Border irregularity loss for automated segmentation of primary brain lymphomas on post-contrast MRI

Rosana El Jurdi, Lucia Nichelli, Agusti Alentorn, Ghislain Vaillant, Guanghui Fu, Khê Hoang-Xuan, Caroline Houillier, Stéphane Lehéricy, Olivier Colliot

Border irregularity loss for automated segmentation of primary brain lymphomas on post-contrast MRI

Rosana El Jurdi[a], Lucia Nichelli[a,c,d], Agusti Alentorn[b,e], Ghislain Vaillant[a], Guanghui Fu[a], Khê Hoang-Xuan[b,e], Caroline Houillier[b,e], Stéphane Lehéricy[b,c,d], and Olivier Colliot[a]

[a]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France
[b]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France
[c]ICM, Centre de NeuroImagerie de Recherche-CENIR, Paris, France.
[d]AP-HP, Pitié Salpêtrière, DMU DIAMENT, Dep. of Neuroradiology, Paris, France
[e]AP-HP, Pitié-Salpêtrière, DMU Neurosciences, Department of Neurology 2, Paris, France

ABSTRACT

Unlike for other brain tumors, there has been little work on the automatic segmentation of primary central nervous system (CNS) lymphomas. This is a challenging task due the highly variable pattern of the tumor and its boundaries. In this work, we propose a new loss function that controls border irregularity for deep learning-based automatic segmentation of primary CNS lymphomas. We introduce a border irregularity loss which is based on the comparison of the segmentation and it smoothed version. The border irregularity loss is combined with a previously proposed topological loss to better control the different connected components. The approach is general and can be used with any segmentation network. We studied a population of 99 patients with primary CNS lymphoma. 40 patients were isolated from the very beginning and formed the independent test set. The segmentations were performed on post-contrast T1-weighted MRI. The MRI were acquired in clinical routine and were highly heterogeneous. The proposed approach substantially outperformed the baseline across the various evaluation metrics (by 6 percent points of Dice, 40mm of Hausdorff distance and 6mm of mean average surface distance). However, the overall performance was moderate, highlighting that automatic segmentation of primary CNS lymphomas is a difficult task, especially when dealing with clinical routine MRI. The code is publicly available here: https://github.com/rosanajurdi/LymphSeg.

Keywords: Lymphoma, Brain, CNS, MRI, segmentation, deep learning, loss function, constraint

1. INTRODUCTION

Primary central nervous system (CNS) lymphoma (PCNSL) is a rare and aggressive type of cancer that primarily affects the brain and spinal cord. It accounts for approximately 4% of newly diagnosed primary CNS tumors and 1% of all non-Hodgkin lymphomas (NHL). Magnetic resonance imaging (MRI) suggests its diagnosis, is needed for tumor biopsy planification, and has a pivotal role in PCNSL post-treatment assessment. Precise tumor quantification is highly desirable, and one first needed step to promote better patient care. It could aid in pre-surgical planning and in objective tumor response evaluation.[1] Manual segmentation is time-consuming and is partly subjective.

A large number of methods have been proposed for automated segmentation of brain tumors.[2–11] However, while many papers have dealt with other types of tumors such as gliomas,[12] only a limited number of studies have focused on brain lymphomas. Some works focused on classification[13,14] or survival analysis.[15] Automatic segmentation of PCNSL from MRI data has been performed in a few publications.[16,17] However, these works

Further author information: (Send correspondence to Rosana El Jurdi.)
Rosana El Jurdi: E-mail: eljurdi.rosana@gmail.com

were not specific to lymphoma and included also gliomas for training or validation. Moreover, they did not report distance-based metrics for evaluation which is critical when dealing with lymphomas which have complex boundaries and multiple components.

PCNSL has an extremely heterogenous MRI appearance. Lesions can be single or multiple. Several morphologies, topographies, and mass effect patterns can (co)exist. As a result, tumor size, shape and boundary can considerably vary between patients.

These variations pose significant challenges in accurately segmenting lymphomas. Convolutional neural network (CNN)-based segmentation approaches are widely used but often produce errors near boundaries. To address this issue, prior knowledge integration methods at the level of the loss function have been explored to enhance the plausibility of automatic segmentations.[18]

In this paper, we propose a novel prior-based loss that integrates border irregularity attributes of the tumor in order to improve segmentation performance. The proposed method was trained and validated for segmentation of PCNSL on post-contrast T1-weighted MR images using a dataset of 99 patients.

