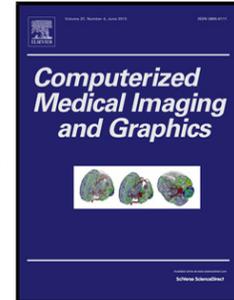


# Journal Pre-proof

Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma

Milan Decuyper, Stijn Bonte, Karel Deblaere, Roel Van Holen



PII: S0895-6111(20)30126-9

DOI: <https://doi.org/10.1016/j.compmedimag.2020.101831>

Reference: CMIG 101831

To appear in: *Computerized Medical Imaging and Graphics*

Received Date: 30 April 2020

Revised Date: 30 October 2020

Accepted Date: 17 November 2020

Please cite this article as: Milan Decuyper, Stijn Bonte, Karel Deblaere, Roel Van Holen, Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma, <![CDATA[Computerized Medical Imaging and Graphics]]> (2020), doi: <https://doi.org/>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

# Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma

Milan Decuyper<sup>a,\*</sup>, Stijn Bonte<sup>a</sup>, Karel Deblaere<sup>b</sup>, Roel Van Holen<sup>a</sup>

<sup>a</sup>*Medical Image and Signal Processing (MEDISIP), Ghent University, Ghent, Belgium*

<sup>b</sup>*Department of Radiology, Ghent University Hospital, Ghent, Belgium*

---

## Abstract

In the WHO glioma classification guidelines grade (glioblastoma versus lower-grade glioma), IDH mutation and 1p/19q co-deletion status play a central role as they are important markers for prognosis and optimal therapy planning. Currently, diagnosis requires invasive surgical procedures. Therefore, we propose an automatic segmentation and classification pipeline based on routinely acquired pre-operative MRI (T1, T1 postcontrast, T2 and/or FLAIR). A 3D U-Net was designed for segmentation and trained on the BraTS 2019 training dataset. After segmentation, the 3D tumor region of interest is extracted from the MRI and fed into a CNN to simultaneously predict grade, IDH mutation and 1p19q co-deletion. Multi-task learning allowed to handle missing labels and train one network on a large dataset of 628 patients, collected from The Cancer Imaging Archive and BraTS databases. Additionally, the network was validated on an independent dataset of 110 patients retrospectively acquired at the Ghent University Hospital (GUH). Segmentation performance calculated on the BraTS validation set shows an average whole tumor dice score of 90% and increased robustness to missing image modalities by randomly excluding input MRI during training. Classification area under the curve scores are 93%, 94% and 82% on the TCIA test data and 94%, 86% and 87% on the GUH data for grade,

---

\*Corresponding author

*Email address:* milan.decuyper@ugent.be (Milan Decuyper)

IDH and 1p19q status respectively. We developed a fast, automatic pipeline to segment glioma and accurately predict important (molecular) markers based on pre-therapy MRI.

*Keywords:* Glioma, Segmentation, Deep Learning, MRI, Molecular markers

## 1. Introduction

Gliomas are the most frequently occurring primary brain tumors and show a large heterogeneity in treatment response and prognosis. The WHO classification system classifies glioma into grades I-IV in order of malignancy based on histopathological and clinical criteria (Louis et al., 2016). Glioblastoma multiforme, WHO grade IV, is the most aggressive type and has a very poor prognosis with a 5-year survival rate of only 5.6%. In contrast, lower-grade glioma (WHO grade II and III) have more favorable survival rates up to 81.6% and 57.6% respectively (Ostrom et al., 2018). In the most recent classification of glioma, the WHO has put increased emphasis on the integration of molecular markers (Louis et al., 2016). As illustrated in Figure 1, two genetic markers play a central role: isocitrate dehydrogenase (IDH) 1 and/or 2 mutation and co-deletion of chromosome arms 1p and 19q.

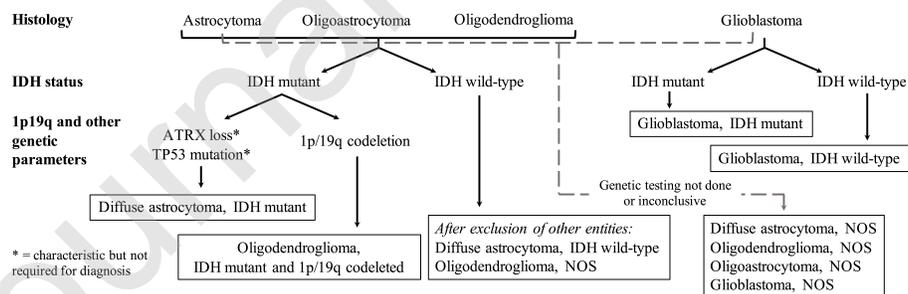


Figure 1: Classification of diffuse gliomas based on histological and genetic features. Not otherwise specified (NOS) designates a group of lesions that cannot be classified into the more narrowly defined groups or for which insufficient information is available. Adapted from (Louis et al., 2016).

IDH mutation occurs in more than 80% of lower-grade glioma cases and ap-

15 proximately 10% of glioblastoma cases, corresponding closely to so-called secondary glioblastoma (Eckel-Passow et al., 2015; Yan et al.). Gliomas with IDH mutation are less aggressive and demonstrate better response to temozolomide chemotherapy than IDH wildtype gliomas. For example, glioblastoma patients with IDH mutation show a longer overall survival (OS) compared to patients  
20 with IDH wildtype glioblastoma (31 versus 15 months) (Yan et al.). Moreover, reported OS of IDH wildtype LGG is only slightly longer than IDH wildtype glioblastoma (The Cancer Genome Atlas Research Network, 2015). Hence IDH mutation is associated with a significantly better prognosis and appears to be a more important predictor than WHO grade as Reuss et al. (2015) reported little  
25 difference in survival between IDH mutant WHO grade II and III astrocytoma.

According to the 2016 WHO classification scheme, diagnosis of oligodendroglioma requires demonstration of both IDH mutation and combined loss of 1p and 19q (see Figure 1). Similarly to IDH mutation, 1p/19q co-deletion is linked to more favorable outcomes and oligodendrogliomas respond well to  
30 combined procarbazine, lomustine and vincristine chemotherapy (Weller et al., 2017).

