RITS 2015

A framework for multimodal imaging-based prognostic model building: Preliminary study on multimodal MRI in Glioblastoma Multiforme

T. Upadhaya [a,b,d,∗], Y. Morvan [a], E. Stindel [b], P.-J. Le Reste [c], M. Hatt [b]

[a] *bcom Institute of Research and Technologies, France*
[b] *LaTIM, INSERM, UMR 1101, Brest, France*
[c] *University Hospital Pontchaillou, Rennes, France*
[d] *University of Western Brittany, Brest, France*

Abstract

In Glioblastoma Multiforme (GBM) image-derived features ("radiomics") could help in individualizing patient management. Simple geometric features of tumors (necrosis, edema, active tumor) and first-order statistics in Magnetic Resonance Imaging (MRI) are used in clinical practice. However, these features provide limited characterization power because they do not incorporate spatial information and thus cannot differentiate patterns. The aim of this work is to develop and evaluate a methodological framework dedicated to building a prognostic model based on heterogeneity textural features of multimodal MRI sequences (T1, T1-contrast, T2 and FLAIR) in GBM. The proposed workflow consists in i) registering the available 3D multimodal MR images and segmenting the tumor volume, ii) extracting image features such as heterogeneity metrics and iii) building a prognostic model by selecting, ranking and combining optimal features through machine learning (Support Vector Machine). This framework was applied to 40 histologically proven GBM patients with the endpoint being overall survival (OS) classified as above or below the median survival (15 months). The models combining features from a maximum of two modalities were evaluated using leave-one-out cross-validation (LOOCV). A classification accuracy of 90% (sensitivity 85%, specificity 95%) was obtained by combining features from T1 pre-contrast and T1 post-contrast sequences. Our results suggest that several textural features in each MR sequence have prognostic value in GBM.

1. Introduction

Glioblastoma multiforme (GBM) is the most malignant grade IV primary intracranial tumor of adults according to the World Health Organization' histological grading system [1]. The prognosis is poor with a median survival of 15 months and occurrence rate is two or three cases per 100,000 per year [2,3]. The current standard treatment of GBM is a surgical resection followed by radiotherapy and chemotherapy [4]. Within this context, multimodal Magnetic Resonance Imaging (MRI) sequences (T1, T1-contrast, T2, FLAIR ...) play a major role for diagnosis, treatment planning, as well as prognosis, on which depend a number of clinical decisions. Thus, image-derived features extracted from standard MRI sequences could potentially be combined into a powerful prognostic tool with impact on patient management through higher stratification.

Although novel contrast agents, tracers and imaging sequences are being developed to investigate various aspects of tumor underlying pathophysiological processes, it is recognized that standard medical images may contain useful yet still unexploited information that could be useful in a clinical setting. Image-derived features (also called "radiomics") extracted from medical images may not only allow to be associated with prognosis but also non-invasively probe molecular and histological profiles of tumors, as a further step towards personalized

∗ Corresponding author.
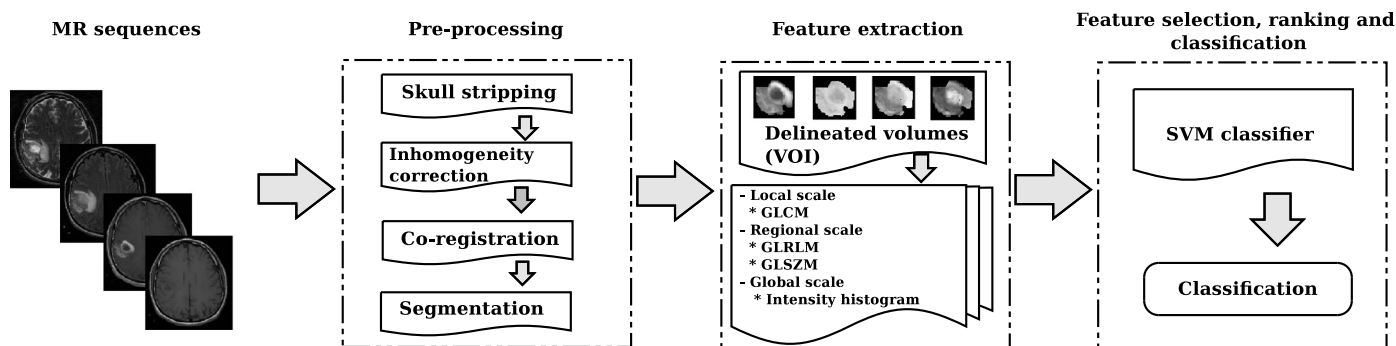*E-mail address:* taman.upadhaya@b-com.com (T. Upadhaya).

