HHS Public Access

Author manuscript

Comput Med Imaging Graph. Author manuscript; available in PMC 2021 April 12.

Published in final edited form as:

Comput Med Imaging Graph. 2021 March ; 88: 101828. doi:10.1016/j.compmedimag.2020.101828.

Analyzing magnetic resonance imaging data from glioma patients using deep learning

Bjoern Menze[a,*], **Fabian Isensee**[b], **Roland Wiest**[c], **Bene Wiestler**[d], **Klaus Maier-Hein**[b], **Mauricio Reyes**[e], **Spyridon Bakas**[f]

[a]Quantitative Biomedicine, University of Zurich, Zurich, Switzerland [b]DKFZ, Heidelberg, Germany [c]Support Center for Advanced Neuroimaging, Institute of Diagnostic and Interventional Neuroradiology, Inselspital, Bern, Switzerland [d]Neuroradiology, TUM, Munich, Germany [e]Data Science Center, Inselspital, Bern, Switzerland [f]Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA, USA

Abstract

The quantitative analysis of images acquired in the diagnosis and treatment of patients with brain tumors has seen a significant rise in the clinical use of computational tools. The underlying technology to the vast majority of these tools are machine learning methods and, in particular, deep learning algorithms. This review offers clinical background information of key diagnostic biomarkers in the diagnosis of glioma, the most common primary brain tumor. It offers an overview of publicly available resources and datasets for developing new computational tools and image biomarkers, with emphasis on those related to the Multimodal Brain Tumor Segmentation (BraTS) Challenge. We further offer an overview of the state-of-the-art methods in glioma image segmentation, again with an emphasis on publicly available tools and deep learning algorithms that emerged in the context of the BraTS challenge.

Keywords

NeuroOncology; Glioma; Brain tumor; Machine learning; Image segmentation; Image quantification; Deep learning; Brain tumor segmentation challenge; BraTS

1. Diagnosing glioma patients from image information

Predicting clinical variables of interest, such as tumor molecular characteristics (Bakas et al., 2020a, 2017b; Fathi Kazerooni et al., 2020b; Binder et al., 2018), treatment response (Akbari et al., 2020), or prognosis (Fathi Kazerooni et al., 2020a; Bakas et al., 2020b) from imaging data has attracted great attention from the neuro-oncology community in recent years. These developments have in part been accelerated by the 2016 WHO classification

[*]Corresponding author. bjoern.menze@uzh.ch (B. Menze).

Author statement

All authors contributed to the manuscript, and approved it.

Conflicts of interest: All authors declare they have no interests conflicting with this manuscript.

protocol for grading glioma patients (Louis et al., 2016), that is based on a neuropathological evaluation of glioma tissue from a biopsy or resection, as it is biologically highly plausible that the driving genomic changes behind the molecular tissue changes can be non-invasively identified in the imaging phenotype.

Compared to the gold standard assessment from neuropathology, radiographic imaging-based evaluation of glioma biomarkers offers two key advantages: (i) Spatial heterogeneity within the tumor can easily be assessed and (ii) longitudinal assessment of changes becomes possible without a need for serial biopsies. To this end, a broad variety of computational tools has been developed, most of them relying on machine learning techniques, that support the increasingly complex visual analysis of the multivariate and longitudinal image data acquired in glioma patients.

1.1. Glioma imaging

Decisions about the diagnosis and therapy of brain tumor require optimal information, as treatment options are limited and those that might be considered, such as radiation therapy or tumor resection, may have life-changing side effects. To this end, a multitude of standard morphological, functional, and metabolic imaging modalities are used (Fig. 1), often in repeated exams with intervals as short as a few month time. Analysing these data adequately poses significant challenges both in clinical practice, when standard imaging protocols are used, as well as when testing advanced imaging sequences and searching for new imaging biomarkers.

Standard imaging protocols and biomarkers—During the last three decades, magnetic resonance imaging (MRI) remained the fundamental imaging technology for the diagnosis and localization of cerebral gliomas. The key objectives of MRI are the characterization of the tumor category and its differential diagnosis (e.g., brain abscess, lymphoma, or metastasis) with implications for clinical decision making about the path of care, planning of the most effective therapy regimen and disease monitoring under therapy. The quality of the MRI exams is dependent on many factors as, e.g., field strength, MRI sequence composition, scanner type, slice thickness and image contrast, with a considerable lack of standardization of scanner protocols. Further, no clear recommendations exist for the application of advanced neuroimaging techniques that encompass metabolic characterization (MR-Spectroscopy), vascularization and blood-brain-barrier deficiency (perfusion imaging), or tissue composition (diffusion-weighted imaging).

Therefore, joint consensus recommendations have been proposed by the United States National Brain Tumor Society (NBTS), the Society for Neuro-oncology (SNO), and the European Organisation for Research and Treatment of Cancer (EORTC). The proposed glioma imaging protocol (EORTC-NBTS) encompasses a standardized set of recommendations for anatomical MRI sequences to assess changes in tumor burden as an imaging endpoint in clinical trials. The minimum requirements for glioma imaging require consistent scanning with an 1.5 T or 3 T MRI scanner with the same imaging parameters are: (1) 3-dimensional T1 weighted imaging before and (2) after the administration of Gadolinium-based contrast enhancements, (3) a 2-D acquisition of fluid attenuated inversion

recovery sequences (FLAIR), (4) an axial 2-D T2-weighted sequence, and (5) an axial 2-D diffusion weighted image (DWI). The acquisition of (6) additional T1-weighted spin echo sequences may further improve the detection of tumor recurrence associated with delayed enhancement (Kaufmann et al., 2020). All sequences display structural properties of the tumor, with T2 weighted scans highlighting the tissue water of the edema, and the T1 contrast-enhanced images showing areas of active tumor growth, where Gad-enriched blood is leaking into the tissue. DWI sequences, that include diffusion tensor imaging, offer insights into areas with modified tissue micro-structure, for example, when additional water or tumor cells increase tissue cellularity, or limit the natural anisotropy of tissue water diffusion. Most data sets discussed in the following (Section 2) encompass standard imaging parameters (1)–(4), but resources for (5)–(6) are also discussed (Section 2.3).

The integration of the 3D imaging sequences into the standard imaging protocol – e.g., (1) and (2), but also optional 3D T2 and diffusion scans – are essential for the volumetric assessment of tumor progression and the application for radiographic response assessment imaging criteria. Conventional 2-D measurements must be considered inadequate to longitudinally assess complex tumor geometry. Slice thickness and rotation induced directly affect the accuracy of bi-dimensional diameter measures errors of >30% (Schmitt et al., 2013; Reuter et al., 2014). In contrast, AI-supported volumetric measurements nowadays yield an excellent accuracy compared to conventional tumor response assessment (see Section 3). Limitations of the (qualitative) response assessment with conventional MRI encompass the lack of sensitivity to distinguish therapy related effects as pseudo-progression and radiation necrosis from true progression at the earliest time that is possible.

Advanced imaging sequences and biomarkers—To improve the diagnostic accuracy, conventional MRI can be complemented by advanced neuroimaging techniques that increase the diagnostic accuracy by evaluating changes in blood flow or in metabolic patterns.

Perfusion-weighted imaging is most frequently used (85%) as an advanced neuroimaging technique to distinguish between low- and high-grade gliomas and between true progression and pseudo-progression. Dynamic susceptibility contrast-enhanced (DSC) MRI is employed in the majority of centers, whereas T1-based dynamic contrast-enhanced (DCE) or a combination of both techniques is frequently restricted to comprehensive cancer centers. Both DSC and DCE-MRI show an excellent diagnostic accuracy to discriminate between high- and low grade gliomas and provide complementary information to discriminate tumor recurrence from therapy related effects. In a recent systematic review, encompassing 27 studies and 298 patients, the pooled sensitivity, specificity and sensitivity for the differentiation between high- and low grade tumors was 0.93, 0.90 and 0.96 and between tumor relapse and treatment-related changes was 0.88; 0.86 and 0.89 (Okuchi et al., 2019). A previous meta-analysis that investigated the discriminative power of DSC-perfusion and DCE perfusion MRI to separate viable tumor from treatment-related effects revealed similar results reported comparable results for both methods (Patel et al., 2017). However, they reported also considerable variability in threshold definitions and a lack of standardisation that hampers the implementation of quantitative perfusion MRI strategies across institutions. Intra-voxel incoherent motion (IVIM) and arterial spin labeling (ASL) perfusion

complement the spectrum of perfusion imaging techniques (Le Bihan, 2019; Paschoal et al., 2018; Grade et al., 2015). The principal advantage of both techniques is their non-invasive character without the need to administer contrast. While IVIM estimates microcirculation in randomly oriented capillaries mimicking a pseudo-diffusion process, ASL uses radiofrequency pulses that saturate water protons to magnetically tag water molecules in the arterial blood. Both methods have shown some potential in tumor grading and outcome prediction, but remain still in an experimental stage.

Magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) are less frequently used in routine glioma imaging protocols, but also represent long-standing application domains for machine learning research, both for automating the signal processing (Menze et al., 2008; Das et al., 2017) and the diagnostic evaluation of the spectral data (Menze et al., 2006). Clinical indications encompass the characterization of a lesion, the differentiation between glial neoplasms and imaging phenotypes that mimic gliomas and the differentiation between true vs. pseudo-progression (Ryken et al., 2014; van Dijken et al., 2017; Hollingworth et al., 2006). More recently, 2-hydroxyglutarate MRS has been proposed as a promising non-invasive method to discriminate between isocitrate-dehydrogenase (IDH)-mutant and IDH- wild type gliomas (van den Bent et al., 2010; Verma et al., 2016). Mutations in IDH are highly prevalent among gliomas of lower grade (70–80%) and carry a better prognosis in grade III gliomas (Verma et al., 2016). A recent meta-analysis reported a higher sensitivity of 2-hydroxyglutarate MRS to differentiate between IDH-mutated and IDH-wild type gliomas than diffusion or perfusion imaging or localization-based features (Suh et al., 2018, 2019). The method has a great potential to overcome limitations of invasive biopsies related to intra-tumoral heterogeneity and subsequent biases of biopsy-based genomic analysis.

Using multi-parametric MRI—Overall, MRI remains a key methodology to diagnose and monitor patients with cerebral gliomas. Conventional structural imaging provides a basis for lesion stratification, treatment planning and monitoring and provides the basis for automated image analysis of lesion progression, predictive monitoring, and radiomic feature extraction (Zwanenburg et al., 2020).

Advanced imaging techniques complement the standard imaging workup and support the analysis and prediction of physiological and molecular characteristics reflected through the burden of disease and the potential of response to therapy. Moreover, with imaging revealing critical information about tumor and surrounding anatomy in the patient, glioma imaging is also always a crucial component in personalizing treatment decisions.

Recent work by Lipkova et al. (Lipkova et al., 2019), for example, used the patient's glioma images together with a tumor growth model and a Bayesian machine learning framework to predict patient-specific tumor cell density with credible intervals from multimodal imaging data as a basis for personalized radiotherapy design.

1.2. Glioma imaging, radiomic profiling, and tumor biology

Classifying glioma by genotypes—Traditionally, gliomas were classified according to their histo-morphological relationship to the glial cell lineages as outlined in the 2007 WHO

classification (Louis et al., 2007). However, this morphology-derived classification had several challenges for its scientific and clinical use. Most importantly, these were the rater-dependence of diagnoses (demonstrated for example by Kros et al. (Kros et al., 2007)) and – closely related – the ambiguity of morphologic appearance, leading to the definition of bucket diagnoses such as “oligo-astrocytoma”, which harbored tumors showing morphologic features reminiscent of both astrocytomas and oligodendrogliomas. With the advent of large-scale, high-throughput techniques for (epi)genome-wide analysis of tumors, several landmark discoveries have been made in many types of cancer, also in gliomas (Brennan et al., 2013; Brat et al., 2015; Sturm et al., 2012; Barthel et al., 2019). Chief among them was the discovery of point mutations in the isocitrate dehydrogenase 1 (IDH1) and to a lesser extent also in the isocitrate dehydrogenase 2 (IDH2) genes (Parsons et al., 2008; Bakas et al., 2018b). While they are rare (<10%) in WHO grade IV glioblastoma, the most malignant gliomas, they are a defining genomic event in WHO grade II and III gliomas (Yan et al., 2009), present in the majority of the tumors. Later analyses found that mutant IDH1 and IDH2 catalyze the production of the onco-metabolite 2-hydroxyglutarate (2HG), which is not present in healthy cells. Subsequently, it was discovered that 2HG is essential for the formation of IDH mutant gliomas (Lai et al., 2011) and leads to dramatic changes in the tumor epigenome (Sturm et al., 2012). Today, it is accepted that IDH-mutant and IDH-wildtype tumors are biologically different tumor entities, despite their similar (and sometimes indistinguishable) histo-morphologic appearance. Importantly, these tumors also significantly differ in their clinical course and treatment response, with IDH mutant tumors carrying a far better prognosis (Yan et al., 2009).

Later genome-wide studies used this knowledge about the impact of epigenetic changes for tumor development and progression and the relative stability of the epigenome (at least compared to mRNA expression analysis) for (un)supervised classification of gliomas (Wiestler et al., 2014; Sturm et al., 2012) and the identification of key molecular features in these subgroups. Collectively, these studies yielded a clear picture of a biology-driven classification of gliomas across WHO grades and entities. These findings have also resulted in a revised WHO classification released in 2016 (Louis et al., 2016), which now groups tumors according to a more integrative schema including their geno-typical characteristics, instead of based solely on their histological phenotype, as previously. This leads to situations where the genotype (e.g., presence of an IDH mutation as well as co-deletion of the short arm of chromosome 1 and the long arm of chromosome 19, 1p/19q codeletion) disagrees with the histological phenotype (in the above example an astrocytoma, which typically do not carry 1p/19q codeletion). In these cases, the genotype “beats” the phenotype and underlines the importance of tumor biology for classification.

Radio(geno)mic correlations—Identifying associations between the glioma genotype and their (MR) imaging phenotype has become an important field of research, often referred to as radiogenomics, and concepts from machine learning have contributed significantly to the analysis of the glioma imaging data. Initial studies focused on identifying key genomic alterations such as IDH mutation or 1p/19q codeletion from preoperative imaging data (see for example the studies by Kickingereder et al. (2015), Eichinger et al. (2017), or Chang et al. (2018)). Meanwhile, the focus of glioma image analysis shifts towards assessing

heterogeneity or predicting clinical course using radiomic image features: In a 2017 paper published in the AJNR (Chang et al., 2017), the authors present a classifier trained to predict cellularity from voxel-wise regression analysis of FLAIR, ADC and T1c data on a set of 91 stereotactically localized biopsies. With this approach, they are able to calculate non-invasive, voxel-wise cellularity maps, which might for example aid the process of selection suitable biopsy locations, or enter a tumor modeling approach as in Menze et al. (2011), Lipkova et al. (2019). Along this line, Hu et al. (2017) extended this concept to predicting spatial heterogeneity of key molecular alterations in glioblastoma. From another data set of 48 targeted biopsies, they built multivariate decision-tree models to predict spatial presence of copy-number alterations of genes like the epidermal growth factor receptor (*EGFR*) in preoperative MR images.

The state-of-the-art analysis of glioma images—Overall, a growing body of research studies in the literature is not aiming only at quantifying tumor structures visible in the multi-parametric MR images, but at stepping beyond analyzing the apparent visual content to identify sub-visual cues using quantitative radiomic features (Suter et al., 2020). Uncovering correlations between image features with biological (molecular, genetic), as well as clinical variables is the ultimate goal of this work. In this, the segmentation and quantification of “semantic” tumor sub-compartments/structures, i.e., regions that can be named and associated with clear properties and even function within the tumor area, remains the first and most crucial step of any subsequent radiomic correlation study.

2. Benchmarking glioma image quantification in the BRATS challenge

Clinical imaging protocols routinely-acquired for the diagnosis of glioma cases reveal information about the anatomical structure, function, vascularization, among other tissue characteristics and standardization of the protocols across centers is only a recent development (Section 1). As a consequence, in the initial efforts for quantifying tumor image information, locally varying imaging datasets were used for developing machine learning algorithms from various researchers. This variation, and the absences of publically available reference data, prohibited a fair comparative algorithmic evaluation. This issue was overcome by the start of the annual “Multimodal Brain Tumor Segmentation Challenge” (BraTS, www.braintumorsegmentation.org) in 2012 (Menze et al., 2014; Bakas et al., 2018a, 2017a), which initiated the creation of a benchmarking environment and the collection, expert annotation, and distribution of a large clinically-acquired multi-institutional dataset of multi-parametric MRI scans from patients diagnosed with gliomas. To the best of our knowledge, the BraTS dataset is one of the largest publicly available curated datasets for glioma imaging and has been heavily used for computer vision and machine learning research. Although it has been initiated with image segmentation in mind, its establishment and recognition of its clinical relevance and potential has offered avenues to a multitude of the glioma image analysis tasks outlined above.

2.1. The public BraTS dataset

Image data—The clinically-acquired and -curated data made available during the BraTS challenges has grown significantly from including a few dozens of cases in 2012 to almost

one thousand cases that are available in 2020. The complete BraTS dataset originated from a handful of institutions while its 2020 edition integrates data from 19 independent international institutions, as well as collections from The Cancer Imaging Archive (TCIA) (Clark et al., 2013). The majority of cases utilized in more recent instances of the BraTS challenge are acquired pre-operatively, but time-series with pre- and post-interventional follow-up scans have also been made available through BraTS in the past (BRATS 2016 data set). Cases originating from TCIA are accompanied by clinical, genetic, and pathological data, available via the Genomic Data Commons (GDC) Data Portal. Further scans are also available from TCIA and can be used in conjunction with the BraTS data.