We can conclude that determination of WHO grade (glioblastoma versus lower-grade glioma), IDH mutation and 1p/19q co-deletion status is necessary for prognosis and optimal therapy planning. Currently, genetic information of  
35 gliomas is derived from the analysis of tumor tissue obtained through biopsy or resection. However, biopsies involve risks, are subject to sampling error and are related to reduced OS compared to a wait-and-scan approach (Jackson et al., 2001; Wijnenga et al., 2017). Tumor resection is standard of care for most glioma types but is not always possible depending on tumor location and accessibility, the patient's clinical condition or when the patient refuses a surgical  
40 procedure. Therefore, non-invasive assessment of clinically relevant markers can aid in characterizing glioma and guide therapy and surgery planning, especially when extraction of tumor tissue is not possible or genetic testing not available.

Correlations between MR phenotypes and glioma subtypes have been widely  
45 investigated. For example, presence of contrast enhancement and necrosis on

T1 contrast enhanced (T1ce) MRI is associated with high-grade glioma (Law et al., 2003). IDH mutant glioma have been reported to demonstrate minimal enhancement, sharp tumor margins and homogeneous signal intensity (Carrillo et al., 2012; Qi et al., 2014). This contrasts with IDH wildtype glioma that is correlated with thick, irregular enhancement with necrosis and infiltrative edema. Furthermore, increased enhancement, poorly circumscribed borders and heterogeneous signal intensity are characteristic MRI features related to 1p/19q co-deletion (Johnson et al., 2017; Sonoda et al., 2015). However, visual interpretation and prediction of tumor properties remains very challenging and inaccurate. For instance, 40-45% of non-enhancing lesions are subsequently found to be highly malignant (Jansen et al., 2012). Conversely, 16% of WHO grade II glioma show contrast enhancement and this percentage is expected to be even higher for low-grade oligodendroglioma (Khalid et al., 2012; Pallud et al., 2009).

To improve speed and accuracy of non-invasive tumor characterization, there is an increasing interest to use machine learning techniques for medical image analysis. A selection of recent studies on the prediction of grade, IDH mutation and 1p/19q co-deletion status of glioma is included in Table 1. Yang et al. (2018) differentiated LGG from glioblastoma with high accuracy (AUC of 0.97) based on T1ce MRI. The tumor was manually segmented followed by slice-level classification through the use of a 2D convolutional neural network (CNN), pre-trained on ImageNet, fine-tuned on 90 patients and evaluated on a test set of 23 patients. State-of-the-art performance on IDH mutation status prediction was reported by Chang et al. (2018a). They predicted IDH mutation based on pre-operative MRI (T1, T1ce, T2 and FLAIR) of 496 patients. Tumors were manually delineated and classified by four 2D CNNs (one for each modality). Through the combination of the four probabilities with age and a logistic regression classifier, an AUC of 0.95 was obtained. Akkus et al. (2017) analyzed T1ce and T2 MRI of 159 LGG patients to predict 1p/19q co-deletion status. The tumor was delineated semi-automatically and each slice was classified using a multi-scale 2D CNN achieving an accuracy of 88%.

Most of the studies included in Table 1 used manual or semi-automatic segmentations which might introduce variability and subjectivity to the classification pipeline and impede clinical adoption. However, encouraged by the annual  
 80 Brain Tumor Segmentation (BraTS) Challenges (Menze et al., 2015), a lot of research is performed on automatic glioma segmentation. In recent years deep learning techniques have surpassed performance of more traditional radiomics methods (Bakas et al., 2018). Myronenko (2019) achieved first place at the  
 2018 BraTS challenge with dice scores of 76.64, 88.39, 81.54 for enhancing tumor (ET), whole tumor (WT), and tumor core (TC) volumes respectively on the  
 85 BraTS 2018 test set. They used an ensemble of 10 encoder-decoder networks.

An additional limitation is that existing studies often train and evaluate their models on a small dataset from one institution. Hence their robustness to data from other clinical centers (with large variations in imaging protocols)  
 90 remains to be evaluated on an independent dataset. Moreover, due to the limited amount of data, often radiomics methods are used where hand-engineered features are extracted depending on expert opinion. Convolutional neural networks, on the other hand, can automatically extract and classify features from complex imaging datasets with increased speed and without requiring human  
 95 interaction resulting in a more objective computer-aided diagnosis tool.

Therefore, in this study we propose a non-invasive fully automatic 3D pipeline to segment glioma and predict clinically relevant markers according to the most recent WHO guidelines based on routinely acquired pre-operative MRI. We collected a large dataset from multiple public databases and an independent  
 100 dataset from the Ghent University Hospital to test the generalization performance. Moreover, our approach is robust to missing T1 and T2 or FLAIR MRI.

Table 1: Overview of recent studies on non-invasive prediction of grade, IDH mutation and 1p/19q codeletion status of glioma.

Author	Task	Dataset	Method	Result
--------	------	---------	--------	--------

Skogen et al. (2016)	predicting WHO grade II, III, IV	95 patients (grade II, III, IV) T1ce	manual 2D segmentation texture features ROC analysis	AUC = 0.91 (II, III vs. IV) AUC = 0.84 (II vs. III) AUC = 0.73 (III vs. IV)
Hsieh et al. (2017)	distinguishing lower-grade (II, III) from high- grade GBM (IV)	107 patients (grade II, III, IV) T1ce	manual 2D segmentation histogram, texture features logistic regres- sion	AUC = 0.89
Yang et al. (2018)	distinguishing lower-grade (II, III) from high- grade GBM (IV)	113 patients (grade II, III, IV) T1ce	Manual ROI segmentation (pre-trained) 2D CNN: AlexNet, GoogleNet	AUC = 0.97
Chang et al. (2018a)	IDH mutant vs. IDH wildtype	496 patients (grade II, III, IV) T1 + T1ce + T2 + FLAIR	Manual ROI segmentation 2D CNN: ResNet34 Logistic regres- sion combining age with proba- bility output	AUC = 0.95
Yu et al. (2017)	IDH1 mutant vs. IDH1 wildtype	140 patients (grade II) FLAIR	Automatic seg- mentation with 2D CNN Location, shape, texture and his- togram features SVM, AdaBoost	AUC = 0.86