Fig. 1. The proposed workflow for GBM prognostic model development and validation.

medicine. Hence, considerable efforts have been carried out in identifying radiomics as a prognostic factor in GBM. Gutman et al. conducted an assessment of GBM tumor size and composition by using 24 visual observations familiar to neuro-radiologists from MR sequences dubbed as VASARI, and found that contrast-enhanced tumor volume and major axis length of tumor were strongly associated with poor survival [5]. Recently, Gevaert et al. evaluated 153 2D MR image-derived features for the development of a radiogenomic map, establishing a link between image features and underlying molecular data. Three features were found to be correlated with survival, 77 with VASARI, and 7 with molecular subgroups [6]. Mazurowski et al. investigated 8 MRI derived features and concluded that proportion enhancing tumor and major axis length were prognostic factors [7]. Clinical features such as Karnofsky Performance Status (KPS), age, extent of resection and degree of necrosis were found to be significant prognostic factors [8]. However, most studies were focused on the geometry of tumors, which do not incorporate the within-tumor heterogeneity characteristics imaged through multimodal MR sequences. Therefore, quantifying heterogeneity could be of interest for prognosis in GBM and it can be achieved through textural features analysis, which has been found to be of interest in various pattern recognition applications in medical imaging [9]. For example textural features have been used to quantify FDG tracer uptake heterogeneity in PET images of tumors [10,11]. Likewise, Zacharaki et al. used Gabor textural features and geometry from brain tumor as quantitative approach to characterize brain tumor type and grade from MRI [12].

The aim of this work[1] is to develop and evaluate the prognostic value of textural features quantifying the heterogeneity of GBM tumors in baseline multimodal MRI sequences. We developed a fully operational workflow for multimodal MR images pre-processing, registration, segmentation of tumors, characterization of tumors heterogeneity, and prognostic model training and validation using support vector machine (SVM) based on recursive feature elimination algorithm for feature selection and classification [13] (Fig. 1). We present preliminary results obtained on 40 patients with leave-one-out cross-validation (LOOCV).

2. Materials and methods

2.1. Patients population and imaging data

In order to simplify the analysis, patients were divided into two groups for the classification task: short and long overall survival (SOS and LOS) with a threshold of 14.8 months corresponding to the median survival ($n = 20$ in each group) with the standard treatment [4]. The MR images of GBM for this study were acquired from the Cancer Imaging Archive,[2] an imaging portal consisting of images from four centers (Henry Ford Hospital, University of California San Francisco, M.D. Anderson Cancer Center, and Emory University). The MR images are anonymized and publicly available. Forty patients with histopathologically diagnosed GBM and treated with radiotherapy and chemotherapy were retrospectively analyzed. Mean age of patients was 59 y (median 60 y, range 31–80 y). There were 25 males and 15 females with median survival of 14.8 months (mean 15.5, range 2.8–51.2). All baseline MRI sequences consist of: 1) T1-weighted pre contrast, 2) T1-weighted post contrast, 3) T2-weighted, and 4) FLAIR available for all 40 patients.

2.2. Pre-processing of MRI sequences

MRI sequences were acquired using various acquisition parameters and pulse sequences thereby with potential impact on the resulting quantification within the images. Standardization for comparison was performed using several sequentially implemented pre-processing steps.

2.2.1. Inhomogeneity correction
The undesirable signal that corrupts the MR images due to magnetic settings, patients' position and other factors is known as inhomogeneity or field bias. These inhomogeneities in a same tissue type degrade the performance of segmentation and classification, based on the assumption of spatial invariance. Therefore all the images included in this study were corrected for inhomogeneity using the N3ITK filter [14].

2.2.2. Mono-modal co-registration

All the MRI sequences (T1 pre-contrast, T2 and FLAIR) were registered to the T1 post-contrast image using mutual information similarity metric [15].