All MRI data distributed through the BraTS challenges have been acquired by scanners with magnetic field strength of 1–3 T, and comprise for all patients and timepoints the basic structural imaging sequences: native T1-weighted (T1); T1 after administration of contrast agent (T1c); T2-weighted (T2); and T2 Fluid-Attenuated Inversion Recovery (FLAIR). To standardize the highly variable resolutions and orientations of the acquired scans, all data have been consistently rigidly aligned to the same brain anatomical template (Rohlfing et al., 2010) and interpolated to 1 mm^3 isotropic resolution. Furthermore, in compliance with institutional requirements for anonymization, all imaging scans have undergone brain extraction prior to their public release.

Expert annotations of semantic structures—These multi-parametric MRI datasets are accompanied by tumor boundary annotations for each of the histologically distinct tumor sub-structures (Menze et al., 2014; Bakas et al., 2018a), generated and approved by clinical experts following a dedicated harmonized protocol. These labels comprise regions reflecting edematous/infiltrated tissue, as well as the potentially resectable tumor core with solid/enhancing, necrotic, and/or fluid filled compartments (Fig. 2). These labels have been considered as three independent sets of binary areas: (i) the whole tumor are (described by the union of all the classes), (ii) the tumor core area (with all tumor core labels, excluding the peritumoral edematous/infiltrated tissue), and (iii) the area of the active part of the tumor (being represented by the contrast-enhancing compartment). As these three areas have a direct translation into “semantically” meaningful tumor quantification tasks, and since assigning weights to inter-class classification errors is difficult in the current medical use case, they are evaluated during the BraTS challenge individually rather than a straightforward multi-class evaluation.

In addition to the experts’ image annotations, a subset of the 2018–2020 datasets have been accompanied by clinical information on patient age and patient overall survival (Bakas et al., 2018a).

Availability—The BraTS data are available for general use under a creative commons license and can be downloaded from links available on the challenge website.[1] A clear distinction should be made to the subset used in different years, e.g., BraTS 2016 or BraTS 2020, as the focus of the evaluation changed in time as did the composition of the training,

[1] www.braintumorsegmentation.org.

validation, and testing partitions of the data. For example, the 2020 data have only pre-operative scans, while the 2016 data have both pre- and post-interventional scans.

2.2. The BRATS benchmarking challenges

Evaluating image segmentation algorithms—The BraTS datasets and the annual BraTS benchmarking challenges, running in conjunction with the conference on Medical Image Computing and Computer-Assisted Interventions (MICCAI), have been instrumental in spear-heading the development of brain tumor segmentation algorithms. Specifically, numerous algorithms (between 10 in 2012 and more than 60 in 2019) have been comparatively evaluated every year. During the challenge workshop data are provided to the BraTS participants for training, validation, as well as for testing, and then the participants are required to submit their predicted segmentation labels for evaluation by the BraTS online evaluation portal. At the same time, brief descriptions of the algorithms appear in the challenge's proceedings of each annual workshop, offering insights into how the algorithms differed in design, training parameters, or implementation configurations. The evaluation system calculates different scores, such as the Dice similarity coefficient, the Hausdorff distance, or volumetric mismatch for the longitudinal task, and the algorithms are scored according to these scores. As the question of how to find the one (or a few) top performing methods using multiple scores and across the three different segmentation tasks remains an open research question, various ranking heuristics have been used during the years.

Top performing algorithms—Among the algorithms that were considered to be top performing since BraTS 2012, different approaches had been employed, reflecting the technological advances in the domain of computer vision during the past 10 years. While the first algorithms from 2012 still employed generative probabilistic models (Menze et al., 2010, 2012a, 2015), or level-set based segmentation (Riklin-Raviv et al., 2010; Raviv et al., 2012), there has been a rapid adoption of machine learning approaches using random forest together with local image features. While already in 2012 a number of algorithms using random forest classifiers had been tested (Zikic et al., 2012; Geremia et al., 2012; Menze et al., 2012b; Stefan Bauer et al., 2012)), a random forest algorithm was not winning the challenge before 2013 (Tustison et al., 2013). The first deep learning methods were used in 2014 (Davy et al., 2014; Havaei et al., 2017; Zikic et al., 2014; Urban et al., 2014), although not yet winning the challenge even during the following year (BraTS 2015), where a hybrid generative-discriminative approach integrative of a biophysical tumor growth model to account for mass-effect was the winner, namely GLISTRboost (Bakas et al., 2015; Zeng et al., 2016). The first winning deep learning algorithm was during BraTS 2016 (Chang, 2016), which was based on a deep neural network with residual connections. From 2014 onwards, a steady improvement of the algorithmic performances could be observed (as measured in the BraTS testing datasets), and a continuously increasing number of the participating algorithms have become available as open source implementations. Algorithms that perform at the top for the BRATS 2020 data set reach Dice scores of up to 90 Beyond benchmarking The BraTS data have been commonly used for benchmarking algorithms beyond the focused settings of the official evaluation schema. Subsets of the BraTS data have been used in the Medical Segmentation Decathlon (Simpson et al., 2019), as well as in the Quantification of Uncertainties in Biomedical Image Quantification (QUBIQ) Challenge – with published

baseline performances and an online evaluation portal – and they can further be used in private evaluation setups. This has fostered the development of brain tumor image quantification in a variety of medical image analysis and computer vision research projects, while glioma segmentation offers an example of early clinical translation of machine learning – and, more specifically deep learning – technology in a quantitative diagnostic evaluation of biomedical images (examples of which will be given in the following section).

2.3. Publicly available data beyond BraTS

Beyond the curated, annotated, publicly-available multi-institutional mpMRI scans of glioma patients released as part of the BraTS challenge, various additional data collections of glioma patients are becoming available, providing radiological, histopathological, molecular, and clinical information. Nevertheless a substantial joint computational-clinical effort would be required to curate and annotate these datasets given a harmonized protocol.

These collections, primarily hosted at TCIA (Clark et al., 2013), describe data of glioblastoma patients (TCGA-GBM[2] (Scarpace et al., 2016), QIN GBM Treatment Response[3] (Prah et al., 2015; Jafari-Khouzani et al., 2015), CPTAC-GBM[4] (CPTAC, 2018), IvyGAP[5] (Puchalski et al., 2018; Nameeta et al., 2016), ACRIN-FMISO-Brain[6] (Gerstner et al., 2016; Ratai et al., 2018; Kinahan et al., 2018), Brain--Tumor-Progression[7] (Schmainda and Prah, 2018), RIDER NEURO MRI[8] (Barboriak, 2015), UPENN-GBM-ICMR[9] (Bakas and Davatzikos, 2020), and ReSPOND (Davatzikos et al., 2020)), as well as lower grade glioma patients (TCGA-LGG[10] (Pedano et al., 2016) and LGG-1p19q-Deletion[11] (Akkus et al., 2017; Erickson et al., 2017)), and datasets combining all grades of gliomas (REMBRANDT[12] (Scarpace et al., 2015), QIN--BRAIN-DSC-MRI[13] (Schmainda et al., 2016), and GLASS[14] (GLASS Consortium, 2018; Bakas et al., 2020a), which is hosted in Synapse[15]).

3. State of the art deep learning segmentation

Offering a concurrent overview of deep learning-based brain segmentation methods related to glioma patients has become increasingly difficult with the advent of easy-to-train approaches based on deep learning architectures, such as the U-Net (Ronneberger et al., 2015) and its variation. To this end, we will focus on a set of methods that performed well in recent BraTS challenges, i.e., those that prove to perform well in a controlled setting. We

[2] https://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM.
[3] https://wiki.cancerimagingarchive.net/display/Public/QIN+GBM+Treatment+Response.
[4] https://wiki.cancerimagingarchive.net/display/Public/CPTAC-GBM.
[5] https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=22515597.
[6] https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=33948305.
[7] https://wiki.cancerimagingarchive.net/display/Public/Brain-Tumor-Progression.
[8] https://wiki.cancerimagingarchive.net/display/Public/RIDER+NEURO+MRI.
[9] https://wiki.cancerimagingarchive.net/display/Public/UPENN-GBM-ICMR.
[10] https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG.
[11] https://wiki.cancerimagingarchive.net/display/Public/LGG-1p19qDeletion.
[12] https://wiki.cancerimagingarchive.net/display/Public/REMBRANDT.
[13] https://wiki.cancerimagingarchive.net/display/Public/QIN-BRAIN-DSC-MRI.
[14] https://www.glass-consortium.org/.
[15] https://www.synapse.org/glass.

attempt to identify – and report – common design choices, as well as differences between methods that may be linked to their strong performance on this type of segmentation task.

3.1. Image pre-processing

Brain tumor MRI typically consists of multiple MRI modalities. The BraTS dataset (Menze et al., 2014; Bakas et al., 2017a, 2018a) provides the most common ones: T1, T1c, T2, and FLAIR. Initial preprocessing steps are image registration followed by brain extraction (Smith et al., 2004; Jenkinson et al., 2012; Isensee et al., 2019; Kleesiek et al., 2014; Thakur et al., 2020).

