



Segmentation of Gliomas and Prediction of Patient Overall Survival: A Simple and Fast Procedure

Elodie Puybareau^(✉), Guillaume Tochon, Joseph Chazalon,
and Jonathan Fabrizio

EPITA Research and Development Laboratory (LRDE), Le Kremlin-Bicêtre, France
elodie.puybareau@lrde.epita.fr

Abstract. This paper proposes, in the context of brain tumor study, a fast automatic method that segments tumors and predicts patient overall survival. The segmentation stage is implemented using a fully convolutional network based on VGG-16, pre-trained on ImageNet for natural image classification, and fine tuned with the training dataset of the MICCAI 2018 BraTS Challenge. It relies on the “pseudo-3D” method published at ICIP 2017, which allows for segmenting objects from 2D color-like images which contain 3D information of MRI volumes. With such a technique, the segmentation of a 3D volume takes only a few seconds. The prediction stage is implemented using Random Forests. It only requires a predicted segmentation of the tumor and a homemade atlas. Its simplicity allows to train it with very few examples and it can be used after any segmentation process. The presented method won the second place of the MICCAI 2018 BraTS Challenge for the overall survival prediction task. A Docker image is publicly available on <https://www.lrde.epita.fr/wiki/NeoBrainSeg>.

Keywords: Glioma · Tumor segmentation · Survival prediction · Fully convolutional network · Random forest

1 Introduction

Gliomas are the most common brain tumors in adults, growing from glial cells and invading the surrounding tissues [9]. Two classes of tumors are observed. The patients with the more aggressive ones, classified as high-grade gliomas (HGG), have a median overall survival of two years or less and imply immediate treatment [13, 16]. The less aggressive ones, the low-grade gliomas (LGG), allow an overall survival of several years, with no need of immediate treatment. Multimodal magnetic resonance imaging (MRI) helps practitioners to evaluate the degree of the disease, its evolution and the response to treatment. Images are analyzed based on qualitative or quantitative measures of the lesion [8, 21]. Developing automated brain tumor segmentation techniques that are able to

analyze these tumors is challenging, because of the highly heterogeneous appearance and shapes of these lesions. Manual segmentations by experts can also be a challenging task, as they show significant variations in some cases. During the past 20 years, different algorithms for segmentation of tumor structures has been developed and reviewed [1, 6, 7]. A fair comparison of those implies a benchmark based on the same dataset, and MICCAI BraTS Challenges [15] serve this purpose.

The work we present here has been done in the context of the MICCAI 2018 Multimodal Brain Tumor Segmentation Challenge (BraTS)¹. The goal of the challenge was to provide a fully automated pipeline for the segmentation of the glioma from multi modal MRI scans without any manual assistance and to predict the patient overall survival. Despite the relevance of glioma segmentation, this task is challenging due to the high heterogeneity of tumors. The development of an algorithm that can perform fully automatic glioma segmentation and overall prediction of survival would be an important improvement for patients and practitioners. A review and results of the 2018 Challenge can be found in [5].

During the challenge, multiple datasets were provided with different volumes (T1, T1ce, T2 and FLAIR):

- a training dataset of 285 patients preprocessed and with ground truth annotated [2–4],
- a dataset without public ground truth but with the possibility to evaluate online our method and obtain preliminary results,
- a final dataset without ground truth, used to rank the participants.

Our contribution is composed of two independent modules: one for tumor segmentation and one for survival time prediction. The tumor segmentation module (Sect. 2.1) blends ideas from two previous publications. It first builds on a work published in the IEEE Intl. Conf. on Image Processing (ICIP) in 2017 [22], which proposed to segment 3D brain MR volumes using fully convolutional network (FCN). It leveraged transfer learning thanks to a VGG network [18] pre-trained on the ImageNet dataset and later fine-tuned on the training set of the challenge. Its input were 2D color-like images composed of 3 consecutive slices of the 3D volume (see Fig. 1). This method used only one modality, and reached good results for brain segmentation with a decent speed. Based on this architecture, we incorporated the ideas of [24] which reused this architecture to take slices from several modalities as input. Our final segmentation solution provides specially designed pre- and post-processing dedicated to the challenge and makes use of both local 3D and multi-modal information. The survival prediction module (Sect. 2.2) we introduce here is based on Random Forests and relies on a very light training to cope with the limited amount of examples available in the challenge. Despite its apparent simplicity, it provides a reasonable survival time estimate as reported in the results (Sect. 3).

