Multimodal Analysis of Vasogenic Edema in Glioblastoma Patients for Radiotherapy Planning

Matthieu Lê[1,2], Hervé Delingette[1], Jayashree Kalpathy-Cramer[3], Elizabeth R. Gerstner[4], Helen A. Shih[2], Tracy Batchelor[4], Jan Unkelbach[2], and Nicholas Ayache[1]

[1] Asclepios Project, INRIA Sophia Antipolis, France
[2] Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[3] Martinos Center for Biomedical Imaging, Harvard–MIT Division of Health Sciences and Technology, Charlestown, MA, USA
[4] Department of Neurology Massachusetts General Hospital, Boston, Massachusetts, USA

Abstract. Glioblastoma (GBM) is the most common type of primary brain tumor, which is characterized by an infiltrative growth pattern. In current practice, radiotherapy planning is primarily based upon T2 FLAIR MRI despite its known lack of specificity in the detection of tumor infiltration. While hyperintensity on T2 FLAIR is widely considered to represent infiltrative tumor, it may also be caused by the presence of vasogenic edema (VE), caused by a leakage of fluid into the brain parenchyma. Distinguishing VE from infiltrative tumor could have impact on improving radiotherapy planning. In this paper we study a data set of 17 GBM patients treated with anti-angiogenic therapy for which a fast decrease of T2 FLAIR hypersignal is observed, which indicates the resolution of VE. We investigate if multimodal MRI acquisitions including diffusion tensor imaging can distinguish between VE and tumor infiltration prior to therapy. Using a random forest classifier, we show that, in this study, morphological information based on the contrast enhanced T1 image explains up to 75% of the extent of VE. The information from different imaging modalities did not significantly improve the classification. We then show that delineating the VE prior to therapy can have substantial impact on radiotherapy target delineation, leading to smaller treatment volumes and reducing potentially harmful radiation dose to normal brain tissue.

1 Introduction

Glioblastoma (GBM) is an infiltrative brain tumor whose cells invade the adjacent brain tissue which is only partially revealed by MRI [1]. Furthermore, the signal abnormality on T2 FLAIR and contrast enhanced T1 (T1Gd) images is only a surrogate for tumor invaded tissue but not per se indicative of the

presence of tumor cells. Indeed, the T2 FLAIR abnormality is a signal resulting from the combination of the bulk of the tumor, tumor cell infiltration, and vasogenic edema (VE) [2]. VE represents an increase in water content in the brain parenchyma, which is a consequence of the disruption of the blood brain barrier. Despite its unspecific nature, for lack of a more accurate tumor infiltration surrogate, clinicians use the T2 FLAIR abnormality to define the treatment volume for radiotherapy. Therefore, discarding VE from the T2 FLAIR abnormality could possibly result in a better surrogate for GBM cell infiltration and, as such, provide a better guide for radiotherapy planning by eliminating targeting of radiation to neighboring regions of normal uninvolved brain and other healthy tissues.

There has been a body of work on characterizing peritumoral edema through different imaging modalities including MRI and PET. For example, it has been investigated whether measures derived from diffusion tensor imaging (DTI) can distinguish between meningiomas and gliomas. The edema surrounding meningiomas is considered to be purely vasogenic while the edema surrounding gliomas is partly infiltrative tumor. It has been suggested that the relationship between the mean diffusivity (MD) and the fractional anisotropy (FA) can help distinguish between the two [3]. Axial and radial diffusivity (AD and RD respectively) were subsequently shown to be alternative markers [4]. At the same time, PET imaging using FDG or amino acid tracers proved to be quite successful in discriminating between meningiomas and gliomas [5]. In such diagnostic applications, the images as a whole are used for tumor classification or staging. However, for radiotherapy planning, the difficulty consists in locally delineating VE from tumor infiltration. To our knowledge, this is the first reported study with this aim.

A fundamental problem in this context is the definition of the VE ground truth. To this end, we use the response of the edema to anti-angiogenic therapy in a dataset containing 17 patients. The treatment is assumed to normalize blood vessels in the tumor, thereby restoring the blood brain barrier. As a consequence, this leads to the resolution of VE within a few weeks of treatment. Residual hyperintensity on T2 FLAIR is assumed to represent infiltrative tumor. The alternative to reliably define the ground truth would be to perform several biopsies, which would be invasive and impractical.