2.2.3. Tumor delineation

Prior to segmentation, a fully-automatic and customized skull-stripping algorithm was used to isolate the brain region from the MRI sequences [16]. Segmentation of the tumor regions was carried-out using all four MRI sequences and a single mask with labels representing edema, necrosis and active tumor was automatically produced. The segmentation algorithm is based on Support Vector Machine exploiting multimodal voxels' intensities with subsequent hierarchical regularization by Conditional Random Fields [17]. An expert clinician in neuro-oncology (P-J Le Reste) validated and manually edited the resulting masks when needed.

2.2.4. Intensity standardization within delineated lesions

As a consequence of using different MRI sequences with the varying intensity ranges, the quantization step which yields finite range of intensities among modalities and across patients was employed using equation (1).

$$V(x) = B \frac{I(x) - \min i_{i \in \Omega}}{\max i_{i \in \Omega} - \min i_{i \in \Omega} + 1} \tag{1}$$

where B is the quantization bins value, usually considered within the range 4–256. We considered 4 different values in the present work for B: 16, 32, 64, and 128. I is the intensity of a voxel in the original image and Ω is the set of voxels in the delineated volume. This quantization step reduces the spatial variability of intensities (noise) and is also necessary to compute relevant values for textural features and facilitate their comparison. All the above pre-processing steps except the quantization were carried out using Brain Tumor Image Analysis software (BraTumIA),[3] a fully dedicated open source software for GBM segmentation using MRI sequences [14–17].

2.3. Textural features extraction

Thirty-four textural features (Table 1) were extracted based on the first-order, second-order and higher-order statistics accounting for intensity distribution in global, local and regional scale respectively, of each delineated tumor volume in each of the four MRI sequences. Global scale textural features were computed from intensity histogram, local scale textural features were computed from Grey-level co-occurrence matrix (GLCM) [18], whereas regional scale textural features were computed from grey-level run-length (GLRLM) [19] and grey-level size-zone matrices (GLSZM) [20]. These texture matrices were calculated in 3D taking into account all 13 directions [21,22]. All these features have been somewhat useful in various pattern recognition applications including medical imaging and were therefore considered in the present work.

---

[3] Nicole et al. Multimodal glioblastoma segmentation. Software available at http://istb-software.unibe.ch/bratumia/requirement.php.

Table 1

34 textural features were computed; 6 local scale, 21 regional scale, 7 global scale quantifying textural properties of delineated tumor volumes.

Order of textural feature	Scale	Description	Features
First	Global	Intensity histogram	Standard deviation Skewness Sum Mean Median Variance Kurtosis
Second	Local	Grey-level co-occurrence matrices (GLCM)	Inertia Energy Entropy Cluster shade Cluster prominence Inverse difference moment
Higher	Regional	Grey-level run length matrix (GLRLM)	Short run emphasis Long run emphasis Grey-level non-uniformity Run length non-uniformity High grey-level run emp. Low grey-level run emp. Short run high grey-level emp. Long run low grey-level emp. Short run low grey-level emp. Long run high grey-level emp.
		Grey-level size zone matrix (GLSZM)	Short zone emphasis Large zone emphasis Grey-level non-uniform sp. hom. Size zone non-uniformity Zone percentage Low grey-level zone emp. High grey-level zone emp. Short zone low grey-level emp. Short zone high grey-level emp Large zone low grey-level emp. Large zone high grey-level emp.

2.4. Features selection and ranking

Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm initially proposed by Guyon et al. for gene selection in a cancer classification problem was employed for the ranking of features [13]. This method selects the features in a linear separation case, based on the vector of weighted magnitude $||w_\psi||^2$ defined as

