



Published in final edited form as:

Brainlesion. 2016 ; 9556: 144–155. doi:10.1007/978-3-319-30858-6_1.

GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation

Spyridon Bakas, Ke Zeng, Aristeidis Sotiras, Saima Rathore, Hamed Akbari, Bilwaj Gaonkar, Martin Rozycki, Sarthak Pati, and Christos Davatzikos

Section of Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA

Abstract

We present an approach for segmenting low- and high-grade gliomas in multimodal magnetic resonance imaging volumes. The proposed approach is based on a hybrid generative-discriminative model. Firstly, a generative approach based on an Expectation-Maximization framework that incorporates a glioma growth model is used to segment the brain scans into tumor, as well as healthy tissue labels. Secondly, a gradient boosting multi-class classification scheme is used to refine tumor labels based on information from multiple patients. Lastly, a probabilistic Bayesian strategy is employed to further refine and finalize the tumor segmentation based on patient-specific intensity statistics from the multiple modalities. We evaluated our approach in 186 cases during the training phase of the BRAIn Tumor Segmentation (BRATS) 2015 challenge and report promising results. During the testing phase, the algorithm was additionally evaluated in 53 unseen cases, achieving the best performance among the competing methods.

Keywords

Segmentation; Brain tumor; Glioma; Multimodal MRI; BRATS challenge; Gradient boosting; Expectation maximization; Brain tumor growth model; Probabilistic model

1 Introduction

Gliomas comprise a group of primary central nervous system (CNS) tumors of neuroglial cells (*e.g.*, astrocytes and oligodendrocytes) that have different degrees of aggressiveness. They are mainly divided into low- and high-grade gliomas (LGGs and HGGs) according to their progression rate and histopathology. LGGs and HGGs exhibit distinct pathophysiological phenotypes and are subject to different treatment options. LGGs are less common than HGGs, constitute approximately 20 % of CNS glial tumors, and almost all of them eventually progress to HGGs [15]. HGGs are rapidly progressing malignancies, divided based on their histopathologic features into anaplastic gliomas and glioblastomas (GBMs) [21].

Gliomas consist of various parts, each of which shows a different imaging phenotype in multimodal magnetic resonance imaging (MRI). Typically, the core of HGGs consists of enhancing, non-enhancing and necrotic parts, whereas the core of LGGs does not necessarily include an enhancing part. Another critical feature, for both understanding and treating gliomas, is the peritumoral edematous region. Edema occurs from infiltrating tumor cells, as well as a biological response to the angiogenic and vascular permeability factors released by the spatially adjacent tumor cells [1].

Quantification of the various parts of gliomas, in multimodal MRI, has an important role in treatment decisions, planning, as well as monitoring in longitudinal studies. The accurate segmentation of these regions is required to allow this quantification. However, tumor segmentation is extremely challenging due to the tumor regions being defined through intensity changes relative to the surrounding normal tissue, and such intensity information being disseminated across various modalities for each region. Additional factors that contribute to the difficulty of brain tumor segmentation task is the motion of the patient during the examination, as well as the magnetic field inhomogeneities. Hence, the manual annotation of such boundaries is time-consuming, prone to misinterpretation, human error and observer bias [3], with intra- and inter-rater variability up to 20 % and 28 %, respectively [16]. Computer-aided segmentation of brain tumor images would thus be an important advancement. Towards this end, we present a computer-aided segmentation method that aims to accurately segment such tumors and eventually allow for their quantification.

The remainder of this paper is organized as follows: Sect. 2 details the provided data, while Sect. 3 presents the proposed segmentation strategy. The experimental validation setting is described in Sect. 4 along with the obtained results. Finally, Sect. 5 concludes the paper with a short discussion and potential future research directions.

