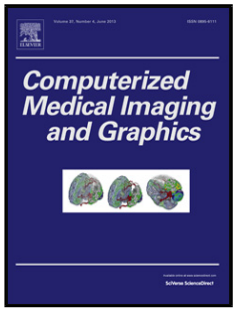

Analyzing Magnetic Resonance Imaging Data from Glioma Patients using Deep Learning

Bjoern Menze, Fabian Isensee, Roland Wiest, Bene Wiestler, Klaus Maier-Hein, Mauricio Reyes, Spyridon Bakas

PII: S0895-6111(20)30123-3

DOI: https://doi.org/10.1016/j.compmedimag.2020.101828

Reference: CMIG 101828

To appear in: *Computerized Medical Imaging and Graphics*

Received Date: 23 August 2020

Revised Date: 29 October 2020

Accepted Date: 18 November 2020

Please cite this article as: Bjoern Menze, Fabian Isensee, Roland Wiest, Bene Wiestler, Klaus Maier-Hein, Mauricio Reyes, Spyridon Bakas, Analyzing Magnetic Resonance Imaging Data from Glioma Patients using Deep Learning, *<![CDATA[Computerized Medical Imaging and Graphics]]>* (2020), doi: https://doi.org/

Analyzing Magnetic Resonance Imaging Data from Glioma Patients using Deep Learning

Bjoern Menze[*] **Fabian Isensee**[†] **Roland Wiest**[‡]
Bene Wiestler[§] **Klaus Maier-Hein**[¶] **Mauricio Reyes**[‖]
Spyridon Bakas[**]

October 29, 2020

Abstract

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.

Submission to Computerized Medical Imaging and Graphics, special issue on 'Cancer Imaging AI'

[*]Quantitative Biomedicine, University of Zurich, Zurich, Switzerland: *bjoern.menze@uzh.ch*

[†]DKFZ, Heidelberg, Germany: *f.isensee@dkfz-heidelberg.de*

[‡]Support Center for Advanced Neuroimaging, Institute of Diagnostic and Interventional Neuroradiology, Inselspital, Bern, Switzerland: *roland.wiest@insel.ch*

[§]Neuroradiology, TUM, Munich, Germany: *b.wiestler@tum.de*

[¶]DKFZ, Heidelberg, Germany: *k.maier-hein@dkfz-heidelberg.de*

[‖]Data Science Center, Inselspital, Bern, Switzerland: *mauricio.reyes@med.unibe.ch*

[**]Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA, USA: *sbakas@upenn.edu*

1 Diagnosing glioma patients from image information

Predicting clinical variables of interest, such as tumor molecular characteristics [14, 39, 20, 12], treatment response [5], or prognosis [38, 8] 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 protocol for grading glioma patients [74], 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 [62]. 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 (subsection 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 % [113, 104]. 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

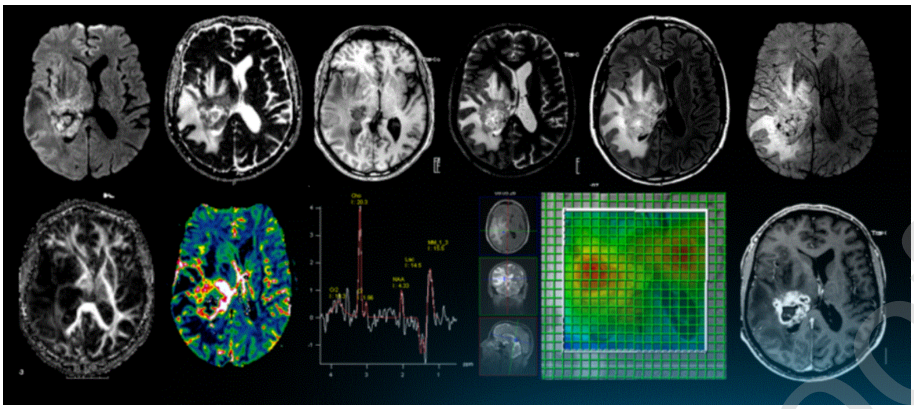

Figure 1: Extended glioma protocol with advanced imaging (exemplarily for the UniBe/SCAN protocol). The basic protocol consists of DWI/ADC, 3-D T1w, T2w, FLAIR and 3-D T1w contrast-enhanced sequences. Protocol extensions encompass DTI, SWI, DSC-perfusion and MR-spectroscopy (single voxel MR spectroscopy and MR spectroscopic imaging).

