Deep Learning-Based Brain Lesion Segmentation Across MRI Registration Templates

Iván Lago Guerreira[2], Ariadna Jiménez-Partinen[1,2,3]*, Beatriz Asenjo[4], Esteban J. Palomo[1,2,3], and Rafael M. Luque-Baena[1,2,3]

[1] ITIS Software, University of Málaga, Málaga, Spain
[2] Department of Computer Languages and Computer Science, University of Málaga, Málaga, Spain
[3] IBIMA Plataforma BIONAND, Málaga, Spain
[4] Department of Radiology, Hospital Regional Universitario de Málaga, Málaga, Spain
{ivanlagog,ariadna,ejpalomo,rmluque}@uma.es,
beatriz.asenjo.sspa@juntadeandalucia.es

Abstract. Automatic lesion segmentation in medical imaging is a key research field supporting the diagnosis and monitoring of neurological disorders. This paper addresses lesion segmentation in magnetic resonance imaging (MRI), considering two neurological conditions: brain metastasis and multiple sclerosis. The aim is to evaluate the performance of deep learning architectures across pathologies and anatomical registration templates. Two publicly available datasets were used, and a preprocessing pipeline was applied to assess the influence of three anatomical templates: MNI152, SRI24, and NMRI225. Using these datasets, two segmentation frameworks were trained and compared: nnU-Netv2 and U-Mamba, representing convolutional-based and Mamba-based segmentation paradigms, respectively. The work provides a detailed description of the preprocessing, training, and evaluation pipeline, along with a comparative analysis using the Dice Similarity Coefficient (DSC). The reported results show that the influence of architecture, registration template, and training dimensionality depends on the pathology considered. Independent of the architecture, 3D configurations consistently outperformed 2D approaches by exploiting volumetric spatial information. Finally, both architectures achieved comparable performance in multiple sclerosis, whereas the Mamba-based model provided clear advantages for brain metastasis segmentation.

Keywords: Magnetic resonance imaging, brain lesion segmentation, multiple sclerosis, brain metastasis, registration templates, nnU-Net, U-Mamba

1 Introduction

Automatic segmentation of lesion-related neurological disorders in magnetic resonance imaging (MRI) remains one of the most challenging tasks in neuroimaging due to the high anatomical complexity of the brain and the strong inter-patient

* Corresponding author: ariadna@uma.es

variability in neurological pathologies [15]. In clinical practice, accurate lesion delineation is essential for diagnosis, disease monitoring, treatment planning, and surgical decision-making. However, manual segmentation is time-consuming, operator-dependent, and highly demanding, even for expert neuroradiologists, especially when lesions are subtle, diffuse, or poorly contrasted. These limitations have motivated the development of computer-aided diagnosis (CAD) systems that support clinicians through robust, reproducible automated analysis [21].

In recent years, deep learning (DL) has significantly transformed medical image segmentation. Convolutional neural networks (CNNs), particularly U-Net-based architectures, have demonstrated remarkable performance by learning hierarchical spatial representations directly from imaging data. Among these approaches, nnU-Net [7] has become a benchmark due to its self-configuring training pipeline, which automatically adapts preprocessing, architecture design, and hyperparameter selection to the properties of each dataset. This high level of automation has enabled state-of-the-art performance across a wide range of biomedical segmentation benchmarks [6]. More recently, alternative architectures have emerged to address the limitations of conventional CNNs in modeling long-range spatial dependencies. In this context, U-Mamba [11] incorporates State Space Model (SSM) modules into the segmentation pipeline, improving the ability to capture global contextual information while maintaining computational efficiency. Such characteristics make this architecture particularly attractive for high-resolution brain MRI volumes, where lesion appearance may depend on both local fine details and broader anatomical context.

In the brain tumor segmentation task, recent publications have explored U-Net variants, including the Modified Recurrent Residual Attention U-Net [22] for diffuse gliomas, U-Net for low-grade gliomas (LGG) [1], and 3D-Fusion-Net for glioblastoma (GBM) [5]. For MS segmentation, 2D strategies were explored using the U-Net [2] and the SAM-based method [17]. Additionally, some 3D approaches using the nnU-Net model were presented, including complementary modalities (FLAIR, Proton Density (PD), and Double Inversion Recovery (DIR) sequences) [14], or common sequences (T1-w, T2-w, and FLAIR), with random convolution as a data augmentation technique [20].

