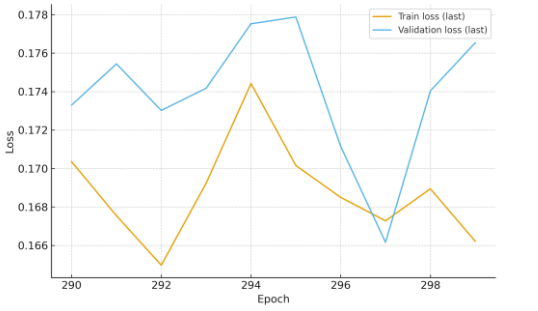
The learning-based model shows a monotonic decrease in both training and validation losses with a small, stable generalization gap toward the end of training (Figure 3a). The best validation loss occurs at epoch 297 (val = 0.1662) and closely matches the training loss (0.1673), suggesting convergence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



(a)



(b)



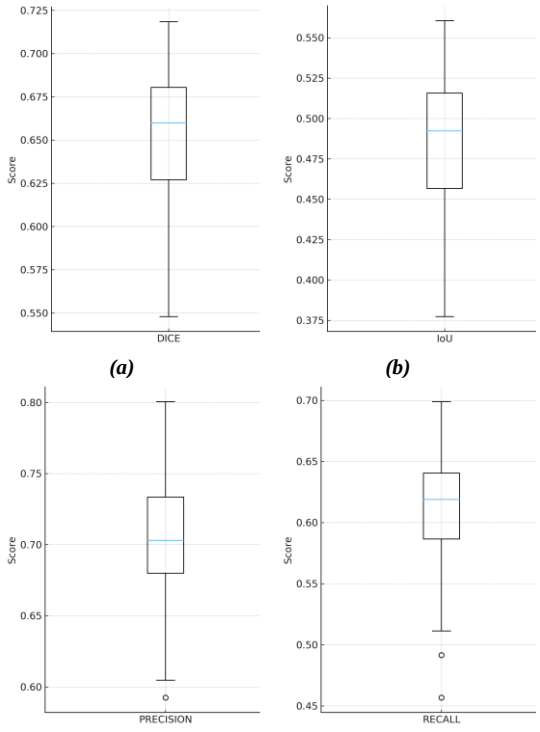
(c)

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

3.3. Segmentation performance

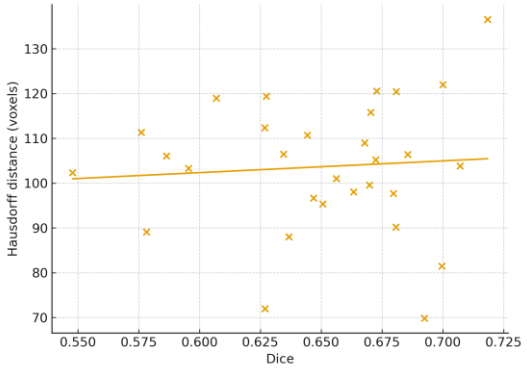
We evaluated the learning-based model on the test set. The model attains Dice 0.650 ± 0.043 (median 0.660; min–max 0.548 – 0.718) and IoU 0.483 ± 0.046 (median 0.492; min–max 0.377 – 0.561), with Precision 0.700 ± 0.046 and Recall 0.610 ± 0.059 (Table 2). The subject-wise distributions (Fig. 4a–d)

show moderate spread with compact interquartile ranges, indicating consistent performance across cases. A complementary analysis against 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.

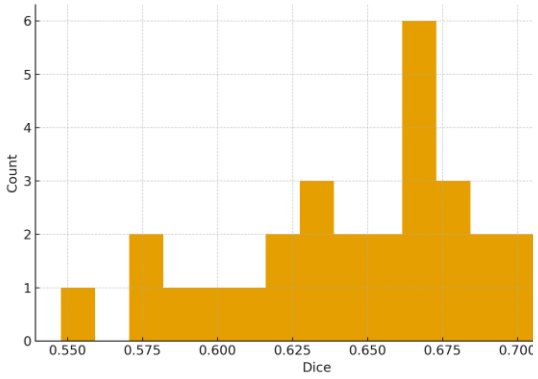


(a) *(b)* *(c)* *(d)*

Fig. 4*. Subject-wise segmentation scores on the test set. (a) Dice; (b) Intersection-over-Union (IoU); (c) Precision; (d) Recall. Boxes show median and interquartile range; whiskers extend to 1.5×IQR; points denote individual subjects. Higher is better for all metrics.*
Source: *own elaboration.*



(a)



(b)

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

In terms of volumetry, the model presents a negative volume bias (mean RVE = −12.54% ± 9.53%, median −11.45%), i.e., a tendency to under-segment gray matter relative to the manual reference (Table 2).

