3D Glioma Segmentation Using Multi-modal Multi-site MRI Data

Shahryar Khorasani **Matthias Norden** **Susanne Ibing**

Abstract

The segmentation of different sub-regions of gliomas in multi-modal imaging is of great clinical relevance, both in diagnosis and treatment. Despite huge advances in automated medical imaging segmentation through deep learning models, accurate segmentation still remains a challenge. In this work, we use a 3D deep convolutional network, based on the V-net for the automated segmentation of different sub-regions of gliomas in multimodal MRI-scans. Using the basic architecture of an autoencoder including residual blocks and skip-connections, our network has only been trained on 104 (36%) cases of BraTS 2018 challenge training dataset. Model evaluation has been done computing the Dice Score for the three different sub-regions in 33 previously selected test cases. After training our best model achieved Dice scores between 0.6078 and 0.787 for the whole tumor, tumor core and the enhancing tumor region which is comparable to similar approaches presented in the BraTS 2018 challenge. However, the accuracy of our model might be improved with more training data, additional image augmentation and using transfer learning methods.

1 Introduction

Tumors of the central nervous system (CNS) are classified by their cells of origin and their location within the brain or spinal cord. Gliomas comprise 70 % of brain tumors and thereby are the most common primary brain malignancies [1]. They are considered highly proliferative cancers, but differ in aggressiveness and prognosis between patients [2]. On a histological level, heterogenous sub-regions can be described for gliomas, such as the the necrotic core, the active and non-enhancing core and the peritumoral edematous/invaded tissue [3].

The volume of a brain tumor is an important factor for diagnosis and treatment decisions, the assessment of therapy response, and the potential patient selection for clinical trials. First step of tumor volume estimation is the tumor segmentation in Magnetic Resonance Imaging (MRI) or Computer Tomography (CT) images [4]. MRI is a medical imaging technique for soft tissue that applies nuclear magnetic resonance. Different MRI modalities in tumor diagnosis enable the visualization of complementary information about tumor sub-regions and provide multi-modal MRI data sets from tumor patients. The main MRI sequences are T1-weighted MRI with and without contrast enhancement (T1/T1c), T2-weighted MRI (T2), and FLuid-Attenuated Inversion Recovery (FLAIR) (Figure 1, [5]). Whereas T1 is well suited for the annotation of healthy tissue, T1c images enhance tumor borders, focusing on the tumor core. T2 and FLAIR highlight the ”whole tumor” region including the peritumoral edema: T2 sequences enhance the edema region, and FLAIR sequences enable the differentiation of the cerebrospinal fluid and the edema region [6, 7].

Manual tumor segmentation, especially when using multi-modal MRI data, is a very labor-intensive process and thus not scalable to large data sets. Therefore, a growing interest arose in the investigation of automated and semi-automated image segmentation throughout the past decades [8]. The benefits of applying automatic segmentation algorithms in the clinics include time efficient, accurate and reproducible analysis of the large amount of generated multi-modal brain tumor images

[6]. However, since tumor shape and appearance greatly differ between patients, and heterogeneity between sub-regions is high, brain tumor segmentation in multi-modal MRI scans is seen as one of the challenging problems in medical image analysis [3].

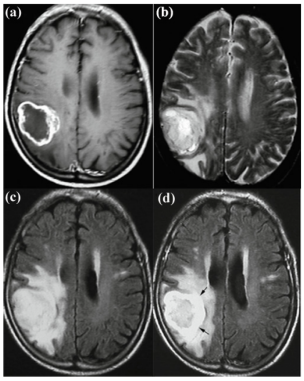

Figure 1: Four different MRI sequences of a glioblastoma patient. (a) T1-weighted MRI, (b) T2-weighted MRI, (c) FLuid-Attenuated Inversion Recovery (FLAIR), and (d) FLAIR with contrast enhancement. *Figure taken from:* [5]

2 Related Work

Current automatic tumor segmentation models are mainly based on generative or discriminative approaches [8]. Generative approaches incorporate domain-specific prior knowledge about different tissue types to model the probability distribution of anatomy and appearance of the tissue. These models often achieve good generalization results when being presented to unseen data [9, 10]. Discriminative directly learn the relationship between segmentation labels and image intensities, focusing on local features. They do not require any prior knowledge, however, manually labelled training data is necessary for the learning [8]. After pixel-/voxel-wise feature extraction, features are usually fed into classification models, for instance based in decision trees or support-vector machines [11, 12].