The rest of the paper is organized as follows. Section 2 describes the dataset and pre-processing. In Section 3, we introduce the proposed loss. Section 4 is devoted to experiments and results. The discussion is provided in Section 5.

2. MATERIALS

2.1 Participants and MRI data

We studied 118 patients with primary CNS lymphoma. The study was approved by the Institutional Ethical Committee (Pitié Salpêtrière Hospital, Ile-de-France VI, n° DC-2009-957) and by the French Data Protection Authority (CNIL, Commission Nationale de l'Informatique et des Libertés, DR 2013-279). According to French regulation, consent was waived as these images were acquired as part of the routine clinical care of the patients. Each patient had a T1-weighted MRI after injection of gadolinium. The images were acquired as part of clinical routine and were thus not harmonized. They were acquired on different scanners and at different field strengths (57 at 3T, 54 at 1.5T, and 7 at 1T). The MRI scan was either 3D or 2D. Manual segmentations were performed by a trained radiologist (L.N.) who also rated the difficulty of the segmentation process. The segmentation was considered difficult when the lymphoma tissue was less extended, when the lesions boundary were difficult to visualize, and/or, in rarer cases, when there were hemorrhagic remnants that are spontaneously T1 hyperintense and that can therefore mimic tumors. Furthermore, the radiologist noted whether the images presented substantial artifacts. This led to partition the dataset into four subsets: D1, D2, D3, D4 according to these two criteria (D1: easy, no artifact; D2: easy, artifacts; D3: difficult, no artifact; D4: difficult, artifacts).

We applied the following preprocessing. All images were converted from DICOM to NIfTI using dicom2niix[19*] and organized according to the Brain Imaging Data Structure (BIDS) standard.[20] Using FSL FLIRT,[21] we linearly registered each image to the MNI-152 template which is 1mm isotropic and of dimensions $181 \times 217 \times 181$.[22,23] We applied a brain mask in order to remove unnecessary information like the skull, nose, and ears. Pydra was used to implement the preprocessing steps.[24] We visually checked the preprocessing results and found that preprocessing failed for 19 patients (9 in D1, 3 in D2, 3 in D3 and 4 in D4). These were excluded from the study. The characteristics of the studied patients and the corresponding subsets are presented in Table 1.

2.2 Splits

From D1 and D3, 26 participants (30%) were isolated for testing (21 participants from D1, 5 from D3). The remainder of D1 and D3 were separated into training (80% of the remaining participants) and validation sets (20%). This resulted in 47 participants for training (39 from D1, 8 from D3) and 12 for validation (10 from D1, 2 from D3). D2 and D4 were used only for testing (n = 14).

*https://github.com/rordenlab/dcm2niix

Subset	Total	Sex	Age	3D acquisition	*1T*	*1.5T*	*3T*
D1	70	40 F (57.14%)	61.64 ± 20.28	50	3	36	31
D2	10	4 F (40.00%)	52.60 ± 29.21	9	-	1	9
D3	15	7 F (46.66%)	63.20 ± 20.81	11	1	8	6
D4	4	2 F (66.66%)	47.75 ± 32.37	3	-	2	2
Total	99	53 F (54.08%)	60.40 ± 21.84	73	4	47	48

Table 1. Characteristics of the study population. Age (in years) is reported as mean±standard-deviation. The age was missing for 8 patients (4 of them were part of D1, 2 of D2, 1 of D3, and 1 of D4). The sex was missing for one patient from D4. The table reports the number of patients for which a 3D acquisition was available (the others had a 2D acquisition). 1T, 1.5T or 3T indicates the MRI magnetic field strength.