Although each of these approaches offers unique contributions, it also exhibits certain limitations, such as only exploring 2D or 3D strategies without comparing them, a lack of work using brain metastasis data, reliance on private cohorts, and failure to evaluate the impact of the registration template on model performance. The present work aims to fill this gap in the literature by exploring, on a broader scale, how the preprocessing step affects DL-based solutions across different architectures and neurological diseases. This work explores the automatic segmentation of two neurological disorders: multiple sclerosis (MS) and brain metastasis (BM) using two publicly available datasets [4, 13] for comparing nnU-Netv2 and U-Mamba under different anatomical template registration settings –MNI152 [12], SRI24 [16], and NMRI225 [8]–, evaluating both validation and test performance in 2D and 3D configurations. The main contributions of this work are:

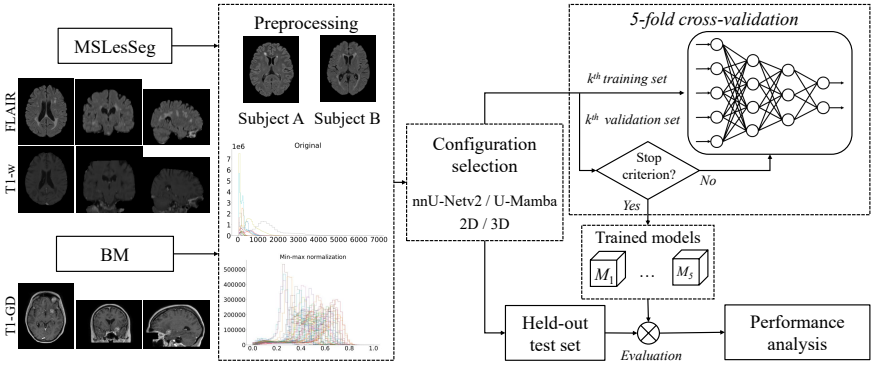


Fig. 1: Overall methodology.

- A comparative evaluation of two state-of-the-art segmentation frameworks: nnU-Net and U-Mamba.
- An experimental analysis across two neurological pathologies: MS and BM.
- An assessment of the impact of anatomical template selection on lesion segmentation performance.
- A reproducible preprocessing and experimental pipeline for heterogeneous brain MRI datasets.

The remainder of this work is structured as follows: first, Section 2 details the proposal; Section 3 describes the experiments and discusses the results attained; finally, Section 4 states the main findings and conclusions.

2 Methodology

In this section, the methodology is described, including the source data description and the approaches used to study input data preprocessing and its impact across two main DL techniques: CNN-based and Mamba-based architectures, with 2D and 3D variants. The overall methodology implemented is depicted in Fig. 1.

2.1 Datasets

Two publicly available brain MRI datasets were used to evaluate the proposed lesion segmentation pipeline across different pathological scenarios: multiple sclerosis (MS), using the MSLesSeg dataset [4], and brain metastasis, using the BM dataset [13]. In both cases, the original studies were provided in NIfTI format (`.nii/.nii.gz`) along with expert-manual annotations used as ground-truth segmentation masks.

2.1.1 Multiple Sclerosis - MSLesSeg

Multiple sclerosis is a chronic autoimmune neurological disorder characterized by demyelinating lesions in the central nervous system (CNS) [18]. These lesions are caused by damage to myelin, the protective sheath surrounding nerve fibers, which disrupts neural signal transmission. From a clinical perspective, lesion

burden, spatial distribution, and temporal evolution are essential biomarkers for diagnosis, disease staging, and longitudinal patient monitoring [9].

The MSLesSeg (*Multiple Sclerosis Lesion Segmentation*) dataset [4] provides a benchmark for the automatic segmentation of MS lesions, including structural brain MRI scans from 75 patients with MS and expert-annotated lesion masks. The main imaging modalities available are T1-weighted (T1w) and FLAIR, both of which are routinely used in clinical neuroimaging. In particular, FLAIR sequences are particularly informative for MS lesion detection, as they suppress cerebrospinal fluid signal while enhancing hyperintense white matter abnormalities associated with inflammation and demyelination [9]. MS lesions are highly diverse in size, shape, location, and intensity, appearing as small, scattered foci or larger, confluent lesions, with variability across patients and sessions. This heterogeneity makes automatic segmentation challenging and serves as a benchmark for model generalization.