In recent years, driven by their success in segmentation, discriminative deep neural networks, mostly represented by convolutional neural networks (CNNs), have increasingly been applied to medical image segmentation problems [3]. 2D CNNs consider each MRI scan independently from each other, whereas 3D CNNs are based on voxel-wise segmentation of the images. Many of the previously developed CNNs are 2D, even though 3D medical data is available [13].

With the aim to assess and compare current brain tumor segmentation methods, the Brain Tumor Image Segmentation Benchmark (BraTS) challenge was first organised in 2012 and 2013. The top rated models from 2017 and 2018 are all based on 3D CNNs, representing state-of-the art models for medical image segmentation.

Kamnitsas *et al.*. proposed a fully connected CNN in 2016 that was amongst the first models to use 3D convolution [14]. In the BraTS 2017 challenge, they extended their model by two further 3D CNN architectures and an ensembling of the different model results to ensure robust performance and generalization of their model [15]. Also Zhou *et al.* employed multiple different CNN architectures and ensembled their prediction results. Their CNNs were specifically trained for each sub-task to ensure overall high performance of the model [16]. Myronenko *et al.* as well as Yang *et al.* based their models on an encoder-decoder network architecture, a shallow network structure where the encoder part is used to extract deep image features and the decoder reconstructs the segmentation mass [17, 18].

Several of the highly rated models were based on the famous U-Net, a convolutional network for 2D image segmentation that was adapted to 3D voxel inputs [19, 15, 20]. The architecture of the U-Net resembles an encoder-decoder architecture where skip connections are introduced that forward features extracted in early layers directly to the decoder part of the neural network [21]. Here, our

network was deeply inspired by the V-Net, a fully convolutional neural network that was developed for volumetric medical image segmentation and is based on the famous U-Net as well [13]. The objective function of the network is based on the Dice coefficient maximization, a similarity measure that compares the overlap of the predicted and true segmentation labels [22].

3 Dataset

For the development of a 3D Glioma Segmentation model we used the training set provided by the BraTS 2018 challenge [8]. Beginning in 2012 the first data set contained 65 clinical MRI scans, 14 from patients with pathologically confirmed low-grade and 51 from patients with high-grade glioma, as well as 65 synthetic images (35 high-grade and 30 low-grade gliomas with same image modalities and resolution as the clinical dataset - generated by TumorSim software). The datasets used in the 2018 BraTS challenge have been updated with more clinically-acquired 3T MRI scans and additional information on overall patient survival [3]. The used dataset for training and validation of our model contains all multimodal MRI scans of the 163 training cases with known survival rates. All cases come with 4 multimodal MRI scans (in total 652 scans) describing native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes and have been segmented manually by four raters. The annotations were approved by experienced neuro-radiologists and comprise the GD-enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor (Figure 2).

The 652 scans were obtained and stored in NIfTI format with the dimensions of 240 x 240 x 155 (155 slices of 240 x 240 resolution). Data was already aligned to the same anatomical template and interpolated to voxel size of 1 x 1 x 1 mm. Subsequently we cropped the images to 192 x 192 x 192 resolution on the fly to reduce empty spaces. We then split the data into 104 cases for training, 26 for validation and 33 for final model evaluation. Furthermore, during training, randomly selected scans were mirrored along the sagittal plane as an augmentation step.