3. PROPOSED METHOD

3.1 Proposed border irregularity loss

In order to include it in a loss function, one needs to be able to quantify the border irregularity of an object. One approach is to smooth the segmentation map, via a Gaussian filter of a kernel size s and standard deviation σ, until it becomes more uniform in shape, and then compare the smoothed segmentation and the original segmentation.[25] A segmentation with a greater degree of irregularity would require stronger smoothing, resulting in a higher index of border irregularity. Conversely, a smoother segmentation would yield a lower index of irregularity. Applying a Gaussian filter with to a segmentation map is a well-established technique in computer vision for achieving smoothness and extracting the global structure of a segmentation map without generating new irregularities, indentations, or protrusions. The irregularity index is defined as $BI = \frac{|A^* \cup A| - |A^* \cap A|}{|A^*|} = \frac{|A^*| + |A| - 2|A^* \cap A|}{|A^*|}$

where $A*$ and A are the smoothed and non-smoothed segmentation maps. The border irregularity index represents the level of dissimilarity between the smoothed and non-smoothed segmentations. Another way to express dissimilarity between two segmentations is via the complement of the Dice coefficient. Specifically, the Dice coefficient can be used to derive a border irregularity index as $I_b = 1 - 2\frac{A^* \cap A}{|A^*| + |A|}$.

If in BI, we normalize by $|A + A^*|$ instead of $|A^*|$, it is equivalent to I_b. In this paper, we will use I_b as a measure of border irregularity at the loss function level, since it leads to a direct implementation of a differentiable and smooth gradient loss function. We optimize the difference between the ground-truth border irregularity index I_b and the predicted border irregularity index $\hat{I}_b$. The impact of the smoothing procedure on the segmentation is determined by two factors: the smoothing level (σ) and the border irregularity. When dealing with a segmentation map that exhibits severe irregularity, a more intense smoothing is needed, leading to a higher σ value, and conversely. Conversely, a smoother segmentation would result in a lower σ measure of irregularity. In this work, the smoothing maps for the ground-truth segmentations are obtained statically prior to the training process and fed to the framework in order to compute the border irregularity of the prediction maps.

The process can be summarized as follows: 1) irregularities are gradually smoothed out in a systematic manner by applying a variable σ. Smaller irregularities are first eliminated, followed by the larger ones. As some indentations or protrusions are smoothed, they may reveal the presence of a larger irregularity in their respective locations. This larger irregularity is considered global irregularity, while the smaller ones are regarded as "local" irregularities. Consequently, a hierarchical structure of irregularities is established given varying σ values.

3.2 Segmentation smoothing module

Implementation of the smoothing module was carried out using a Gaussian kernel of size $k = \{5, 10\}$ and a standard deviation of σ. A variable standard deviation $\sigma = 2^x \quad \text{for} \quad x \in \{0, 1, 2, 3, ...10\}$ was iteratively applied. The smoothed segmentation map is obtained when the value obtained for I_b saturates reaching a fixed value over two iterations. This means that the smoothing process has reached a stable state, and further iterations may not improve the result. The stopping conditions ensure that the smoothing process is performed until a desirable outcome is achieved or until further iterations do not significantly affect the result. They help control the iterations and prevent unnecessary computations, improving the efficiency of the smoothing process.

3.3 Model and implementation details

We zero-padded the data to a size of (184, 220, 184).[26] The MRI scan intensities were normalized between 0 and 1. The border irregularity loss was either used in conjunction with the Dice loss only (corresponding results are denoted as BIL) or in combination with our previously proposed topological loss[27] and the Dice loss (results denoted as BIL-Topo). The proposed approaches were compared to a baseline which used the Dice loss (corresponding results are denoted as Baseline). Each 3D volume processed as a stack of independent 2D images. The network architecture was a 2D U-net[28] which architecture has been used in previous publications.[29,30]

The architecture is a 3-stage structure composed of convolutional, de-convolutional blocks, bottleneck and skip connections. The encoder part is composed of an ensemble of convolutional and batch normalization layers, whereas the decoder part is composed of 2 consecutive convolutional blocks and an upsampling layer in each of the 3 stages. The bottleneck is composed of 2 convolutional blocks separated by a residual block.[31]

The optimizer was Adam and the learning rate was 0.001. The learning rate was halved if the validation performances did not improve over 20 epochs as proposed by.[32] We used batches of 8.

At inference, we predict for each slice independently and then stack the slices belonging to the same patient to form a 3D prediction.