Intensity harmonization—MRI intensities are notoriously non-standardized (Russell et al., 2014; Nyúl et al., 2000). In the pre-deep-learning era, methods were susceptible to potentially nonlinear intensity shifts resulting in a need for histogram matching or more sophisticated normalization methods (Kleesiek et al., 2014; Russell et al., 2014; Nyúl et al., 2000) to be applied. Bias field correction (Tustison et al., 2010) was regularly used to remove inhomogeneities in the images. Interestingly, this issue was alleviated with the emergence of deep learning techniques. While the previous random forest approaches used local image features, which limits the 'receptive field' that contributes to the inference at a voxel, an appropriately chosen network architecture 'sees' the full image and may be better suited to recognize – and ignore – large scale intensity changes that result from bias fields. To this end one might argue that deep learning techniques are more robust with respect to non-standardized intensity values and inhomogeneities, possibly due to their superior capacity, their end-to-end training forcing the extraction of robust feature representations and the application of data augmentation techniques. The now by far most prevalent intensity normalization technique is z-scoring, where the brain region of the images is normalized by subtracting its mean and dividing by its standard deviation (Kamnitsas et al., 2016, 2017b; Isensee et al., 2017, 2018; Myronenko, 2018; Jiang et al., 2019; McKinley et al., 2019; Jungo et al., 2018b).

Spatial harmonization—In particular when working with multi-institutional data and data originating from different MRI scanners, the voxel spacings can be heterogeneous and must be homogenized for processing with convolutional neural networks. Selecting a proper target spacing for resampling can be crucial for downstream performance. A large spacing will result in lower image resolution, making it easier to capture sufficient contextual information at the cost of reduced details in the resulting segmentation maps. A lower voxel spacing retains more details in images and segmentations but make the segmentation problem inherently more difficult due to the required larger input sizes. The training cases provided by the BraTS challenge were resampled to a common voxel spacing of 1 mm^3 isotropic resolution, which strikes a good balance between image size and resolution.

3.2. Network architectures

According to our experience, the quality of a segmentation architecture is related to the expressiveness of the features it can learn, the amount of contextual information it can encode and how well it can upscale semantically rich low resolution representations to a full resolution segmentation.

Early architectures—DeepMedic (Kamnitsas et al., 2017a, 2016) was arguably one of the first tremendously successful brain tumor segmentation architectures and the first to be applied in 3D. As compared to newer architectures, it has a substantially smaller receptive field and processes patches of, for example, only 25 × 25 × 25 voxels. It makes up for the missing contextual information by using an additional feature extraction stream that processes downscaled image patches. Features of low and high resolution patches are recombined shortly before the segmentation decision is made. Other early architectures include Pereira's (Pereira et al., 2016) and Kleesiek's CNNs (Urban et al., 2014), and Dvorak's patch-based prediction of structured labels (Dvorak and Menze, 2015).

Current U-Net architectures—Since the introduction of the U-Net in 2015 (Ronneberger et al., 2015), its derivatives define the state of the art in medical image segmentation, not just in brain tumor segmentation. U-Net consists of an encoder and a decoder network which are inter-connected by skip connections. The encoder follows a similar construction as classical image classification networks (Krizhevsky et al., 2012; Simonyan et al., 2014; He et al., 2016; Huang et al., 2017) by alternating feature extraction (convolutional layers) with down-sampling, thus successively aggregating the semantic information necessary for generating good segmentations. When passing through the encoder, feature representations need to undergo spatial pooling to enable subsequent convolutional kernels to effectively cover greater areas of the image. As a side effect, the feature maps suffer from an increasingly low spatial resolution. The purpose of the decoder is to upscale this semantic information under consideration of higher resolution feature maps forwarded to it from the encoder via skip connections. U-Net thus elegantly recombines spatial with semantic information at multiple locations throughout the decoder.

Even though standard U-Net-like architectures are competitive in brain tumor segmentation (Kamnitsas et al., 2017a; Isensee et al., 2018), most recent approaches make use of a variety of modifications. Common U-Net variations include the addition of residual blocks (He et al., 2016; Myronenko, 2018; Jiang et al., 2019; Zhou et al., 2018; Isensee et al., 2017; Wang et al., 2017) or densely connected convolutional layers (Huang et al., 2017; Zhao et al., 2019; McKinley et al., 2018, 2019; Jungo et al., 2018b). Attention mechanisms (Vaswani et al., 2017) can guide the network towards focusing of relevant spatial locations (McKinley et al., 2019), whereas squeeze and excitation (Hu et al., 2018; Zhou et al., 2018) can adaptively recalibrate feature responses. Various successful approaches (Wang et al., 2017; McKinley et al., 2018) replace some pooling operations in the encoder (and their mirrored up-sampling operations in the decoder) with dilated convolutions (Chen et al., 2017) to bypass down- and up-sampling and thus potential interpolation artifacts. The winning contributions to BraTS 2018 and 2019 used additional decoder branches with auxiliary tasks (Jiang et al., 2019; Myronenko, 2018) to regularize the networks.

3.3. Training scheme

Loss functions—The Dice loss (Milletari et al., 2016; Drozdzal et al., 2016) is the most popular loss for brain tumor segmentation. It directly optimizes the Dice score, the most widely used evaluation metric. Recently, the focal loss (Lin et al., 2017) has gained in popularity and was used by multiple highly successful algorithms (Zhao et al., 2019;

McKinley et al., 2019, 2018). With the target regions of BraTS being the whole tumor, tumor core, and enhancing tumor rather than the three semantic classes, better results are typically achieved by optimizing these regions directly. Thus, many algorithms use the sigmoid function as the final nonlinearity in their architecture and compute the loss relative to the respective ground truth regions computed from the provided labels (Jiang et al., 2019; Zhao et al., 2019; McKinley et al., 2019, 2018; Isensee et al., 2018; Myronenko, 2018) or optimize the regions one after the other in a cascaded approach (Wang et al., 2017).

Regularization—Even though BraTS provides a large publicly available dataset for training, regularization plays a pivotal role in obtaining a good performing brain tumor segmentation algorithm. Data augmentation techniques are widespread, with mirroring and additive, as well as multiplicative intensity shifts being most prominent (Jiang et al., 2019; Zhao et al., 2019; McKinley et al., 2019; Isensee et al., 2018, 2017). Auxiliary tasks seem to be particularly effective on the BraTS dataset, with the winning contribution from 2018 using a separate VAE decoder branch to reconstruct the network input (Myronenko, 2018) and the winning entry in 2019 utilizing two decoders with different up-sampling strategies (Jiang et al., 2019). The authors of Zhao et al. (2019) report that concurrent learning of the target regions, as well as the semantic classes improve the segmentation performance. The loss function can also be used to improve the robustness of the network. For example, (Isensee et al., 2018) have demonstrated that the segmentation performance increases by using the combination of the cross-entropy and Dice loss as learning objective.

Considering uncertainty—The generation of manual reference annotations for glioma segmentation is tedious and can contain small errors that the networks should ideally ignore. To tackle this problem, inspired by developments in uncertainty estimation for segmentation in brain imaging (Jungo et al., 2018a; Jungo and Reyes, 2019), while (McKinley et al., 2018) developed a label-flip uncertainty loss, which enables the network to recognize potentially mislabeled voxels and reduce their influence during training.

3.4. Inference

Ensemble predictions—Ensembling plays a fundamental role in competitive segmentation performance, as powerfully demonstrated by Kamnitsas et al. (2017b) in their BraTS 2017 entry. Virtually all concurrent segmentation methods make use of some sort of ensembling, whether that is training multiple times with different random seeds (Myronenko, 2018; Jiang et al., 2019), using the models resulting from cross-validation (Isensee et al., 2017, 2018; Jiang et al., 2019), using weights from different epochs of the same network training (Jiang et al., 2019) or applying the network in different orientations (McKinley et al., 2019). Test time data augmentation in the form of mirroring and applying the model to different crops of the image and accumulating the predictions can also be used to boost the performance at the cost of an increased inference time.

Postprocessing rules—Postprocessing is often used in the BraTS dataset to boost a models performance with the challenges specific validation scheme in mind. For example, Isensee et al. (2018) observed that the accumulated Dice score for the enhancing tumor region improved when removing the enhancing tumor entirely from a predicted image if less

than some volume threshold of enhancing tumor was predicted. This is due to BraTS awarding a Dice score of 1, if no false positive voxels are predicted in an image that does not contain the label in its reference segmentation. Even though this strategy also removed some true positive predictions, the net gain outweights the losses. This strategy was also applied by the BraTS 2019 winning contribution (Jiang et al., 2019). It should be noted that such a postprocessing, while useful for BraTS, may not be desired in a clinical setup when false positive predictions are less consequential than false negatives. As an alternative post-processing, McKinley et al. (2019) proposed further heuristic modifications of connected components, some of which leverage the uncertainty estimation of their label-flip loss.

3.5. Clinical translation beyond BraTS

Therapy response assessment is a critical aspect of monitoring treatment success in glioma patients. Due to inter-observer variability in the two-dimensional nature of the clinical state of the art (Wen et al., 2010), the variance of two independent measurements is high. Kickingereder et al. (2019) developed a segmentation method that, even though it was trained on 455 MRI scans from only a single institution, generalized well not only to other MRI scans from that same institution but more importantly across a large multi-institutional cohort (34 different institutions comprising a total of 2034 MRI scans). Quantifying tumor response based on their segmentation masks proved to be significantly more reliable and more reproducible than the clinical state of the art. Their pretrained segmentation model is publicly available[16] and can be used as fully functioning segmentation tool (Fig. 3).