Zhang et al. (2017)	IDH mutant vs. IDH wildtype	120 patients (grade III, IV) T1, T1ce, T2, FLAIR, DWI (ADC)	Semi-automatic segmentation  Anatomical, shape, texture and histogram features  Random forest classification	AUC = 0.92
Arita et al. (2018)	IDH mutant vs. IDH wildtype	199 patients (grade II, III) T1, T1ce, T2, FLAIR	Manual segmen- tation  Location, shape, texture features  LASSO regres- sion	Accuracy = 87%
Chang et al. (2018b)	IDH mutant vs. IDH wildtype 1p/19q co- deleted vs. 1p/19q Intact MGMT methy- lated vs. un- methylated	259 patients (grade II, III, IV) T1, T1ce, T2, FLAIR	Automatic seg- mentation with 2D CNN  2D CNN: resid- ual network	AUC = 0.91 (IDH) AUC = 0.88 (1p/19q) AUC = 0.81 (MGMT)
Choi et al. (2019)	IDH mutant vs. IDH wildtype 1p/19q co- deleted vs. 1p/19q Intact	463 patients (grade II, III, IV) T1, T1ce, T2, FLAIR, DSC perfusion MRI	Automatic seg- mentation with CNN followed by manual correction  2D convolu- tional LSTM	AUC = 0.95 (IDH) AUC = 0.78 (1p/19q)

Zhou et al. (2019)	IDH mutant vs. IDH wildtype IDH mutant: 1p/19q co- deleted vs. 1p/19q Intact	744 patients (grade II, III, IV) T1ce, FLAIR	Manual segmen- tation Histogram, shape, texture and age features Random forest classification	AUC = 0.92 (IDH) AUC = 0.72 (1p/19q)
Akkus et al. (2017)	1p/19q co- deleted vs. 1p/19q Intact	159 patients (grade II, III) T1ce, T2	Semi-automatic 2D segmenta- tion 2D CNN	Accuracy = 88%
Kim et al. (2019)	1p/19q co- deleted vs. 1p/19q Intact	167 patients (grade II, III, IV) T1, T1ce, T2, FLAIR	Manual segmen- tation Texture, topo- logical and pre-trained CNN features Random forest classification	AUC = 0.71
van der Voort et al. (2019)	1p/19q co- deleted vs. 1p/19q Intact	284 patients + 129 from TCIA (grade II, III) T1ce, T2	Manual segmen- tation Intensity, tex- ture, shape, texture, age and sex features SVM classifier	AUC = 0.72 (TCIA)

## 2. Materials and methods

The pipeline designed in this study consists of a segmentation stage and a  
 105 subsequent classification stage as illustrated in Figure 2. In this section, the  
 collected data and methodology are explained for both stages.

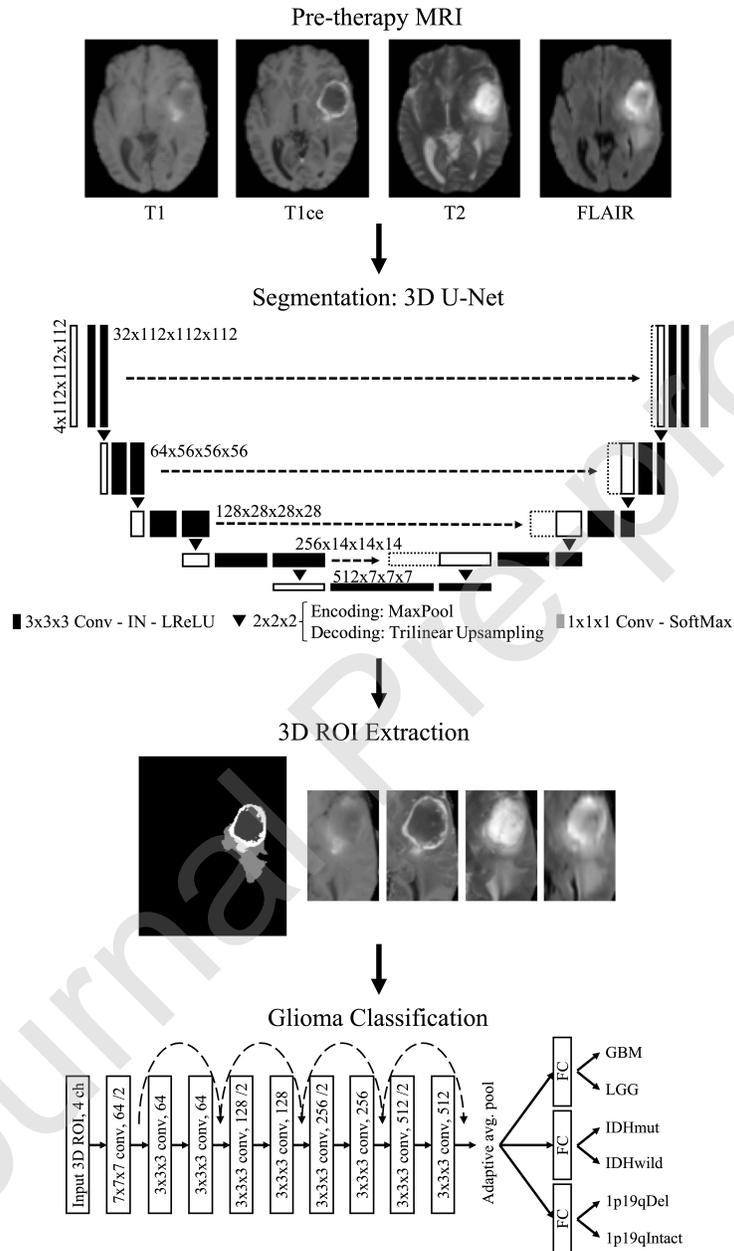


Figure 2: Schematic overview of the pipeline presented in this study. Both the segmentation and classification network architectures are illustrated

## 2.1. Glioma Segmentation

### 2.1.1. Patient data

For the segmentation, we used 335 patients (glioma WHO grade II, III and  
 110 IV) from the BraTS 2019 training dataset (Bakas et al., 2017; Menze et al.,  
 2015). This dataset includes routine clinically acquired pre-operative T1, T1ce,  
 T2 and FLAIR MRI from multiple institutions together with manual segmen-  
 tation maps denoting the GD-enhancing, peritumoral edema and the necrotic  
 and non-enhancing tumor core regions. All MRI were co-registered to the same  
 115 anatomical template, interpolated to  $1\text{ mm}^3$  voxel sizes, skull-stripped and in-  
 dependently normalized by subtracting the mean and dividing by the standard  
 deviation.