Subset	Method	Dice	HD	MASD
D1+D3	BIL-Topo	**64.97 [56.23, 73.08]**	**20.15 [13.61, 27.77]***	**4.07 [2.37, 6.77]***
+D2+D4	BIL	63.99 [56.68, 70.85]	61.99 [53.12, 70.72]	10.39 [8.26, 12.72]
(n = 40)	Baseline	58.63 [50.83, 66.19]	57.28 [47.05, 67.99]	10.41 [7.99, 13.08]
D1	BIL-Topo	**68.0 [58.42, 77.02]**	**19.86 [10.41, 31.98]***	**2.81 [1.99, 3.81]***
(n = 21)	BIL	67.85 [60.22, 75.15]	64.12 [50.5, 77.2]	10.53 [7.54, 13.93]
	Baseline	60.37 [52.24, 68.44]	61.32 [44.7, 77.77]	11.68 [7.96, 15.91]
D2	BIL-Topo	**80.44 [69.15, 89.16]**	**8.32 [4.89, 11.97]***	**1.33 [0.78, 1.99]***
(n = 10)	BIL	75.7 [65.99, 84.6]	53.23 [42.31, 64.88]	7.47 [4.69, 11.21]
	Baseline	72.86 [57.65, 85.34]	46.51 [34.46, 58.58]	6.16 [3.38, 10.59]
D3	BIL-Topo	**28.57 [9.94, 47.19]**	**29.93 [16.75, 44.58]**	**6.61 [4.07, 9.0]**
(n = 5)	BIL	36.99 [18.76, 52.41]	54.97 [29.65, 78.28]	11.56 [4.74, 18.38]
	Baseline	27.96 [9.46, 44.61]	42.1 [24.54, 57.5]	11.16 [5.11, 17.2]
D4	BIL-Topo	**55.9 [19.03, 84.38]**	**39.06 [15.49, 61.27]***	**14.41 [1.88, 36.32]**
(n = 4)	BIL	48.21 [21.25, 75.16]	81.44 [60.44, 104.64]	15.42 [12.8, 17.15]
	Baseline	52.23 [24.46, 76.14]	81.98 [68.06, 106.17]	13.39 [9.4, 17.37]

Table 2. Results on the whole test set and separately for D1, D2, D3 and D4 test sets. Results are presented as mean and confidence interval (computing using bootstrapping with 15,000 resamplings). D1+D2+D3+D4 refers to the union of the test sets of D1 and D3 together with D2 and D4. HD: 95% 3D Hausdorff distance. MASD: mean average surface distance. n is the number of samples. Best result in each case is in bold. * indicates that the improvement over the baseline is statistically significant.

4. EXPERIMENTS AND RESULTS

4.1 Evaluation framework

We chose the following performance metrics, based on the recommendations of Reinke et al,[33] the 3D Dice coefficient, the 95% 3D Hausdorf distance, and the mean average surface distance (MASD)[†]. The results are reported as the mean values along with their corresponding confidence intervals, which were computed using bootstrapping with 15,000 resamplings.

Since the dataset is constituted of 3D and 2D MRI scans, we report the mean and the 95 % confidence interval obtained via bootstrapping via 15000 resamples for each of the 2D and 3D patients both combined and separately.

[†]Performance metrics were computed using this code: https://github.com/deepmind/surface-distance

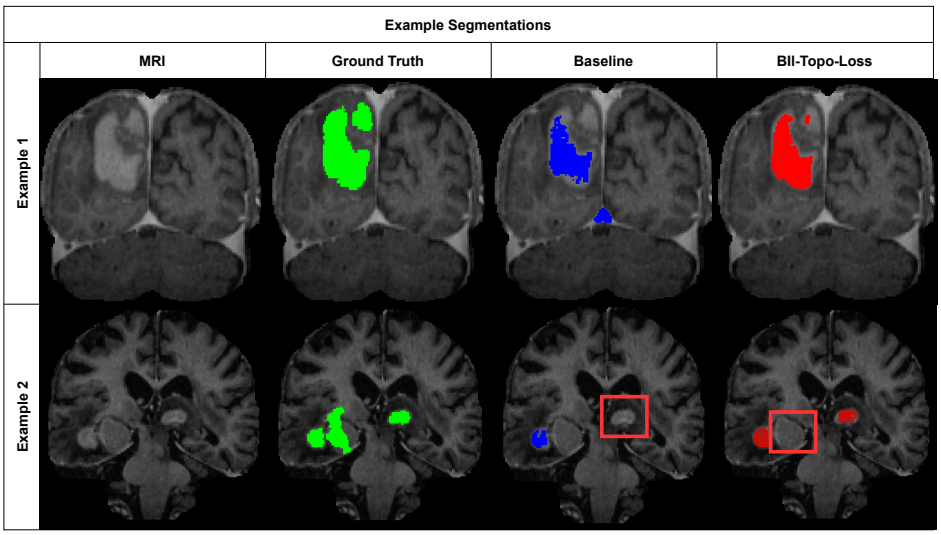

Figure 1. Examples of segmentations (ground truth, baseline and proposed methods). Red boxes are tumor parts that have been missed by the models.