$$w = \sum_{i \in \zeta} y_i \alpha_i x_i \tag{2}$$

where ζ is the number of training sets, ψ the total number of features, y is the class label, x is a feature and α is the corresponding solution of the SVM classifier. The features corresponding to smallest weight magnitudes are recursively eliminated. The whole procedure follows the nested k-fold cross-validation in training set in order to correct for the selection bias. The features most frequently occurring on a given particular rank are assigned said ranking.
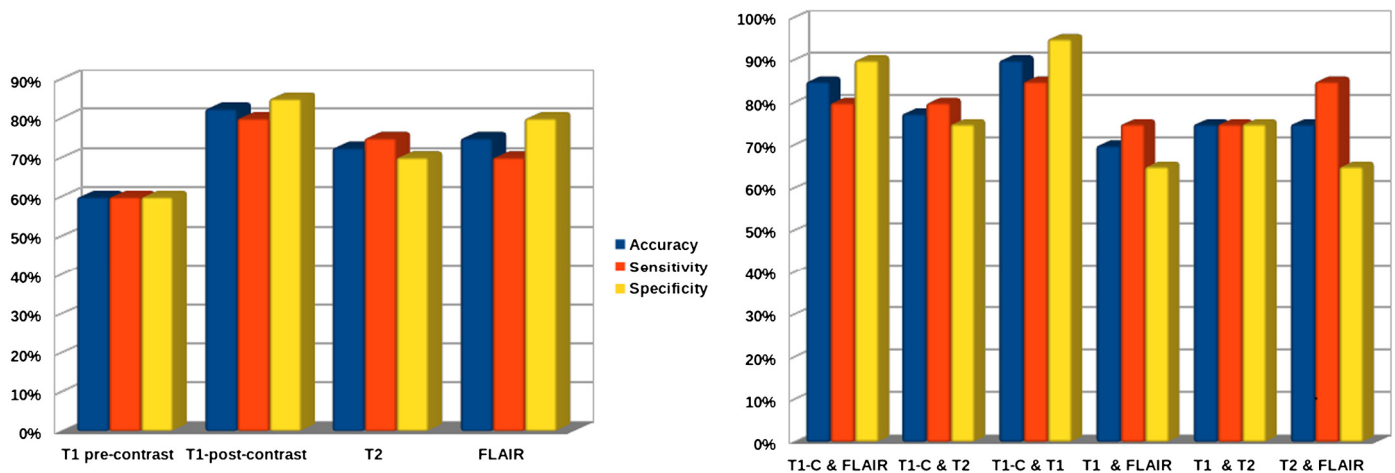
Fig. 2. Model accuracy, specificity and sensitivity assessed by LOOCV. Left: using 1 modality only, right: all combinations of 2 modalities.

2.5. Classification for prognosis

We used multivariate pattern analysis methods "Support Vector Machine (SVM)"[4] to solve the classification problem, as this approach has been intensively studied and benchmarked against a variety of other techniques [13]. As a preliminary step in our research, we chose to simplify the problem by considering a binary classification problem, in which the two classes are 'long overall survival' (above the median OS of 14.8 months) and 'short overall survival' (below or equal to 14.8 months). The concatenation of multi textural features from each MRI sequence was fed as input to the algorithm. Four different models, one for each modality, and six different models corresponding to all combinations of a maximum of 2 different MRI sequences were trained, tuned and cross-validated using LOOCV on the 40 patient datasets. Model parameters selection for the SVM classifier with a linear kernel was performed using grid-based cross-validation on the training data. The desired number of features to build the model were collected in a forward selection manner based on higher cross-validation accuracy in the training set from outcome of feature selection and ranking procedure of Section 2.4.

3. Results

Fig. 2 presents the results for ten different models (one for each MRI sequence, as well as the combination of all the two MRI sequences) that were trained and validated using LOOCV. The best results were consistently obtained in all modalities using textural features calculated with quantization value of $B = 64$, so only these results are provided. The best classification results considering only one MRI sequence at a time were obtained using a combination of five features (cluster shade, high grey-level zone emphasis, large zone high grey-level emphasis, grey-level non-uniform spectral homogeneity and size zone non-uniformity) extracted from the T1 post-contrast se-

quence, with an accuracy of 82.5%. Models based on T1 pre-contrast, T2 and FLAIR sequence exploited cluster prominence, large zone emphasis, cluster shade and grey-level non-uniformity features, with slightly lower associated accuracy of 62.5%, 72.5% and 75% respectively. Regarding the combination of two different MRI sequences, the highest accuracy was obtained by combining the T1 pre-contrast and T1 post-contrast sequences, with an associated accuracy of 90% (sensitivity 85% and specificity 95%). Two features only (high grey-level zone emphasis and variance) from T1 pre-contrast and four features (cluster prominence, short run high grey-level emphasis, size zone non-uniformity and sum) from T1 post-contrast contributed in this model. Note that for this best model, features from first, second and third orders were retained. The feature subset selection method showed that four features from local and regional scales (cluster prominence, cluster shade, grey-level non-uniformity and size zone non-uniformity) were consistently considered as relevant. Overall, the T1 post-contrast MRI sequence seemed to offer the most relevant information regarding prognosis. In another set of experiments, we built a model with the contextual clinical variables such as age, gender, Karnofsky and treatment modality (radiotherapy, chemotherapy and surgery). The variables gender and treatment modality were categorized using 1-of-k coding scheme. The model built using contextual clinical variables only reached a limited accuracy of 58%. In the multivariate selection when included with image-derived features, none of these clinical variables were retained for building the model.