Some patients in the repository contain multiple longitudinal acquisitions. To avoid data leakage and ensure sample independence, only one study per patient was systematically selected. Specifically, the first available acquisition was retained, using the paired T1w and FLAIR volumes together with the corresponding lesion mask. This strategy ensures that each case represents an independent subject and prevents bias during comparative evaluation. The final split used in this work consisted of 60 training cases and 15 test cases.

2.1.2 Brain Metastasis - BM

The second dataset used in this work corresponds to the MetaS-RD brain metastasis dataset, described in [13]. Unlike primary brain tumors, BMs are secondary lesions caused by the hematogenous spread of a primary cancer—most commonly lung, breast, or melanoma— to the brain parenchyma. From a clinical perspective, their accurate detection and delineation are essential for treatment planning, including radiosurgery and surgical resection, as well as for longitudinal disease monitoring [10].

The dataset contains T1-contrast-enhanced (T1-GD) structural brain MRI studies in NIfTI format, along with expert manual segmentation masks from 75 patients. Each patient is organized into an individual folder that may contain multiple longitudinal follow-up acquisitions. Although this temporal variability is valuable for progression studies, including multiple scans from the same subject could introduce sample dependency and bias the comparative evaluation of segmentation architectures. To ensure case independence, only one acquisition per patient was selected.

The original annotations may contain multiple labels representing different tumor-related subregions. However, to ensure direct comparability with the MSLesSeg dataset, the task was reformulated as a binary lesion segmentation problem (*tumor vs. background*). Therefore, all labels were merged into a single lesion class. This dataset provides a complementary pathological scenario to MS, characterized by more focal lesions, stronger contrast enhancement, and a markedly different spatial distribution. As in the previous dataset, the final split consisted of 60 training cases and 15 test cases.

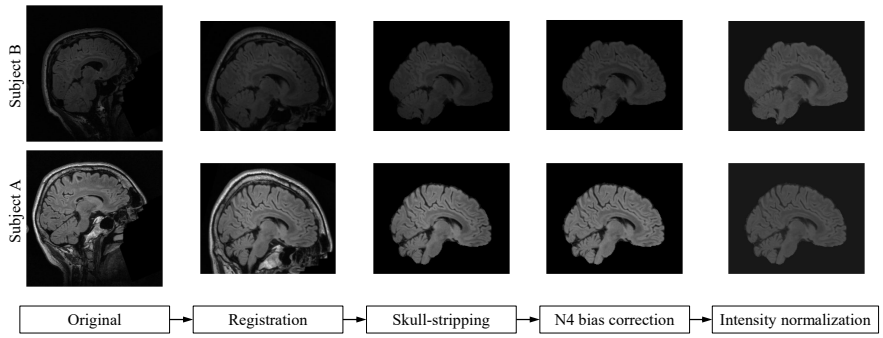
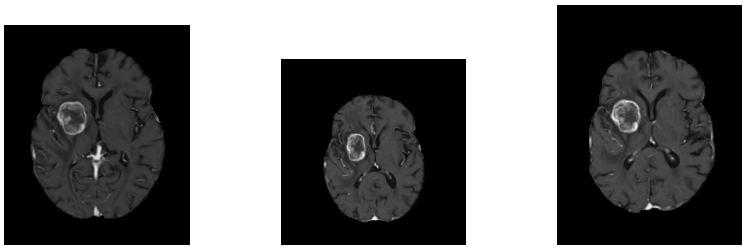


Fig. 2: Preprocessing pipeline.