4.2 Results

Table 2 presents the results on the whole test set as well as on the different subsets. The proposed method (BIL-Topo) outperformed the baseline across all performance metrics (improvement of about 6 percent points of Dice, 40mm of HD and 6mm of MASD). Confidence intervals are quite wide due to the relatively small size of the test set (n=40)[29] but the difference was statistically significant for the boundary-based metrics (HD, MASD). For the BIL alone, the mean Dice over the entire test set was substantially higher than for the baseline (5 points) and similar to that of the BIL-Topo. On the other hand, the boundary-based metrics were substantially better with the BIL-Topo, demonstrating the added value of the topological loss. Results in the different subsets are consistent with those on the entire set. One can observe that performances tend to be lower in subsets for which the segmentation was considered difficult by the rater (D3 and D4). Some examples of segmentation are shown on Figure 1.

Table 3 presents separately the results obtained on 2D vs 3D acquisitions. For both 2D and 3D acquisitions, the BIL-Topo method achieved higher Dice, lower HD and MASD a compared to the baseline. The improvement was statistically significant for HD and MASD.

Subset	Method	DSC	HD	MASD
2D (n = 7)	BIL-Topo	**73.03 [59.9, 85.42]**	**14.0 [3.49, 30.43]***	2.38 **[1.11, 3.99]***
	BIL	65.78 [52.78, 76.06]	78.65 [62.66, 95.43]	13.28 [8.76, 18.39]
	Baseline	63.23 [51.14, 74.48]	84.68 [66.57, 102.08]	14.32 [10.04, 19.07]
3D (n = 33)	BIL-Topo	**63.26 [53.25, 72.58]**	**21.45 [14.17, 30.16]***	**4.43 [2.42, 7.59]**
	BIL	63.61 [55.24, 71.42]	58.45 [48.75, 68.11]	9.77 [7.49, 12.3]
	Baseline	57.65 [48.45, 66.44]	51.47 [40.88, 62.46]	9.58 [6.86, 12.59]

Table 3. Results on test set depending on whether the acquisition was 2D or 3D. Results are presented as mean and confidence interval (computing using bootstrapping with 15,000 resamplings). HD: 95% 3D Hausdorff distance. MASD: mean average surface distance. n is the number of samples. Best result in each case is in bold. * indicates that the improvement over the baseline is statistically significant.

5. DISCUSSION

The paper introduces a novel border irregularity loss for automatic segmentation of brain lymphomas. It considers border information and combines it with a topological loss to better handle multiple connected components.

Results demonstrate the usefulness of the proposed approaches. The BIL and topological losses capture border characteristics, improve boundary delineation, and enhance segmentation performance. They both resulted in improvement in terms of Dice score over the baseline. The topological loss provided additional improvements in boundary metrics. Nevertheless, the overall performances remain moderate. The average Dice is around 65% which corresponds to a moderate spatial agreement. The relatively high Hausdorff distance mainly reflects the fact that in several cases, some tumor components are missed while erroneous connected components are detected. This highlights that automatic segmentation of brain lymphomas is a very difficult task, in particular when dealing with clinical routine MRI data of heterogeneous quality. Therefore, further work in needed on this application. The present work remains preliminary and has several limitations. Firstly, the BIL exhibits sensitivity towards small connected components. Also, the smoothing parameters are chosen in an ad-hoc manner. Future work should propose more general ways to set these parameters. Finally, it will be necessary to assess the impact of the BIL when associated with other segmentation architectures.

In future work, the aim is to address these limitations. Specifically, efforts will be made to decouple the BIL loss from its reliance on the topological loss, enabling it to effectively handle very small objects and multi-connected components.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the French government under the management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6).

REFERENCES

[1] Abdel Razek, A. A. K., Alksas, A., Shehata, M., AbdelKhalek, A., Abdel Baky, K., El-Baz, A., and Helmy, E., "Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging," *Insights Imaging* **12**, 152 (Oct. 2021).

