Comparison of manual and semi-manual delineations for classifying glioblastoma multiforme patients based on histogram and texture MRI features

Adrian Ion-Mărgineanu [1,2,*], Sofie Van Cauter[3], Diana M Sima[1,2], Frederik Maes[4], Stefan Sunaert[3], Uwe Himmelreich[5] and Sabine Van Huffel[1,2 *]

1- KU Leuven - ESAT - STADIUS, Leuven, Belgium

2- iMinds - Medical IT, Belgium

3- University Hospitals of Leuven, Department of Radiology, Leuven, Belgium

4- KU Leuven - ESAT - PSI, Leuven, Belgium

5- KU Leuven, Biomedical MRI/MoSAIC, Department of Imaging and Pathology, Leuven, Belgium

Abstract. In this paper we study the task of classifying the follow-up course of brain tumour patients that had surgery. Multiple magnetic resonance imaging brain scans were taken for each patient. We propose a simple method of delineating the contrast enhancing tumour lesion based on the total tumour region. We compare balanced accuracy values after tuning SVM-lin and SVM-rbf on histogram and 3-D texture features extracted from semi-manual and manual delineations. Results show that our proposed delineating method outperforms the classical method.

1 Introduction

Glioblastoma multiforme (GBM) is the most common and malignant intracranial tumour, representing as much as 30% of primary brain tumours. The patients have a median survival of only 10 to 14 months after diagnosis with only 3 to 5% of patients surviving more than two years. Recurrence is universal, and at the time of relapse, the median survival is only five to seven months despite therapy. The current standard of care is surgical resection followed by radiotherapy and concomitant adjuvant temozolomide chemotherapy. Magnetic resonance imaging (MRI) is the most widely used medical imaging technique for identifying the location and size of brain tumours. However, conventional MRI (cMRI) has a limited specificity in determining the underlying type and grade of the brain tumour [4]. More advanced techniques like perfusion weighted MRI (PWI) and

*This work was funded by: FWO G.0869.12N; Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, 'Dynamical systems, control and optimization', 2012-2017); EU MC ITN TRANSACT 2012 (no. 316679) and the ERC Advanced Grant BIOTENSORS nr.339804. EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

diffusion kurtosis MRI (DKI) might improve the physiological information of brain tumours [13].

The goal of this study is to find a map between multi-parametric MRI data acquired during the follow-up of GBM patients and the relapse of brain tumour after surgery, as described by the clinically accepted Response Assessment in Neuro-Oncology (RANO) criteria [14]. In a previous study done on 18 patients [6], we could differentiate between progressive and responsive follow-up GBM patients with 100% accuracy one month before the patients were labelled, based only on PWI features. In this study we extend our analysis in three directions: (1) we increase the dataset to 29 patients, (2) we extract multiple histogram and texture features, (3) we present a simple method to semi-manually delineate contrast enhancing regions of interest (CE ROIs) and compare it to the manually delineated CE ROIs.

2 Materials and methods

2.1 Patient population

Twenty-nine patients were included in this study, out of which sixteen patients were treated according to the HGG-IMMUNO-2003 protocol, and thirteen patients were treated according to the HGG-IMMUNO-2010 protocol [12]. The local ethics committee approved this study and informed consent was obtained from every patient before the first imaging time point. Based on radiological evaluation of the follow-up MRI scans using the current guidelines for response assessment of high grade glioma [14], each patient was assigned to one of two clinical groups: *progressive disease* or *complete response*. We consider that prior to this assignment, patients are not categorized to any definite form, therefore their MRI acquisitions are unlabelled, and after the assignment their MRI acquisitions are labelled either as 'progressive' or 'responsive'.

2.2 MRI acquisition and processing

The MR images were acquired on a clinical 3 Tesla MR imaging system (Philips Achieva, Best, The Netherlands), using a body coil for transmission and a 32-channel head coil for signal reception. The imaging protocol consisted of cMRI, PWI, and DKI, and their acquisition parameters can be found in [11]. Four types of conventional MR images were acquired : T1, T2, T1 post contrast (T1pc), and FLAIR. PWI were processed using the DSCoMAN plugin [1] for ImageJ (http://rsb.info.nih.gov/ij/). For each PWI acquisition, seven parameter maps were extracted: corrected cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), time to peak (TTP), R^2 (Rsquare), a blood volume correction parameter K1, and a leakage correction parameter K2. For each DKI acquisition, seven parameters maps were derived as described in [11]: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), mean kurtosis (MK), axial kurtosis (AK), and radial kurtosis (RD).

2.3 Manual and semi-manual delineations

Three ROIs were manually drawn by a radiologist with 5 years experience: first one around the Total tumour lesion, avoiding areas of necrosis or cystic components such as the surgical cavity; second one around the contrast enhancing (CE) lesion inside the tumour; third one around the contra-lateral normal appearing white matter (NAWM) to standardize measurements extracted from the tumour region. By substracting the CE ROI from the Total ROI, the non-enhancing (NE) region of the tumour was obtained.