### 2.1.2. Network architecture and training

In recent BraTS challenges, U-Nets have shown state-of-the-art performance  
 120 for brain tumor segmentation. A U-Net is an encoder-decoder network that com-  
 bines semantic and spatial information through the use of skip connections from  
 the encoder to the decoder which allows to segment fine structures very well.  
 We therefore implemented a 3D U-Net similar to the architecture proposed by  
 Isensee et al. (2019) (see Figure 2). The network has four input channels (one  
 125 for each modality), 32 feature maps at the highest resolution, five levels (depths  
 in the U shape) and four output channels (background, necrosis, edema and  
 enhancing tissue). In the decoding part, the number of filters is reduced right  
 before trilinear upsampling instead of transposed convolutions to limit the num-  
 ber of parameters and memory consumption and to allow for a suitable number  
 130 of feature maps. During training we process patches of  $112 \times 112 \times 112$  and a  
 small batch size of two due to the high memory consumption of 3D convolutions.  
 Leaky ReLU activation and instance instead of batch normalization was used as  
 the exponential moving averages of mean and variance within small batches are  
 unstable. The network was trained using the ADAM optimizer with an initial  
 135 learning rate  $lr_{init} = 10^{-4}$  which is halved if the validation loss has not im-  
 proved in the last 50 epochs and an L2 weight decay of  $10^{-6}$ . A combination of

cross-entropy and multi-class soft dice loss was used as the optimization metric. The network was implemented using PyTorch and trained on an 11GB NVIDIA GTX 1080 Ti GPU.

140 Sixty patients were held out for validation and the network was trained on the remaining 275 patients. To prevent overfitting, data augmentations such as flipping and random axial rotations were applied on the fly during training.

In clinical practice, not all four MRI sequences (T1, T1ce, T2 and FLAIR) are always available. This is also the case for our dataset that is used to train 145 the classification network (see Section 2.2.1). For some patients, a good quality T1, T2 or FLAIR MRI is lacking. To increase robustness of the segmentation network to missing T1 and T2 or FLAIR modalities, channels were randomly set to zero during training. We made sure that at least the T1ce and a T2 or FLAIR sequence was available at the input.

150 The network is finally evaluated on the BraTS 2019 validation dataset containing 125 patients and the reported dice scores and Hausdorff distances were computed by the online evaluation platform (<https://ipp.cbica.upenn.edu>). The dice score denotes the percentage of overlap between the predicted and groundtruth segmentation volumes. The robust Hausdorff distance reports the 95% quantile 155 over all surface distances.

## 2.2. Glioma Classification

### 2.2.1. Patient data

To acquire a large dataset, we collected data from multiple public databases: the TCGA-GBM (Scarpance et al., 2016), TCGA-LGG (Pedano et al., 2016) and 160 LGG-1p19qDeletion (Erickson et al., 2017) collections on The Cancer Imaging Archive (TCIA) (Clark et al., 2013) and the BraTS 2019 dataset. Inclusion criteria were: a histologically proven glioma of WHO grade II, III or IV, the availability of pre-operative T1ce MRI together with a T2 and/or FLAIR sequence of sufficient quality and information on WHO grade, IDH mutation and 165 1p19q co-deletion status. In total 628 patients were included: 164 patients from TCGA-GBM, 121 from TCGA-LGG, 141 from 1p19qDeletion and 202

from BraTS 2019 (only patients that were not already included in the TCGA collections). The LGG-1p19qDeletion collection only includes a T1ce and T2 sequence. Hence the required robustness of the segmentation network to lacking  
 170 T1 and FLAIR MRI. For all patients, WHO grade information was available (337 GBM vs. 291 LGG). IDH mutation status was known for 380 patients (212 mutated vs. 168 wildtype) and 1p19q co-deletion status for 280 LGG patients (133 co-deleted vs. 147 intact). Molecular data of patients in the TCGA-GBM and TCGA-LGG collections were obtained from (Ceccarelli et al., 2016).

175 Additionally, data was retrospectively acquired at the Ghent University Hospital, with permission of the local ethics committee, informed consent was waived (Belgian registration number B670201838395 2018/1500). Using the same inclusion criteria, we collected data from 110 patients with known WHO grade (61 GBM vs. 49 LGG). For 86 patients IDH status was determined  
 180 (32 IDH mutant vs. 54 IDH wildtype) through immunohistochemistry (IHC) and for 40 LGG patients (12 co-deleted vs. 28 intact) 1p19q co-deletion status was known by fluorescence in-situ hybridization (FISH). The same pre-processing steps were performed as for the segmentation data using SPM12 (version 7219, Wellcome Trust Centre for Neuroimaging, University College  
 185 London) and MATLAB R2018b (The MathWorks, Inc., Natick, MA).

### 2.2.2. Network architecture and training

Using the segmentation mask, a tumor region of interest (ROI) is extracted from the MRI and subsequently fed into the classification network as illustrated in Figure 2. A similar architecture design is used as described in the original  
 190 ResNet paper (He et al., 2016). The architecture starts with a convolutional layer with  $64 \times 7 \times 7 \times 7$  filters and stride two followed by four residual blocks. Each residual block consists of two  $3 \times 3 \times 3$  convolutional layers with a skip connection via addition. The convolutional layers in the first block have 64 filters without downsampling of the feature maps. In the following three residual blocks the  
 195 number of filters are doubled and downsampling is directly performed by the first convolutional layer that has a stride of two. This results in 512 features maps

after the last convolutional layer. To allow identity shortcuts, input and output of a residual block must have the same dimensions. This is not the case in the last three residual blocks where the input is matched to the output dimensions through a  $1 \times 1 \times 1$  convolutional layer with stride two. Every convolutional layer is succeeded with instance normalization and Leaky ReLU activation (negative slope of 0.01). The adaptive average pool layer after the last convolutional layer allows the network to process different ROI input sizes hence no resizing to a fixed shape is required. In the end the network splits into three separate fully connected (FC) layers to simultaneously predict WHO grade, IDH mutation and 1p19q co-deletion. This so-called multi-task learning helps the network to learn features that are relevant for multiple tasks, reduces the risk of overfitting and allows a better generalization (Caruana, 1997). As explained in the introduction, MRI features describing enhancing regions and tumor margins are important to predict grade, IDH and 1p19q status which shows that these tasks are very much related and that knowledge on one characteristic is informative for the other markers as well. Moreover, not all ground truth labels are available for every patient in our dataset. Multi-task learning allows us to deal with missing labels and train one network on all data instead of training separate networks for each task on a smaller dataset. The 1p19q co-deletion classifier is only trained for LGG patients as all GBM patients in the dataset are 1p19q intact and the 2016 WHO classification system does not include 1p19q status for GBM cases (see Figure 1).

