Ischemic Stroke Segmentation from a Cross-Domain Representation in Multimodal Diffusion Studies

Santiago Gómez[1], Daniel Mantilla[2], Brayan Valenzuela[1], Andres Ortiz[2], Daniela D Vera[2], Paul Camacho[2], and Fabio Martínez[1(✉)]

[1] BIVL2ab, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia
famarcar@saber.uis.edu.co

[2] Clínica FOSCAL, Floridablanca, Santander, Colombia

Abstract. Localization and delineation of ischemic stroke are crucial for diagnosis and prognosis. Diffusion-weighted MRI studies allow to associate hypoperfused brain tissue with stroke findings, observed from ADC and DWI parameters. However, this process is expensive, time-consuming, and prone to expert observational bias. To address these challenges, currently, deep representations are based on deep autoencoder representations but are limited to learning from only ADC observations, biased also for one expert delineation. This work introduces a multimodal and multi-segmentation deep autoencoder that recovers ADC and DWI stroke segmentations. The proposed approach learns independent ADC and DWI convolutional branches, which are further fused into an embedding representation. Then, decoder branches are enriched with cross-attention mechanisms and adjusted from ADC and DWI findings. In this study, we validated the proposed approach from 82 ADC and DWI sequences, annotated by two interventional neuroradiologists. The proposed approach achieved higher mean dice scores of 55.7% and 57.7% for the ADC and DWI annotations by the training reference radiologist, outperforming models that only learn from one modality. Notably, it also demonstrated a proper generalization capability, obtaining mean dice scores of 60.5% and 61.0% for the ADC and DWI annotations of a second radiologist. This study highlights the effectiveness of modality-specific pattern learning in producing cross-domain embeddings that enhance ischemic stroke lesion estimations and generalize well over annotations by other radiologists.

Keywords: Ischemic stroke · Difussion-weighted MRI · Cross-domain learning · Multi-segmentation strategies

Ministry of science, technology and innovation of Colombia (MINCIENCIAS).

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 776–785, 2023.
https://doi.org/10.1007/978-3-031-43901-8_74

1 Introduction

Stroke is the second leading cause of death worldwide, with over twelve million new cases reported annually. Projections estimate that one in four individuals will experience a stroke in their lifetime [3]. Ischemic stroke, caused by blood vessel occlusion, is the most common stroke, accounting for approximately 80% of all cases. These strokes carry a high risk of morbidity, emphasizing the need for timely and accurate diagnosis and treatment [4].

Lesion delineation and quantification are critical components of stroke diagnosis, guiding clinical intervention and informing patient prognosis [7]. Diffusion-weighted MRI studies are the primary tool for observing and characterizing stroke lesions. Among these sequences, the B-1000 (DWI) and the apparent diffusion coefficient (ADC) are biomarkers of cytotoxic injury that predicts edema formation and outcome after ischemic stroke [1]. Typically, the standard lesion analysis requires neuroradiologists to observe both modalities to derive a comprehensive lesion characterization. However, this observational task is challenging, time-consuming (taking approximately 15 min per case), and susceptible to biased errors [9,11].

To address these challenges, computational approaches have supported stroke lesion segmentation, including multi-context approaches from convolutional architectures and standard training schemes to achieve a latent lesion representation [5]. Other strategies have tackled the class imbalance between lesion regions and healthy tissue by using mini-batches of image patches, ensuring stratified information but overlooking global lesion references [2]. More recent architectures have included attention modules that integrate local and contextual information to enhance stroke-related patterns [8,12]. These approaches in general are however adjusted from only one modality reference, losing complementary radiological findings. Furthermore, many of these approaches are calibrated to one-expert annotations, introducing the possibility of learning expert bias.

To the best of our knowledge, this work is the first effort to explore and model complementary lesion annotations from DWI and ADC modalities, utilizing a novel multimodal cross-attentional autoencoder. The proposed architecture employs modality-specific encoders to extract stroke patterns and generate a cross-domain embedding that enhances lesion segmentations for both modalities. To mitigate noise propagation, cross-attention mechanisms were integrated with skip connections between encoder and decoder blocks with similar spatial dimensions. The approach was evaluated on 82 ADC and DWI sequences, annotated by two neuroradiologists, evidencing generalization capabilities and outperforming typical unimodal models. The proposed method makes significant contributions to the field, as outlined below.