4. Discussion–conclusion

Our results suggest that using multivariate methods such as SVM with the combination of global, local and regional textural features quantifying heterogeneity in all four MR modalities available in routine clinical acquisitions can provide a model with high prognostic value. However, one drawback of multivariate methods using high dimensional features is the lack of a framework for estimating the statical interpretation because unlike univariate methods, SVM model does not compute statistical tests (and corresponding p-values) associated with the

[4] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/.
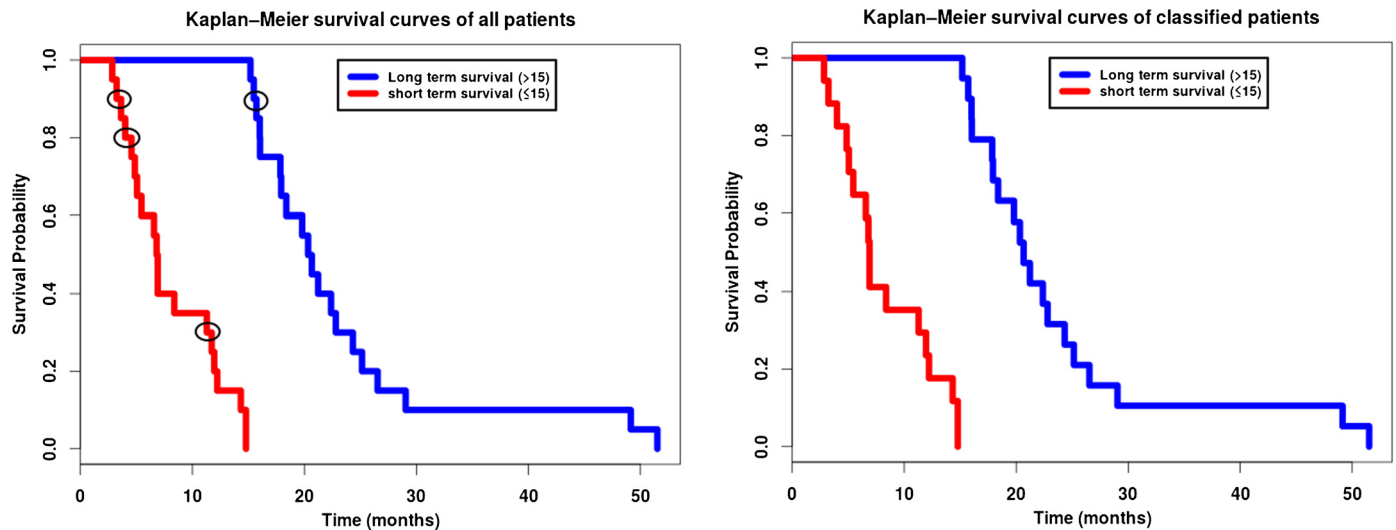
Fig. 3. Kaplan–Meier survival curves for patients according to left: classification using the median survival and right: classification obtained by the model. Misclassified patients are indicated by black circles.