2 Materials

The data used in this study comprise 186 preoperative multimodal MRI scans of patients with gliomas (54 LGGs and 132 HGGs) that were provided as the training set for the multimodal BRATS 2015 challenge, from the Virtual Skeleton Database (VSD) [12]. Specifically, these data are a combination of the training set (10 LGGs and 20 HGGs) used in the BRATS 2013 challenge [17], as well as 44 LGG and 112 HGG scans provided from the National Institutes of Health (NIH) Cancer Imaging Archive (TCIA). The data of each patient consists of native and contrast-enhanced (CE) T1-weighted, as well as T2-weighted and T2 Fluid-attenuated inversion recovery (FLAIR) MRI volumes. The volumes of the various modalities were, co-registered to the same anatomical template and interpolated to 1 mm³ voxel resolution. In addition to the training set, 53 multimodal volumetric images were provided as the testing set for the challenge, comprising both preoperative and after initial therapy scans.

Finally, ground truth (GT) segmentations for the training set were also provided. Specifically, the data from BRATS 2013 were manually annotated, whereas data from TCIA were automatically annotated by fusing the approved by experts results of the segmentation

algorithms that ranked high in the BRATS 2012 and 2013 challenges [17]. The GT segmentations comprise the enhancing part of the tumor (ET), the tumor core (TC), which is described by the union of necrotic, non-enhancing and enhancing parts of the tumor, and the whole tumor (WT), which is the union of the TC and the peritumoral edematous region. Note that the testing sets have been segmented manually by one to four rates, but the GT segmentations were not provided to the participating teams, allowing for their evaluation only by the challenge organizers.

3 Methods

The provided skull-stripped and co-registered MRI volumes were initially smoothed using a low-level image processing method, namely Smallest Univalued Segment Assimilating Nucleus (SUSAN) [20], to reduce intensity noise in regions of uniform intensity profile. The intensity histograms of all modalities of all patients were then matched to the corresponding modality of a single reference patient.

A modified version of the GLioma Image Segmentation and Registration (GLISTR) software [10] was subsequently used to delineate the boundaries of healthy tissues (*i.e.*, white and gray matter, cerebrospinal fluid, vessels and cerebellum), as well as tumor tissues (*i.e.*, edema, necrosis, non-enhancing and enhancing parts of the tumor). Although GLISTR was inspired by a sequential approach of segmentation of the input brain scans followed by the registration of the outcome to a given healthy atlas [8], it was originally proposed in [9,10] as a tool that jointly performs segmentation and registration, but handles only scans with solitary HGGs. It was then conceptually improved in [14] to target broader brain tumor appearances, including multifocal masses and complex shapes with heterogeneous textures (*e.g.*, LGGs), enabling it to also participate in the BRATS 2014 challenge [13]. The version of GLISTR used here, was modified in terms of using multiple seed-points for each brain tissue label, in order to model the exact intensity distribution (*i.e.*, mean and variance) of each, whereas [14] uses a single seed-point for each label, assuming it is representative of each label's mean intensity value, and the variance is described by a fixed value for all labels. Note that both previous and current versions of GLISTR do not depend on the coordinates of the initialization seed-points, but on the intensity value of the corresponding voxel on each modality. Therefore, even if different seed-points are initialized across two independent segmentation attempts, GLISTR output segmentation results should be identical, if the intensity distributions modeled during these attempts are the same. The whole framework of GLISTR is based on a probabilistic generative model that relies on Expectation-Maximization (EM), to recursively refine the estimates of the posteriors for all tissue labels, the deformable mapping to the atlas, and the parameters of the incorporated brain tumor growth model [11].

This modified version of GLISTR requires as input a single seed-point and a radius for each apparent tumor, as well as multiple seed-points for each brain tissue label. These seed-points were initialized using BrainTumorViewer¹, which has been primarily developed for this purpose. Given the single seed-point and the radius inputs, the center and the bulk volume of

¹Available on: <https://www.cbica.upenn.edu/sbia/software/braintumorviewer/>.

each tumor are approximated by a sphere (Fig. 1). The parametric model of the sphere is used to initiate the tumor growth model for each apparent tumor. This growth model is used to modify the healthy atlas into one with tumor and edema tissues matching the input scans, whilst it approximates the deformation occurred to the surrounding brain tissues, due to the effect of the tumors' mass. A tumor shape prior is also estimated separately, by a random-walk-based generative model, which uses multiple tumor seed-points as initial foreground cues. This tumor shape prior is systemically incorporated into the EM framework via an empirical Bayes model, as described in [14]. Furthermore, a minimum of three seed-points are initialized for each brain tissue label, with the intention of capturing the intensity variation of each tissue label, and modeling each label's intensity distribution. This provides a better initialization to the EM framework, resulting to more accurate delineation of all tissue labels, when compared to [14] that uses a single seed-point for each label. The output of GLISTR is a posterior probability map for each tissue label, as well as a label map, which is a very good initial segmentation of all different tissues within a patient's brain.