4. Going beyond tumor boundaries

In an automated workflow of glioma patients, automated segmentation of tumor targets, as well as other structures such as organs at risk, and resection cavities holds the potential to reduce interobserver variability and accelerate the delineation process, leading to a more efficient and effective clinical workflow. Current state of the art approaches based on deep learning technologies have largely focused on brain tumor segmentation from multisequence MRI, tailored to neuroradiology tasks, and mainly on pre-operative scenarios (Menze et al., 2014; Bakas et al., 2018a; Kickingereder et al., 2019). While the automated segmentation offers great avenues for an objective neuro-radiogical evaluation of disease activity, beside glioblastoma (Kickingereder et al., 2019) is it has established itself in a similar evaluation of Multiple Sclerosis patients (Gabr et al., 2020), the image segmentation algorithms are also well suited to support radiation therapy planning.

Delineating gliomas for radiotherapy planning

The European Organization for Radiation Therapy in Cancer (EORTC) is a large and active network of researchers and clinicians working towards improvement of cancer patient treatment As part of a study performed by the EORTC RTQA group and the Emmanuel van der Schueren Fellowship Program for Quality Assurance in Clinical Trials, the delineation review process in an ongoing multicenter phase III trial was conducted with an accrual goal of 750 glioblastoma patients. Before participating centers could enter patients into the trial,

[16] https://github.com/NeuroAI-HD/HD-GLIO.

each center had to complete a glioblastoma delineation benchmark case exercise, which was used to assess the interobserver variability of experts in a clinical trial. Results of this ongoing work indicate that despite the availability of delineation guidelines, glioblastoma delineation is subject to significant interobserver variability (Kaidar-Person et al., 2019). Studies have shown that non-adherence to protocol-specified radiotherapy requirements is frequent in prospective clinical trials, with major deviation rates ranging from 11.8% to 48.0%. Similarly, retrospective analyses of EORTC intergroup trials on low grade glioma and anaplastic glioma patients revealed that erroneous delineation of target volumes and organs at risk is a common cause of protocol deviations in clinical trials on brain tumor patients (Fairchild et al., 2012; André et al., 2018)).

Although radiotherapy for glioblastoma is considered a standard treatment, there is considerable delineation variability even among experts: Only moderate agreement, with a mean kappa of 0.58 among GTVs was observed in a delineation study, including fifteen panels of radiation oncologists from independent institutions (Wee et al., 2016). In radiotherapy planning, combining deep learning-based auto--segmentation and manual contour adjustment resulted in superior accuracy, consistency and efficiency for CTV delineation in a retrospective study of post-operative lung cancer patients (Bi et al., 2019). Similarly, deep learning-based auto-segmentation helped to improve CTV consistency for breast (Men et al., 2018) and rectal cancer (Men et al., 2017), and may thus streamline the radiotherapy workflow. However, no similar studies have been conducted for glioblastoma. Previous work has either focused on a mathematical description to take presumed microscopic tumor growth into account (Unkelbach et al., 2014a,b) or used non standard MRI, like DTI (Peeken et al., 2019) to derive novel representations of a CTV definition. Both approaches are still highly investigational and not applicable to current standard of care radiotherapy for glioblastoma in clinical trials.

Localizing organs at risk nearby the tumor

A first deep learning-based segmentation of Organs at Risk (OAR) from MRI has recently been proposed by Mlynarski et al., who presented preliminary results of a CNN-based deep learning approach to segment organs at risk from a single T1-weighted MR image. This indicates that a deep learning-based segmentation will also be capable of auto-segmenting organs at risk in multi-parametric MRI (Mlynarski et al., 2020). In line with the findings of Mlynarski et al. (2020), an effective OAR auto-segmentation approach needs to provide anatomically consistency results where shape prior of segmented of organs is incorporated. A different approach, presented in Orasanu et al. (2018), demonstrates the use of a classical atlas-based approach, with results refined via a deep learning-based contour detector working on triangular mesh representations. Similarly, the authors in Agn et al. (2019) proposed an approach that utilizes a deformable tetrahedral atlas of the brain and structures within a contrast-adaptive generative model for whole-brain segmentation and OAR segmentation. The approach also incorporates a tumor regularization using a conditional restricted Boltzmann machine. Differently from other approaches, the method in Agn et al. (2019) is based on a generative model and is designed to handle differences in imaging protocols. Interestingly, the evaluation of the approach not only relied on standard metrics such as Dice and Hausdorff distances, but also on metrics derived from dose volume

histograms. Overall, the challenge when segmenting OAR is to attain a good balance between injected shape prior and image content driving the segmentation process. This is particularly important when segmenting cases where tumor mass effect can dramatically lower shape prior information, learned by a deep learning model.

In summary, tumor segmentation methods in radiation therapy have offer assistance in gross tumor delineation, as well as probabilistic margin definition and their conformance to natural anatomical barriers such as optic chiasm/nerve, brainstem interface, and falx cerebelli, cerebri, as well as the skull, as these preclude or limit the spread of glioblastoma. This is a future direction of research and a future application domain of brain tumor image segmentation.

Acknowledgments

This work was partly supported by the National Institutes of Health (NIH) under award numbers NIH/NINDS: R01NS042645, NIH/NCI: U24CA189523, NIH/NCI: U01CA242871. The content of this publication is solely the responsibility of the authors and does not represent the official views of the NIH. We gladly acknowledge the support of the Swiss Cancer League (grant number KFS-3979-08-2016), and of the Swiss Personalized Health Network (SPHN) Imagine project.

References

Agn M, et al., 2019. A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning. Med. Image Anal 54, 220–237. [PubMed: 30952038]

Akbari H, et al., 2020. Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma. Cancer 126 (11), 2625–2636. [PubMed: 32129893]

Akkus Z, et al., 2017. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. J. Digit. Imaging 30 (4), 469–476. [PubMed: 28600641]

Abrunhosa-Branquinho André, N., et al., 2018. Radiotherapy quality assurance for the RTOG 0834/EORTC 26053–22054/NCIC CTG CEC. 1/CATNON intergroup trial “concurrent and adjuvant temozolomide chemotherapy in newly diagnosed non-1p/19q deleted anaplastic glioma”: individual case review analysis. Radiother. Oncol 127 (2), 292–298. [PubMed: 29606522]

Bakas S, et Int, Davatzikos C, 2020. The university of pennsylvania glioblastoma imaging, clinical, molecular, and radiomics (UPenn-GBM-ICMR) collection. Sci. Data (in press).

Bakas S, et al., 2015. GLISTRboost: combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. BrainLes 2015 144–155.

Bakas S, et al., 2017a. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data 4, 170117. [PubMed: 28872634]

Bakas S, et al., 2017b. In vivo detection of EGFRvIII in glioblastoma via perfusion magnetic resonance imaging signature consistent with deep peritumoral infiltration: the φ-index. Clin. Cancer Res 23 (16), 4724–4734. [PubMed: 28428190]

Bakas S, et al., 2018a. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge (arXiv preprint). arXiv:1811.02629.

Bakas S, et al., 2018b. Non-invasive in vivo signature of IDH1 mutational status in high grade glioma, from clinically-acquired multi-parametric magnetic resonance imaging, using multivariate machine learning. Neuro-Oncology 20 (Suppl. 6), vi184–vi185.

Bakas S, et al., 2020a. iGLASS: imaging integration into the Glioma Longitudinal AnalySiS Consortium. Neuro-Oncology

Bakas S, et al., 2020b. Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): advanced radiomic features may compensate for lack of advanced MRI modalities. J. Med. Imaging 7 (3), 031505.

Barboriak D, 2015. Data from rider neuro MRI. Cancer Imaging Arch

Barthel FP, et al., 2019. Longitudinal molecular trajectories of diffuse glioma in adults. Nature 576 (7785), 112–120. [PubMed: 31748746]

Bi N, et al., 2019. Deep learning improved clinical target volume contouring quality and efficiency for postoperative radiation therapy in non-small cell lung cancer. Front. Oncol 9, 1192. [PubMed: 31799181]

Binder ZA, et al., 2018. Epidermal growth factor receptor extracellular domain mutations in glioblastoma present opportunities for clinical imaging and therapeutic development. Cancer Cell 34 (1), 163–177. [PubMed: 29990498]

Brat DJ, et al., 2015. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Engl. J. Med 372 (26), 2481–2498. [PubMed: 26061751]

Brennan CW, et al., 2013. The somatic genomic landscape of glioblastoma. Cell 155 (2), 462–477. [PubMed: 24120142]

Chang PD, 2016. Fully convolutional deep residual neural networks for brain tumor segmentation. International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries 108–118.