¹ <https://www.med.upenn.edu/sbia/brats2018.html>.

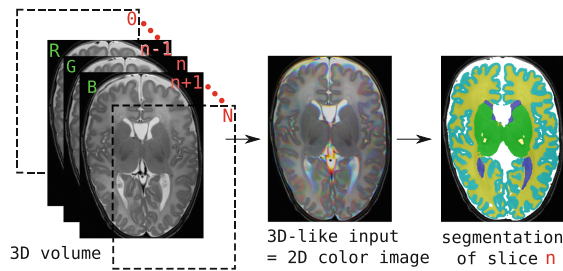


Fig. 1. Illustration of 3D-like color image and associated segmentation used in [22]. (Color figure online)

2 Method

This section describes the method we submitted to the MICCAI 2018 Multimodal Brain Tumor Segmentation Challenge (BraTS). As it was previously mentioned, it is composed of a tumor segmentation module and a survival prediction module, mapping the two tasks of the challenge.

2.1 Tumor Segmentation

An overview of the proposed segmentation method is given in Fig. 2. The method is fully automatic, and takes pseudo-3D images as input. It is really fast as about 10s are needed to process a complete volume with a GPU-equipped machine. It consists in three sub-stages: a data pre-processing, a deep network inference, and a segmentation post-processing.

Pre-processing. We first normalize the input data to fit in the range imposed by the original network (before fine tuning). Let n, m be respectively the minimum non-null and maximum gray-level value of the histogram. We requantize all voxel values using a linear function so that the gray-level range $[n, m]$ is mapped to $[-127, 127]$.

Then, as our inference network processes 2D color-like images (3 channels of 2D slices), the question amounts to how to prepare appropriate inputs given that a brain MR image is a 3D volume. Our second step is therefore to stack successive 2D slices: for each n^{th} slice of the volume to segment, we consider three images corresponding to the $(n - 1)^{\text{th}}$, n^{th} , and $(n + 1)^{\text{th}}$ slices of the original volume. These three gray-level 2D images are assembled to form a 2D color-like image (one image per channel). Each 2D color-like image thus forms a representation of a part (a small volume) of the MR volume. This image is the input of the FCN to obtain a 2D segmentation of the n^{th} slice. This process is depicted in Fig. 2 (left).

To combine information from different modalities, we complete this process. The n^{th} slice is taken from one modality and its $(n - 1)^{\text{th}}$ and $(n + 1)^{\text{th}}$ from another one. This combination brings not only 3D information but also multi-modality information. Figure 3 illustrates this variant.

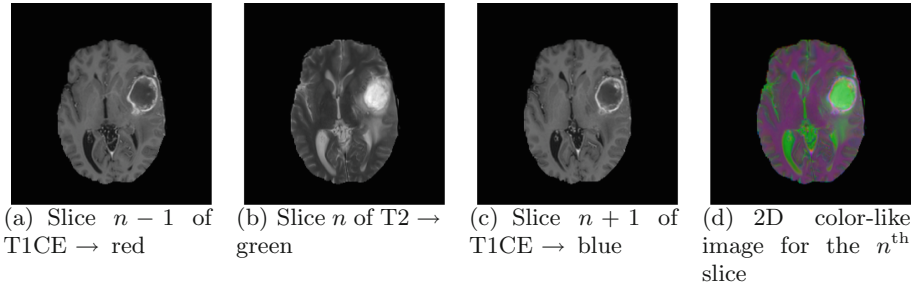


Fig. 3. Three successive slices (a–c) are used to build a 2D color-like image (d) from for example T1CE and T2 images. (Color figure online)

also very similar with the one used in [14] for retinal image analysis, where the retinal images are already 2D color-like images. Using such a 2D representation avoids the expensive computational and memory requirements of fully 3D FCN.