2.2 Preprocessing

A dedicated preprocessing pipeline was applied to MRI scans to reduce acquisition-related variability while preserving clinically relevant lesion information. Brain MRI studies often exhibit substantial heterogeneity due to differences in scanners, acquisition protocols, spatial resolution, intensity distributions, and patient positioning, which may hinder comparisons between intra- and inter-subject analyses. As shown in Fig. 2, the complete pipeline consisted of four main stages:

1. Registration: image registration is a key process in digital image processing that involves aligning multiple images of the same scene by matching features such as intensities or structures to determine the optimal transformation. This process aims to geometrically modify an input image (moving) to produce an output that aligns with a reference (fixed) image. Spatial alignment through non-linear registration to a common anatomical reference space for guaranteeing intra- and inter-subject comparison. For this purpose, three templates –**MNI152** [12], **SRI24** [16], and **NMRI225** [8] – were chosen to analyze their influence on the model performance. Without registration, voxels at the same grid position in different modalities do not correspond to the same anatomical location. Let $I(\boldsymbol{x}')$ and $F(\boldsymbol{x})$ denote the moving and fixed images on domains $\mathbb{D}_I$ and $\mathbb{D}_F$, respectively. Image registration seeks a transformation $\mathcal{T} : \mathbb{D}_F \rightarrow \mathbb{D}_I$ such that the transformed moving image
$$\tilde{I}(\boldsymbol{x}) = I(\mathcal{T}(\boldsymbol{x})), \tag{1}$$
is aligned with the fixed image such that $\tilde{I}(\boldsymbol{x}) \approx F(\boldsymbol{x})$. For this study, the ANTs library with the "SyN" transformation was employed. The MSLesSeg has already been registered to the MNI152 atlas. Fig. 3 depicts a BM sample with each atlas template registration applied.
2. Skull-stripping: non-brain tissue removal from MRI scans to reduce areas where lesions cannot appear and to ensure anonymity. The ANTs library was used to apply the brain extraction. This step was already performed in the released MSLesSeg dataset.
3. N4 bias field correction: because of low spatial-frequency image-intensity variations, MRI scans are affected by the phenomenon known as *bias field*.



(a) MNI152 (b) SRI24 (c) NMRI225

Fig. 3: T1-GD axial sample from the BM dataset for each considered template.

The bias field appears as a slight non-homogeneity in the intensity values for the same tissue across different locations. To counteract it, the N4 algorithm [19] was applied, improving intra-tissue intensity consistency.

4. Intensity normalization: the lack of a standard intensity scale across MRI scans from different institutions, scanners, acquisition protocols, or software versions makes it difficult to compare in complex multimodality, multicenter, or longitudinal imaging studies. Therefore, intensity normalization is a crucial preprocessing step that aims to reduce image intensity variability, align histograms, and enable comparison. In this work, all volumes were intensity-normalized using *Min-Max* scaling to the range $[0, 1]$ which is defined as:

$$I_{minmax} = \frac{I - I_{min}}{I_{max} - I_{min}}, \tag{2}$$

where I corresponds to the voxel intensity value, and $Imax$ and $Imin$ are the maximum and minimum intensity values of the image, respectively.

This preprocessing pipeline was applied identically to both datasets, enabling a controlled evaluation of how anatomical template selection affects the performance of DL methods across MS and BM lesion segmentation tasks.

2.3 Methods

This work compares two segmentation frameworks in a controlled, fully reproducible experimental setup: nnU-Netv2 [7] and U-Mamba [11]. These two architectures represent complementary paradigms for medical imaging analysis, comparing convolutional and Mamba-based approaches. Both approaches were trained using the same datasets, preprocessing pipeline, data splits, and evaluation protocol, ensuring a controlled comparison across architectures, dimensionalities, and registration templates.

The first evaluated framework was nnU-Netv2 [7], a self-configuring biomedical image segmentation framework based on U-Net-derived encoder–decoder architectures. Rather than proposing a novel network design, nnU-Netv2 focuses on automatically optimizing the entire segmentation workflow. This includes experiment planning, target spacing selection, patch size estimation, batch size definition, intensity normalization, data augmentation, and inference strategy, all of which are automatically adapted to the intrinsic characteristics of each

dataset. In this work, the `nnUNetPlannerResEncM` planner was employed, which uses a residual encoder variant that improves feature propagation and stabilizes training for complex segmentation tasks. This configuration was selected to maximize robustness across both neurological pathologies and across the different anatomical template spaces evaluated.