Chang PD, et al., 2017. A multiparametric model for mapping cellularity in glioblastoma using radiographically localized biopsies. AJNR Am. J. Neuroradiol 38 (5), 890–898. [PubMed: 28255030]

Chang P, et al., 2018. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. AJNR Am. J. Neuroradiol 39 (7), 1201–1207. [PubMed: 29748206]

Chen L-C, et al., 2017. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell 40 (4), 834–848. [PubMed: 28463186]

Clark K, et al., 2013. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging 26 (6), 1045–1057. [PubMed: 23884657]

CPTAC, 2018. National cancer institute clinical proteomic tumor analysis consortium (CPTAC). Radiology data from the clinical proteomic tumor analysis consortium glioblastoma multiforme [CPTAC-GBM] collection [Data set]. Cancer Imaging Arch 10.7937/k9/tcia.2018.3rje41q1.

Das D, et al., 2017. Quantification of metabolites in magnetic resonance spectroscopic imaging using machine learning. International Conference on Medical Image Computing and Computer-Assisted Intervention 462–470.

Davatzikos C, et al., 2020. AI-based prognostic imaging biomarkers for precision neuro-oncology: the ReSPOND consortium. Neuro-Oncology 22 (6), 886–888. [PubMed: 32152622]

Davy A, et al., 2014. Brain tumor segmentation with deep neural networks. Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge)

Drozdzal M, et al., 2016. The importance of skip connections in biomedical image segmentation. Deep Learning and Data Labeling for Medical Applications Springer, pp. 179–187.

Dvorak P, Menze B, 2015. Conditional value-at-risk for general loss distributions. Proc MICCAI-BRATS 13–24.

Eichinger P, et al., 2017. Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas. Sci. Rep 7 (1), 13396. [PubMed: 29042619]

Erickson B, et al., 2017. Data from LGG-1p19qDeletion. Cancer Imaging Arch

Fairchild A, et al., 2012. EORTC Radiation Oncology Group quality assurance platform: establishment of a digital central review facility. Radiother. Oncol 103 (3), 279–286. [PubMed: 22633815]

Fathi Kazerooni A, et al., 2020a. Cancer imaging phenomics via CaPTk: multi-institutional prediction of progression-free survival and pattern of recurrence in glioblastoma. JCO Clin. Cancer Inform 4, 234–244. [PubMed: 32191542]

Fathi Kazerooni A, et al., 2020b. Imaging signatures of glioblastoma molecular characteristics: a radiogenomics review. J. Magn. Reson. Imaging 52 (1), 54–69. [PubMed: 31456318]

Gabr RE, et al., 2020. Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: a large-scale study. Mult. Scler 10, 1217–1226.

Geremia E, Menze BH, Ayache N, et al., 2012. Spatial decision forests for glioma segmentation in multi-channel MR images. MICCAI Challenge on Multimodal Brain Tumor Segmentation 34.

Gerstner ER, et al., 2016. ACRIN 6684: assessment of tumor hypoxia in newly diagnosed glioblastoma using 18F-FMISO PET and MRI. Clin. Cancer Res 22 (20), 5079–5086. [PubMed: 27185374]

GLASS Consortium, 2018. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the glioma longitudinal analysis consortium. Neuro-Oncology 20 (7), 873–884. [PubMed: 29432615]

Grade M, et al., 2015. A neuroradiologist's guide to arterial spin labeling MRI in clinical practice. Neuroradiology 57 (12), 1181–1202. [PubMed: 26351201]

Havaei M, et al., 2017. Brain tumor segmentation with deep neural networks'. Med. Image Anal 35, 18–31. [PubMed: 27310171]

He K, et al., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778.

Hollingworth W, et al., 2006. A systematic literature review of magnetic resonance spectroscopy for the characterization of brain tumors. AJNR Am. J. Neuroradiol 27 (7), 1404–1411. [PubMed: 16908548]

Hu LS, et al., 2017. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. Neuro-Oncology 19 (1), 128–137. [PubMed: 27502248]

Hu J, Shen L, Sun G, 2018. Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 7132–7141.

Huang G, et al., 2017. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 4700–4708.

Isensee F, et al., 2017. Brain tumor segmentation and radiomics survival prediction: contribution to the brats 2017 challenge. International MICCAI Brainlesion Workshop 287–297.

Isensee F, et al., 2018. No new-net. International MICCAI Brainlesion Workshop 234–244.

Isensee F, et al., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. Hum. Brain Mapp 40 (17), 4952–4964. [PubMed: 31403237]

Jafari-Khouzani K, et al., 2015. Repeatability of cerebral perfusion using dynamic susceptibility contrast MRI in glioblastoma patients. Transl. Oncol 8 (3), 137–146. [PubMed: 26055170]

Jenkinson M, et al., 2012. Fsl. Neuroimage 62 (2), 782–790. [PubMed: 21979382]

Jiang Z, et al., 2019. Two-stage cascaded U-net: 1st place solution to BraTS challenge 2019 segmentation task. International MICCAI Brainlesion Workshop 231–241.

Jungo A, Reyes M, 2019. Assessing reliability and challenges of uncertainty estimations for medical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019. Vol (in press). Lecture Notes in Computer Science

Jungo A, et al., 2018a. On the effect of inter-observer variability for a reliable estimation of uncertainty of medical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018. Vol (in press). Lecture Notes in Computer Science

Jungo A, et al., 2018b. Towards uncertainty-assisted brain tumor segmentation and survival prediction. In: Crimi A (Ed.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries Springer International Publishing, pp. 474–485. ISBN: 978-3-31975238-9.

Kaidar-Person O, et al., 2019. ESTRO ACROP consensus guideline for target volume delineation in the setting of postmastectomy radiation therapy after implant-based immediate reconstruction for early stage breast cancer. Radiother. Oncol 137 (August), 159–166. [PubMed: 31108277]

Kamnitsas K, et al., 2016. DeepMedic for brain tumor segmentation. International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries 138–149.

Kamnitsas K, et al., 2017a. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal 36, 61–78. [PubMed: 27865153]

Kamnitsas K, et al., 2017b. Ensembles of multiple models and architectures for robust brain tumour segmentation. International MICCAI Brainlesion Workshop 450–462.

Kaufmann TJ, et al., 2020. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. Neuro-Oncology 22 (6), 757–772. [PubMed: 32048719]

Kickingereder P, et al., 2015. IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. Sci. Rep 5, 16238. [PubMed: 26538165]

Kickingereder P, et al., 2019. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. Lancet Oncol 20 (5), 728–740. [PubMed: 30952559]

Kinahan P, et al., 2018. Data from ACRIN-FMISO-Brain. Cancer Imaging Arch 10.7937/K9/TCIA.2018.vohlekok.

Kleesiek J, et al., 2014. Ilastik for multi-modal brain tumor segmentation. Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge) 12–17.

Krizhevsky A, Sutskever I, Hinton GE, 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst 1097–1105.

Kros JM, et al., 2007. Panel review of anaplastic oligodendroglioma from European organization for research and treatment of cancer trial 26951: assessment of consensus in diagnosis, influence of 1p/19q loss, and correlations with outcome. J. Neuropathol. Exp. Neurol 66 (6), 545–551. [PubMed: 17549014]

Lai A, et al., 2011. Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin. J. Clin. Oncol 29 (34), 4482–4490. [PubMed: 22025148]

Le Bihan D, 2019. What can we see with IVIM MRI? Neuroimage 187, 56–67. [PubMed: 29277647]

Lin T-Y, et al., 2017. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision 2980–2988.

Lipkova J, et al., 2019. Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and Bayesian inference. IEEE Trans. Med. Imaging 38 (8), 1875–1884. [PubMed: 30835219]

Louis DN, et al., 2007. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 114 (2), 97–109. [PubMed: 17618441]

Louis DN, et al., 2016. The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 131 (6), 803–820. [PubMed: 27157931]

McKinley R, Meier R, Wiest R, 2018. Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation. International MICCAI Brainlesion Workshop 456–465.

McKinley R, et al., 2019. Triplanar ensemble of 3D-to-2D CNNs with label-uncertainty for brain tumor segmentation. International MICCAI Brainlesion Workshop 379–387.

Men K, Dai J, Li Y, 2017. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. Med. Phys 44 (12), 6377–6389. [PubMed: 28963779]

Men K, et al., 2018. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. Phys. Med 50, 13–19. [PubMed: 29891089]

Menze BH, et al., 2006. Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. NMR Biomed 19 (5), 599–609. [PubMed: 16642460]

Menze BH, et al., 2008. Mimicking the human expert: pattern recognition for an automated assessment of data quality in MR spectroscopic images. Magn. Reson. Med 59 (6), 1457–1466. [PubMed: 18421692]

Menze BH, et al., 2010. A generative model for brain tumor segmentation in multi-modal images. International Conference on Medical Image Computing and Computer-Assisted Intervention 151–159.

Menze BH, et al., 2011. A generative approach for image-based modeling of tumor growth. Biennial International Conference on Information Processing in Medical Imaging 735–747.

Menze BH, et al., 2012a. Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation. Proc MICCAIBRATS (Multimodal Brain Tumor Segmentation Challenge) 8.

Menze BH, et al., 2012b. Segmenting glioma in multi-modal images using a generative-discriminative model for brain lesion segmentation. Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge) 8.