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 [92]. 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 [96]. 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 [70, 95, 44]. 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 spectro-

scopic 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 [79, 29] and the diagnostic evaluation of the spectral data [80]. 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 [108, 32, 47]. 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 [18, 129]. Mutations in IDH are highly prevalent among gliomas of lower grade (70–80%) and carry a better prognosis in grade III gliomas [129]. 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 [119, 120]. 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 [140].

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. [72], 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 [73]. 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. [68]) and - closely related - the ambiguity of morphologic appearance, leading to the definition of bucket diagnoses such as "oligoastrocytoma", 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 [22, 21, 118, 16]. 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 [94, 13]. 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 [134], 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 [69] and leads to dramatic changes in the tumor epigenome [118]. 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 [134].

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 [133, 118] 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 [74], which now groups tumors according to a more integrative schema including their genotypical 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. [63], Eichinger et al. [35], or Chang et al. [24]). 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 [23], 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 Lipkova et al. [83, 72]. Along this line, Hu et al. [49] 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 [121]. 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 [82, 7, 10], 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) [27]. 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 1-3T, 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 [106] and interpolated to 1mm^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 [82, 7], 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 [7].

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, 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 spearheading 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 [84, 85, 81], or level-set based segmentation [105, 103], 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 [138, 41, 86, 17]), a random forest algorithm was not winning the challenge before 2013 [124]. The first deep learning methods were used in 2014 [31, 45, 139, 127], 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 [11, 135]. The first winning deep learning algorithm was during BraTS 2016 [25],

[1] www.braintumorsegmentation.org

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 [116], 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 [27], describe data of glioblastoma patients (TCGA-GBM[2] [110], QIN GBM Treatment Response[3] [100, 54], CPTAC-GBM[4] [28], IvyGAP[5] [101, 90], ACRIN-FMISO-Brain[6] [42, 102, 65], Brain-Tumor-Progression[7] [111], RIDER NEURO MRI[8] [15], UPENN-GBM-ICMR[9] [9], and ReSPOND [30]), as well as lower grade glioma patients (TCGA-LGG[10] [97] and LGG-1p19q-Deletion[11] [6, 36]), and datasets combining all grades of gliomas (REMBRANDT[12] [109], QIN-BRAIN-DSC-MRI[13] [112], and

[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

GLASS[14] [43, 14], which is hosted in Synapse[15]).

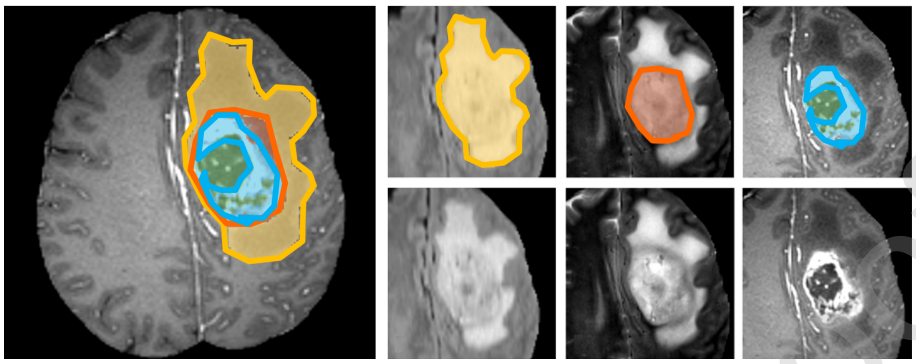

Figure 2: Semantic annotations available in the BRATS data set: Labels (shown in the left) summarize three semantic regions: whole tumor as visible from hyper-intense areas in T2w and FLAIR images (left column, yellow), the tumor core visible heterogenous signals in T2w MRI (central column, red), and the active tumor visible from intensity enhancements in post-Gd T1w scans (right column, blue).

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 [107] 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 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 [82, 10, 7] provides the most common ones: T1, T1c, T2, and FLAIR. Initial preprocessing steps are image registration followed by brain extraction [117, 55, 51, 66, 122].

Intensity harmonization MRI intensities are notoriously non-standardized [114, 91]. In the pre-deep-learning era, methods were susceptible to potentially nonlinear intensity shifts resulting in a need for histogram matching or

[14] https://www.glass-consortium.org/
[15] https://www.synapse.org/glass

more sophisticated normalization methods [66, 114, 91] to be applied. Bias field correction [123] 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 [60, 59, 52, 53, 89, 56, 76, 57].

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 1mm^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 [61, 60] 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 $25 \times 25 \times 25$ 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 [99] and Kleesiek's CNNs [127], and Dvorak's patch-based prediction of structured labels [34].