Based on multimodal imaging and the delineation of VE, feature selection and classification are performed to locally distinguish between VE and infiltrative tumor prior to therapy. More precisely, we consider the following features: standard MRI, morphological information (distance from the abnormalities), and DTI based information. We show that, within the approach taken in this work, morphological information is the most important input to define the VE. Surprisingly, DTI based features did not refine the classification. Finally, we show how a segmentation of the VE prior to therapy can change the radiotherapy plan: being able to detect VE prior to therapy leads to smaller treatment volumes and reduce potentially harmful radiation dose to normal brain tissue.

2 Material and Methods

Database. 40 patients were treated with the anti-angiogenic drug cediranib. Out of those 40, 17 were considered as responsive patients, i.e. the T2 FLAIR abnormality significantly shrinks during the early course of treatment. The remaining 23 patients were excluded because they appeared to not respond to the therapy. Figure 1 shows two responsive patients. The patient in the upper row shows extended edema surrounding the gross tumor which is almost completely resolved within 6 weeks. The patient in the bottom row shows both persistent hyperintensity (in the contralateral hemisphere) and resolved VE anterior to the gross tumor. For each patient, we have access to 5 structural images (T1, T1Gd, T1Gd High Resolution (T1 HR), T2, T2 FLAIR); in addition DTI was available from which we derived 4 images (FA, ADC, axial diffusivity (AD) and radial diffusivity (RD)) (Figure 3).

Pre-processing of the data. Each patient went through the following pipeline: bias field correction, rigid registration of the images on the pre-treatment T1Gd MRI, extraction of the brain and segmentation of the white matter, grey matter and cerebrospinal fluid. Structural images were normalized such that each modality has the same manually fixed mean intensity in the white matter tissue across patients.

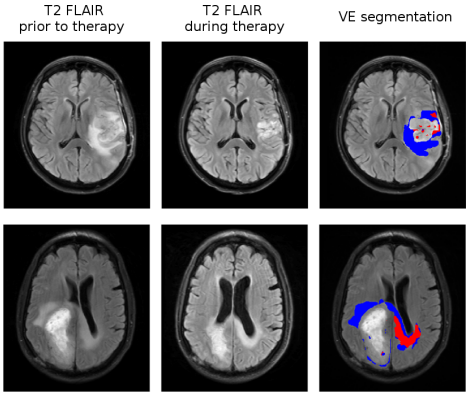

Fig. 1: Response to anti-angiogenic therapy for two patients. The smallest T2 FLAIR abnormality (middle) can be observed 42 days after the beginning of therapy for the first patient (first row) and 100 days for the second patient(second row). The VE corresponds to the responsive voxels (blue). The non responsive voxels are tumor related (red).

Definition of the ground truth. For each patient and for each acquisition, the T2 FLAIR and T1Gd abnormalities were manually segmented by clinicians. The imaging time points corresponding to the largest and smallest T2 FLAIR abnormalities were used to define the VE. The T1Gd abnormality was excluded from these volumes. VE was defined as the voxels included in the largest T2 FLAIR abnormality but not in the smallest T2 FLAIR abnormality. Accordingly, the class of non responsive voxels was defined as all voxels that are within

the T2 FLAIR abnormality at both time points. (Figure 1).

Feature Definition. For each voxel, we define 56 features. We use two morphological features which are defined as the logit function of the signed distance from the pre-treatment T2 FLAIR and T1Gd abnormalities (log-odds map) [6]. In addition, we use 5 structural images (T1, T1Gd, T1 HR, T2, T2 FLAIR) and 4 DTI based images (FA, MD, RD, AD). For each of these 9 images, we derive 6 features: the image intensity, two Gaussian convolutions at two different scales as well as the fractional anisotropy, the mean diffusivity and the determinant of the structure tensor of the original image.