A machine-learning approach was then used to refine GLISTR results by utilizing information across multiple patients. Specifically, the gradient boosting algorithm [5] is employed for voxel-level multi-label classification. Gradient boosting is an ensemble method that produces a prediction model by combining weak learners in a stage-wise fashion. It generalizes other boosting techniques by allowing the optimization of an arbitrary differentiable loss function. We used the Python package scikit-learn [18] for the implementation, choosing deviance as the loss function. At each iteration, a weak learner, specifically a decision tree of maximum depth 3, was added to the decision function, approximating the current negative gradient of the objective. Randomness was introduced when constructing each tree [6]. Each decision tree was fit to a sub-sample of the training set, with the sampling rate set equal to 0.6. The split was also determined stochastically by sampling a subset of features at each node, with the number of sampled features set equal to the square root of the total number of features. The algorithm was terminated after 100 such iterations.

The features used for training our model consist of five components; image intensity, image derivative, geodesic information, texture features, and the GLISTR posterior probability maps. The intensity component comprises the raw intensity value of each voxel ($I(v_i)$), as well as their differences among all four modalities (*i.e.*, T1, T1-CE, T2, T2-FLAIR). The image derivative component comprises the Laplacian of Gaussian and the image gradient magnitude. Note that prior to calculating any intensity-based feature, intensity normalization was performed based on the median intensity value of the GLISTR segmented cerebrospinal fluid. The geodesic information at voxel v_i was given by the geodesic distance from the seed-point used in GLISTR as the tumor center, at voxel v_s . Specifically, the geodesic distance between v_i and v_s was defined as $\min_{\gamma} \int_{\gamma} P(\chi(s)) ds$, where γ is a path connecting v_i to v_s . Similar to the approach taken in [7], we set the weight P at each voxel to be proportional to its gradient magnitude, and the optimization was solved using the fast marching method [4,19]. Furthermore, the texture features describe the first and second order texture statistics computed from a gray-level co-occurrence matrix. Specifically, the first order statistics comprise the mean and variance of the intensities from each modality within a radius of 2 voxels for each voxel. For the second order statistics, the image volumes

were firstly normalized to 64 different gray levels, and then a bounding box of 5-by-5 pixels was used for all the pixels of each image slice. Subsequently, a graylevel co-occurrence matrix was filled with the intensity values within a radius of 2 pixels for all eight main directions (*i.e.*, $\{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$) to extract the energy, entropy, dissimilarity, homogeneity (*i.e.*, inverse difference moment of order 2), and inverse difference moment of order 1. It should also be mentioned that our model was trained using both LGG and HGG training samples simultaneously using a 54-fold cross-validation setting (given that 54 LGGs were present in the training data, *i.e.*, allowing for using a single LGG within each fold). The cross-validation setting is necessary in order to avoid over-fitting.

Finally, a patient-wise refinement was performed by assessing the local intensity distribution of the current segmentation labels and updating their spatial configuration based on a probabilistic model, inspired by [2]. Firstly, the intensity distribution of voxels with GLISTR posterior probability equal to 1 for the tissue classes of white matter, edema, necrosis, non-enhancing and enhancing tumor, were populated separately. Note that in the current segmentation goal, there is no distinction between the non-enhancing and the necrotic parts of the tumor. A normalization to the histograms of pair-wise distributions was then applied. The class-conditional probability densities ($Pr(I(v_i)/Class_1)$ and $Pr(I(v_i)/Class_2)$) were modeled by fitting distinct Gaussian models, using Maximum Likelihood Estimation to find the mean and standard deviation for each class. There are three pair-wise distributions considered here; the edema voxels opposed to the white matter voxels in the T2-FLAIR volume, the ET voxels opposed to the edema voxels in the T1-CE volume, and the ET voxels opposed to the union of the necrosis and the non-enhancing tumor in the T1-CE volume. In all cases, the former intensity population is expected to have much higher (*i.e.*, brighter) values. Hence, voxels of each class with small spatial proximity (namely 3 voxels) to the opposing tissue class were evaluated based on their intensity. Specifically, the intensity $I(v_i)$ of each of these voxels was assessed and $Pr(I(v_i)/Class_1)$ was compared with $Pr(I(v_i)/Class_2)$. The voxel v_i was then classified into a tissue class according to the larger of the two conditional probabilities. This is equivalent to a classification based on Bayes' Theorem with equal priors for the two classes, *i.e.*, $Pr(Class_1) = Pr(Class_2) = 0.5$.