Current U-Net architectures Since the introduction of the U-Net in 2015 [107], 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 interconnected by skip connections. The encoder follows a similar construction as classical image classification networks [67, 115, 46, 50] 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 [61, 53], most recent approaches make use of a variety of modifications. Common U-Net variations include the addition of residual blocks [46, 89, 56, 137, 52, 130] or densely connected convolutional layers [50, 136, 75, 76, 57]. Attention mechanisms [128] can guide the network towards focusing of relevant spatial locations [76], whereas squeeze and excitation [48, 137] can adaptively recalibrate feature responses. Various sucessful approaches [130, 75, 75] replace some pooling operations in the encoder (and their mirrored up-sampling operations in the decoder) with dilated convolutions [26] 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 [56, 89] to regularize the networks.

3.3 Training scheme

Loss functions The Dice loss [87, 33] 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 [71] has gained in popularity and was used by multiple highly successful algorithms [136, 76, 75]. 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 [56, 136, 76, 75, 53, 89] or optimize the regions one after the other in a cascaded approach [130].

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 [56, 136, 76, 53, 52]. 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 [89] and the winning entry in 2019 utilizing two decoders with different up-sampling strategies [56]. The authors of [136] 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, [53] 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 [2, 1], while [75] 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 Kamnitsats et al. [59] 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 [89, 56], using the models resulting from cross-validation [52, 53, 56], using weights from different epochs of the same network training [56] or applying the network in different orientations [76]. 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. [53] 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 [56]. 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. [76] proposed further heuristic modifications of connected components, some of which leverage the uncertainty estimation of their label-flip loss.

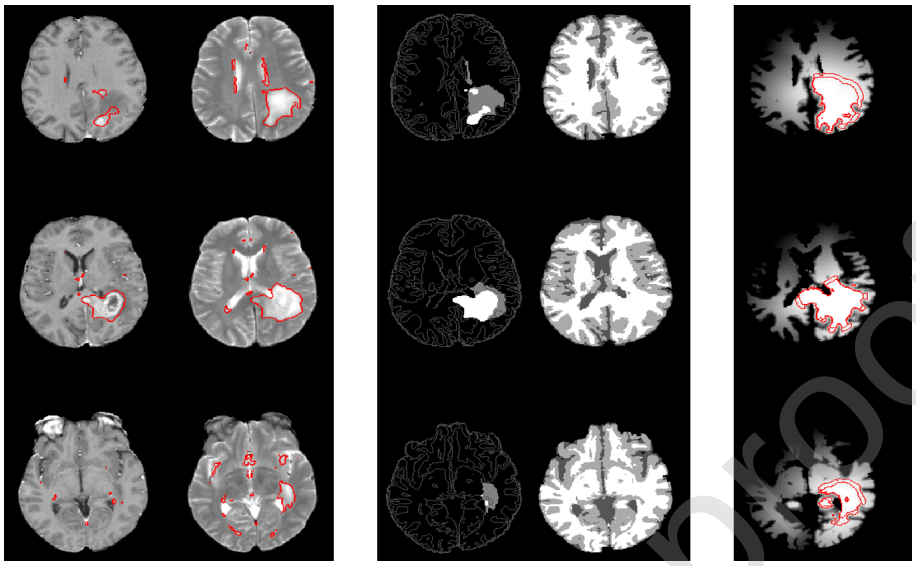

Figure 3: Segmentations of both the tumor and its sub-structures visible in the different image modalities (T1c and T2 images; left column) and the surrounding brain tissues (tumor and tissue segmentations, central column) are used as input to radiation treatment planning. Information of the patient specific brain anatomy are required to infer directions of most likely tumor cell infiltration using methods such as those by Lipkova et al. [72] (infiltration maps with 5% and 20% infiltration isolines in red, right column).

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 [132], the variance of two independent measurements is high. Kikingereder et al. [64] 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.

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

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 [82, 7, 64]. While the automated segmentation offers great avenues for an objective neuroradiogical evaluation of disease activity, beside glioblastoma [64] is it has established itself in a similar evaluation of Multiple Sclerosis patients [40], 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, 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 [58]. 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 [37, 3]).

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 [131]. 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 [19]. Similarly, deep learning-based auto-segmentation helped to improve CTV consistency for breast [78] and rectal cancer [77], 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 [126, 125] or used non standard MRI, like DTI [98] to de-

rive 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 [88]. In line with the findings of [88], 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 [93], 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 [4] 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 [4] 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

[1] Jungo A. and Reyes M. "Assessing Reliability and Challenges of Uncertainty Estimations for Medical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019*. Vol. In Press. Lecture Notes in Computer Science. 2019.