The second evaluated framework was U-Mamba [11], a recent extension of nnU-Netv2 that replaces part of the purely convolutional processing with blocks based on State Space Models (SSMs). These Mamba modules enhance the network's ability to capture long-range contextual dependencies, which is especially relevant in brain MRI volumes, where lesions may be distributed across distant anatomical regions. A major advantage of U-Mamba is its seamless integration into the nnU-Netv2 ecosystem through a dedicated trainer. This allows the architecture to reuse the same preprocessing, planning, augmentation, cross-validation, inference, and postprocessing stages as the baseline framework. Consequently, the architectural contribution of the Mamba blocks can be analyzed in isolation, without confounding factors introduced by differences in the experimental pipeline.

2.4 Experimental setup

For both frameworks, experiments were conducted under two standard nnU-Netv2 configurations: `2D` and `3D full-resolution`. The 2D setup processes MRI slices independently and is computationally efficient, while the 3D full-resolution setup operates on volumetric patches, preserving inter-slice spatial context and allowing richer contextual modeling.

Datasets were split into train and test subsets. A 5-fold cross-validation strategy was applied over the training partition. Each fold was trained independently and later used for multi-fold inference on the held-out test subset, reducing variance and improving robustness. The optimization procedure followed the default nnU-Netv2 setup, based on a combined Dice and Cross-Entropy loss, stochastic gradient descent with momentum, and on-the-fly data augmentation. The experiments were conducted on a workstation equipped with an NVIDIA RTX 4070 Ti SUPER GPU (16 GB of VRAM), an Intel Core i9-14900KF CPU, and 64 GB of RAM.

To evaluate segmentation performance, the similarity between the predicted segmentation mask, Pr, and the ground truth, Gt, is quantified using the Dice Similarity Coefficient (DSC) [3], computed as follows:

$$\mathrm{DSC} = \frac{2 \cdot |Pr \cap Gt|}{|Pr| + |Gt|}, \tag{3}$$

where $|Pr|$ and $|Gt|$ denote the cardinalities of the predicted and ground truth segmentation masks, respectively, $|Pr \cap Gt|$ is the number of voxels correctly classified as *lesion*. A DSC of 1 indicates perfect overlap, while a value of 0 denotes no overlap.

3 Results

Tables 1 and 2 summarize the Dice Similarity Coefficient (DSC) obtained on both datasets across anatomical templates, architectures, and training dimensionalities. Results are reported as *mean* $\pm$ *standard deviation* of the five folds

Table 1: DSC obtained in the MSLesSeg dataset across architectures and templates. The highest value per subset is indicated in **bold**.

		nnU-Netv2		U-Mamba	
	Template	**Val**	**Test**	**Val**	**Test**
2D	MNI152	0.641 ± 0.040	0.640 ± 0.006	0.639 ± 0.049	0.641 ± 0.005
	SRI24	0.638 ± 0.047	0.637 ± 0.006	0.641 ± 0.052	0.643 ± 0.004
	NMRI225	0.638 ± 0.042	0.638 ± 0.005	0.640 ± 0.049	0.642 ± 0.006
3D	MNI152	0.721 ± 0.049	**0.714 ± 0.008**	0.724 ± 0.043	0.709 ± 0.005
	SRI24	**0.722 ± 0.050**	0.707 ± 0.005	0.724 ± 0.042	**0.714 ± 0.002**
	NMRI225	0.721 ± 0.043	0.712 ± 0.007	**0.727 ± 0.032**	0.711 ± 0.003

Table 2: DSC obtained in the BM dataset across architectures and templates. The highest value per subset is indicated in **bold**.

		nnU-Netv2		U-Mamba	
	Template	**Val**	**Test**	**Val**	**Test**
2D	MNI152	0.508 ± 0.102	0.455 ± 0.035	0.532 ± 0.108	0.494 ± 0.040
	SRI24	0.516 ± 0.119	0.460 ± 0.019	0.551 ± 0.130	0.453 ± 0.032
	NMRI225	0.476 ± 0.062	0.434 ± 0.043	0.475 ± 0.082	0.532 ± 0.031
3D	MNI152	0.591 ± 0.065	0.522 ± 0.063	0.676 ± 0.097	0.720 ± 0.049
	SRI24	0.551 ± 0.063	0.385 ± 0.035	0.692 ± 0.088	0.619 ± 0.026
	NMRI225	**0.618 ± 0.136**	**0.545 ± 0.027**	**0.694 ± 0.060**	**0.732 ± 0.025**

of validation partitions and over the held-out test subset. The Fig. 4 depicts the held-out test results as a heatmap for both pathologies across architectures and configurations.