- A multi-context and multi-segmentation deep attentional architecture that retrieves lesion annotations from ADC and DWI, sequences.
- An architecture generalization study by training the model with one reference radiologist but evaluating with respect to two radiologist annotations.

2 Proposed Approach

This work introduces a multi-path encoder-decoder representation with the capability to recover stroke segmentations from ADC and DWI parameters. Independent branches are encoded for each modality and fused into an embedding representation. Then, independent decoder paths are evolved to recover stroke lesions, supported by cross-attention modules from the respective encoders. A multi-segmentation task is herein implemented to update the proposed deep representation. The general description of the proposed approach is illustrated in Fig. 1.

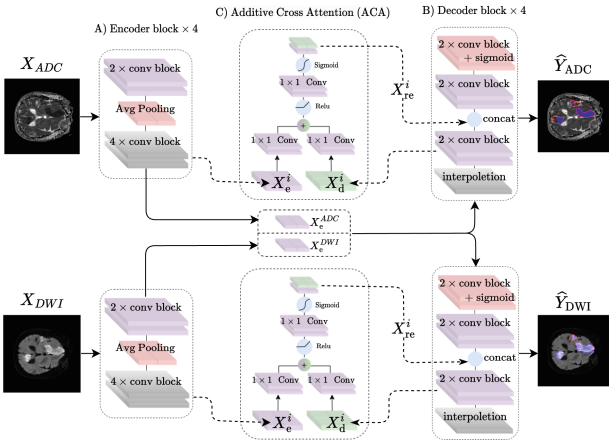

Fig. 1. Graphical representation of the proposed multimodal network for accurate segmentation of lesions in ADC and DWI modalities. The architecture comprises three main parts: A) the encoder, which characterizes the most relevant information in the input image and transforms it into a low-dimensional embedding; B) the decoder, which recovers high-dimensional information captured in the embedded representation; and C) the attention module, which filters and relates the most important features in both modalities.

2.1 Multimodal Attentional Mechanisms

In this work, a dual path encoder-decoder is introduced to recover stroke shape findings, observed from ADC and DWI. From the decoder section, two independent convolutional paths build a deep representation of each considered parameter. Both paths project studies into a low-dimensional embedding space that code stroke associated findings (see Fig. 1a).

Then, a single and complementary latent representation is achieved by the concatenation of both encoders. We hypothesize that such late fusion from embedding space allows a more robust integration of salient features related to stroke findings in both modalities.

Hence, independent decoder branches, for each modality, are implemented from the fused embedding representation. These decoder paths are updated at each level from skip connections, implemented as cross-attentional mechanisms. Particularly, these connections control the contribution of the encoder X_e^i into the decoder X_d^i representation, at each convolutional level i. The refined encoder features are computed as:

$$X_{re}^i = \sigma_2(W_{re}^T \sigma_1(W_e^T X_e^i + W_d^T X_d^i)) \cdot X_e^i$$

where σ_1 and σ_2 are ReLU and Sigmoid activations, respectively (see in Fig. 1c). Consequently, the refined features X_{re}^i preserve highly correlated features between encoder and decoder branches, at the same level of processing. These cross-attention mechanisms enhance the quality of lesion estimations as a result of complementing the decoder representation with early representations of the encoder.