[2] Jungo A. et al. "On the Effect of Inter-observer Variability for a Reliable Estimation of Uncertainty of Medical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018*. Vol. In Press. Lecture Notes in Computer Science. 2018.

[3] André N Abrunhosa-Branquinho et al. "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". In: *Radiotherapy and Oncology* 127.2 (2018), pp. 292–298.

[4] Mikael Agn et al. "A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning". In: *Medical image analysis* 54 (2019), pp. 220–237.

[5] Hamed Akbari et al. "Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma". In: *Cancer* 126.11 (2020), pp. 2625–2636.

[6] Zeynettin Akkus et al. "Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence". In: *Journal of digital imaging* 30.4 (2017), pp. 469–476.

[7] Spyridon Bakas et al. "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge". In: *arXiv preprint arXiv:1811.02629* (2018).

[8] Spyridon Bakas et al. "Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): advanced radiomic features may compensate for lack of advanced MRI modalities". In: *Journal of Medical Imaging* 7.3 (2020), p. 031505.

[9] Spyridon Bakas, et Int, and Christos Davatzikos. "The University of Pennsylvania Glioblastoma Imaging, Clinical, Molecular, and Radiomics (UPenn-GBM-ICMR) collection". In: *Scientific data* In Press (2020).

[10] Spyridon Bakas et al. "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Scientific data* 4 (2017), p. 170117.

[11] Spyridon Bakas et al. "GLISTRboost: combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation". In: *BrainLes 2015*. Springer. 2015, pp. 144–155.

[12] Spyridon Bakas et al. "In vivo detection of EGFRvIII in glioblastoma via perfusion magnetic resonance imaging signature consistent with deep peritumoral infiltration: the φ-index". In: *Clinical Cancer Research* 23.16 (2017), pp. 4724–4734.

[13] Spyridon Bakas et al. "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". In: *Neuro-oncology* 20.suppl_6 (2018), pp. vi184–vi185.

[14] S Bakas et al. "iGLASS: Imaging integration into the Glioma Longitudinal AnalySiS Consortium". In: *Neuro-oncology* ().

[15] Daniel Barboriak. "Data from RIDER_NEURO_MRI". In: *The Cancer Imaging Archive* (2015).

[16] Floris P Barthel et al. "Longitudinal molecular trajectories of diffuse glioma in adults". In: *Nature* 576.7785 (2019), pp. 112–120.

[17] Stefan Bauer et al. "Segmentation of brain tumor images based on integrated hierarchical classification and regularization". In: *MICCAI BraTS Workshop. Nice: Miccai Society*. 2012, p. 11.

[18] M. J. van den Bent et al. "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". In: *Clin. Cancer Res.* 16.5 (2010), pp. 1597–1604.

[19] Nan Bi et al. "Deep Learning Improved Clinical Target Volume Contouring Quality and Efficiency for Postoperative Radiation Therapy in Non-small Cell Lung Cancer". In: *Frontiers in oncology* 9 (2019), p. 1192.

[20] Zev A Binder et al. "Epidermal growth factor receptor extracellular domain mutations in glioblastoma present opportunities for clinical imaging and therapeutic development". In: *Cancer cell* 34.1 (2018), pp. 163–177.

[21] D. J. Brat et al. "Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas". In: *N. Engl. J. Med.* 372.26 (2015), pp. 2481–2498.

[22] C. W. Brennan et al. "The somatic genomic landscape of glioblastoma". In: *Cell* 155.2 (2013), pp. 462–477.

[23] P. D. Chang et al. "A Multiparametric Model for Mapping Cellularity in Glioblastoma Using Radiographically Localized Biopsies". In: *AJNR Am J Neuroradiol* 38.5 (2017), pp. 890–898.

[24] P. Chang et al. "Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas". In: *AJNR Am J Neuroradiol* 39.7 (2018), pp. 1201–1207.

[25] Peter D Chang. "Fully convolutional deep residual neural networks for brain tumor segmentation". In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer. 2016, pp. 108–118.

[26] Liang-Chieh Chen et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 834–848.

[27] Kenneth Clark et al. "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository". In: *Journal of digital imaging* 26.6 (2013), pp. 1045–1057.

[28] CPTAC. "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]". In: *The Cancer Imaging Archive* (2018). DOI: 10.7937/k9/tcia.2018.3rje41q1.