For the MSLesSeg dataset, the most consistent trend is the clear superiority of **3D models** over their 2D counterparts. Across all anatomical templates, 3D configurations improved both validation and test DSC by approximately 0.07–0.08, confirming the importance of volumetric context for MS lesion segmentation. The influence of the anatomical template was comparatively limited. Differences among MNI152, SRI24, and NMRI225 remained below 0.01 DSC in nearly all settings, suggesting that once spatial normalization is applied, the specific choice of template has only a minor effect on performance. Regarding the comparison between architectures, both methods achieved very similar DSC scores in MSLesSeg, with no consistent advantage on the held-out test subset. The best test performance was achieved by the 3D configurations, with nnU-Netv2 on MNI152 and U-Mamba on SRI24 both reaching a DSC of 0.714. Fig. 5 illustrates the qualitative result of an MS subject across strategies.

The BM dataset exhibited a markedly different behavior. In this case, the performance gap between 2D and 3D was substantially larger, increasing by 0.1-0.2, indicating a stronger dependence of metastasis segmentation on full volumetric context. Template comparison suggests that NMRI225 is the most suitable template, followed by MNI152 and SRI24. Fig. 6 illustrates the qualitative result of

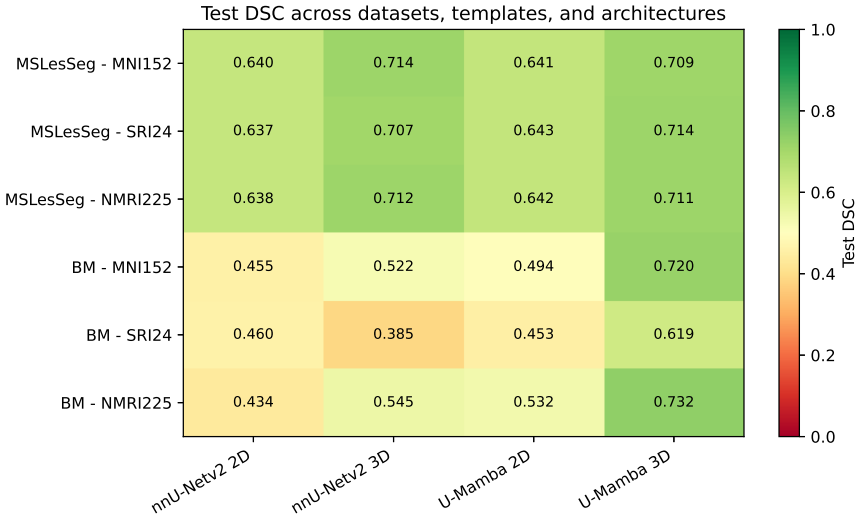


Fig. 4: Test DSC heatmap on test subset for MSLesSeg and BM datasets across architectures, 2D/3D configurations, and templates.

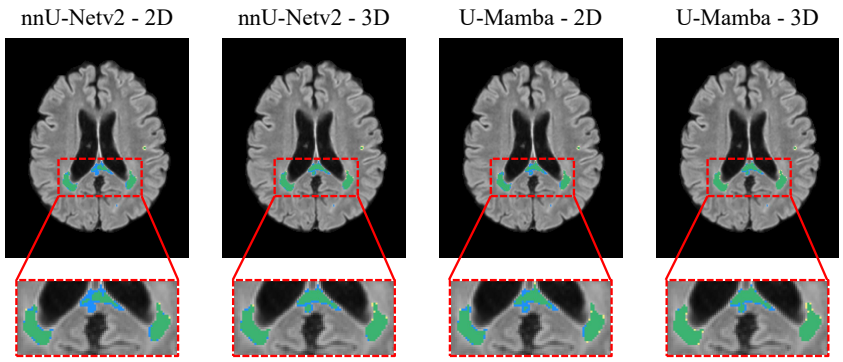


Fig. 5: Axial FLAIR prediction sample from the MSLesSeg dataset across strategies with the MNI152 template. The ground truth mask, the predicted mask, and their intersection are shown in blue, yellow, and green, respectively.

a BM subject across templates. More importantly, U-Mamba showed a substantial improvement over nnU-Netv2, especially in the 3D setting. While the best 3D nnU-Netv2 configuration achieved a test DSC of 0.545, U-Mamba attained 0.732 using the NMRI225 template. This large gain suggests that Mamba blocks' ability to capture long-range contextual dependencies is particularly beneficial for compact, spatially structured lesions such as BMs.