2.2 Multi-segmentation Minimization

A multi-segmentation loss function is here introduced to capture complementary textural patterns from both MRI parameters and improve lesion segmentation. This function induces the learning of highly correlated features present in ADC and DWI images. Specifically, the loss function is defined as:

$$\mathcal{L} = -\alpha(1 - \hat{\mathbf{Y}}_{\text{ADC}})^{\gamma} \log(\hat{\mathbf{Y}}_{\text{ADC}})\mathbf{C} - \alpha(1 - \hat{\mathbf{Y}}_{\text{DWI}})^{\gamma} \log(\hat{\mathbf{Y}}_{\text{DWI}})\mathbf{C} \qquad (1)$$

Here, γ is a focusing parameter, α is a weight balancing factor, and $\hat{\mathbf{Y}}_{\text{ADC}}$ and $\hat{\mathbf{Y}}_{\text{DWI}}$ are the annotations for ADC and DWI, respectively. It should be noted, that ischemic stroke segmentation is highly imbalanced, promoting background reconstruction. To overcome this issue, each modality loss is weighted by a set of class weight maps ($\mathbf{C} \in \mathbb{N}^{H \times W}$) that increase the contribution of lesion voxels and counteract the negative impact of class imbalance. From this multi-segmentation loss function, the model can more effectively capture complementary information from both ADC and DWI, leading to improved lesion segmentation.

2.3 ADC-DWI Data and Experimental Setup

A retrospective study was conducted to validate the capability of the proposed approach, estimating ischemic stroke lesion segmentation. The study collected 82 studies of patients with stroke symptoms at a clinical center between October 2021 and September 2022. Each study has between 20 to 26 slices with resolutions ranging in each axis as x, y $= [0.83 - 0.94]$ and z $= [5.50 - 7.20]$ mm. After reviewing imaging studies, the studies were categorized into control (n $= 7$) or ischemic stroke (n $= 75$) studies. Control patients with stroke symptoms were included to diversify tissue samples, potentially enhancing our models' ability

to segment stroke lesion tissue. Each study included ADC and DWI sequences obtained using a Toshiba Vantage Titan MRI scanner. Two neuro-interventional radiologists, each with more than five years of experience, individually annotated each medical sequence based on clinical records using the MRIcroGL software [6]. To meet the inclusion criteria, patients had to be older than 18 years and show no signs of a cerebral hemorrhage. Patients who showed evidence of partial reperfusion were not excluded from the study.

The dataset was stratified into two subsets for training and testing. The training set consisted of 51 studies with annotations from a single expert (R1), while the test set contained 31 studies with manual delineations from two experts. In addition, a validation set was created from 11 studies of the training set. This validation set was used to measure the model's capability to segment ischemic stroke lesions on unseen data during training and to save the best weights before predicting on the test set. The test set was specifically designed to evaluate the model's performance on quantifying differences between lesion annotations. It allowed for a comparison of annotations between modalities of the same radiologist and between radiologists with the same modality. Furthermore, the test set was used to assess the generalizability of the model when compared to annotations made by the two radiologists. Segmentation capability was measured from the standard Dice score (DSC), precision, and recall.

Regarding the model, the proposed symmetric architecture consists of five convolutional levels (32, 64, 128, 256, and 512 filters for both networks and 1024 filters in the bottleneck) with residual convolutional blocks (3×3, batch normalization, and ReLU activation). For training, the architecture was updated over 300 epochs with a batch size of 16, and a Focal loss with an AdamW optimizer with $\gamma = 2$ and $\alpha = 0.25$. A linear warm-up strategy was used in the first 90 epochs (learning rate from 5e−3 and decay 1e−5), followed by 210 epochs of cosine decay. The class weight map values were set to 1 for non-lesion voxels and 3 for lesion voxels. The whole studies were normalized and resized to 224×224 pixels. Registration and skull stripping were not considered. Furthermore, data augmentation such as random brightness and contrast, flips, rotations, random elastic transformations, and random grid and optical distortions were applied to the slices. As a baseline, the proposed architecture was trained with a single modality (ADC, DWI). The code for our work can be found in https://gitlab.com/bivl2ab/research/2023-cross-domain-stroke-segmentation.

3 Evaluation and Results

From a retrospective study, we conducted an analysis to establish radiologist agreement in both ADC and DWI, following the Kappa coefficient from lesion differences. In Fig. 2a and Fig. 2b are summarized the distribution agreement between radiologist annotations, observing the ADC and DWI studies, respectively. In Fig. 2c is reported the agreement distribution between modalities, observed for each radiologist.