For the training phase, we use the multinomial logistic loss function for a one-of-many classification task, passing real-valued predictions through a softmax to get a probability distribution over classes. During training, we use the classical data augmentation strategy by scaling and rotating. We rely on the ADAM optimization procedure [10] (AMSGrad variant [17]) to minimize the loss of the network. The relevant parameters of the methods are the following: the learning rate is set to 0.002 (we did not use learning rate decay), the beta_1 and beta_2 are respectively set to 0.9 and 0.999, and we use a fuzz factor (epsilon) of 0.001.

At test time, after having pre-processed the 3D volume, we prepare the set of 2D color-like images and pass every image through the network. We run the train and test phase on an NVIDIA GPU. The testing one lasts less than 10s for a complete volume.

Post-processing. The output of the network for one slice during the inference phase is a 2D segmented slice. After treating all the slices of the volume, all the segmented slices are stacked to recover a 3D volume with the same shape as the initial volume, and containing only the segmented lesions.

This segmentation procedure is repeated three times as we slice the initial volume three times (along the three axis). We get three different segmentations and we merge them to get the final segmentation by a majority voting procedure.

Then, as a final step, we regularize the segmented volumes using a morphological closing to fill small holes lying within tumor regions.

2.2 Patient Survival Prediction

The second task of the MICCAI 2018 BraTS challenge is concerned with the prediction of patient overall survival from pre-operative scans (only for subjects with gross total resection (GTR) status). Note that, to comply with the evaluation framework, the classification procedure is conducted by labeling subjects

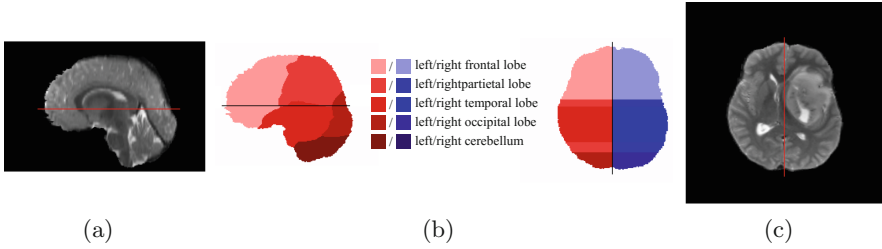


Fig. 4. (a) Sagittal and (c) axial slices from the T2 modality of a brain and (b) corresponding rescaled brain atlas.

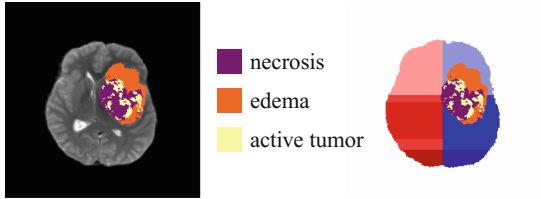





Fig. 5. T2 slice (left) and corresponding atlas slice (right) with segmented tumor overlaid. (Color figure online)

into three classes: short-survivors (less than 10 months), mid-survivors (between 10 and 15 months) and long-survivors (greater than 15 months).

Definition and Extraction of Relevant Features. The first step of the prediction task is the definition and extraction of relevant features impacting the survival of patients. Beside the patient age, we decided to focus on the tumor size and its localization within the brain. More specifically, we denote by S_i the segmented volume predicted by our Deep FCN architecture, as described in Sect. 2.1 for the i^{th} patient. Voxels in S_i are labeled by 1, 2 and 4 (corresponding to ,  and  in Fig. 5, respectively), depending whether they were classified as necrosis, edema or active tumor, respectively.