Table 2: *Test-set segmentation performance: 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 subject-to-subject morphological variability, suggesting the model is conservative at tissue interfaces—consistent with the Precision > Recall pattern. The best subject (by Dice) attains 0.718 while the worst achieves 0.548, illustrating the observed spread.

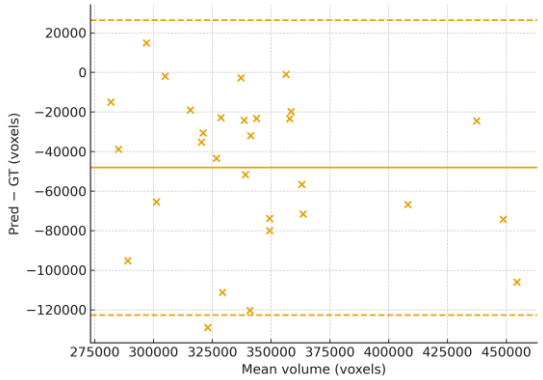
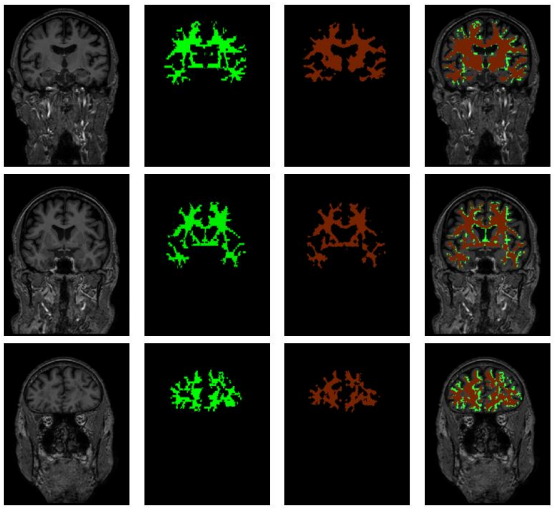


Fig. 6. *Bland–Altman analysis of gray-matter volume (Pred − GT) with mean bias and 95% limits of agreement*
Source: *own elaboration.*

The results shows that the encoder–decoder with global context yields stable overlap metrics across subjects while tending to under-estimate GM volume (some examples are shown in Figure 7). Future refinements could target recall and boundary adherence to reduce the observed volume bias without sacrificing precision.



(a) ***(b)*** ***(c)*** ***(d)***

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

To contextualize the performance of the DL model, we computed baseline results for two classical approaches: multi-threshold Otsu 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% relative to the best baseline, evidencing the effectiveness of learned global context and skip-based reconstruction for gray-matter delineation.

Table 3: *Comparative segmentation performance of baseline and proposed methods.*

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

Source: *own elaboration.*

4. CONCLUSIONS

We presented a streamlined pipeline for gray-matter segmentation that combines a minimal preprocessing strategy with a 3D encoder–decoder network integrating global self-attention and multi-scale 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 showed no monotonic association, reflecting the sensitivity of boundary metrics to localized errors even when global overlap is adequate. The Dice histogram reveals a centered distribution with a small fraction of lower-performing cases, consistent with subject-to-subject variability.

From a volumetric perspective, the model exhibits a negative bias (mean RVE = −12.54% ± 9.53%), i.e., a tendency to under-segment 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 tissue interfaces.

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

Strengths of the approach include (i) a lightweight, reproducible preprocessing path that preserves native voxel geometry, (ii) stable training dynamics with limited overfitting, and (iii) cross-subject consistency in overlap metrics. Limitations include a binary label setting, a measurable under-

segmentation bias, and the use of a single public dataset.

Future work will target boundary adherence and recall, volume calibration, and broader validation across scanners and cohorts. Extensions to multi-class segmentation (GM/WM/CSF) and uncertainty modeling could further enhance robustness, while weak anatomical priors or domain adaptation may reduce 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.