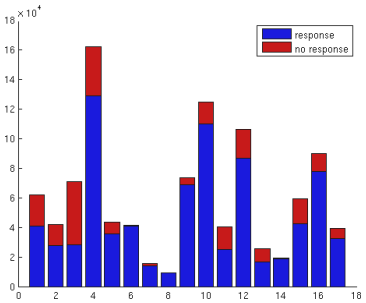

Fig. 2: The distribution of responsive (blue) and non responsive (red) voxels among the 17 selected patients

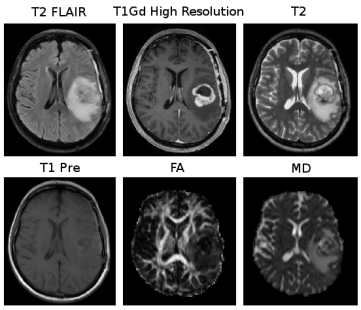

Fig. 3: Selected MRI modalities used to define the features.

Experiments. We want to analyze the performance of a classification algorithm in three different conditions: using all the features, using only morphological information, and using only DTI based information. For these three experiments, a random forest (RF) classifier [7] was used for the final classification. The design of the experiments differs in the feature selection prior to training the RF:

Experiment 1. An l_1-penalized support vector machine (l_1-SVM) was trained on a small bootstrapped training sample for a repeated number of times using the 56 features. The features that were selected by the l_1-SVM at every iteration were used for the classification. The regularization term of the l_1-SVM was set such that 10 features were selected in the end.

Experiment 2. We restricted ourselves to using only the morphological information, i.e. the 2 log-odds based features. In order to compare the results with the first experiment, 8 among the 54 remaining features were randomly selected. Those features were then randomized: within one feature, we randomly permuted the value of this feature among the different samples.

Experiment 3. The 24 different DTI based features were used.

The training data set is imbalanced as 80% of voxels belong to the VE class, while 20% are non responsive (Figure 2). We drew an equal number of samples from each class to re-balance the data set for the l_1-SVM. For the RF, the

minority class (non responsive voxels) was oversampled using bootstrap while keeping the dominant class untouched. This ensures that the classification error do not lean toward the minority class [8]. For each experiment, a leave-one-out approach was used by running the experiments 17 times, each time leaving one patient out of the training set.

Radiotherapy planning. We compare the radiotherapy plan based on the initial T2 FLAIR image with the plan that discards VE from the tumor delineation. 9 equally spaced coplanar photon beams were used to compute the plans for intensity modulated radiotherapy (IMRT). Dose calculation was performed with CERR [9], and an L-BFGS quasi-newton method was used to optimized the IMRT treatment plans [10].

3 Results

3.1 Comparison between the Three Experiments

ROC curves were computed by changing the voting threshold of the prediction output of the RF. The areas under the curve were then averaged over all the iterations of the leave-one-out process (Figure 4). The training set corresponds to 16 patients while the testing set is the left out patient. When considering the training set, the first experiment yields the best results since it involves all features. However, on the testing set, experiment 2 yields comparable results with an AUC of 0.75, compared to an AUC of 0.77 for experiment 1. The third experiment involving only DTI based features yields the worst results with an AUC of 0.68 and 0.54 for the training and testing sets, respectively.

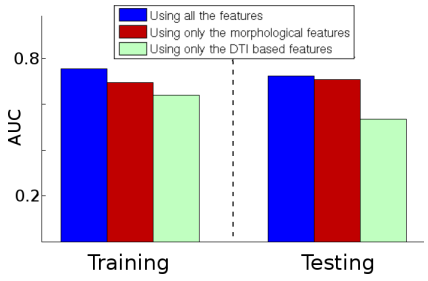

Fig. 4: Area under the ROC curve for the three different experiments for the training and testing sets.

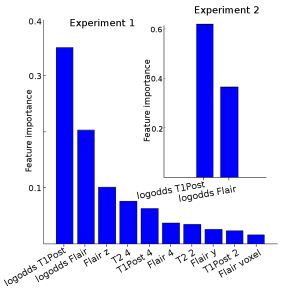

Fig. 5: Feature importance measured as the mean decrease impurity of the RF for all the features (left) and only the morphological features (right).