Note that our challenge winning methodology has been made publicly available on the Online Image Processing Portal (IPP)² of the Center for Biomedical Image Computing and Analytics (CBICA), of the University of Pennsylvania. CBICA's IPP allows users to perform their data analysis using the integrated algorithms, without any software installation, whilst also using CBICA's High Performance Computing resources.

4 Experiments and Results

In order to assess the segmentation performance of our method, we evaluated the overlap between the proposed tumor labels and the GT in three regions, *i.e.*, WT, TC and ET, as suggested in [17]. Figure 2 showcases example segmentation results along with the respective GT segmentations for eight patients (four HGGs and four LGGs). These correspond to the two most and least successful segmentation results for each glioma grade.

²Available on: <https://ipp.cbica.upenn.edu/>.

We observe high agreement between the generated results and the provided labels. We note that the highest overlap is observed for edema, while there is some disagreement between the segmentations of the enhancing and non-enhancing parts of the tumor.

To further appraise the performance of the proposed method, we quantitatively validated the per-voxel overlap between respective regions, in the training set, using the DICE coefficient (see Fig. 3 and Table 1). This metric takes values between 0 and 1, with higher values corresponding to increased overlap. Moreover, aiming to understand fully the obtained results, we stratified them based on the labeling protocol of the GT segmentation. In particular, data with manually annotated GT (*i.e.*, BRATS 2013 data) was evaluated separately from data with automatically defined GT (*i.e.*, TCIA data). The reason behind this distinction is twofold. First, only manual segmentation can be considered as gold standard, thus allowing us to evaluate the potential of our approach when targeting an interactive clinical work-flow. Second, results validated using automatically defined GT should be interpreted with caution because of the inherently introduced bias towards the employed automated methods, which also influences visually inspecting experts [3]. As a consequence, our method may be negatively impacted since it may learn to reproduce the systematic mistakes of the provided annotations. Furthermore, since LGGs are characterized by a distinct pathophysiological phenotype (*i.e.*, lack of enhancing tumor part), we also divided the obtained results in terms of the tumors' grade (*i.e.*, LGG and HGG). This allows the performance assessment of the proposed approach on the distinct imaging phenotype of each grade separately.

Figure 3 reports the distributions of the cross-validated DICE score across patients of the training set, for each step of the proposed method and for each tissue label (WT, TC and ET) while Table 1 reports the respective mean and median values. The results are presented following the previously described stratifications. Figure 3 shows a clear step-wise improvement in both the mean and median values of all tissue labels when considering the complete set of data, the automatically annotated, the LGGs and the HGGs. On the contrary, we observe a step-wise deterioration of both the mean and median values for the TC label when assessing the manually annotated subset of the data (see Table 1 for the exact values). This is probably the effect of learning systematically mislabeled voxels present in the automatically generated GT annotations (see mislabeled ET in GT of the second HGG in Fig. 2(a)). Furthermore, we note the segmentation results for the ET label to vary significantly between LGGs and HGGs, with the former showing lower and less consistent results. This seems to be the effect of training our learning model using both classes simultaneously, when LGGs typically show a different pathophysiological phenotype marked by the lack of an enhancing part. Nevertheless, the segmentation of the WT label in the LGGs is comparable to this of the HGGs.