The network is trained with AdamW optimization ( $lr_{init} = 10^{-5}$ ), L2 weight decay of  $10^{-2}$ , a batch size of eight and focal binary cross-entropy loss. Focal loss weights the contribution of each sample based on the classification error and thereby reduces the contribution of already correctly classified samples. This is especially useful to deal with class imbalance. The loss is calculated for each task separately on all samples in the batch with known ground truth labels and averaged to a global loss which is backpropagated through the network. If the validation loss did not improve in the last 10 epochs the learning rate is halved and early stopping occurs after no improvement for 30 epochs. In the last fully

connected layer, dropout is applied with probability of 10%. Different hyperparameters of the network were tuned based on the validation set. The network  
230 was implemented with the PyTorch deep learning framework and trained on an 11GB NVIDIA GeForce RTX 2080 Ti.

The 628 patients are split into a training set of 458 (264 GBM vs. 194 LGG, 123 IDH mutant vs. 87 IDH wildtype and 83 1p19q co-deleted vs. 100 1p19q intact), a validation set of 70 (27 GBM vs. 43 LGG, 41 IDH mutant vs. 29 IDH  
235 wildtype and 20 1p19q co-deleted vs. 23 1p19q intact) and a test set of 100 (46 GBM vs. 54 LGG, 48 IDH mutant vs. 52 wildtype and 30 1p19q co-deleted vs. 24 1p19q intact) patients. For patients in the validation and test set, all ground truth labels were available and test patients were not used in the training set of the segmentation network in order to evaluate the system on new cases that  
240 both the segmentation and classification stages have never seen before. The dataset was augmented with random flipping, axial rotations, intensity scaling, elastic transform and setting input channels to zero as was done to train the segmentation network. Data from the Ghent University Hospital was used to evaluate the performance of the classification pipeline on an entirely independent  
245 dataset.

### 3. Results

#### 3.1. Glioma Segmentation

Segmentation results on the BraTS 2019 validation data are summarized in Table 2 and Table 3. Dice scores and Hausdorff distances are reported for the  
250 enhancing tumor (ET), whole tumor (WT) and tumor core (TC) regions and for different available modalities: all four sequences, only T1ce and FLAIR or only T1ce and T2. To illustrate the increased robustness to missing modalities, the results are included with (Table 2) and without (Table 3) randomly setting input channels to zero during training. A whole tumor dice score of 90% is achieved  
255 which lowers to 89% and 87% when only the T1ce and FLAIR and T1ce and T2 MRI are available respectively. Without randomly removing channels while

Table 2: Segmentation results on the BraTS2019 validation data with randomly setting input channels to zero while training. Metrics were computed by the online evaluation platform.

Available modalities	Dice Score (%)			Hausdorff distance (mm)		
	ET	WT	TC	ET	WT	TC
T1, T1ce, T2, FLAIR	75.71	89.81	83.18	5.08	4.99	6.66
T1ce, FLAIR	74.35	89.37	82.74	4.34	5.12	6.82
T1ce, T2	74.09	86.98	82.20	5.58	7.15	7.37

Table 3: Segmentation results on the BraTS2019 validation data without randomly setting input channels to zero while training. Metrics were computed by the online evaluation platform.

Available modalities	Dice Score (%)			Hausdorff distance (mm)		
	ET	WT	TC	ET	WT	TC
T1, T1ce, T2, FLAIR	76.33	90.02	79.68	3.89	5.72	6.97
T1ce, FLAIR	64.37	82.77	69.33	51.88	15.09	26.13
T1ce, T2	62.46	60.98	59.86	9.02	23.03	23.27

training, the difference in performance is larger with WT dice scores of 90%, 83% and 61%.

The network achieves a high average WT specificity of 99% and a sensitivity of 87%. Figure 3 shows an example segmentation of the patient from the validation set with the lowest whole tumor dice score of 32% (when only providing T1ce and T2). Some of the edema surrounding the segmented tumor core is missed. When providing all four MRI the WT dice increases to 54%. On the Nvidia 1080 Ti GPU, a patient’s MRI can be segmented in around 0.3 seconds.

### 3.2. Glioma Classification

In Table 4 the results are presented of the multi-task classification network. For each task (WHO grade, IDH mutation and 1p19q co-deletion status) the AUC, Matthews Correlation Coefficient (MCC), accuracy, sensitivity and specificity scores are included. MCC is a balanced performance measure for binary classifications that considers all four components of the confusion matrix and

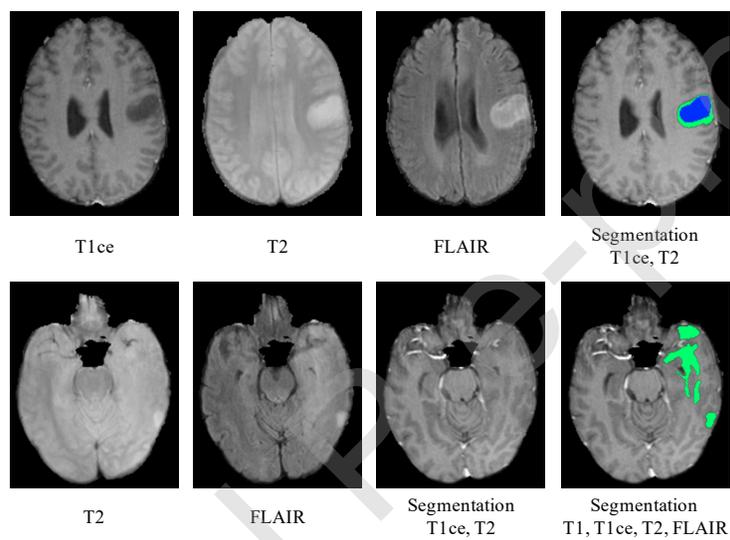


Figure 3: Example MRI and segmentation (overlaid on T1ce) for two different slices of the patient with lowest whole tumor dice score. Blue denotes the necrotic and non-enhancing tissue, green indicates the peritumoral edema. The predicted segmentations are included when only providing the T1ce and T2 sequences and when providing all sequences (for the bottom slice).