The morphological information based on the log-odds features gives results comparable to using all 56 features defined via multimodal MRI images. This is supported through Figure 5 which shows the feature importance for the RF

training as the mean decrease impurity [11] at each node for experiment 1 and 2. It appears that, regardless of using all the features, the log-odds are largely dominant. More specifically, the distance from the T1Gd abnormality seems to be the single best feature to segment the VE.

To obtain a final segmentation, the threshold for the RF is selected by imposing a cost of 2 for predicting tumor infiltration as VE and a cost of 1 for predicting VE as tumor infiltration was set. This reflects the idea that the radiotherapy target should be conservative and enclose all the tumor infiltration. DICE coefficient between RF segmentations and ground truth were calculated (Figure 6). This confirms the previous findings that DTI alone yields poor results with a mean DICE of 0.48. Using the RF with all features gives similar results to using solely the morphological features (mean DICE of 0.63 and 0.64).

For comparison, we analyzed how well the distance from the T1Gd abnormality alone could be used to delineate the VE. To that end, we computed the median distance that encloses 90% of the persistent T2 FLAIR abnormality, which yields 4.8 mm for this subset of patients. VE was then defined as the voxels in the T2 FLAIR abnormality that are further than 4.8 mm away from the T1Gd abnormality. The mean DICE coefficient for this segmentation is 0.75, i.e. it outperforms the RF (Figure 6). Figure 7 shows the segmentation of the RF using all the features, and the segmentation based on the distance from the T1Gd abnormality for one of the patients. Using the distance from the T1Gd abnormality gives overall good results. The RF output yields irregular contours and fails to improve on the distance based segmentation. It should be noted that, while the distance from the T1Gd abnormality yields comparatively good results on average, it fails to identify persistent hyperintensity in some patients. For the patient shown in the bottom row of Figure 1, the hyperintensity in the contralateral hemisphere is classified as VE by this method.

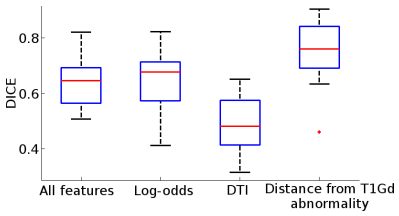

Fig. 6: Distribution of the DICE coefficient for the four final VE segmentations.

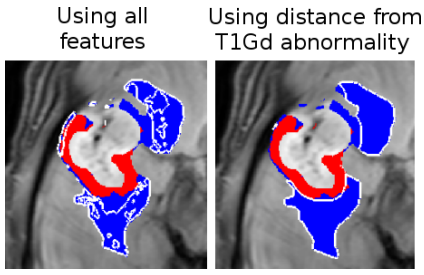

Fig. 7: Segmentation of VE using the RF with all features (left) and the distance based criteria (right). The ground truth for tumor infiltration is shown in red, VE in blue. The prediction is outlined in white.

One limitation of our work is the inability to decipher whether persistent T2 FLAIR hyperintensity represents persistent disease, or VE without disease that

was unresponsive to anti-angiogenic therapy. In our study, these patients were excluded, which introduced a patient selection bias.

3.2 Application to radiotherapy

In current clinical practice, the clinical target volume (CTV1) for radiotherapy is often defined as a 2-2.5 cm isotropic expansion of the T2 FLAIR abnormality. Using the above result, we defined an alternative CTV2 based on a 2.48 cm isotropic expansion of the T1Gd abnormality. Such a target would enclose the tumor infiltration with the 4.8 mm margin, to which we add a 2 cm expansion. The CTV2 is then solely based on the delineation of the T1Gd abnormality. IMRT plans based on the two targets have been calculated for the patient in figure 8.

The patient shows extensive edema, which extends posteriorly more than anteriorly (8, left). This leads to a large CTV1 and a high dose delivered to most of the left parietal lobe (8, middle). The residual T2 FLAIR abnormality after resolution of the VE is located more symmetrically around the initial T1Gd abnormality. This leads to a CTV2 that extends less far posteriorly, which translates into a lower dose delivered to posterior region of the parietal lobe. The dose difference plot (8, right) shows a dose reduction of more than 30 Gy in this region. In total, the plan based on CTV2 delivers 24% less dose to the brain. Considering the reduction of the T2 FLAIR abnormality after the resolution of VE, the inclusion of almost the entire left parietal lobe in CTV1 does not seem warranted.