[29] Dhritiman Das et al. "Quantification of metabolites in magnetic resonance spectroscopic imaging using machine learning". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2017, pp. 462–470.

[30] Christos Davatzikos et al. "AI-based prognostic imaging biomarkers for precision neuro-oncology: the ReSPOND consortium". In: *Neuro-oncology* 22.6 (2020), pp. 886–888.

[31] Axel Davy et al. "Brain tumor segmentation with deep neural networks". In: *Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge)* (2014).

[32] B. R. J. van Dijken et al. "Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation in patients with high-grade glioma, a systematic review and meta-analysis". In: *Eur Radiol* 27.10 (2017), pp. 4129–4144.

[33] Michal Drozdzal et al. "The importance of skip connections in biomedical image segmentation". In: *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 179–187.

[34] Pavel Dvorak and Bjoern Menze. "Conditional value-at-risk for general loss distributions". In: *Proc MICCAI-BRATS* (2015), pp. 13–24.

[35] P. Eichinger et al. "Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas". In: *Sci Rep* 7.1 (2017), p. 13396.

[36] Bradley Erickson et al. "Data from LGG-1p19qDeletion". In: *The Cancer Imaging Archive* (2017).

[37] Alysa Fairchild et al. "EORTC Radiation Oncology Group quality assurance platform: establishment of a digital central review facility". In: *Radiotherapy and Oncology* 103.3 (2012), pp. 279–286.

[38] Anahita Fathi Kazerooni et al. "Cancer Imaging Phenomics via CaPTk: Multi-Institutional Prediction of Progression-Free Survival and Pattern of Recurrence in Glioblastoma". In: *JCO Clinical Cancer Informatics* 4 (2020), pp. 234–244.

[39] Anahita Fathi Kazerooni et al. "Imaging signatures of glioblastoma molecular characteristics: A radiogenomics review". In: *Journal of Magnetic Resonance Imaging* 52.1 (2020), pp. 54–69.

[40] RE Gabr et al. "Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study." In: *Mult Scler* 10 (2020), pp. 1217–1226.

[41] Ezequiel Geremia, Bjoern H Menze, Nicholas Ayache, et al. "Spatial decision forests for glioma segmentation in multi-channel MR images". In: *MICCAI Challenge on Multimodal Brain Tumor Segmentation* 34 (2012).

[42] Elizabeth R Gerstner et al. "ACRIN 6684: assessment of tumor hypoxia in newly diagnosed glioblastoma using 18F-FMISO PET and MRI". In: *Clinical Cancer Research* 22.20 (2016), pp. 5079–5086.

[43] "Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium". In: *Neuro-oncology* 20.7 (2018), pp. 873–884.

[44] M. Grade et al. "A neuroradiologist's guide to arterial spin labeling MRI in clinical practice". In: *Neuroradiology* 57.12 (2015), pp. 1181–1202.

[45] Mohammad Havaei et al. "Brain tumor segmentation with deep neural networks". In: *Medical image analysis* 35 (2017), pp. 18–31.

[46] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[47] W. Hollingworth et al. "A systematic literature review of magnetic resonance spectroscopy for the characterization of brain tumors". In: *AJNR Am J Neuroradiol* 27.7 (2006), pp. 1404–1411.

[48] Jie Hu, Li Shen, and Gang Sun. "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.

[49] L. S. Hu et al. "Radiogenomics to characterize regional genetic heterogeneity in glioblastoma". In: *Neuro-oncology* 19.1 (2017), pp. 128–137.

[50] Gao Huang et al. "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.

[51] Fabian Isensee et al. "Automated brain extraction of multisequence MRI using artificial neural networks". In: *Human brain mapping* 40.17 (2019), pp. 4952–4964.

[52] Fabian Isensee et al. "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge". In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 287–297.

[53] Fabian Isensee et al. "No new-net". In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 234–244.

[54] Kourosh Jafari-Khouzani et al. "Repeatability of cerebral perfusion using dynamic susceptibility contrast MRI in glioblastoma patients". In: *Translational oncology* 8.3 (2015), pp. 137–146.

[55] Mark Jenkinson et al. "Fsl". In: *Neuroimage* 62.2 (2012), pp. 782–790.

[56] Zeyu Jiang et al. "Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task". In: *International MICCAI Brainlesion Workshop*. Springer. 2019, pp. 231–241.

[57] Alain Jungo et al. "Towards Uncertainty-Assisted Brain Tumor Segmentation and Survival Prediction". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi et al. Springer International Publishing, 2018, pp. 474–485. ISBN: 978-3-319-75238-9.