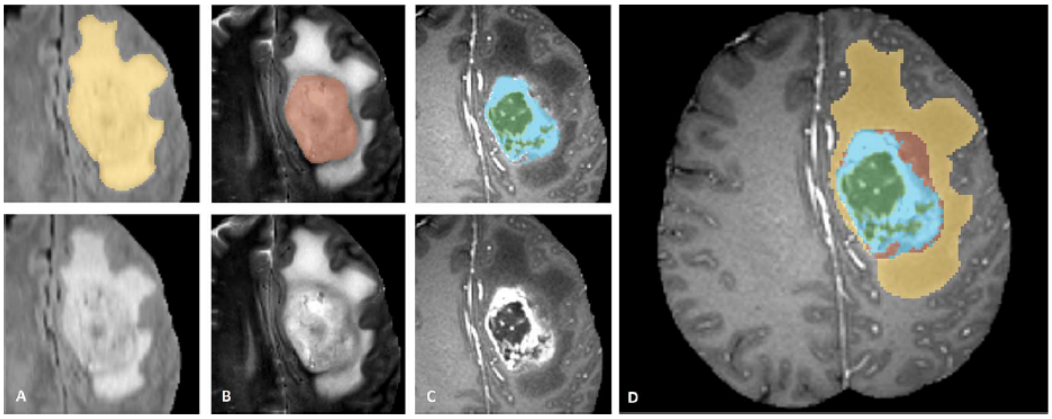

Figure 2: "Manual annotation through expert raters. Shown are image patches with the tumor structures that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). Image patches show from left to right: the whole tumor visible in FLAIR (A), the tumor core visible in T2 (B), the enhancing tumor structures visible in T1c (blue), surrounding the cystic/necrotic components of the core (green) (C). Segmentations are combined to generate the final labels of the tumor structures (D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue)." *Figure and Description taken from:* [8]

4 Architecture and Training

This work utilizes a 3-dimensional CNN based on the V-net architecture proposed by Milletari *et al.* to perform the segmentation task [13]. The V-net uses the general idea of the famous 2D U-net CNN and applies it to 3D image data as it is given in MRI scans, taking advantage of the third dimension spatial properties. It therefore compromises an autoencoder basic architecture with residual blocks

and skip-connections. In Figure 3 we provide a schematic overview of our V-net based CNN. The model input included four channels representing the four MRI modalities. The final model output was the probability predictions of each voxel to belong to one of the four classes of background, the GD-enhancing tumor, the peritumoral edema and the necrotic/non-enhancing tumor.

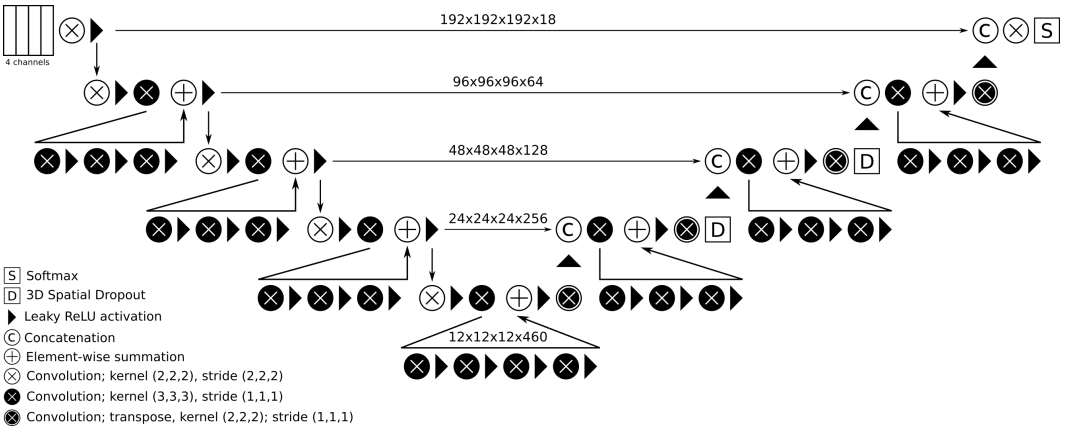

Figure 3: Schematic network overview. Our convolutional neural network was based on the V-Net described by Milletary *et al.* [13]. 3D voxels are from four different MRI sequences are used as input. The architecture comprises a fully convolutional autoencoder-like structure with skip connections and residual blocks. Class probability of the four different segmentation labels is computed using a softmax function.

