Tumour relapse prediction using multi-parametric MR data recorded during follow-up of GBM patients

Adrian Ion-Margineanu[1,2], Sofie Van Cauter[3], Diana M Sima[1,2], Frederik Maes[2,4], Stefaan W Van Gool[5], Stefaan Sunaert[3], Uwe Himmelreich[6], and Sabine Van Huffel[1,2]
[1]STADIUS, KU Leuven - ESAT, Leuven, Belgium, Belgium, [2]iMinds Medical IT, Leuven, Belgium, [3]Department of Radiology, University Hospitals of Leuven, Leuven, Belgium, [4]PSI, KU Leuven - ESAT, Belgium, [5]Department of Pedriatic Neuro-oncology, University Hospitals of Leuven, Belgium, [6]Department of Imaging and Pathology - Biomedical MRI/ MoSAIC, KU Leuven, Belgium

Target audience: Researchers and doctors that use multi-parametric magnetic resonance imaging (MRI) data for brain tumor follow-up.

Purpose: The focus of our study is finding a relation between the multi-parametric MR data acquired during the follow-up of glioblastoma multiforme (GBM) patients and the relapse of the brain tumour after surgery, as described by the clinically accepted RANO criteria. Conventional MRI (cMRI) has a limited specificity in determining the underlying type of brain tumour and tumour grade. More advanced MR techniques like diffusion-kurtosis MRI (DKI), perfusion-weighted MRI (PWI), and MR spectroscopy are promising in the characterization of brain tumours as they give potentially more physiological information.

Methods: ***Acquisition:*** Twenty-nine GBM patients who underwent surgery were scanned using a 3 Tesla MRI unit (Philips Achieva, Best, The Netherlands). The protocol consisted of conventional imaging (T1-weighted MRI after contrast administration, T2-weighted MRI and FLAIR (fluid attenuated inversion recovery) MRI) and advanced imaging (DKI, dynamic-susceptibility weighted contrast (DSC) - MRI, and chemical shift imaging (CSI)). High resolution anatomical reference images were acquired with T1-weighted 3D spoiled gradient echo scan with fast field echo, TR/TE: 9.7/4.6 msec, flip angle: 8°, turbo field echo factor: 180, acquisition voxel size: 0.98x0.98x1 mm^3, 118 contiguous partitions, inversion time: 900 msec. Axial spin echo T2 images were acquired with TR/TE: 3000/80 msec, slice/gap: 4/1 mm, field of view (FOV): 230x184 mm^2, turbo factor: 10, acquisition matrix: 400x300. Axial FLAIR images were acquired with TR/TE/IR: 11000/120/2800 msec, slice/gap: 4/1 mm, FOV: 230x184 mm^2, acquisition matrix: 240x134. Regions of interest (ROI) were manually drawn around the solid contrast-enhancing (CE) region and the entire lesion (TO), while the ROI containing perilesional oedema (ED) was obtained by subtracting CE from TO. Another ROI was drawn around the contralateral normal appearing white matter (NAWM) to standardize the hemodynamic measurements of PWI. DKI data were acquired with a spin-echo echo-planar-imaging diffusion weighted imaging (SE-EPI-DWI) sequence with TR/TE: 3200/90 msec, δ/Δ: 20/48.3 msec, FOV: 240x240 mm^2, matrix: 96x96, 1 signal average acquired, section thickness/gap: 2.5/0, b-values: 700, 1000, and 2800 sec/mm^2 in 25, 40 and 75 uniformly distributed directions. Additionally, 10 images without diffusion sensitization (b=0) were obtained. The DKI data were processed as described in [1]. Fractional anisotropy (FA), mean diffusivity (MD) and mean kurtosis (MK) were derived from the tensors. After a nonlinear registration of the parameter maps to the anatomical MR images, average values of MK, MD and FA were determined in the CE and ED regions. PWI was obtained with a DSC-MRI protocol consisting of a gradient echo-EPI sequence, TR/TE: 1350/30 msec, section thickness/gap: 3/0 mm, dynamic scans: 60, FOV: 200x200 mm^2, matrix: 112x109. PWI data were analyzed using DPTools (www.fmritools.org), as described in [1]. The average values of the considered perfusion parameters were computed in the CE, ED, and NAWM regions. We compute relative regional Cerebral Blood Volume (rrCBV), relative regional Cerebral Blood Flow (rrCBF) and relative Decrease Ratio (rDR) by using the corresponding parameter value in the NAWM region as internal reference. A 2D-CSI short echo time protocol was used as validated in [2]. MR spectra were processed using the MATLAB 2010b environment (MathWorks, Massachusetts, U.S.A.) with SPID graphical user interface, as described in detail in [2]. Nine metabolites were quantified using AQSES-MRSI: N-acetyl aspartate (NAA), glutamine (Gln), glutamate (Glu), total creatine (Cre), phosphorylcholine (PCh), glycerophosphorylcholine (GPC), myo-inositol (Myo), and lipids (Lips) at 0.9 and 1.3ppm, referred to as Lip1 and Lip2 respectively. Glu+Gln and PCh+GPC were reported as Glx and tCho (total choline), respectively. Good quality voxels were selected in the CE region based on Cramer-Rao Lower Bounds and spectral quality. The following metabolite ratios are reported: NAA/tCho, NAA/sum, tCho/sum, NAA/Cre, Lips/tCho, tCho/Cre, Myo/sum, Cre/sum, Lips/Cre and Glx/sum.