To automatically split the Total region in these two ROIs (CE&NE), a threshold was set at the 90th percentile of T1pc Total voxels. In this way, two semi-manual ROIs were made for each patient based on the T1pc intensities selected from Total: one containing very high T1pc intensity Total voxels, and another one containing the rest of Total voxels. The 90th percentile threshold was selected after visually inspecting T1pc maps of multiple patients. Figure 1 shows a comparison between two modalities of delineating CE&NE for a randomly selected patient.

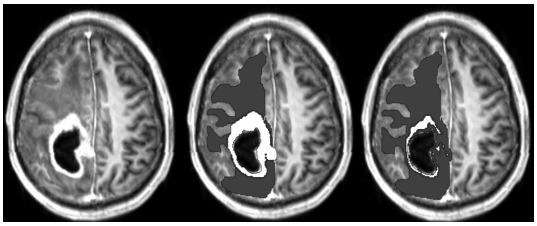

Fig. 1: Left - T1pc. Center - Manual delineations on top of T1pc. Right - Semi-manual delineations on top of T1pc. CE is white and NE is dark gray.

2.4 Feature extraction

Co-registration of all 17 MR parameter maps (3 cMRI, 7 PWI, 7 DKI) to T1pc was done in two steps: first, skull-stripping was performed on the raw MR images using FSL-BET with default parameters [9], and then an affine co-registration of the skull-stripped images to T1pc was done using NiftyReg [7] with default parameters.

After co-registering all maps to T1pc, the four ROIs (manual CE&ED and semi-manual CE&ED) are used as separate 3-D masks on each map to extract histogram and texture features. Voxel intensities inside each mask are normalized to the average value computed from the corresponding NAWM ROI. Six histogram measures are computed for each mask and map: mean, coefficient of variation, 90th percentile, 10th percentile, skewness, and kurtosis. Texture features are extracted from the 3-D Gray Level Co-occurring Matrix (GLCM). To compute the GLCM, each map has been rescaled such that the voxel intensities are integers varying from 1 to 32. The computation was done using the function *graycomatrix* (Matlab R2015a, MathWorks, Massachusetts, U.S.A.) with

distance set to 1, the *'Symmetric'* flag set to *true*, and 4 values of *'Offset'* set to four directions: 0°, 45°, 90°, and 135°. Twenty 3-D texture features, as described in [5, 10, 3], were extracted from GLCM: autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares: variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation 1 and 2, inverse difference normalized, and inverse difference moment normalized.

In the end 468 features are extracted from each ROI: 26 histogram and texture features for each of the 18 maps.

2.5 Datasets comparison

In total there are 183 time points, 56 are labelled and 127 are unlabelled. We perform our analysis on two datasets containing only labelled data points.

The first set contains 43 labelled data points that have manual delineations and complete MRI acquisitions. In this set there are 9 data points from responsive patients, and 34 data points from progressive patients.

The second dataset contains 55 labelled data points, 21 from responsive patients and 34 from progressive patients. Compared to the first dataset, the supplementary 12 points were imputed using our semi-manually delineating method only from responsive patients, because there was no contrast enhancement visible to the radiologist in the Total ROI.

Therefore, we train classifiers on the first dataset with features extracted from manual and semi-manual delineations, while on the second dataset we train classifiers with features extracted only from semi-manual delineations.

2.6 Classification approach

In order to have robust classification results and practical clinical interpretation, a leave-one-patient-out cross-validation (LOPO-CV) setup is preferred. In this way 29 separate folds are created in which the test patient is independent of the training patients: data points from one patient are considered test points, while data points from the remaining 28 patients are used for training a classifier.

Because of the large number of features, random forests [2] are used in each fold to rank features based on their mean decrease in Gini. Classifiers will be trained on the best ranked five features, and then predict the label of test data points. The predicted labels are compared to the true labels by measuring the balanced accuracy rate (BAR) on all points from a dataset.

We tuned Support Vector Machines (SVM) with linear (lin) and radial basis function (rbf) kernel, over a grid-search with 5-fold internal CV. We tune the misclassification cost (C) parameter of SVM-lin and both C and γ parameters of SVM-rbf by searching through a logarithmic grid between 0.00001 and 100000. This process was done in Python 2.7.11 with scikit-learn 0.17.1 [8]. To automatically adjust for class unbalance, we set the *class_weight* parameter to *balanced*.