The downward path transforms the data into a smaller but high dimensional space via convolutional operations with kernel size of 2 x 2 x 2 and strides of 2 x 2 x 2. The upward path transforms the features in the bottle-neck layer back to their original size through transpose convolution operations with kernel size of 2 x 2 x 2 and strides of 1 x 1 x 1. Both paths include residual blocks with kernel size of 3 x 3 x 3 and strides of 1 x 1 x 1. The features captured in each level of the downward path are transferred directly to the upward path through skip connections and concatenations. All convolutions are applied with appropriate padding. In the two intermediate levels of the upward path, 3D spatial dropout was used to reduce overfitting. The output of the model had the same size as the input (192 x 192 x 192 x 4) with the fourth dimension representing the classes. All convolutional operations were followed by Leaky ReLU non-linearity except for the output which was a softmax function (Figure 3).

The network predictions consist of 4 volumes with the same dimensions as the input data and are processed through a softmax layer in order to compute the probability of each voxel to belong to one of the 4 classes of background, the GD-enhancing tumor, the peritumoral edema and the necrotic/non-enhancing tumor.

4.1 Dice Loss

auch For model training we used the Dice Loss as proposed by Milletari *et al.* [13]. The Dice Loss has been used successfully for medical segmentation problems in which the field of interest usually occupies only a very small region of the volume and obtained better results than other objective functions. The Generalized Dice Loss can be used to evaluate segmentation tasks with multiple output classes. It was computed following the equation by Sudre *et al.*:

$$GDL = 1 - 2\frac{\sum_{l=1}^{2} w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^{2} w_l \sum_n r_{ln} + p_{ln}}$$

with r_{ln} being the voxel values of the reference ground truth segmentation and p_n the predicted class probability over N image elements. w_l is used to provide invariance to different label set properties [22].

4.2 Training Details

The experiments were performed on the HPI-DHC local server with Ubuntu 18.04.4 LTS, TITAN RTX graphics processing unit, 24 GB memory, CUDA 10.2. Implementation was realised in Python 3.6.8, using the Tensorflow and Keras.[1]

We trained the model from scratch for 500 epochs with batch size of 1. We used the RMSProp optimizer with a learning rate of 1.5×10^{-5} and weight decay of 5×10^{-5}. To mitigate the small sample size we applied k-fold cross-validation with k=5. The hyperparameter tuning involved changing the regularization rate, the number of filters, the depth of residual blocks, the learning rate, the optimizer and the weight decay.

5 Evaluation

To evaluate the performance of our final model we computed the model predictions on 33 cases which were chosen for final evaluation and been kept from the model during training and hyperparameter tuning. The accuracy of the predictions were calculated using the average Dice Score for each sub-regions of the tumor.

Table 1 reports the preliminary quantitative evaluation results obtained by our proposed segmentation model on the 33 final evaluation cases of BraTS 2018 training data set.

Table 1: Model accuracy for tumor sub-regions

Average Dice Loss		
Whole Tumor	Tumor Core	Enhancing Tumor
0.678	0.787	0.706

In Figure 4 we provide an one-slice example comparison of the model predictions for the different sub-regions with the ground truth sub-regions of the tumor.