Table 4: Classification performance on the TCIA and Ghent University Hospital (GUH) test data. AUC, Matthews Correlation Coefficient (MCC), accuracy, sensitivity and specificity scores are reported for all three tasks: WHO grade, IDH mutation and 1p19q co-deletion status. A case is classified as Glioblastoma (WHO grade IV), IDH mutant and 1p19q co-deleted respectively if the predicted probability is higher than 0.5.

Dataset	Task	AUC	MCC	Acc.	Sens.	Spec.
TCIA test data	GBM vs. LGG	93.28	80.26	90.00	93.48	87.04
	IDH mutation	94.03	78.00	89.00	89.58	88.46
	1p19q co-deletion	82.08	66.16	83.33	86.67	79.17
GUH data	GBM vs. LGG	93.98	79.81	90.00	90.16	89.80
	IDH mutation	86.23	52.92	75.58	84.38	70.37
	1p19q co-deletion	86.61	40.48	75.00	58.33	82.14

has a value between -1 and 1 (with zero corresponding to random prediction). The sensitivity scores indicate the percentage of GBM, IDH mutant and 1p19q co-deleted cases that are correctly classified as such. The results on the unseen TCIA test data show high classification performances with AUC scores of 93% and 94% for grade and IDH status respectively. Predicting 1p19q co-deletion status for lower-grade glioma is harder but still an AUC of 82% is achieved. The performance was also evaluated on the completely independent dataset from the Ghent University Hospital. The resulting AUC scores on the GUH data are 94%, 86% and 87% for grade, IDH and 1p19q status respectively.

#### 280 4. Discussion

The segmentation results on the BraTS 2019 validation set show that very good dice scores are achieved. With an average whole tumor dice score of 90%, our segmentation algorithm matches the performance of state-of-the-art algorithms of the BraTS 2019 challenge with the top three winning algorithms obtaining a mean WT dice score of 91% according to the validation leaderboard 285 (Bakas and Sako, 2019). We believe that the obtained performance with the implemented U-Net is sufficient for the current task, as small variations be-

tween manual and predicted segmentations won't have a strong influence on the tumour ROI. Additionally, the network shows increased robustness to missing modalities when randomly setting the T1 and T2 or FLAIR sequence to zero while training. There is only a small decrease in performance when only providing the T1ce and FLAIR or T1ce and T2 scans compared to all four MRI as input. Without randomly excluding image modalities during training, the performance with only two modalities is much lower. This is especially useful as not all four MRI modalities are available for all patients in both our public dataset and the Ghent University Hospital dataset. This way an accurate segmentation and ROI extraction could still be obtained for these patients. The very high specificity and slightly lower sensitivity indicates that segmentation inaccuracies are due to parts of the surrounding edema that are not detected by the network. This is illustrated by the example shown in Figure 3 with the lowest WT dice score. Based only on the T1ce and T2 MRI, the network is able to segment the tumor core but misses the surrounding edema which is not clearly visible on the T2 sequence. When adding the FLAIR sequence where the edema is more evident, the network is able to detect more of the edema tissue.

With the trained segmentation network, a 3D tumor ROI is extracted and used as input to the subsequent 3D CNN that predicts binary tumor grade, IDH mutation and 1p19q co-deletion status. For binary grade prediction, very high accuracies of 90% on both the TCIA and GUH test data are achieved. This shows that the network is able to accurately distinguish glioblastoma from lower-grade glioma and generalizes well to unseen data from different institutions. The IDH mutation prediction performance is high on the TCIA test set (AUC of 94%). On the GUH data, the performance is lower with an AUC of 86%. Especially a lower specificity of 70% compared to 88% is observed. This difference in performance might be because immunohistochemistry was used to determine IDH status for the GUH data while for the TCGA data IDH status was assessed using gene sequencing (Ceccarelli et al., 2016). However, a negative IDH status using IHC does not necessarily mean an IDH wildtype tumor and if no sequencing is available the resulting diagnosis suggested by the WHO

is astrocytoma, not otherwise specified (NOS) (Louis et al., 2016). The GUH  
320 database contains 14 IDH wildtype astrocytoma while this diagnosis should be  
very rare according to the WHO. Hence some IDH mutant astrocytoma might  
be missed with IHC resulting in more false positives of the model and thus a  
lower specificity. An additional limitation to the dataset is that there are only  
two IDH mutant glioblastoma in the TCIA training set making it very unlikely  
325 that the network will predict this combination of classes. Therefore, the four  
GBM with IDH mutation in the GUH database were predicted as GBM, IDH  
wildtype. In terms of 1p19q co-deletion status prediction of lower-grade glioma,  
a good performance is achieved on both the TCIA and GUH datasets (AUC of  
82% and 87% respectively). Although 1p19q status was known for the GBM  
330 cases in the TCGA-GBM collection (all 1p19q intact), we only included LGG  
patients as this marker is only considered for those patients according to the  
WHO guidelines (Figure 1). Including the GBM cases would increase the overall  
prediction accuracy of 1p19q status but would introduce a large data imbalance  
and thereby decrease the performance for LGG cases. Results on the GUH  
335 dataset show a lower sensitivity compared to the results on the more balanced  
TCIA test set. In the GUH dataset, 1p19q status was only available for 40  
LGG patients with just 12 1p19q co-deleted cases which might be too small to  
obtain reliable performance estimations. Depending on the classification thresh-  
old, the sensitivity can also be optimized. For example, with a threshold of 0.45  
340 the sensitivity on the GUH dataset increased to 75% with the same specificity.

In this study, we trained a 3D classification network to classify the entire  
tumor ROI. Current applications of CNNs for brain tumor classification are  
mostly 2D, taking only a small part of the tumor into account while brain  
tumors have a very heterogeneous appearance with strong variations between  
345 different slices. Furthermore, extracting only the tumor ROI allows the network  
to focus on this region but context information on surrounding tissues and  
location is excluded. Including this information may further improve diagnostic  
performance.

The clinical translation potential of the developed pipeline is strong as only

350 routinely acquired MRI are necessary as input and no further human inter-  
action is required. Data preprocessing is minimal, and the segmentation and  
classification takes less than 5 seconds on an NVIDIA GeForce RTX 2080 Ti  
GPU.