As expected, a substantial level of agreement among radiologists was found in both ADC and DWI sequences, with an average agreement of 65% and 70%,

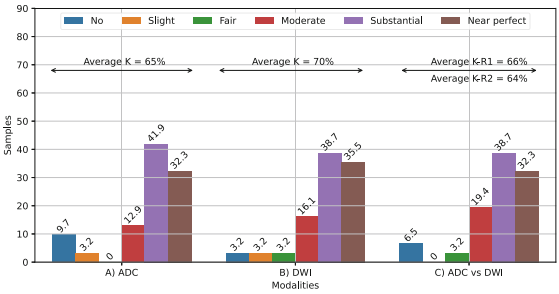

Fig. 2. Percentage of samples and average kappa values stratified in the different levels of agreement. The kappa values for inter-modality are given for both radiologists.

respectively. It should be noted however that in 25% of studies there exist between fair and moderate agreement. These discrepancies may be associated with acute strokes, where radiological findings are negligible and challenging even for expert radiologists. In such a sense, it should be noted that computational approaches may be biased for expert annotations, considering the substantial agreement. Also, among modalities is reported a kappa value of around 65% that evidences the importance to complement findings to achieve more precise stroke delineations.

Hence, an ablation study was then carried out to measure the capabilities of the proposed approach to recover segmentation, regarding multi-segmentation segmentation and the contribution of cross-attention mechanisms. The capability of architecture was measured with respect to the two radiologist delineations and with respect to each modality. Figure 3 illustrates four different variations of the proposed approach. Firstly, a unimodal configuration was implemented from decoder cross-attention mechanisms but adjusted with only one modality. The Dual V1 configuration corresponds to multi-segmentation proposed strategy that uses cross-attention mechanisms at decoder paths. In this version, each cross-attention is learned from the encoder and previous decoder levels. The version (Dual V2) implements cross-attention mechanisms into the encoder path, fusing both modalities at each processing level. Conversely, the Dual V3 integrates encoder and decoder cross-attention modules to force multimodal fusion at different levels of architecture.

Table 1 summarizes the achieved results for each architecture variation and with respect to each radiologist reference at ADC and DWI modalities. Interestingly, the Dual V1 architecture, following an embedding fusion, achieves consistent dice scores in both ADC (V1 = 0.58 vs V3 = 0.53) and DWI (V1 = 0.60 vs V3 = 0.62) for both radiologists, taking advantage of multi-segmentation segmentation learning. Notably, the Dual V1 outperforms most of the other Dual models by up to 10% and the typical approximation by up to 30%. It should be noted that the score achieved for unimodal architecture is coherent with the achieved by other strategies in public datasets [5,10]. The other dual versions, despite of multilevel fusion, have a trend to filter stroke features and reduce the

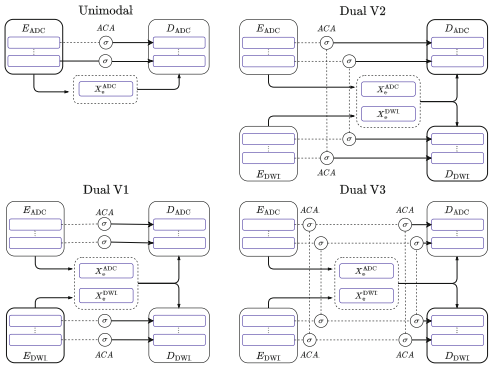

Fig. 3. Baseline unimodal models and different variations of the proposed approach.

quality of the estimated segmentations. Also, the proposed approach achieves remarkable results regarding expert radiologists, even with respect to R2 annotations, unseen during training. These results evidence the potential capabilities of the architecture to generalize segmentations, avoiding expert bias.

Table 1. Estimated segmentation metrics for the four proposed models. The best results are highlighted in gray.