[2] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (2014).

[3] Cabral Jr, J. E., White, K. S., Kim, Y., and Effmann, E. L., "Interactive segmentation of brain tumors in MR images using 3D region growing," in [*Medical Imaging 1993: Image Processing*], **1898**, 171–181, SPIE (1993).

[4] Liu, J.-G., Udupa, J. K., Hackney, D., and Moonis, G., "Brain tumor segmentation in MRI by using the fuzzy connectedness method," in [*Medical Imaging 2001: Image Processing*], **4322**, 1455–1465, SPIE (2001).

[5] Prastawa, M., Bullitt, E., Ho, S., and Gerig, G., "A brain tumor segmentation framework based on outlier detection," *Medical Image Analysis* **8**(3), 275–283 (2004).

[6] Khotanlou, H., Colliot, O., Atif, J., and Bloch, I., "3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy Sets and Systems* **160**(10), 1457–1473 (2009).

[7] Bauer, S., Wiest, R., Nolte, L.-P., and Reyes, M., "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine & Biology* **58**(13), R97 (2013).

[8] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., and Larochelle, H., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis* **35**, 18–31 (2017).

[9] Pereira, S., Pinto, A., Alves, V., and Silva, C. A., "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging* **35**(5), 1240–1251 (2016).

[10] Rahimpour, M., Bertels, J., Vandermeulen, D., Maes, F., Goffin, K., and Koole, M., "Improving T1w MRI-based brain tumor segmentation using cross-modal distillation," in [*Medical Imaging 2021: Image Processing*], **11596**, 261–268, SPIE (2021).

[11] Yerachmiel, M. and Greenspan, H., "Weakly supervised brain tumor segmentation via semantic affinity deep neural network," in [*Medical Imaging 2022: Image Processing*], **12032**, 872–878, SPIE (2022).

[12] Chow, D. S., Khatri, D., Chang, P. D., Zlochower, A., Boockvar, J. A., and Filippi, C. G., "Updates on deep learning and glioma: use of convolutional neural networks to image glioma heterogeneity," *Neuroimaging Clinics* **30**(4), 493–503 (2020).

[13] Miao, X., Shao, T., Wang, Y., Wang, Q., Han, J., Li, X., Li, Y., Sun, C., Wen, J., and Liu, J., "The value of convolutional neural networks-based deep learning model in differential diagnosis of space-occupying brain diseases," *Front Neurol* **14**, 1107957 (Feb. 2023).

[14] Hsu, S. P. C., Hsiao, T.-Y., Pai, L.-C., and Sun, C.-W., "Differentiation of primary central nervous system lymphoma from glioblastoma using optical coherence tomography based on attention ResNet," *Neurophotonics* **9**, 015005 (Mar. 2022).

[15] She, Z., Marzullo, A., Destito, M., Spadea, M. F., Leone, R., Anzalone, N., Steffanoni, S., Erbella, F., Ferreri, A. J. M., Ferrigno, G., Calimeri, T., and De Momi, E., "Deep learning-based overall survival prediction model in patients with rare cancer: a case study for primary central nervous system lymphoma," *Int J Comput Assist Radiol Surg* **18** (Apr. 2023).

[16] Pennig, L., Hoyer, U. C. I., Goertz, L., Shahzad, R., Persigehl, T., Thiele, F., Perkuhn, M., Ruge, M. I., Kabbasch, C., Borggrefe, J., Caldeira, L., and Laukamp, K. R., "Primary central nervous system lymphoma: Clinical evaluation of automated segmentation on multiparametric MRI using deep learning," *J Magn Reson Imaging* **53**, 259–268 (July 2020).

[17] Conte, G. M., Weston, A. D., Vogelsang, D. C., Philbrick, K. A., Cai, J. C., Barbera, M., Sanvito, F., Lachance, D. H., Jenkins, R. B., Tobin, W. O., et al., "Generative adversarial networks to synthesize missing t1 and flair mri sequences for use in a multisequence brain tumor segmentation model," *Radiology* **299**(2), 313–323 (2021).

[18] El Jurdi, R., Petitjean, C., Honeine, P., Cheplygina, V., and Abdallah, F., "High-level prior-based loss functions for medical image segmentation: A survey," *Computer Vision and Image Understanding* **210**, 103248 (2021).