Lastly, the hereby proposed method was also quantitatively evaluated during the testing phase of the BRATS 2015 challenge along other 12 participating teams, using the DICE score and the robust Hausdorff distance (95 % quantile), similar to [17]. Each team had only 48 hours for producing their segmentation labels, from the time the testing set was made available, until the submission of the results to VSD. The limited time of the testing phase was considered essential to minimize the chance of optimizing the proposed algorithms on

the given data. According to the results presented during the challenge, our semi-automatic approach performed best when compared to the other competing methods.

5 Discussion

We presented an approach that combines generative and discriminative methods towards providing a reliable and highly accurate segmentation of LGGs and HGGs in multimodal MRI volumes. Our proposed approach is built upon the brain segmentation results provided by a modified version of GLISTR. GLISTR segments the brain into tumor and healthy tissue labels by means of a generative model encompassing a tumor growth model and a probabilistic atlas of healthy individuals. GLISTR tumor labels are subsequently refined taking into account population-wide tumor label appearance statistics that were learned by employing a gradient boosting multi-class classifier. The final results are produced by adapting the segmentation labels based on patient-specific label intensity distributions from the multiple modalities.

Our approach was able to deliver high quality tumor segmentation results, eventually performing best among the competing methods in BRATS 2015 challenge, by significantly improving GLISTR results [13] through the adopted post-processing strategies. This improvement was evident for both manually and automatically segmented data, as well as for both LGGs and HGGs. The only case where the post-processing resulted in a decrease of the performance is for the TC label when considering only the manually segmented data. This could be probably attributed to the fact that the supervised gradient boosting model learned consistent errors present in the automatically generated segmentations and propagated them when refining GLISTR results. While pooling information for more patients seems to be benefiting the learning algorithm, it also introduces a bias towards the more numerous automatically generated data. Accounting for this bias by weighting accordingly manually and automatically segmented samples could possibly allow for harnessing the additional information without compromising quality. Moreover, the proposed approach performed best in the WT label, which is clinically considered of the highest importance since it allows for: (i) assessment and evaluation of the heterogeneity of the peritumoral edematous region [1], (ii) estimates of diffuse tumor infiltration, rather than a binary tumor/no-tumor classification, and (iii) guidance to spatially-precise treatment decisions.

The proposed approach segmented the whole tumor and the tumor core with high accuracy for both LGGs and HGGs. However, the segmentation results for the enhancing tumor varied importantly between the two classes of tumors, with the performance of our method in the case of LGGs being significantly lower and less consistent. This is due to the fact that LGGs are characterized by a distinct pathophysiological phenotype that is often marked by the lack of an enhancing part, hence not having the same imaging phenotype with the HGGs. In addition, the segmentation of the enhancing tumor could be further improved considering that gliomas can be distinguished into two distinct imaging phenotypes, which are not necessarily consistent with their clinical grade (*i.e.*, LGG/HGG). These distinct imaging signatures could be possibly exploited in a machine learning framework that considers separately radiologically defined HGGs and LGGs, *i.e.*, tumors with and without a

distinctive enhancing part. By modeling separately these distinct imaging phenotypes, it is possible to capture better the imaging heterogeneity and improve label prediction.