Menze BH, et al., 2014. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34 (10), 1993–2024. [PubMed: 25494501]

Menze BH, et al., 2015. A generative probabilistic model and discriminative extensions for brain lesion segmentation-with application to tumor and stroke. IEEE Trans. Med. Imaging 35 (4), 933–946. [PubMed: 26599702]

Milletari F, Navab N, Ahmadi S-A, 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV) IEEE, pp. 565–571.

Mlynarski P, et al., 2020. Anatomically consistent CNN-based segmentation of organs-at-risk in cranial radiotherapy. J. Med. Imaging 7 (1), 014502.

Myronenko A, 2018. 3D MRI brain tumor segmentation using autoencoder regularization. International MICCAI Brainlesion Workshop 311–320.

Nameeta S, Feng X, Lankerovich M, et al., 2016. Data from Ivy GAP. The cancer imaging archive. Cancer Imaging Arch 10.7937/K9/TCIA.2016.XLwaN6nL.

Nyúl LG, Udupa JK, Zhang X, 2000. New variants of a method of MRI scale standardization. IEEE Trans. Med. Imaging 19 (2), 143–150. [PubMed: 10784285]

Okuchi S, et al., 2019. Diagnostic accuracy of dynamic contrast-enhanced perfusion MRI in stratifying gliomas: a systematic review and metaanalysis. Cancer Med 8 (12), 5564–5573. [PubMed: 31389669]

Orasanu E, et al., 2018. Organ-at-risk segmentation in brain MRI using model-based segmentation: benefits of deep learning-based boundary detectors. International Workshop on Shape in Medical Imaging 291–299.

Parsons DW, et al., 2008. An integrated genomic analysis of human glioblastoma multiforme. Science 321 (5897), 1807–1812. [PubMed: 18772396]

Paschoal AM, et al., 2018. Intravoxel incoherent motion MRI in neurological and cerebrovascular diseases. Neuroimage Clin 20, 705–714. [PubMed: 30221622]

Patel P, et al., 2017. MR perfusion-weighted imaging in the evaluation of highgrade gliomas after treatment: a systematic review and meta-analysis. Neuro-Oncology 19 (1), 118–127. [PubMed: 27502247]

Pedano N, et al., 2016. Radiology data from the cancer genome atlas low grade glioma [TCGA-LGG] collection. Cancer Imaging Arch 2.

Peeken JC, et al., 2019. Deep learning derived tumor infiltration maps for personalized target definition in Glioblastoma radiotherapy. Radiother. Oncol 138, 166–172. [PubMed: 31302391]

Pereira S, et al., 2016. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging 35 (5), 1240–1251. [PubMed: 26960222]

Prah MA, et al., 2015. Repeatability of standardized and normalized relative CBV in patients with newly diagnosed glioblastoma. Am. J. Neuroradiol 36 (9), 1654–1661. [PubMed: 26066626]

Puchalski RB, et al., 2018. An anatomic transcriptional atlas of human glioblastoma. Science 360 (6389), 660–663. [PubMed: 29748285]

Ratai E-M, et al., 2018. ACRIN 6684: multicenter, phase II assessment of tumor hypoxia in newly diagnosed glioblastoma using magnetic resonance spectroscopy. PLOS ONE 13 (6), e0198548. [PubMed: 29902200]

Raviv TR, Van Leemput K, Menze BH, 2012. Multi-modal brain tumor segmentation via latent atlases. Proc. MICCAIBRATS 64.

Reuter M, et al., 2014. Impact of MRI head placement on glioma response assessment. J. Neurooncol 118 (1), 123–129. [PubMed: 24566765]

Riklin-Raviv T, et al., 2010. Segmentation of image ensembles via latent atlases. Med. Image Anal 14 (5), 654–665. [PubMed: 20580305]

Rohlfing T, et al., 2010. The SRI24 multichannel atlas of normal adult human brain structure. Hum. Brain Mapp 31 (5), 798–819. [PubMed: 20017133]

Ronneberger O, Fischer P, Brox T, 2015. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 234–241.

Shinohara RT, et al., 2014. Statistical normalization techniques for magnetic resonance imaging. NeuroImage: Clin 6, 9–19. [PubMed: 25379412]

Ryken TC, et al., 2014. The role of imaging in the management of progressive glioblastoma: a systematic review and evidence-based clinical practice guideline. J. Neurooncol 118 (3), 435–460. [PubMed: 24715656]

Scarpace L, et al., 2015. Data from REMBRANDT. Cancer Imaging Arch 10, K9.

Scarpace L, et al., 2016. Radiology data from the cancer genome atlas glioblastoma multiforme [TCGA-GBM] collection. Cancer Imaging Arch 11 (4), 1.

Schmainda KM, Prah M, 2018. Data from brain-tumor-progression. Cancer Imaging Arch 10.7937/K9/TCIA.2018.15quzvnb.

Schmainda KM, et al., 2016. Glioma DSC-MRI perfusion data with standard imaging and ROIs. Cancer Imaging Arch 9 http://doi.org/10.7937K.

Schmitt P, et al., 2013. Effects of slice thickness and head rotation when measuring glioma sizes on MRI: in support of volume segmentation versus two largest diameters methods. J. Neurooncol 112 (2), 165–172. [PubMed: 23397270]

Simonyan K, Zisserman A, 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition (arXiv preprint). arXiv:1409.1556.

Simpson AL, et al., 2019. A Large Annotated Medical Image Dataset for the Development and Evaluation of Segmentation Algorithms (arXiv preprint). arXiv: 1902.09063.

Smith SM, et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219. [PubMed: 15501092]

Bauer Stefan, et al., 2012. Segmentation of brain tumor images based on integrated hierarchical classification and regularization. MICCAI BraTS Workshop. Nice: Miccai Society 11.

Sturm D, et al., 2012. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell 22 (4), 425–437. [PubMed: 23079654]

Suh CH, et al., 2018. 2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and metaanalysis using individual patient data. Neuro-Oncology 20 (12), 1573–1583. [PubMed: 30020513]

Suh CH, et al., 2019. Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: a systemic review and meta-analysis. Eur. Radiol 29 (2), 745–758. [PubMed: 30003316]

Suter Y, et al., 2020. Radiomics for glioblastoma survival analysis in pre-operative MRI: exploring feature robustness, class boundaries, and machine learning techniques. Cancer Imaging (in press).

Thakur S, et al., 2020. Brain extraction on MRI scans in presence of diffuse glioma: multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training. NeuroImage 117081. [PubMed: 32603860]

Tustison NJ, et al., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320. [PubMed: 20378467]

Tustison N, et al., 2013. Ants andarboles. Multimodal Brain Tumor Segmentation 47.

Unkelbach J, et al., 2014a. Radiotherapy planning for glioblastoma based on a tumor growth model: improving target volume delineation. Phys. Med. Biol 59 (3), 747. [PubMed: 24440875]

Unkelbach J, et al., 2014b. Radiotherapy planning for glioblastoma based on a tumor growth model: implications for spatial dose redistribution. Phys. Med. Biol 59 (3), 771. [PubMed: 24440905]

Urban G, et al., 2014. Multi-modal brain tumor segmentation using deep convolutional neural networks. MICCAI BraTS (Brain Tumor Segmentation) Challenge Proceedings, Winning Contribution 31–35.

van den Bent MJ, et al., 2010. IDH1 and IDH2 mutations are prognostic but not predictive for outcome in anaplastic oligodendroglial tumors: a report of the European organization for research and treatment of cancer brain tumor group. Clin. Cancer Res 16 (5), 1597–1604. [PubMed: 20160062]

van Dijken BRJ, et al., 2017. Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation in patients with high-grade glioma, a systematic review and meta-analysis. Eur. Radiol 27 (10), 4129–4144. [PubMed: 28332014]

Vaswani A, et al., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst 5998–6008.

Verma G, et al., 2016. Non-invasive detection of 2-hydroxyglutarate in IDHmutated gliomas using two-dimensional localized correlation spectroscopy (2D L-COSY) at 7 Tesla. J. Transl. Med 14 (1), 274. [PubMed: 27659543]

Wang G, et al., 2017. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. International MICCAI Brainlesion Workshop 178–190.

Wee CW, et al., 2016. Evaluation of variability in target volume delineation for newly diagnosed glioblastoma: a multi-institutional study from the Korean Radiation Oncology Group. Radiat. Oncol 10 (1), 137.

Wen PY, et al., 2010. Updated response assessment criteria for highgrade gliomas: response assessment in neuro-oncology working group. J. Clin. Oncol 28 (11), 1963–1972. [PubMed: 20231676]

Wiestler B, et al., 2014. Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma. Acta Neuropathol 128 (4), 561–571. [PubMed: 25008768]

Yan H, et al., 2009. IDH1 and IDH2 mutations in gliomas. N. Engl. J. Med 360 (8), 765–773. [PubMed: 19228619]

Zeng K, et al., 2016. Segmentation of gliomas in pre-operative and postoperative multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries 184–194.

Zhao Y-X, Zhang Y-M, Liu C-L, 2019. Bag of tricks for 3D MRI brain tumor segmentation. International MICCAI Brainlesion Workshop 210–220.