[19] Li, X., Morgan, P. S., Ashburner, J., Smith, J., and Rorden, C., "The first step for neuroimaging data analysis: Dicom to nifti conversion," *Journal of neuroscience methods* **264**, 47–56 (2016).

[20] Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., et al., "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," *Scientific data* **3**(1), 1–9 (2016).

[21] Jenkinson, M., Bannister, P., Brady, M., and Smith, S., "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage* **17**(2), 825–841 (2002).

[22] Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E. E., Markiewicz, C. J., Halchenko, Y. O., Ghosh, S. S., Gorgolewski, K. J., Poldrack, R. A., and Esteban, O., "Templateflow: Fair-sharing of multi-scale, multi-species brain models," *Nature Methods* **19**, 1568–1571 (Dec 2022).

[23] Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., and Lancaster, J., "A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (icbm)," *NeuroImage* **2**(2, Part A), 89–101 (1995).

[24] Jarecka, D., Goncalves, M., Markiewicz, C. J., Esteban, O., Lo, N., Kaczmarzyk, J., and Ghosh, S., "Pydra-a flexible and lightweight dataflow engine for scientific analyses," in [*Proceedings of the 19th python in science conference*], **132**, 139 (2020).

[25] Lee, T. K. and Atkins, M. S., "New approach to measure border irregularity for melanocytic lesions," in [*Medical Imaging 2000: Image Processing*], **3979**, 668–675, Spie (2000).

[26] Pérez-García, F., Sparks, R., and Ourselin, S., "Torchio: A python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning," *Computer Methods and Programs in Biomedicine* **208**, 106236 (2021).

[27] Fu, G., Jurdi, R. E., Chougar, L., Dormont, D., Valabregue, R., Lehéricy, S., and Colliot, O., "Introducing Soft Topology Constraints in Deep Learning-based Segmentation using Projected Pooling Loss," in [*SPIE Medical Imaging 2023*], (Feb. 2023).

[28] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional networks for biomedical image segmentation," in [*MICCAI*], 234–241, Springer (2015).

[29] Jurdi, R. E. and Colliot, O., "How precise are performance estimates for typical medical image segmentation tasks?," in [*Proc. IEEE International Symposium on Biomedical Imaging - IEEE ISBI 2023*], (2023).

[30] El Jurdi, R., Petitjean, C., Cheplygina, V., and Abdallah, F., "A Surprisingly Effective Perimeter-based Loss for Medical Image Segmentation," (2021).

[31] Zhang, Z., Liu, Q., and Wang, Y., "Road extraction by deep residual U-Net," *IEEE Geoscience and Remote Sensing Letters* **15**, 749–753 (May 2018).

[32] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., and Ben Ayed, I., "Boundary loss for highly unbalanced segmentation," in [*Medical Imaging with Deep Learning*], **102**, 285–296 (July 2019).

[33] Reinke, A., Maier-Hein, L., Christodoulou, E., Glocker, B., Scholz, P., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M. A., Wiesenfarth, M., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Kavur, A. E., Rädsch, T., Tizabi, M. D., Acion, L., Antonelli, M., Arbel, T., Bakas, S., Bankhead, P., Benis, A., Cardoso, M. J., Cheplygina, V., Cimini, B. A., Collins, G. S., Farahani, K., van Ginneken, B., Hamprecht, F. A., Hashimoto, D. A., Hoffman, M. M., Huisman, M., Jannin, P., Kahn, C., Karargyris, A., Karthikesalingam, A., Kenngott, H., Kopp-Schneider, A., Kreshuk, A., Kurc, T., Landman, B. A., Litjens, G., Madani, A., Maier-Hein, K., Martel, A., Mattson, P., Meijering, E., bjoern menze, Moher, D., Moons, K. G., Müller, H., Nichyporuk, B., Nickel, F., Petersen, J., Rajpoot, N., Rieke, N., Saez-Rodriguez, J., Sánchez, C. I., Shetty, S., van Smeden, M., Sudre, C. H., Summers, R. M., Taha, A. A., Tsaftaris, S. A., Calster, B. V., Varoquaux, G., and Jaeger, P. F., "Metrics reloaded - a new recommendation framework for biomedical image analysis validation," in [*Medical Imaging with Deep Learning*], (2022).