References

1. Akbari H, Macyszyn L, Da X, Wolf RL, Bilello M, Verma R, O'Rourke DM, Davatzikos C. Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity. *Radiology*. 273(2):502–510.2014; [PubMed: 24955928]
2. Bakas S, Chatzimichail K, Hunter G, Labbe B, Sidhu PS, Makris D. Fast semi-automatic segmentation of focal liver lesions in contrast-enhanced ultrasound, based on a probabilistic model. *Comput Methods Biomech Biomed Eng: Imaging Vis.* :1–10.2015; DOI: 10.1080/21681163.2015.1029642
3. Deeley MA, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, Malcolm AW, Moretti L, Jaboin J, Niermann K, Yang ES, Yu DS, Yei F, Koyama T, Ding GX, Dawant BM. Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Phy Med Biol.* 56(14):4557–4577.2011;
4. Deschamps T, Cohen LD. Fast extraction of minimal paths in 3D images and applications to virtual endoscopy. *Med Image Anal.* 5(4):281–299.2001; [PubMed: 11731307]
5. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 29(5):1189–1232.2001;
6. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal.* 38(4):367–378.2002;
7. Gaonkar B, Macyszyn L, Bilello M, Sadaghiani MS, Akbari H, Attiah MA, Ali ZS, Da X, Zhan Y, O'Rourke D, Grady SM, Davatzikos C. Automated tumor volumetry using computer-aided image segmentation. *Acad Radiol.* 22(5):653–661.2015; [PubMed: 25770633]
8. Gooya A, Biros G, Davatzikos C. Deformable registration of glioma images using EM algorithm and diffusion reaction modeling. *IEEE Trans Med Imaging.* 30(2):375–390.2011; [PubMed: 20876010]
9. Gooya A, Pohl KM, Bilello M, Biros G, Davatzikos C. Joint segmentation and deformable registration of brain scans guided by a tumor growth model. *Med Image Comput Comput-Assist Interv.* 14(2):532–540.2011;
10. Gooya A, Pohl KM, Bilello M, Cirillo L, Biros G, Melhem ER, Davatzikos C. GLISTR: glioma image segmentation and registration. *IEEE Trans Med Imaging.* 31(10):1941–1954.2012; [PubMed: 22907965]
11. Hoge C, Davatzikos C, Biros G. An image-driven parameter estimation problem for a reaction diffusion glioma growth model with mass effects. *J Math Biol.* 56(6):793–825.2008; [PubMed: 18026731]
12. Kistler M, Bonaretti S, Pfahrer M, Niklaus R, Büchler P. The virtual skeleton database: an open access repository for biomedical research and collaboration. *J Med Internet Res.* 15(11):e245.2013; [PubMed: 24220210]
13. Kwon D, Akbari H, Da X, Gaonkar B, Davatzikos C. Multimodal brain tumor image segmentation using GLISTR. *MICCAI Brain Tumor Segmentation (BraTS) Challenge Manuscripts.* :18–19.2014
14. Kwon D, Shinohara RT, Akbari H, Davatzikos C. Combining generative models for multifocal glioma segmentation and registration. *Med Image Comput Comput-Assist Interv.* 17(1):763–770.2014; [PubMed: 25333188]
15. Louis DN. Molecular pathology of malignant gliomas. *Annu Rev Pathol - Mech Dis.* 1:97–117.2006;
16. Mazzara GP, Velthuisen RP, Pearlman JL, Greenberg HM, Wagner H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentations. *Int J Radiat Oncol - Biol - Phy.* 59(1):300–312.2004;
17. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp C, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A,

- Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Riklin-Raviv T, Reza SMS, Ryan M, Sarikaya D, Schwartz L, Shin H-C, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. 34(10):1993–2024.2015; DOI: 10.1109/TMI.2014.2377694 [PubMed: 25494501]
18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in python. *J Mach Learn Res*. 12:2825–2830.2011;
19. Sethian JA. A fast marching level set method for monotonically advancing fronts. *Proc Nat Acad Sci USA*. 93(4):1591–1595.1996; [PubMed: 11607632]
20. Smith SM, Brady JM. SUSAN - a new approach to low level image processing. *Int J Comput Vis*. 23(1):45–78.1997;
21. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med*. 359(5):492–507.2008; [PubMed: 18669428]