[58] O. Kaidar-Person et al. "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". In: *Radiother Oncol* 137 (Aug. 2019), pp. 159–166.

[59] Konstantinos Kamnitsas et al. "Ensembles of multiple models and architectures for robust brain tumour segmentation". In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 450–462.

[60] Konstantinos Kamnitsas et al. "DeepMedic for brain tumor segmentation". In: *International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*. Springer. 2016, pp. 138–149.

[61] Konstantinos Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical image analysis* 36 (2017), pp. 61–78.

[62] T. J. Kaufmann et al. "Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases". In: *Neuro-oncology* 22.6 (2020), pp. 757–772.

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

[64] Philipp Kickingereder et al. "Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study". In: *The Lancet Oncology* 20.5 (2019), pp. 728–740.

[65] Paul Kinahan et al. "Data from ACRIN-FMISO-Brain". In: *The Cancer Imaging Archive* (2018). DOI: 10.7937/K9/TCIA.2018.vohlekok.

[66] Jens Kleesiek et al. "Ilastik for multi-modal brain tumor segmentation". In: *Proceedings MICCAI BraTS (brain tumor segmentation challenge)* (2014), pp. 12–17.

[67] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[68] J. M. Kros et al. "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". In: *J. Neuropathol. Exp. Neurol.* 66.6 (2007), pp. 545–551.

[69] A. Lai et al. "Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin". In: *J. Clin. Oncol.* 29.34 (2011), pp. 4482–4490.

[70] D. Le Bihan. "What can we see with IVIM MRI?" In: *Neuroimage* 187 (2019), pp. 56–67.

[71] Tsung-Yi Lin et al. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[72] J. Lipkova et al. "Personalized Radiotherapy Design for Glioblastoma: Integrating Mathematical Tumor Models, Multimodal Scans, and Bayesian Inference". In: *IEEE Trans Med Imaging* 38.8 (2019), pp. 1875–1884.

[73] David N Louis et al. "The 2007 WHO classification of tumours of the central nervous system". In: *Acta neuropathologica* 114.2 (2007), pp. 97–109.

[74] David N Louis et al. "The 2016 World Health Organization classification of tumors of the central nervous system: a summary". In: *Acta neuropathologica* 131.6 (2016), pp. 803–820.

[75] Richard McKinley, Raphael Meier, and Roland Wiest. "Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation". In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 456–465.

[76] Richard McKinley et al. "Triplanar Ensemble of 3D-to-2D CNNs with Label-Uncertainty for Brain Tumor Segmentation". In: *International MICCAI Brainlesion Workshop*. Springer. 2019, pp. 379–387.

[77] Kuo Men, Jianrong Dai, and Yexiong Li. "Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks". In: *Medical physics* 44.12 (2017), pp. 6377–6389.

[78] Kuo Men et al. "Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning". In: *Physica Medica* 50 (2018), pp. 13–19.

[79] B. H. Menze et al. "Mimicking the human expert: pattern recognition for an automated assessment of data quality in MR spectroscopic images". In: *Magn Reson Med* 59.6 (2008), pp. 1457–1466.

[80] B. H. Menze et al. "Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors". In: *NMR Biomed* 19.5 (2006), pp. 599–609.

[81] Bjoern H Menze et al. "A generative probabilistic model and discriminative extensions for brain lesion segmentation—with application to tumor and stroke". In: *IEEE transactions on medical imaging* 35.4 (2015), pp. 933–946.

[82] Bjoern H Menze et al. "The multimodal brain tumor image segmentation benchmark (BRATS)". In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024.

[83] Bjoern H Menze et al. "A generative approach for image-based modeling of tumor growth". In: *Biennial International Conference on Information Processing in Medical Imaging*. Springer. 2011, pp. 735–747.

[84] Bjoern H Menze et al. "A generative model for brain tumor segmentation in multi-modal images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2010, pp. 151–159.

[85] Bjoern H Menze et al. "Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation". In: *Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge)* 8 (2012).

[86] Bjoern H Menze et al. "Segmenting glioma in multi-modal images using a generative-discriminative model for brain lesion segmentation". In: *Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge)* 8 (2012).

[87] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation". In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 2016, pp. 565–571.

[88] Pawel Mlynarski et al. "Anatomically consistent CNN-based segmentation of organs-at-risk in cranial radiotherapy". In: *Journal of Medical Imaging* 7.1 (2020), p. 014502.

[89] Andriy Myronenko. "3D MRI brain tumor segmentation using autoencoder regularization". In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 311–320.