Thus, we define the relative size of each class in S_i with respect to the total brain size (the number of non-zero voxels in the patient T2 modality) as the features related to the tumor size.

In order to describe the tumor position, we created a crude brain atlas divided in 10 regions accounting for the frontal, parietal, temporal and occipital lobes and the cerebellum for each hemisphere, as displayed by Fig. 4(b). The 3D atlas was first shaped to the average bounding box dimensions of all patients with GTR status, *i.e.* $170 \times 140 \times 140$ pixels. It is then adjusted to each patient bounding box dimensions by nearest-neighbors interpolation, and finally masked by all non-zero voxels in the patient T2 modality. Finally, we retrieve the centroid coordinates of the region within the atlas that is affected the most by the necrosis

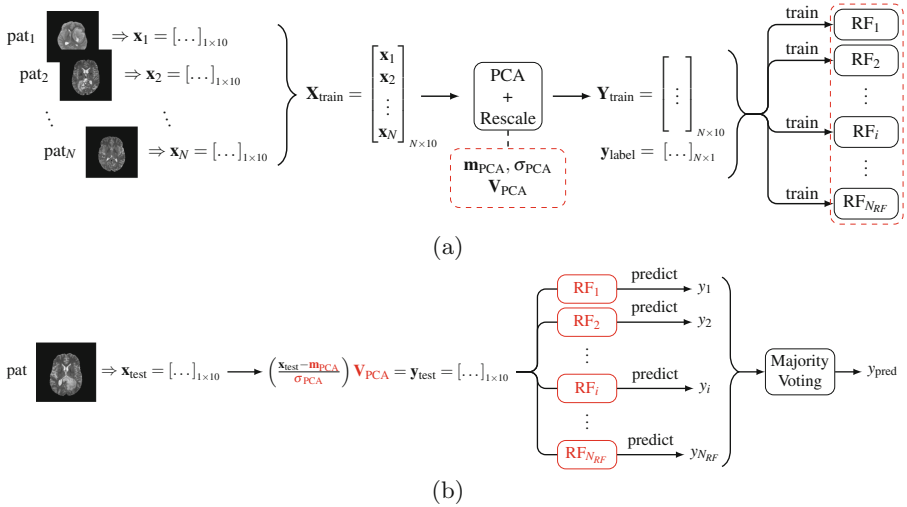


Fig. 6. (a) Training and (b) test procedures. The stored information after the training phase is encircled in dashed red in the training workflow (a). (Color figure online)

(i.e., the region that has the most voxels labeled as necrosis in S_i with respect to its own size) relatively to the brain bounding box as well as the relative centroid coordinates of the necrosis + active tumor and defined those as the relevant features accounting for the tumor position.

In summary, each patient is defined by the following 6 criteria:

1. the patient age.
2. the relative size of necrosis with respect to brain size.
3. the relative size of edema with respect to brain size.
4. the relative size of active tumor with respect to brain size.
5. the relative centroid coordinates of the region in the atlas that is the most affected by necrosis with respect to the brain bounding box.
6. the relative centroid coordinates of the binarized tumor (only considering necrosis and active tumor) with respect to the brain bounding box.

This leads to a total of **10 features per patient** (since both centroids coordinates are 3-dimensional).

Training Phase. For the training phase, we first extract the feature vector $\mathbf{x}_i \in \mathbb{R}^{10}$ of each of the N patients in the training set (with $N = 59$), as described in Sect. 2.2 above. All those feature vectors are stacked in a $N \times 10$ feature matrix $\mathbf{X}_{\text{train}}$ on which a principal component analysis (PCA) is performed. The feature-wise mean \mathbf{m}_{PCA} and standard deviation $\boldsymbol{\sigma}_{\text{PCA}}$ vectors computed during the scaling phase of the PCA, as well as the projection matrix \mathbf{V}_{PCA} are stored for further use. Finally, the PCA output is normalized again, yielding the $N \times 10$ matrix $\mathbf{Y}_{\text{train}}$. Finally, we train N_{RF} random forest (RF) classifiers [20]