Machine Learning: In total, from 29 patients, we have 178 data points and each of these has 27 features (3 volumes from cMRI, 6 from PWI, 6 from DKI, 10 from CSI and 2 parameters for indicating total resection of the tumor and another one for describing the treatment). After quality control for each advanced modality, we removed 30% of PWI data, 44% of DKI data and 66% of CSI data. This drop-out of data resulted in a subset of 18 patients with 45 measurements with complete features. We developed an imputation method to fill in the missing data. For each patient a label (i.e. responsive treatment or progressive disease) has been put by the doctors according to the RANO criteria. Each patient has been scanned at least twice, so we can divide every patient's time points into two classes: before the label was put (unlabeled data) and after the label was put (labeled data). We used several supervised classifiers (k-Nearest Neighbours (kNN), diagonal Linear Discriminant Analysis (dLDA), Support Vector Machine (SVM), Least Squares SVM (LSSVM), Random forests (RF), Classification Tree (CT), Boost ensembles, Neural networks (NN)) with the goal of testing whether the unlabeled data could have been reliably labeled before the actual labeling was performed in the clinic. We test these classifiers on all features, but also on subsets of features pertaining to a single advanced MR modality (PWI, DKI, CSI). We used a leave-one-patient-out testing method, where all time points that belong to the test patient are, in turn, excluded when training the classifier so that they can be used as test data. We compute the balanced error rate (BER) at each time point, using the assigned label as expected label for all time points of a patient. For each classifier we have a total of 17 time points, because there are patients with up to 6 time points after the labeling point and there are others with up to 11 time points before the labeling point. The overall performance of each classifier is reported as a weighted average of the BER values obtained at all time points, weighted BER (wBER).

Table 1. BER values before and after the labeling time point for the best six classifiers using complete features.

BER	RF	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+2	0	0	0	0	0	0
L+1	0	0	0	0	0	0
L	0	0.1	0.217	0	0	0.1
L-1	0	0.125	0	0	0	0.125
L-2	0.25	0.25	0.5	0.25	0.25	0.25

$$wBER = \frac{\sum W_i^p \cdot W_i^t \cdot BER_i}{\sum W_i^p \cdot W_i^t}$$

$$W_i^t = 1, i \geq \text{labeling time point}$$

$$W_i^t = 1 - \frac{0.5}{11} \cdot i, i < \text{labeling time point}$$

$$W_i^p = \frac{\text{Number of patients at time point } i}{\text{Total number of patients}}$$

Table 2. Overall performance (wBER) of the best six classifiers when using complete and imputed features.

wBER	RF	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
Complete features (PWI/DKI/CSI)	0.15 (0.15/0.36/0.57)	0.17 (0.26/0.26/56)	0.28 (0.22/0.26/0.6)	0.15 (0.15/0.37/0.61)	0.15 (0.15/0.37/0.62)	0.14 (0.19/0.35/0.63)
Imputed features (PWI/DKI/CSI)	0.29 (0.29/0.28/0.41)	0.22 (0.31/0.33/0.4)	0.24 (0.28/0.32/0.42)	0.34 (0.29/0.28/0.42)	0.33 (0.27/0.28/0.41)	0.35 (0.28/0.38/0.41)

Results: Table 1 shows the classifiers performance for several time points before and after the labeling time point. It contains the detailed BER performance of the best 6 classifiers, from a total of 23 different classifiers, when using complete features. Table 2 reports wBER of the best 6 classifiers when using complete and imputed features on the whole dataset, but also on the small subsets pertaining to a single advanced MR modality (PWI, DKI, CSI).

Discussion: By using either RF, LogitBoost or RobustBoost classifier on all complete features we can predict with 100% accuracy the outcome of clinical labeling with one time point (ca. 1 month) in advance. The same results are obtained when using only complete perfusion features. Classifiers trained only on complete diffusion or spectroscopy features yield worse results. If we use imputed data, linear classifiers (dLDA, SVM-lin) perform better than RF and Boost ensembles.

Conclusion: In this study we proved that it is possible to accurately predict at least one month earlier than doctors if a patient is responsive to the treatment or not, by using all complete features or just complete perfusion features. For future work we plan on integrating the temporal evolution of the features when classifying different MR sessions and also allow updating of the class labels in time.

Acknowledgements: FWO G.0869.12N; IUAP P7/19/ ; EU FP7/2007-2013, MC ITN TRANSACT 2012 n. 316679.

Refererecens: [1] S. Van Cauter et al., *Neuro-Oncology*, 16(7), 1010-1021, 2014; [2] S. Van Cauter et al., *Journal of MRI*, vol. 37, no. 2, 445-456, 2013.