## 5. Conclusion

355 In conclusion, we developed a fully automatic 3D pipeline to segment glioma  
and non-invasively predict important (molecular) markers according to the WHO  
classification guidelines with high diagnostic performance. The segmentation al-  
gorithm shows increased robustness to missing image modalities by randomly  
excluding input MRI during training. Through the use of multi-task learning to  
360 handle missing labels, one classification network could be trained on a large  
multi-institutional database. Evaluation on an independent private dataset  
demonstrated the generalizability of the algorithm. The non-invasive assess-  
ment of clinically relevant genetic mutations can help to characterize glioma  
and thereby guide therapy and surgery planning.

## 365 6. Acknowledgements

This work was supported by the Ghent University Special Research Fund  
(BOF, project number 01N04817).

## References

370 Akkus, Z., Ali, I., Sedlář, J., Agrawal, J.P., Parney, I.F., Gian-  
nini, C., Erickson, B.J., 2017. Predicting Deletion of Chromo-  
somal Arms 1p/19q in Low-Grade Gliomas from MR Images Us-  
ing Machine Intelligence. *Journal of digital imaging* 30, 469–  
476. URL: <http://www.ncbi.nlm.nih.gov/pubmed/28600641>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5537096>,  
375 doi:10.1007/s10278-017-9984-3.

- Arita, H., Kinoshita, M., Kawaguchi, A., Takahashi, M., Narita, Y., Terakawa, Y., Tsuyuguchi, N., Okita, Y., Nonaka, M., Moriuchi, S., Takagaki, M., Fujimoto, Y., Fukai, J., Izumoto, S., Ishibashi, K., Nakajima, Y., Shofuda, T., Kanematsu, D., Yoshioka, E., Kodama, Y., Mano, M., Mori, K., Ichimura, K.,  
380 Kanemura, Y., 2018. Lesion location implemented magnetic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. *Scientific Reports* 8, 11773. URL: <http://www.nature.com/articles/s41598-018-30273-4>, doi:10.1038/s41598-018-30273-4.
- Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C., 2017. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* 4, 170117. URL: <http://www.ncbi.nlm.nih.gov/pubmed/28872634><http://www.nature.com/articles/sdata2017117>, doi:10.1038/sdata.2017.117.  
385
- Bakas, S., Reyes, M., Jakab, A., et al., 2018. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *CoRR* abs/1811.0. URL: <http://arxiv.org/abs/1811.02629>, arXiv:1811.02629.  
390
- Bakas, S., Sako, C., 2019. MICCAI-BraTS 2019 Leaderboard. URL: <https://www.cbica.upenn.edu/BraTS19/lboardValidation.html>.  
395
- Carrillo, J.A., Lai, A., Nghiemphu, P.L., Kim, H.J., Phillips, H.S., Kharbanda, S., Moftakhar, P., Lalaezari, S., Yong, W., Ellingson, B.M., Cloughesy, T.F., Pope, W.B., 2012. Relationship between Tumor Enhancement, Edema, IDH1 Mutational Status, MGMT Promoter Methylation, and Survival in Glioblastoma. *American Journal of Neuroradiology* 33, 1349–1355.  
400 URL: <http://dx.doi.org/10.3174/ajnr.A2950>, doi:10.3174/ajnr.A2950.
- Caruana, R., 1997. Multitask Learning. *Machine Learning* 28, 41–75. URL: <http://link.springer.com/10.1023/A:1007379606734>, doi:10.1023/A:1007379606734.

405 Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., Anjum, S., Wang, J., Manyam, G., Zoppoli, P., Ling, S., Rao, A.A., Grifford, M., Cherniack, A.D., Zhang, H., Poisson, L., Carlotti, C.G., Junior, Tirapelli, D.P.d.C., Rao, A., Mikkelsen, T., Lau, C.C., Yung, W.A., Rabadan, R., Huse, J., Brat, 410 D.J., Lehman, N.L., Barnholtz-Sloan, J.S., Zheng, S., Hess, K., Rao, G., Meyerson, M., Beroukhi, R., Cooper, L., Akbani, R., Wrensch, M., Haussler, D., Aldape, K.D., Laird, P.W., Gutmann, D.H., Network, T.R., Nushmeh, H., Iavarone, A., Verhaak, R.G., 2016. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 415 550. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4754110/>, doi:10.1016/J.CELL.2015.12.028.

Chang, K., Bai, H.X., Zhou, H., Su, C., Bi, W.L., Agbodza, E., Kavouridis, V.K., Senders, J.T., Boaro, A., Beers, A., Zhang, B., Capellini, A., Liao, W., Shen, Q., Li, X., Xiao, B., Cryan, J., Ramkissoon, S., Ramkissoon, L., Ligon, 420 K., Wen, P.Y., Bindra, R.S., Woo, J., Arnaout, O., Gerstner, E.R., Zhang, P.J., Rosen, B.R., Yang, L., Huang, R.Y., Kalpathy-Cramer, J., 2018a. Residual Convolutional Neural Network for the Determination of IDH Status in Low-and High- Grade Gliomas from MR Imaging. *Clin Cancer Res* 24, 1073-1081. URL: <http://clincancerres.aacrjournals.org/content/clincanres/early/2018/02/01/1078-0432.CCR-17-2236.full.pdf>, 425 doi:10.1158/1078-0432.CCR-17-2236.

Chang, P., Grinband, J., Weinberg, B.D., Bardis, M., Khy, M., Cadena, G., Su, M.Y., Cha, S., Filippi, C.G., Bota, D., Baldi, P., Poisson, L.M., Jain, R., Chow, D., 2018b. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *AJNR. American journal of* 430 *neuroradiology* doi:10.3174/ajnr.A5667.

Choi, K.S., Choi, S.H., Jeong, B., 2019. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro-Oncology* URL:

435 <https://academic.oup.com/neuro-oncology/advance-article/doi/10.1093/neuonc/noz095/5498696>, doi:10.1093/neuonc/noz095.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F., 2013. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging* 26, 1045–1057. 440 URL: <http://link.springer.com/10.1007/s10278-013-9622-7>, doi:10.1007/s10278-013-9622-7.