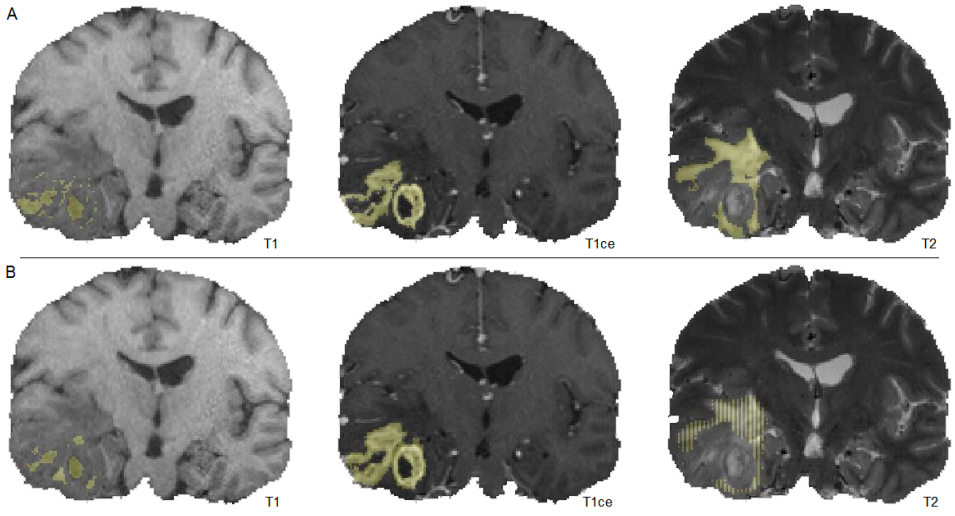

Figure 4: Ground truth vs model prediction. The yellow area depicts the ground truth for enhancing tumor, non enhancing solid tumor core and edema from left to right respectively (A). The yellow area depicts model predictions for enhancing tumor, non enhancing solid tumor core and edema from left to right respectively (B). The MRI modalities from left to right include T1-weighted, contrast-enhancing T1-weighted and T2-weighted, respectively. Note that the FLAIR modality is not shown here, although it was included in the model inputs.

[1] Implementation available at https://github.com/shahryarkhorasani/brats_tumor_segmentation_vnet

6 Discussion

In this work, we present a V-Net based CNN for brain tumor segmentation that achieves good accuracy results within the average accuracy of the top models from the BraTS challenge 2018. Even though we used only 36% of the training data provided by BraTS, our most successful model performs with a dice score which lies between 0.6078 and 0.787 for the three sub-regions of the whole tumor, tumor core and the enhancing tumor region (Table 1). In comparison, the top 54 of 63 models from BraTS 2018 achieved an average dice coefficient between 0.61 and 0.77 [3].

Since the first BraTS challenge in 2012, it came apparent that leading state-of-the art models for image segmentation tasks like this are all convolutional neural networks. The model architecture of our choice contains an autoencoder structure to extract deep features of the image, and additionally skip connections and residual blocks that maintain and pass on the information extracted at more shallow layers. Several of the top models from BraTS 2017 or 2018 were based on the U-Net as well [19, 15, 20]. However, they are often outperformed by ensemble approaches where multiple models are trained and the averaged prediction results are chosen [15, 16]. Such ensemble approaches usually show higher robustness and generalization compared to individual CNNs. On the downside, they require a much higher number of GPUs and training time, which is why we did not use an ensemble approach.

In previous years, the dice coefficient of the whole tumor prediction task was most robust with a median of 0.9 across most teams [3]. With 0.6078, the whole tumor dice coefficient of our model shows least accuracy of the three sub-regions. There are several ideas how the accuracy of our model might be improved to achieve comparable performance also regarding the whole tumor segmentation: First of all, as for all deep learning models, accuracies generally rise with the increase of the data size. Our model was trained on only 104 and validated with 26 data samples. Using 5-fold cross-validation, the RMSProp optimizer and image augmentation through flipping, overfitting of the model was prevented as much as possible. More augmentation techniques such as geometric transformations, color space augmentations, kernel filters, cropping and rotation might further improve the model accuracy by artificially increasing the sample size and heterogeneity. In case more unlabeled data was available, one could also apply transfer learning techniques. In this approach, the model would be pre-trained in an unsupervised manner. After freezing some of the first layers, the rest of the layers would be retrained in a supervised manner using the labelled training data available for this task.