significance level. Rather, evaluating the SVM models' maximum weight coefficients recursively helps in ranking the significant features. In addition, selecting an optimal number of significant features and building the SVM models based on forward selection of features gives the highest cross-validation accuracy. We emphasize that these results are only preliminary and the present work has a few limitations: i) the analysis included only 40 patients (Fig. 3) and validation was carried out using LOOCV, which is more optimistic than validation in an independent dataset; ii) prognosis analysis was simplified by considering a binary classification problem (SOS and LOS) instead of a continuous variable; iii) we did not investigate the combination of more than two MR sequences among the four available due to the limited number of patients. We emphasize however that the analysis is currently ongoing, and that additional patients will be included in the future, which will allow for a more rigorous evaluation of the accuracy, by testing the trained models on an additional group of patients not used in the model training. In addition, we expect that by projecting these image features in a higher dimension using a "kernel trick" (for e.g. Radial Basis Kernel) for non-linear multivariate analysis and by combining features from all four MRI sequences, an even more robust and accurate model for feature selection and classification can be developed. However, for a reliable analysis, we will have to wait for more patients datasets to be processed through the entire described workflow (pre-processing and segmentation of the images constitutes the bottleneck for producing more results). Future work will consist in i) adding more patients for validation using training and testing groups, ii) improving the pre-processing steps of voxels spatial resampling, standardization and quantization, as recently suggested [23], as well as updating the bias correction using N4ITK instead of N3ITK, iii) considering additional features to improve the proposed model, iv) building a fully multimodal MRI model by combining features from more than two sequences, v) considering survival as a continuous variable using survival SVM method [24] and vi) building and comparing models with vari-ous other machine learning based classifier techniques such as random forest or artificial neural network.

References

[1] Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009. Neuro-Oncol 2012;14(suppl 5):v1–49.

[2] Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 2007;114(2):97–109.

[3] Krex D, Klink B, Hartmann C, von Deimling A, Pietsch T, Simon M, et al. Long-term survival with glioblastoma multiforme. Brain 2007;130(10):2596–606.

[4] Preusser M, de Ribaupierre S, Wohrer A, Erridge SC, Hegi M, Weller M, et al. Current concepts and management of glioblastoma. Ann Neurol 2011;70(1):9–21.

[5] Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao J, Aurora TD, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology 2013;267(2):560–9.

[6] Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology 2014.

[7] Mazurowski MA, Zhang J, Peters KB, Hobbs H. Computer-extracted MR imaging features are associated with survival in glioblastoma patients. J Neuro-Oncol 2014:1–6.

[8] Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. J Neurosurg 2001;95(2):190–8.

[9] Alic L, Niessen WJ, Veenland JF. Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review. PLoS ONE 2014;9(10):e110300.

[10] Tixier F, Hatt M, Valla C, Fleury V, Lamour C, Ezzouhri S, et al. Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer. J Nucl Med 2014;55(8):1235–41.

[11] Tixier F, Cheze Le Rest C, Hatt M, Albarghach N, Pradier O, Metges J-P, et al. Intratumor heterogeneity characterized by textural features on base-line 18F-FDG PET images predicts response to

concomitant radiochemotherapy in esophageal cancer. J Nucl Med 2011;52(3):369–78.

[12] Zacharaki EI, Wang S, Chawla S, Yoo DS, Wolf R, Melhem ER, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn Reson Med 2009;62(6):1609–18.

[13] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1–3):389–422.

[14] Ibanez L, Schroeder W, Ng L, Cates J. The ITK software guide. Second edition. Kitware, Inc.; ISBN 1-930934-15-7, 2005. http://www.itk.org/ItkSoftwareGuide.pdf.

[15] Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 2003;22(8):986–1004.

[16] Porz N, Bauer S, Pica A, Schucht P, Beck J, Kumar Verma R, et al. Multi-modal glioblastoma segmentation: man versus machine. PLoS ONE 2014;9(5):e96873.

[17] Bauer S, Nolte L-P, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Medical image computing and computer-assisted intervention – MICCAI 2011. Springer; 2011. p. 354–61.

[18] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;6:610–21.

[19] Loh H-H, Leu J-G, Luo RC. The analysis of natural textures using run length features. IEEE Trans Ind Electron 1988;35(2):323–8.

[20] Thibault G, Fertil B, Navarro C, et al. Texture indexes and gray level size zone matrix: application to cell nuclei classification. Pattern Recognit Inf Process 2009:140–5.

[21] Gutierrez, Rodriguez D, et al. Metrics and textural features of MRI diffusion to improve classification of pediatric posterior fossa tumors. Am J Neuroradiol 2014;35(5):1009–15.

[22] Xian G. An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. Expert Syst Appl 2010;37(10):6737–41.

[23] Vallieres M, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol 2015;60(14):5471.

[24] Evers L, Messow C-M. Sparse kernel methods for high-dimensional survival data. Bioinformatics 2008;24(14):1632–8.