Modality	Model	Radiologist 1			Radiologist 2		
		Dice ↑	Prec ↑	Sens ↑	Dice ↑	Prec ↑	Sens ↑
ADC	Unimodal	0.28 ± 0.30	0.66 ± 0.43	0.15 ± 0.18	0.32 ± 0.30	0.60 ± 0.46	0.17 ± 0.19
	Dual V1	0.56 ± 0.28	0.71 ± 0.28	0.44 ± 0.30	0.60 ± 0.29	0.69 ± 0.33	0.51 ± 0.32
	Dual V2	0.49 ± 0.29	0.68 ± 0.32	0.36 ± 0.27	0.53 ± 0.29	0.66 ± 0.37	0.41 ± 0.28
	Dual V3	0.51 ± 0.28	0.71 ± 0.29	0.43 ± 0.29	0.55 ± 0.29	0.69 ± 0.33	0.49 ± 0.30
DWI	Unimodal	0.45 ± 0.34	0.57 ± 0.37	0.45 ± 0.37	0.45 ± 0.34	0.59 ± 0.38	0.40 ± 0.34
	Dual V1	0.58 ± 0.30	0.61 ± 0.32	0.54 ± 0.36	0.61 ± 0.29	0.67 ± 0.33	0.51 ± 0.31
	Dual V2	0.52 ± 0.29	0.62 ± 0.30	0.56 ± 0.34	0.53 ± 0.28	0.67 ± 0.30	0.52 ± 0.28
	Dual V3	0.62 ± 0.26	0.67 ± 0.29	0.55 ± 0.34	0.61 ± 0.28	0.71 ± 0.31	0.50 ± 0.31

Taking advantage of the substantial agreement between the annotations of both radiologists in our study, we carried out a t-test analysis between different versions of the proposed approach, finding statistically significant differences between Unimodal and Dual models, while the distributions among different Dual models have a similar statistical performance. This analysis was extended to compare the masks of the models with respect to expert radiologists. In this study, the masks of the Dual V1 model have the same distribution ($p > 0.05$) as the reference dice scores over both ADC and DWI. Similarly, the Dual V3 model masks follow the same distribution as the reference dice scores, but only on DWI. Complementary, from Bland-Altman plots in Fig. 4 it can be seen that Dual models exhibit small and clustered differences around the mean, with a

mean difference closer to zero. In contrast, single models show more spread-out differences and a higher mean difference compared to Dual models.

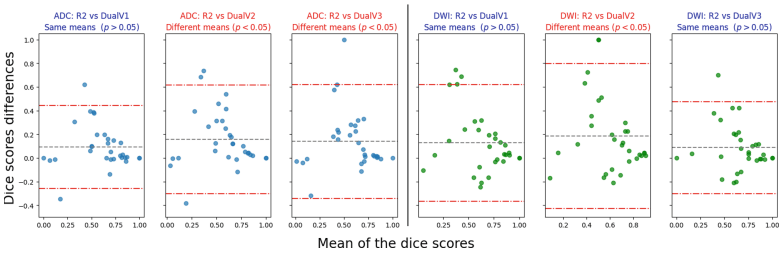

Fig. 4. Bland-Altmans between reference dice scores (R1 vs R2) and dice scores from the proposed dual models predictions.

Figure 5 illustrated the segmentations generated in each of the experiments for the proposed approach and the unimodal baseline. In top, the 3D brain rendering illustrates the stroke segmentations achieved by the dual architecture (red contours), the unimodal baseline (blue areas) and the reference of a radiologist (green areas). At the bottom, the slices illustrate each estimated segmentation as red contours, while the radiologist delineation is represented as the blue area. As expected, the proposed approach reports a remarkable capability to correctly define the limits of the ischemic lesion. Additionally, the proposed approach achieves adequate localization without overestimating the lesion, which is a key issue in identifying treatments and estimating patient diagnosis.