[90] S Nameeta, X Feng, M Lankerovich, et al. "Data from Ivy GAP. The Cancer Imaging Archive". In: *The Cancer Imaging Archive* (2016). DOI: 10.7937/K9/TCIA.2016.XLwaN6nL.

[91] László G Nyúl, Jayaram K Udupa, and Xuan Zhang. "New variants of a method of MRI scale standardization". In: *IEEE transactions on medical imaging* 19.2 (2000), pp. 143–150.

[92] S. Okuchi et al. "Diagnostic accuracy of dynamic contrast-enhanced perfusion MRI in stratifying gliomas: A systematic review and meta-analysis". In: *Cancer Med* 8.12 (2019), pp. 5564–5573.

[93] Eliza Orasanu et al. "Organ-at-risk segmentation in brain MRI using model-based segmentation: benefits of deep learning-based boundary detectors". In: *International Workshop on Shape in Medical Imaging*. Springer. 2018, pp. 291–299.

[94] D. W. Parsons et al. "An integrated genomic analysis of human glioblastoma multiforme". In: *Science* 321.5897 (2008), pp. 1807–1812.

[95] A. M. Paschoal et al. "Intravoxel incoherent motion MRI in neurological and cerebrovascular diseases". In: *Neuroimage Clin* 20 (2018), pp. 705–714.

[96] P. Patel et al. "MR perfusion-weighted imaging in the evaluation of high-grade gliomas after treatment: a systematic review and meta-analysis". In: *Neuro-oncology* 19.1 (2017), pp. 118–127.

[97] Nancy Pedano et al. "Radiology data from the cancer genome atlas low grade glioma [TCGA-LGG] collection". In: *The Cancer Imaging Archive* 2 (2016).

[98] Jan C Peeken et al. "Deep learning derived tumor infiltration maps for personalized target definition in Glioblastoma radiotherapy". In: *Radiotherapy and Oncology* 138 (2019), pp. 166–172.

[99] Sérgio Pereira et al. "Brain tumor segmentation using convolutional neural networks in MRI images". In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1240–1251.

[100] MA Prah et al. "Repeatability of standardized and normalized relative CBV in patients with newly diagnosed glioblastoma". In: *American Journal of Neuroradiology* 36.9 (2015), pp. 1654–1661.

[101] Ralph B Puchalski et al. "An anatomic transcriptional atlas of human glioblastoma". In: *Science* 360.6389 (2018), pp. 660–663.

[102] Eva-Maria Ratai et al. "ACRIN 6684: Multicenter, phase II assessment of tumor hypoxia in newly diagnosed glioblastoma using magnetic resonance spectroscopy". In: *PloS one* 13.6 (2018), e0198548.

[103] T Riklin Raviv, K Van Leemput, and Bjoern H Menze. "Multi-modal brain tumor segmentation via latent atlases". In: *Proceeding MICCAI-BRATS* 64 (2012).

[104] M. Reuter et al. "Impact of MRI head placement on glioma response assessment". In: *J. Neurooncol.* 118.1 (2014), pp. 123–129.

[105] Tammy Riklin-Raviv et al. "Segmentation of image ensembles via latent atlases". In: *Medical image analysis* 14.5 (2010), pp. 654–665.

[106] Torsten Rohlfing et al. "The SRI24 multichannel atlas of normal adult human brain structure". In: *Human brain mapping* 31.5 (2010), pp. 798–819.

[107] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[108] T. C. Ryken et al. "The role of imaging in the management of progressive glioblastoma : a systematic review and evidence-based clinical practice guideline". In: *J. Neurooncol.* 118.3 (2014), pp. 435–460.

[109] Lisa Scarpace et al. "Data from REMBRANDT". In: *The Cancer Imaging Archive* 10 (2015), K9.

[110] Lisa Scarpace et al. "Radiology data from the cancer genome atlas glioblastoma multiforme [TCGA-GBM] collection". In: *The Cancer Imaging Archive* 11.4 (2016), p. 1.

[111] KM Schmainda and M Prah. "Data from Brain-Tumor-Progression". In: *The Cancer Imaging Archive* (2018). DOI: 10.7937/K9/TCIA.2018.15quzvnb.

[112] KM Schmainda et al. "Glioma DSC-MRI perfusion data with standard imaging and ROIs". In: *The Cancer Imaging Archive. http://doi. org/10.7937 K* 9 (2016).

[113] P. Schmitt et al. "Effects of slice thickness and head rotation when measuring glioma sizes on MRI: in support of volume segmentation versus two largest diameters methods". In: *J. Neurooncol.* 112.2 (2013), pp. 165–172.