3 Results and Discussion

Table 1 shows BAR values after tuning SVM-lin and SVM-rbf on the two previously described datasets, one with 43 points and one with 55 points, with feature extracted from manual and semi-manual delineations. Eight separate subsets of MR features were used for training classifiers to analyze which combination of acquisition modality (cMRI, PWI, DKI, all together) and feature extraction method (histogram or texture) performs best. The highest BAR values (93-94) are obtained after training SVM-lin on cMRI or CPD texture features extracted from semi-manually delineated CE&NE.

BAR (%)		Histogram				Texture			
		C	P	D	CPD	C	P	D	CPD
MD (43p)	SVM-lin	62	36	42	55	42	72	44	62
	SVM-rbf	55	49	47	51	35	56	62	54
SMD (43p)	SVM-lin	76	43	41	70	94	70	79	93
	SVM-rbf	76	45	44	52	75	52	72	63
SMD (55p)	SVM-lin	78	81	80	86	69	76	74	65
	SVM-rbf	78	85	59	83	64	84	61	59

Table 1: BAR values between 80 and 90 have a light gray background, while BAR values higher than 90 have a medium gray background. Abbreviations: MD - manual delineations; SMD - semi-manual delineations; C - conventional MRI features; P - perfusion MRI features; D - diffusion MRI features; CPD - conventional, perfusion, and diffusion MRI features.

BAR values obtained after training classifiers on the first dataset suggest two ideas: (1) overall, features extracted from semi-manual delineations are better than features extracted from manual delineations for discriminating progressive from responsive GBM patients, and (2) texture features are better than histogram features at the same discrimination problem.

BAR values obtained after training classifiers on the second dataset with histogram features are always higher than BAR values obtained after training classifiers on the first dataset with histogram features. Although the same is not true for texture features, after training SVM-rbf on PWI texture features we obtain a BAR value of 84, which is comparable to 86, the highest BAR value after training classifiers on CPD histogram features. This finding suggests that our method of imputing MR features based on semi-manual delineations is helpful when training classifiers on histogram features, but it may not be suitable for texture features.

4 Conclusions

In this paper we studied the classification between responsive and progressive GBM patients based on multi-parametric MRI. We proposed a simple method

of delineating contrast enhancing ROIs starting from the total tumour ROI. We showed that classifiers trained on features extracted from our delineations perform better than classifiers trained on features extracted from manual delineations. We compared histogram to texture features and we obtained the maximum BAR using cMRI texture features. We extended the first dataset by imputing data points using our semi-manually delineating method and we show that histogram features yield consistently higher BAR values than texture features do.

References

[1] J. Boxerman, K. Schmainda, and R. Weisskoff. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *American Journal of Neuroradiology*, 27(4):859–867, 2006.

[2] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[3] D. A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[4] B. L. Dean, B. P. Drayer, C. R. Bird, R. A. Flom, J. A. Hodak, S. W. Coons, and R. G. Carey. Gliomas: classification with MR imaging. *Radiology*, 174(2):411–415, 1990.

[5] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6):610–621, 1973.

[6] A. Ion-Margineanu, S. Van Cauter, D. M. Sima, F. Maes, S. W. Van Gool, S. Sunaert, U. Himmelreich, and S. Van Huffel. Tumour relapse prediction using multiparametric mr data recorded during follow-up of GBM patients. *BioMed research international*, 2015.

[7] S. Ourselin, R. Stefanescu, and X. Pennec. *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part II*, chapter Robust Registration of Multi-modal Images: Towards Real-Time Clinical Applications, pages 140–147. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[9] S. M. Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, 2002.

[10] L.-K. Soh and C. Tsatsoulis. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(2):780–795, 1999.

[11] S. Van Cauter, F. De Keyzer, D. M. Sima, A. C. Sava, F. D'Arco, J. Veraart, R. R. Peeters, A. Leemans, S. Van Gool, G. Wilms, et al. Integrating diffusion kurtosis imaging, dynamic susceptibility-weighted contrast-enhanced mri, and short echo time chemical shift imaging for grading gliomas. *Neuro-oncology*, 16(7):1010–1021, 2014.

[12] S. Van Gool, W. Maes, H. Ardon, T. Verschuere, S. Van Cauter, and S. De Vleeschouwer. Dendritic cell therapy of high-grade gliomas. *Brain Pathology*, 19(4):694–712, 2009.

[13] M. Vrabec, S. Van Cauter, U. Himmelreich, S. W. Van Gool, S. Sunaert, S. De Vleeschouwer, D. Šuput, and P. Demaerel. MR perfusion and diffusion imaging in the follow-up of recurrent glioblastoma treated with dendritic cell immunotherapy: a pilot study. *Neuroradiology*, 53(10):721–731, 2011.

[14] P. Y. Wen, D. R. Macdonald, D. A. Reardon, T. F. Cloughesy, A. G. Sorensen, E. Galanis, J. DeGroot, W. Wick, M. R. Gilbert, A. B. Lassman, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of Clinical Oncology*, 28(11):1963–1972, 2010.