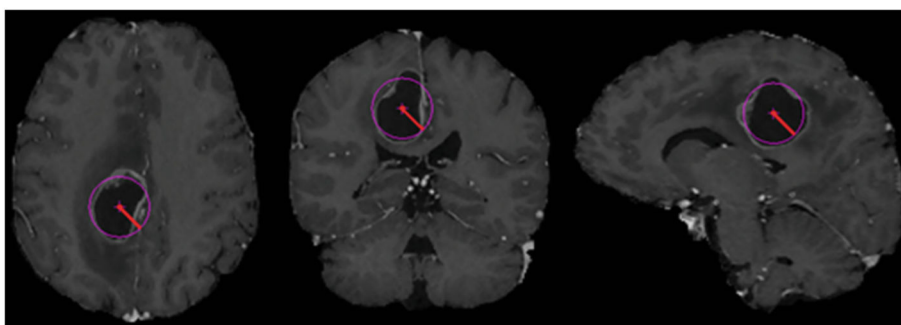
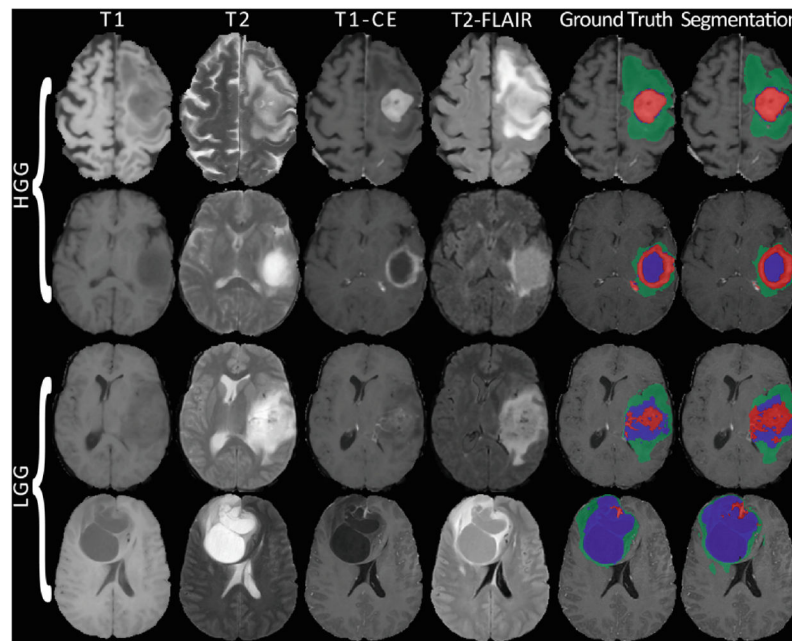
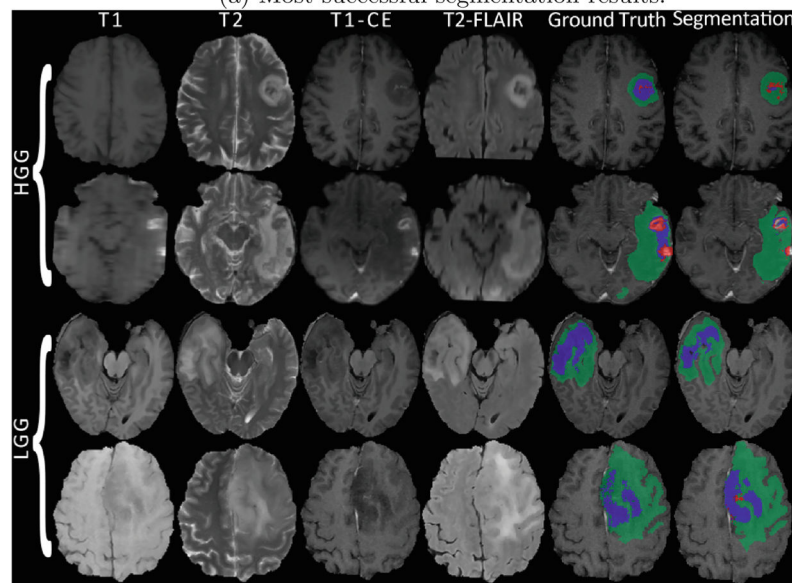


Fig. 1. Example of using a single seed-point and a radius to approximate the center and the bulk volume of a tumor by a sphere. The figures illustrate (from left to right) the axial, coronal and sagittal view of the same patient.



(a) Most successful segmentation results.



(b) Least successful segmentation results.

Fig. 2.

Examples for four LGG and four HGG patients. Green, red and blue masks denote the edema, the enhancing tumor and the union of the necrotic and non-enhancing parts of the tumor, respectively (Color figure online).