Zhou C, et al., 2018. Learning contextual and attentive information for brain tumor segmentation. International MICCAI Brainlesion Workshop 497–507.

Zikic D, et al., 2012. Context-sensitive classification forests for segmentation of brain tumor tissues. Proc. MICCAI-BRATS 1–9.

Zikic D, et al., 2014. Segmentation of brain tumor tissues with convolutional neural networks. Proceedings MICCAI-BRATS 36–39.

Zwanenburg A, et al., 2020. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295 (2), 328–338. [PubMed: 32154773]

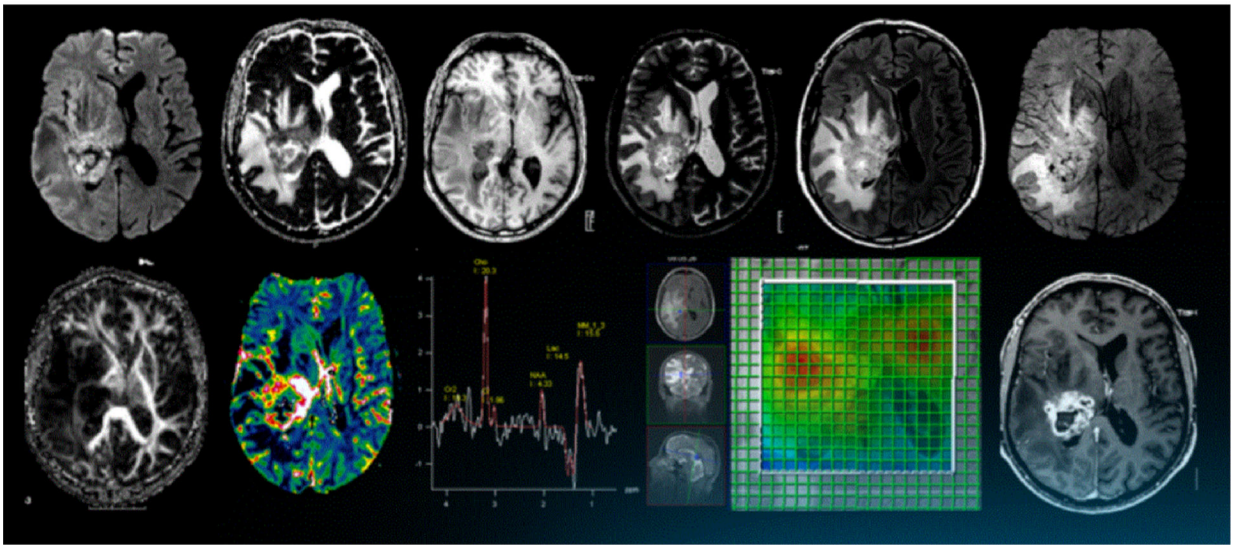

Fig. 1.

Extended glioma protocol with advanced imaging (exemplarily for the UniBe/SCAN protocol). The basic protocol consists of DWI/ADC, 3-D T1w, T2w, FLAIR and 3-D T1w contrast-enhanced sequences. Protocol extensions encompass DTI, SWI, DSC-perfusion and MR-spectroscopy (single voxel MR spectroscopy and MR spectroscopic imaging).

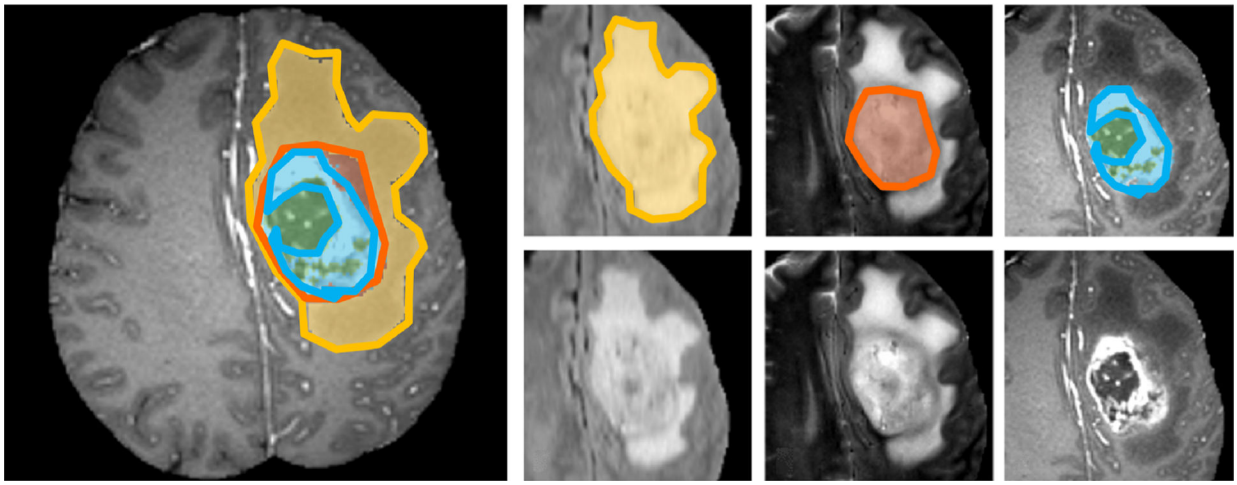

Fig. 2.

Semantic annotations available in the BRATS data set: Labels (shown in the left) summarize three semantic regions: whole tumor as visible from hyper-intense areas in T2w and FLAIR images (left column, yellow), the tumor core visible heterogenous signals in T2w MRI (central column, red), and the active tumor visible from intensity enhancements in post-Gd T1w scans (right column, blue).

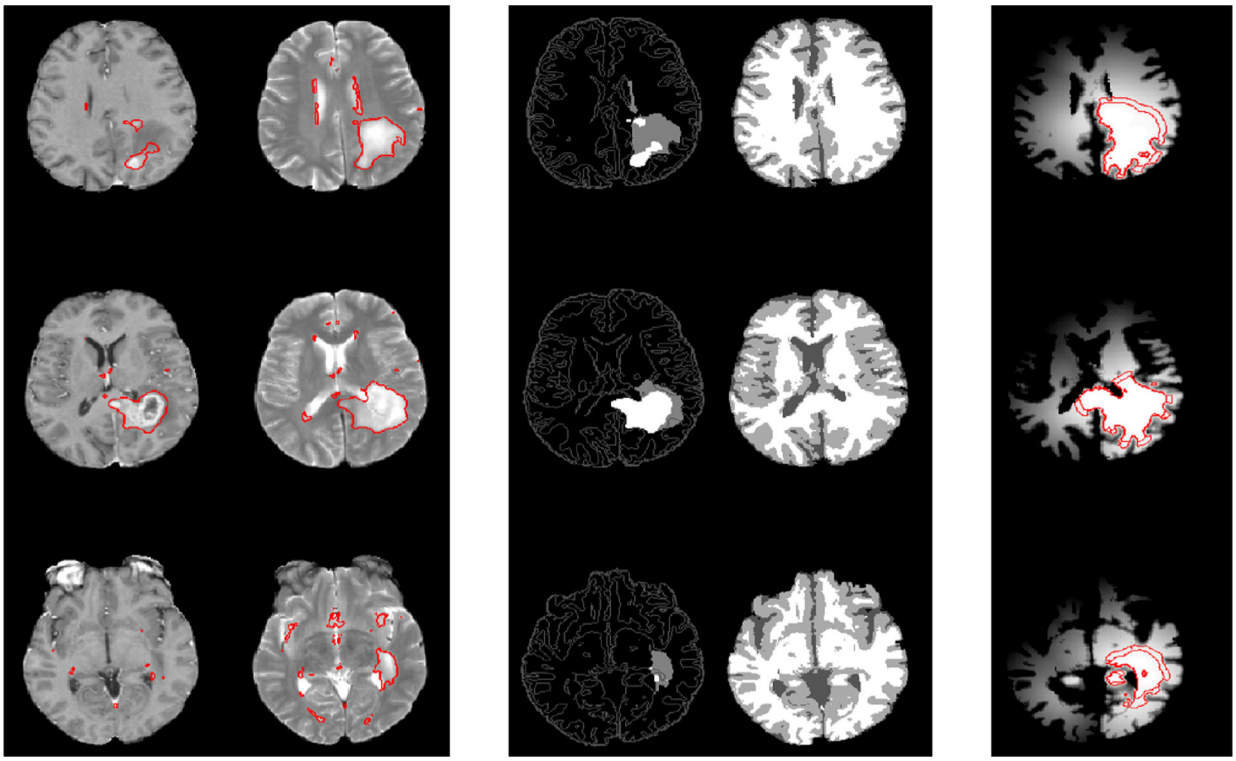

Fig. 3.

Segmentations of both the tumor and its sub-structures visible in the different image modalities (T1c and T2 images; left column) and the surrounding brain tissues (tumor and tissue segmentations, central column) are used as input to radiation treatment planning. Information of the patient specific brain anatomy are required to infer directions of most likely tumor cell infiltration using methods such as those by Lipkova et al. (2019) (infiltration maps with 5% and 20% infiltration isolines in red, right column).