[114] Russell T Shinohara et al. "Statistical normalization techniques for magnetic resonance imaging". In: *NeuroImage: Clinical* 6 (2014), pp. 9–19.

[115] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[116] Amber L Simpson et al. "A large annotated medical image dataset for the development and evaluation of segmentation algorithms". In: *arXiv preprint arXiv:1902.09063* (2019).

[117] Stephen M Smith et al. "Advances in functional and structural MR image analysis and implementation as FSL". In: *Neuroimage* 23 (2004), S208–S219.

[118] D. Sturm et al. "Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma". In: *Cancer Cell* 22.4 (2012), pp. 425–437.

[119] C. H. Suh et al. "2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and meta-analysis using individual patient data". In: *Neuro-oncology* 20.12 (2018), pp. 1573–1583.

[120] C. H. Suh et al. "Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: a systemic review and meta-analysis". In: *Eur Radiol* 29.2 (2019), pp. 745–758.

[121] Yannick Suter et al. "Radiomics for Glioblastoma Survival Analysis in Pre-operative MRI: Exploring Feature Robustness, Class Boundaries, and Machine Learning Techniques". In: *Cancer Imaging* In Press (2020).

[122] Siddhesh Thakur et al. "Brain extraction on MRI scans in presence of diffuse glioma: Multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training". In: *NeuroImage* (2020), p. 117081.

[123] Nicholas J Tustison et al. "N4ITK: improved N3 bias correction". In: *IEEE transactions on medical imaging* 29.6 (2010), pp. 1310–1320.

[124] Nick Tustison et al. "Ants andarboles". In: *Multimodal Brain Tumor Segmentation* (2013), p. 47.

[125] Jan Unkelbach et al. "Radiotherapy planning for glioblastoma based on a tumor growth model: implications for spatial dose redistribution". In: *Physics in Medicine & Biology* 59.3 (2014), p. 771.

[126] Jan Unkelbach et al. "Radiotherapy planning for glioblastoma based on a tumor growth model: improving target volume delineation". In: *Physics in Medicine & Biology* 59.3 (2014), p. 747.

[127] Gregor Urban et al. "Multi-modal brain tumor segmentation using deep convolutional neural networks". In: *MICCAI BraTS (brain tumor segmentation) challenge. Proceedings, winning contribution* (2014), pp. 31–35.

[128] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.

[129] G. Verma et al. "Non-invasive detection of 2-hydroxyglutarate in IDH-mutated gliomas using two-dimensional localized correlation spectroscopy (2D L-COSY) at 7 Tesla". In: *J Transl Med* 14.1 (2016), p. 274.

[130] Guotai Wang et al. "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks". In: *International MICCAI brainlesion workshop*. Springer. 2017, pp. 178–190.

[131] Chan Woo Wee et al. "Evaluation of variability in target volume delineation for newly diagnosed glioblastoma: a multi-institutional study from the Korean Radiation Oncology Group". In: *Radiation Oncology* 10.1 (2016), p. 137.

[132] Patrick Y Wen et al. "Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group". In: *Journal of clinical oncology* 28.11 (2010), pp. 1963–1972.

[133] B. Wiestler et al. "Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma". In: *Acta Neuropathol.* 128.4 (2014), pp. 561–571.

[134] H. Yan et al. "IDH1 and IDH2 mutations in gliomas". In: *N. Engl. J. Med.* 360.8 (2009), pp. 765–773.

[135] Ke Zeng et al. "Segmentation of gliomas in pre-operative and post-operative multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework". In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer. 2016, pp. 184–194.

[136] Yuan-Xing Zhao, Yan-Ming Zhang, and Cheng-Lin Liu. "Bag of Tricks for 3D MRI Brain Tumor Segmentation". In: *International MICCAI Brainlesion Workshop*. Springer. 2019, pp. 210–220.

[137] Chenhong Zhou et al. "Learning contextual and attentive information for brain tumor segmentation". In: *International MICCAI Brainlesion Workshop*. Springer. 2018, pp. 497–507.

[138] D Zikic et al. "Context-sensitive classification forests for segmentation of brain tumor tissues". In: *Proc. MICCAI-BRATS*. 2012, pp. 1–9.

[139] Darko Zikic et al. "Segmentation of brain tumor tissues with convolutional neural networks". In: *Proceedings MICCAI-BRATS* (2014), pp. 36–39.

[140] Alex Zwanenburg et al. "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping". In: *Radiology* 295.2 (2020), pp. 328–338.