Across both datasets, the results indicate that training dimensionality is the dominant factor, with 3D models achieving the strongest overall performance. The impact of anatomical template selection was minor for MS and more pronounced for metastases, with NMRI225 yielding the strongest results. From an architectural perspective, U-Mamba provided only marginal gains in MS but de-



Fig. 6: Axial T1-GD prediction sample from the BM dataset across templates with the nnU-Netv2 - 2D model. The ground truth mask, the predicted mask, and their intersection are shown in blue, yellow, and green, respectively.

livered a substantial improvement in BMs, suggesting that its enhanced global context modeling is especially advantageous for focal lesions with stronger spatial structure.

Analysis of training times reveals an interesting pattern: 3D configurations are not slower than 2D ones in nnU-Netv2. On average, 3D epochs are completed about 20–25% faster. This occurs because of the framework's sampling strategy: the 2D setup processes each slice separately, resulting in more iterations per epoch. Conversely, 3D uses volumetric patches that cover larger regions, reducing the total number of batches needed. Although 3D has a higher computational cost per forward pass, the fewer total iterations mean shorter epochs. When comparing architectures, U-Mamba consistently trains longer than nnU-Netv2. With MSLesSeg, 2D training takes nearly twice as long per epoch, while 3D is 70–80% longer due to the increased complexity of the State Space Model blocks, which add extra global contextual operations. Overall, there's a clear trade-off: nnU-Net v2 balances performance and training time best, while U-Mamba achieves higher performance, especially in complex tasks like metastasis, at the expense of increased computational cost.

4 Conclusions

This study compared two medical image segmentation architectures: the CNN-based nnU-Netv2 and the Mamba-based U-Mamba, for automatic brain lesion segmentation in two clinically distinct scenarios: multiple sclerosis and brain metastases. It also evaluated the influence of registration preprocessing by considering three atlas templates: MNI152, SRI24, and NMRI225.

The results suggest that lesion morphology and spatial organization strongly modulate the relative benefit of each architecture. For MS, where lesions are small, diffuse, and distributed across white matter, both models achieved highly similar performance. This suggests that local convolutional representations are sufficient to capture the relevant lesion patterns. In contrast, for BM, U-Mamba provided a substantial improvement over nnU-Netv2, particularly in the 3D setting. This finding highlights the benefit of long-range contextual modeling when

lesions are compact, spatially structured, and strongly dependent on surrounding anatomical context. Across both datasets, 3D training generally outperformed 2D configurations, confirming that volumetric continuity is a dominant factor in brain MRI lesion segmentation. Overall, the main contribution of this work is to demonstrate that the effectiveness of global context modeling depends on the underlying lesion topology, and to analyze which configuration outperforms for each pathology considered. Conventional CNNs remain highly competitive for diffuse local lesions, whereas architectures with enhanced global dependency modeling are especially advantageous for focal, structurally coherent pathologies.

Acknowledgments. This work is partially supported by the Autonomous Government of Andalusia (Spain) under project PPRO-TIC163-G-2023 (TIC163-G-FEDER), and under grant number DGP_PIDI_2024_00462; also by the Ministry of Science, Innovation, and Universities of Spain, grant number PID2022-136764OA-I00. It includes funds from the European Regional Development Fund (ERDF). It is also partially supported by the Instituto de Salud Carlos III, project code PI25/02129 (co-financed by the European Union). The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. They also gratefully acknowledge the support of NVIDIA Corporation with the donation of an RTX A6000 GPU with 48Gb. The authors also thankfully acknowledge the grant of the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND.

Data availability. The open-access MSLesSeg [4] dataset is available on https://doi.org/10.6084/m9.figshare.27919209 and the open-access BM dataset [13] on https://molab.es.

Disclosure of Interests. The authors declare no conflicts of interest.