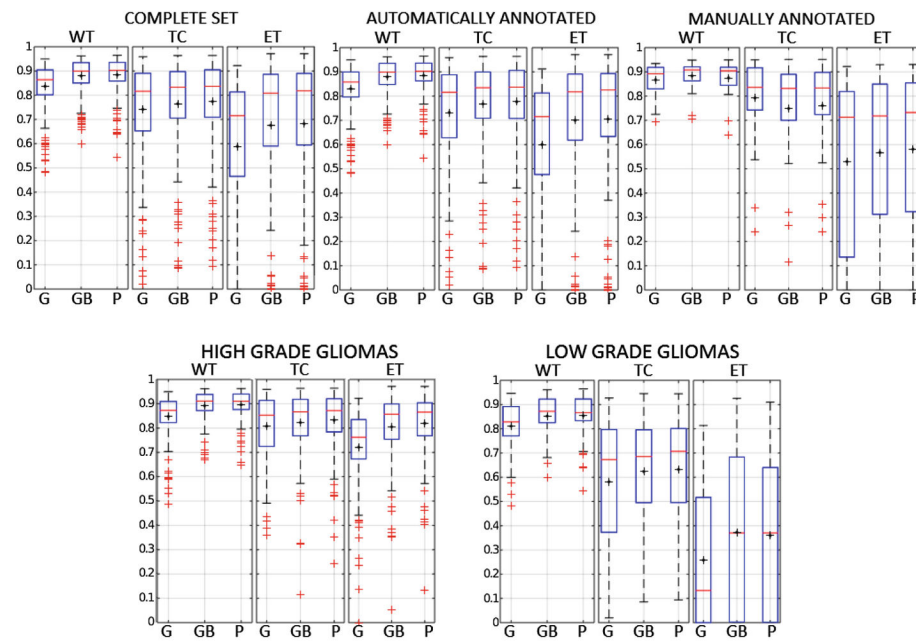


Fig. 3. Distributions of the DICE score across patients for each step (G: GLISTR, GB: gradient boosting, P: proposed) of the proposed method, each tissue label and different groupings of data. The black cross and the red line inside each box denote the mean and median values, respectively (Color figure online).

Table 1

Mean and median values of the DICE score for each step of the proposed method, each tissue label and different groupings of data.

| Data | Method | Dice score (mean) | | | Dice score (median) | | |
|---------------------------------|-----------|-------------------|---------------|---------------|---------------------|---------------|---------------|
| | | WT | TC | ET | WT | TC | ET |
| Complete training set (n=186) | GLISTR | 83.7 % | 74.2 % | 58.6 % | 86.4 % | 81.6 % | 71.6 % |
| | GLISTR+GB | 87.9 % | 76.5 % | 67.6 % | 89.9 % | 83.3 % | 80.9 % |
| | Proposed | 88.4 % | 77.4 % | 68.2 % | 90.3 % | 83.7 % | 82 % |
| Automatically annotated (n=156) | GLISTR | 83.1 % | 73.2 % | 60.1 % | 85.8 % | 81.6 % | 71.6 % |
| | GLISTR+GB | 87.9 % | 76.8 % | 70.5 % | 89.9 % | 83.5 % | 82.6 % |
| | Proposed | 88.5 % | 77.7 % | 71 % | 90.3 % | 83.7 % | 82.8 % |
| Manually annotated (n=30) | GLISTR | 86.7 % | 79.2 % | 52.9 % | 89.2 % | 83.6 % | 71.3 % |
| | GLISTR+GB | 88.3 % | 74.8 % | 56.7 % | 90.8 % | 83.2 % | 72.6 % |
| | Proposed | 87.6 % | 76.1 % | 58.1 % | 90.5 % | 83.4 % | 75.7 % |
| All HGGs (n=132) | GLISTR | 84.7 % | 80.8 % | 72.2 % | 87.3 % | 85.3 % | 76.2 % |
| | GLISTR+GB | 89.1 % | 82.3 % | 80.5 % | 91 % | 86.7 % | 85.7 % |
| | Proposed | 89.6 % | 83.2 % | 82 % | 91 % | 87.2 % | 86.5 % |
| All LGGs (n=54) | GLISTR | 81.1 % | 58.1 % | 25.8 % | 82.8 % | 67.3 % | 12.8 % |
| | GLISTR+GB | 85.2 % | 62.4 % | 37.3 % | 87.2 % | 68.6 % | 36.9 % |
| | Proposed | 85.4 % | 63.2 % | 35.9 % | 86.9 % | 70.7 % | 37.3 % |