on all rows of $\mathbf{Y}_{\text{train}}$, using the true label vector $\mathbf{y}_{\text{label}}$ as target values, and store those RFs. The whole training phase is depicted by the workflow in Fig. 6(a). Each RF is composed of 10 decision trees, for which splits are performed using 3 features randomly selected among the 10 available, and based on the Gini impurity criterion [19]. Here, we arbitrary fixed $N_{\text{RF}} = 50$ in order to account for the stochastic behavior of RF classifiers.

Test Phase. The test phase is summarized by the workflow in Fig. 6(b). Features are computed in a similar manner for a patient belonging to the test data set as they are for the training set. The feature vector \mathbf{x}_{test} is then normalized using \mathbf{m}_{PCA} and $\boldsymbol{\sigma}_{\text{PCA}}$ and further projected in the PC space with \mathbf{V}_{PCA} , learned during the PCA step of the training stage. The resulting vector \mathbf{y}_{test} is then fed to the N_{RF} RF classifiers, leading to N_{RF} independent class label predictions. The final label prediction y_{pred} (1, 2 and 3 for short-, mid- and long-survivors, respectively) is eventually obtained by majority voting.

3 Setup and Results

This section presents the setup of the experiments and results obtained during the development of our method (using the training dataset), and the final ranking during the challenge.

3.1 Setup and Experiments for Tumor Segmentation

In this part, we used only the training scans provided during the challenge.

Modalities. Instead of using only one modality to form the pseudo-3D color images (the input of the network), we formed multi-modality pseudo-3D images using T1ce and T2 modalities: for each slice n , we combined the slice n of T2 with the slices $n - 1$ and $n + 1$ from T1ce.

Axis and Combination. Our method deals with 2D color-like images that are pseudo-3D. To take advantage of the entire volume, we associated three networks, each network being trained on a particular axis (axial, sagittal and coronal), and combined their results to obtain the final segmentation. We trained 3 networks, one for each axis. The inference was done according to the axis, so for one volume we obtained 3 segmentation. These segmentations are then combined: for each voxel, the final segmentation is the result of the majority voting procedure.

Training and Testing. To train our model we select randomly 90% of scans from Brats challenge training dataset. The model was trained using the parameters described in the Sect. 2.1. We tested on the 10% remaining scans.

3.2 Results

Tumor Segmentation. The results on the training dataset evaluate to a dice of 0.82 for the whole tumor segmentation (evaluation on 10% of the training set). More precisely, we obtained for the 3 classes: 0.6 for the GD-enhancing tumor, 0.63 for the peritumoral edema, and 0.56 for the necrotic and non-enhancing tumor core. We did not achieved a good ranking during the challenge for this task. Precise results can be found in [5].

Survival Time Prediction. For the survival prediction task we obtained an accuracy of 0.54 on the training dataset. During the challenge, we obtained an accuracy of 0.61. This allowed us to reach the **2nd place of the challenge for the survival prediction task.**

3.3 Discussion

The prediction task is the final aim of the entire pipeline. The segmentation task is a basis for the prediction task but is not a finality. We developed the prediction procedure using the ground truth segmentations, so that our prediction method is independent from our segmentation method. We can notice that the prediction procedure can deal with precise segmentations (i.e. ground truth segmentations) and coarser ones (such as our segmentation results).

This is the main advantage of our prediction method as it does not require a lot of data: it relies on a coarse segmentation and a brain atlas. Simple descriptors are extracted to perform the prediction. A strong point which differentiates our method from others is that our method does not rely on a specific modality to work: the method relies on a segmentation result, regardless of how it has been obtained, and not directly on a modality. It can be used without the constraints of working on one modality or an other. Furthermore the segmentation does not need to be precise to permits the prediction as illustrated by our results during the challenge.