References

1. Amri, Y., Slama, A.B., Mbarki, Z., Selmi, R., Trabelsi, H.: Automatic glioma segmentation based on efficient u-net model using mri images. Intelligence-Based Medicine **11**, 100216 (2025)
2. Belwal, P., Singh, S.: U-net approach to mri-based segmentation of multiple sclerosis lesions. Procedia Computer Science **259**, 1316–1325 (2025)
3. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
4. Guarnera, F., Rondinella, A., Crispino, E., Russo, G., Di Lorenzo, C., Maimone, D., Pappalardo, F., Battiato, S.: Mslesseg: baseline and benchmarking of a new multiple sclerosis lesion segmentation dataset. Scientific Data **12**(1), 1–10 (2025)
5. Guo, X., Zhang, B., Peng, Y., Chen, F., Li, W.: Segmentation of glioblastomas via 3d fusionnet. Frontiers in Oncology **14**, 1488616 (2024)
6. Huang, L., Miron, A., Hone, K., Li, Y.: Segmenting medical images: from unet to res-unet and nnunet. In: 2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS). pp. 483–489. IEEE (2024)
7. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)

8. Kreilkamp, B.A., Martin, P., Bender, B., la Fougère, C., Van de Velden, D., Stier, C., Ethofer, S., Kotikalapudi, R., Marquetand, J., Rauf, E.H., et al.: Big field of view mri t1w and flair template-nmri225. Scientific Data **10**(1), 211 (2023)
9. Lassmann, H., Van Horssen, J., Mahad, D.: Progressive multiple sclerosis: pathology and pathogenesis. Nature reviews neurology **8**(11), 647–656 (2012)
10. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H., Pfister, S.M., Reifenberger, G., et al.: The 2021 who classification of tumors of the central nervous system: a summary. Neuro-oncology **23**(8), 1231–1251 (2021)
11. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
12. Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al.: A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **356**(1412), 1293–1322 (2001)
13. Ocaña-Tienda, B., Pérez-Beteta, J., Villanueva-García, J.D., Romero-Rosales, J.A., Molina-García, D., Suter, Y., Asenjo, B., Albillo, D., Ortiz de Mendivil, A., Pérez-Romasanta, L.A., et al.: A comprehensive dataset of annotated brain metastasis mr images with clinical and radiomic data. Scientific data **10**(1), 208 (2023)
14. Preložnik, D., Špiclin, Ž.: Cross-modality white matter lesion segmentation by modality de-indentification. Pattern Recognition Letters (2025)
15. Rasool, N., Bhat, J.I.: A critical review on segmentation of glioma brain tumor and prediction of overall survival. Archives of Computational Methods in Engineering **32**(3), 1525–1569 (2025)
16. Rohlfing, T., Zahr, N.M., Sullivan, E.V., Pfefferbaum, A.: The sri24 multichannel atlas of normal adult human brain structure. Human brain mapping **31**(5), 798–819 (2010)
17. Salanitri, F.P., Bellitto, G., Calcagno, S., Bagci, U., Spampinato, C., Pennisi, M.: Sam-guided prompt learning for multiple sclerosis lesion segmentation. Pattern Recognition Letters (2025)
18. Thompson, A.J., Baranzini, S.E., Geurts, J., Hemmer, B., Ciccarelli, O.: Multiple sclerosis. The Lancet **391**(10130), 1622–1636 (2018). https://doi.org/https://doi.org/10.1016/S0140-6736(18)30481-1, https://www.sciencedirect.com/science/article/pii/S0140673618304811
19. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)
20. Varma, A., Scholz, D., Erdur, A.C., Peeken, J.C., Rueckert, D., Wiestler, B.: Improving out-of-domain generalization in multiple sclerosis detection and segmentation using random convolutions. Pattern Recognition Letters (2025)
21. Xia, Q., Zheng, H., Zou, H., Luo, D., Tang, H., Li, L., Jiang, B.: A comprehensive review of deep learning for medical image segmentation. Neurocomputing **613**, 128740 (2025)
22. Yadav, A.C., Kolekar, M.H., Zope, M.K.: Modified recurrent residual attention u-net model for mri-based brain tumor segmentation. Biomedical Signal Processing and Control **102**, 107220 (2025)