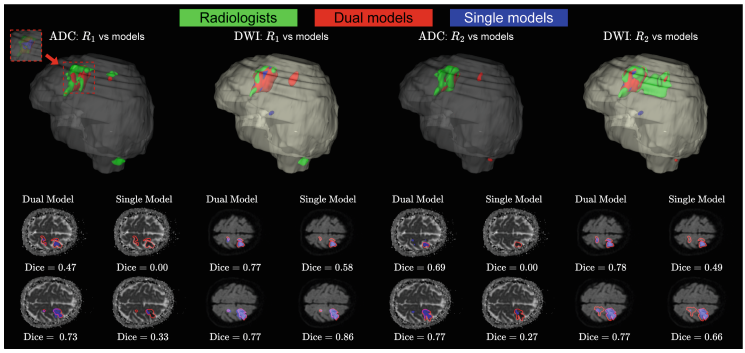

Fig. 5. Comparison of the ischemic stroke lesion predictions for the Dual V1 and unimodal baseline models against the two radiologists. (Color figure online)

4 Conclusions and Perspectives

The proposed approach demonstrates the potential of a multi-context and multi-segmentation cross-attention encoder-decoder architecture for recovering stroke segmentations from DWI and ADC images. The proposed approach has the potential to improve the accuracy and reliability of stroke segmentation in clinical practice, nevertheless, it introduces potential challenges in scalability. Future research directions include the development of attention mechanisms that can fuse additional MRI modalities to improve the accuracy of stroke segmentation. Furthermore, the study could benefit from a more exhaustive preprocessing for the medical images. Finally, an extension of the agreement study to include a larger number of radiologists and studies could provide valuable insights into the reliability and consistency of lesion annotation in MRI studies.

Acknowledgements. The authors thank Ministry of science, technology and innovation of Colombia (MINCIENCIAS) for supporting this research work by the project "Mecanismos computacionales de aprendizaje profundo para soportar tareas de localización, segmentación y pronóstico de lesiones asociadas con accidentes cerebrovasculares isquémicos", with code 91934.

References

1. Bevers, M.B., et al.: Apparent diffusion coefficient signal intensity ratio predicts the effect of revascularization on ischemic cerebral edema. Cerebrovasc. Dis. **45**(3–4), 93–100 (2018)
2. Clèrigues, A., et al.: Acute and sub-acute stroke lesion segmentation from multimodal MRI. Comput. Meth. Programs Biomed. **194**, 105521 (2020). https://doi.org/10.1016/j.cmpb.2020.105521
3. Feigin, V.L., et al.: World stroke organization (WSO): global stroke fact sheet 2022. Int. J. Stroke **17**(1), 18–29 (2022)
4. Hinkle, J.L., Guanci, M.M.: Acute ischemic stroke review. J. Neurosci. Nurs. **39**(5), 285–293 (2007)
5. Hu, X., et al.: Brain SegNet: 3D local refinement network for brain lesion segmentation. BMC Med. Imaging **20**(1), 1–10 (2020). https://doi.org/10.1186/s12880-020-0409-2
6. Li, X., et al.: The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Meth. **264**, 47–56 (2016)
7. Liew, S.L., et al.: A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci. Data **5**(1), 1–11 (2018)
8. Liu, P.: Stroke lesion segmentation with 2D novel CNN pipeline and novel loss function. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11383, pp. 253–262. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_25
9. Martel, A.L., Allder, S.J., Delay, G.S., Morgan, P.S., Moody, A.R.: Measurement of infarct volume in stroke patients using adaptive segmentation of diffusion weighted MR images. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 22–31. Springer, Heidelberg (1999). https://doi.org/10.1007/10704282_3

10. Pinto, A., et al.: Enhancing clinical MRI perfusion maps with data-driven maps of complementary nature for lesion outcome prediction. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 107–115. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_13
11. Rana, A.K., Wardlaw, J.M., Armitage, P.A., Bastin, M.E.: Apparent diffusion coefficient (ADC) measurements may be more reliable and reproducible than lesion volume on diffusion-weighted images from patients with acute ischaemic stroke-implications for study design. Magn. Reson. Imaging **21**(6), 617–624 (2003)
12. Wang, G., et al.: Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. Med. Image Anal. **65**, 101787 (2020). https://doi.org/10.1016/j.media.2020.101787