4 Conclusion

We proposed in this article a method to first segment glioma in few seconds based on transfer learning from VGG-16, a pre-trained network used to classify natural images, and then to predict the survival time of the patients. This segmentation method takes the advantage of keeping 3D information of the MRI volume and the speed of processing only 2D images, thanks to the pseudo-3D concept while the prediction method uses only a segmentation result and a homemade brain atlas.

This method can also deal with multi-modality, and can be applied to other segmentation problems, such as in [24], where a similar method is proposed to segment white matter hyperintensities, but pseudo-3D has been replaced by an association of multimodality and mathematical morphology pre-processing to

improve the detection of small lesions. Hence, we might also try to modify our inputs thanks to some highly non-linear filtering to help the network segment tumors, precisely some mathematical morphology operators [23].

The strength of this method is its modularity and its simplicity. It is easy to implement, fast, and does not need a huge amount of annotated data for training (in the work on brain segmentation [22], there is only 2 images for training for some cases).

From a segmentation result, we introduced a simple and efficient method to predict the patient overall survival, based on Random Forests. This method only needs as input a segmentation, a brain atlas and a brain volume for atlas registration. This method is not only fast; it is also easy to train with few samples and can be used after any tumor segmentation module.

Finally, we made a Docker image of the overall method publicly available at <https://www.lrde.epita.fr/wiki/NeoBrainSeg>.

Acknowledgments. The authors would like to thank the organizers of the BraTS 2018 Challenge and the MICCAI Brainles Workshop, and Dr. Marie Donzel from Claude Bernard University Lyon 1 medical school for the useful discussions regarding the definition of relevant brain features for the survival prediction. The GPU card “Quadro P6000” used for the work presented in this paper was donated by NVIDIA Corporation.

References

1. Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications. *Curr. Med. Imaging Rev.* **3**(4), 262–276 (2007)
2. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *Cancer Imaging Arch.* (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
3. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
4. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 170117 (2017)
5. Bakas, S., Reyes, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629* (2018)
6. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* **58**(13), R97 (2013)
7. Bonnín Rosselló, C.: Brain lesion segmentation using Convolutional Neuronal Networks. B.S. thesis, Universitat Politècnica de Catalunya (2018)
8. Eisenhauer, E.A., et al.: New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**(2), 228–247 (2009)
9. Holland, E.C.: Progenitor cells and glioma formation. *Curr. Opin. Neurol.* **14**(6), 683–688 (2001)

10. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR abs/1412.6980 (2014)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
12. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
13. Louis, D.N., et al.: The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**(2), 97–109 (2007)
14. Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L.: Deep retinal image understanding. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016. LNCS*, vol. 9901, pp. 140–148. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_17
15. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993 (2015)
16. Ohgaki, H., Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J. Neuropathol. Exp. Neurol.* **64**(6), 479–489 (2005)
17. Reddi, S.J., Kale, S., Kumar, S.: On the convergence of Adam and beyond. In: *International Conference on Learning Representations* (2018)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
19. Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T.: Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform.* **8**(1), 25 (2007)
20. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P.: Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **43**(6), 1947–1958 (2003)
21. Wen, P.Y., et al.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J. Clin. Oncol.* **28**(11), 1963–1972 (2010)
22. Xu, Y., Géraud, T., Bloch, I.: From neonatal to adult brain MR image segmentation in a few seconds using 3D-like fully convolutional network and transfer learning. In: *Proceedings of the 23rd IEEE International Conference on Image Processing (ICIP)*, Beijing, China, pp. 4417–4421, September 2017
23. Xu, Y., Géraud, T., Najman, L.: Connected filtering on tree-based shape-spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(6), 1126–1140 (2016)
24. Xu, Y., Géraud, T., Puybareau, É., Bloch, I., Chazalon, J.: White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) *BrainLes 2017. LNCS*, vol. 10670, pp. 501–514. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_42