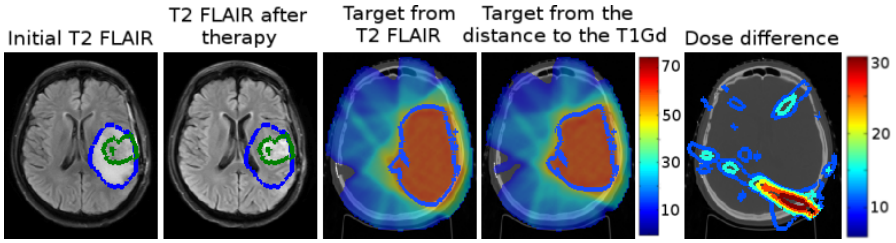

Fig. 8: Radiotherapy plans for one patient. The initial (blue) and residual (green) T2 FLAIR abnormalities are outlined on the initial T2 FLAIR (first image) and the T2 FLAIR after radiotherapy (second image). The dose distributions (in Gy) based on CTV1 and CTV2 are overlaid on the CT image (third and fourth images). The target is outlined in blue. The dose difference between the two plans is shown on the right.

4 Discussion

Many GBM patients present extensive T2 FLAIR hyperintensity on brain imaging that is known in part to represent peritumoral edema and to less well understood extent, represent infiltrative tumor cells. Excluding VE that does not

harbor tumor cells from the RT target delineation would enable reduction in target volumes and with the potential of exposing less radiation to surrounding normal brain tissue. This could lead to less toxicity and may leave more opportunity for a possible re-irradiation after recurrence. Our dataset of patient treated with anti-angiogenic therapy shows that a substantial part of the T2 FLAIR hyperintensity disappears during the initial weeks of treatment, providing support for the idea of excluding parts of the T2 FLAIR abnormality from the delineation of gross disease. We investigated if multimodal MR imaging can identify VE prior to RT. In our approach, the distance from the contrast enhancing tumor is the single best feature to segment the VE, reflecting the observation that for most patients, infiltrative tumor is adjacent to the T1Gd abnormality. Improving on this distance based segmentation is a difficult task. DTI measures and image intensity features did not yield an improvement in conformal tumor target definition. Future improvements may be possible by including MR spectroscopy data and more contextual as well as texture features.

Acknowledgements

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080.

References

1. Kelly, P.J.: Computed tomography and histologic limits in glial neoplasms: tumor types and selection for volumetric resection. Surg. neurology **39**(6) (1993) 458–465
2. Coons, S.: Anatomy and growth patters of diffuse gliomas. In: The gliomas. W.B. Saunders Company, Philadelphia, PA, USA (1999) 210–225
3. Lu, S., et al.: Dti of intracranial neoplasia and associated peritumoral edema: introduction of the tumor infiltration index. Radiology **232**(1) (2004) 221–228
4. Min, Z.G., et al.: Differentiation of pure vasogenic edema and tumor-infiltrated edema in patients with peritumoral edema by analyzing the relationship of axial and radial diffusivities on 3.0 t mri. Clinical Neurology and Neurosurgery (2013)
5. Kinoshita, M., et al.: Imaging 18f-fluorodeoxy glucose/11c-methionine uptake decoupling for identification of tumor cell infiltration in peritumoral brain edema. Journal of Neuro-Oncology **106**(2) (2012) 417–425
6. Pohl, K., et al.: Logarithm odds maps for shape representation. In: MICCAI 2006. Springer (2006) 955–963
7. Pedregosa, F., et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
8. Zhang, D., Tsai, J.J.: Advances in machine learning applications in software engineering. Igi Global (2007)
9. Deasy, J.O., Blanco, A.I., Clark, V.H.: Cerr: a computational environment for radiotherapy research. Medical physics **30**(5) (2003) 979–985
10. Unkelbach, J., et al.: Radiotherapy planning for glioblastoma based on a tumor growth model: improving target volume delineation. Physics in Medicine and Biology **59**(3) (2014) 747–770
11. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and regression trees. CRC press (1984)