Currently, the estimation of the brain tumor volume for diagnosis and therapy is based on rectangular 2D area estimation of the MRI slice with biggest tumor fraction in most clinics. Compared to semi-automatic and automatic approaches, the manual approach often underestimates the tumor size [23]. Additionally, the differences in manual segmentation, even between experts, is high [8]. Both to save time of clinicians, but also hoping to improve the quality of the size estimation, automated models based on CNNs could be a relevant improvement in the clinical context. Ensembling segmentation models of different CNNs might be the preferred way to start applying automated brain tumor segmentation. Compared to individual modes, they showed a superior accuracy and even outperformed the robustness of the ground-truth segmentation labelling from experts [8, 3].

References

[1] Hiroko Ohgaki and Paul Kleihues. Genetic alterations and signaling pathways in the evolution of gliomas. *Cancer science*, 100(12):2235–2241, 2009.

[2] Y Rajesh, Ipsita Pal, Payel Banik, et al. Insights into molecular therapy of glioma: current challenges and next generation blueprint. *Acta Pharmacologica Sinica*, 38(5):591–613, 2017.

[3] Spyridon Bakas et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. 2018.

[4] Bilwaj Gaonkar, Luke Macyszyn, Michel Bilello, et al. Automated tumor volumetry using computer-aided image segmentation. *Academic radiology*, 22(5):653–661, 2015.

[5] Antonios Drevelegas. *Imaging of brain tumors with histological correlations*. Springer Science & Business Media, 2010.

[6] Jin Liu, Min Li, Jianxin Wang, et al. A survey of mri-based brain tumor segmentation methods. *Tsinghua science and technology*, 19(6):578–595, 2014.

[7] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI brainlesion workshop*, pages 178–190. Springer, 2017.

[8] Bjoern H. Menze, Andras Jakab, Stefan Bauer, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.

[9] Bjoern H Menze, Koen Van Leemput, Danial Lashkari, et al. A generative model for brain tumor segmentation in multi-modal images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 151–159. Springer, 2010.

[10] Michael R Kaus, Simon K Warfield, Arya Nabavi, et al. Automated segmentation of mr images of brain tumors. *Radiology*, 218(2):586–591, 2001.

[11] Darko Zikic, Ben Glocker, Ender Konukoglu, et al. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 369–376. Springer, 2012.

[12] Chi-Hoon Lee, Mark Schmidt, Albert Murtha, et al. Segmenting brain tumors with conditional random fields and support vector machines. In *International Workshop on Computer Vision for Biomedical Image Applications*, pages 469–478. Springer, 2005.

[13] 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)*, pages 565–571. IEEE, 2016.

[14] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, et al. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

[15] Konstantinos Kamnitsas, Wenjia Bai, Enzo Ferrante, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation. In *International MICCAI Brainlesion Workshop*, pages 450–462. Springer, 2017.

[16] Chenhong Zhou, Shengcong Chen, Changxing Ding, and Dacheng Tao. Learning contextual and attentive information for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 497–507. Springer, 2018.

[17] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*, pages 311–320. Springer, 2018.

[18] T Yang, Y Ou, and T Huang. Automatic segmentation of brain tumor from mr images using segnet: selection of training data sets. In *Proc. 6th MICCAI BraTS Challenge*, pages 309–312, 2017.

[19] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, et al. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.

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

[21] 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*, pages 234–241. Springer, 2015.

[22] Carole H Sudre, Wenqi Li, Tom Vercauteren, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 240–248. Springer, 2017.

[23] LA Gilligan, MD DeWire-Schottmiller, M Fouladi, et al. Tumor response assessment in diffuse intrinsic pontine glioma: Comparison of semiautomated volumetric, semiautomated linear, and manual linear tumor measurement strategies. *American Journal of Neuroradiology*, 41(5):866–873, 2020.