Eckel-Passow, J.E., Lachance, D.H., Molinaro, A.M., Walsh, K.M., Decker, P.A., Sicotte, H., Pekmezci, M., Rice, T., Kosel, M.L., Smirnov, I.V., Sarkar, G., Caron, A.A., Kollmeier, T.M., Praska, C.E., Chada, A.R., Halder, C., 445 Hansen, H.M., McCoy, L.S., Bracci, P.M., Marshall, R., Zheng, S., Reis, G.F., Pico, A.R., O'Neill, B.P., Buckner, J.C., Giannini, C., Huse, J.T., Perry, A., Tihan, T., Berger, M.S., Chang, S.M., Prados, M.D., Wiemels, J., Wiencke, J.K., Wrensch, M.R., Jenkins, R.B., 2015. Glioma Groups Based on 1p/19q, 450 IDH , and TERT Promoter Mutations in Tumors. *New England Journal of Medicine* 372, 2499–2508. URL: <http://www.nejm.org/doi/10.1056/NEJMoa1407279>, doi:10.1056/NEJMoa1407279.

Erickson, B., Akkus, Z., Sedlar, J., Korfiatis, P., 2017. Data From LGG-1p19qDeletion. The Cancer Imaging Archive doi:10.7937/K9/TCIA.2017. 455 dwehtz9v.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90.

Hsieh, K.L.C., Lo, C.M., Hsiao, C.J., 2017. Computer-aided grading of gliomas based on local and global MRI features. *Computer Methods and Programs in Biomedicine* 139, 31–38. URL: <https://www.sciencedirect.com/science/article/pii/S0169260716303406>, doi:10.1016/J.CMPB.2016.10.021. 460

- Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.,  
2019. No New-Net, in: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F.,  
465 Reyes, M., van Walsum, T. (Eds.), *Brainlesion: Glioma, Multiple Sclerosis,  
Stroke and Traumatic Brain Injuries*. Springer, Cham, pp. 234–244. URL:  
[http://link.springer.com/10.1007/978-3-030-11726-9\\_{\\_}21](http://link.springer.com/10.1007/978-3-030-11726-9_{_}21), doi:10.  
1007/978-3-030-11726-9\_21.
- Jackson, R.J., Fuller, G.N., Abi-Said, D., Lang, F.F., Gokaslan, Z.L., Shi,  
470 W.M., Wildrick, D.M., Sawaya, R., 2001. Limitations of stereotactic  
biopsy in the initial management of gliomas. *Neuro-oncology* 3, 193–  
200. URL: <http://www.ncbi.nlm.nih.gov/pubmed/11465400>[http://www.  
pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1920616](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1920616).
- Jansen, N.L., Graute, V., Armbruster, L., Suchorska, B., Lutz, J., Eigenbrod,  
475 S., Cumming, P., Bartenstein, P., Tonn, J.C., Kreth, F.W., la Fougère,  
C., 2012. MRI-suspected low-grade glioma: is there a need to perform  
dynamic FET PET? *European Journal of Nuclear Medicine and Molec-  
ular Imaging* 39, 1021–1029. URL: [http://link.springer.com/10.1007/  
s00259-012-2109-9](http://link.springer.com/10.1007/s00259-012-2109-9), doi:10.1007/s00259-012-2109-9.
- 480 Johnson, D.R., Diehn, F.E., Giannini, C., Jenkins, R.B., Jenkins, S.M., Parney,  
I.F., Kaufmann, T.J., 2017. Genetically Defined Oligodendroglioma Is Char-  
acterized by Indistinct Tumor Borders at MRI. *AJNR. American journal of  
neuroradiology* 38, 678–684. URL: [http://www.ncbi.nlm.nih.gov/pubmed/  
28126746](http://www.ncbi.nlm.nih.gov/pubmed/28126746), doi:10.3174/ajnr.A5070.
- 485 Khalid, L., Carone, M., Dumrongpisutikul, N., Intrapromkul, J., Bonekamp,  
D., Barker, P.B., Yousem, D.M., 2012. Imaging Characteristics of Oligoden-  
drogliomas That Predict Grade. *American Journal of Neuroradiology* 33, 852–  
857. URL: <http://dx.doi.org/10.3174/ajnr.A2895>, doi:10.3174/ajnr.  
A2895.
- 490 Kim, D., Wang, N., Ravikumar, V., Raghuram, D.R., Li, J., Patel, A.,  
Wendt, R.E., Rao, G., Rao, A., 2019. Prediction of 1p/19q Codele-

tion in Diffuse Glioma Patients Using Pre-operative Multiparametric Magnetic Resonance Imaging. *Frontiers in Computational Neuroscience* 13, 52. URL: <https://www.frontiersin.org/article/10.3389/fncom.2019.00052/full>, doi:10.3389/fncom.2019.00052.

495

Law, M., Yang, S., Wang, H., Babb, J.S., Johnson, G., Cha, S., Knopp, E.A., Zagzag, D., 2003. Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging. *AJNR American journal of neuroradiology* 24, 1989–1998. URL: <papers3://publication/uuid/C58DFDF1-444E-440B-BD8E-55C3C32585DE>, doi:10.3174/ajnr.a3686.

500

Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica* 131, 803–820. URL: <http://link.springer.com/10.1007/s00401-016-1545-1>, doi:10.1007/s00401-016-1545-1.

505

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekeley, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K., 2015. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*

515

520

34, 1993–2024. URL: <http://ieeexplore.ieee.org/document/6975210/>,  
doi:10.1109/TMI.2014.2377694.

Myronenko, A., 2019. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization, in: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F.,  
525 Reyes, M., van Walsum, T. (Eds.), *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer, Cham. pp. 311–320. URL:  
[http://link.springer.com/10.1007/978-3-030-11726-9\\_28](http://link.springer.com/10.1007/978-3-030-11726-9_28), doi:10.1007/978-3-030-11726-9\_28.

Ostrom, Q.T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., Barnholtz-  
530 Sloan, J.S., 2018. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015. *Neuro-Oncology* 20, iv1–iv86. URL: [https://academic.oup.com/neuro-oncology/article/20/suppl\\_4/iv1/5090960](https://academic.oup.com/neuro-oncology/article/20/suppl_4/iv1/5090960), doi:10.1093/neuonc/noy131.

535 Pallud, J., Capelle, L., Taillandier, L., Fontaine, D., Mandonnet, E., Guillevin, R., Bauchet, L., Peruzzi, P., Laigle-Donadey, F., Kujas, M., Guyotat, J., Baron, M.H., Mokhtari, K., Duffau, H., 2009. Prognostic significance of imaging contrast enhancement for WHO grade II gliomas. *Neuro-Oncology* 11, 176–182. URL: <http://academic.oup.com/neuro-oncology/article/11/2/176/1028590/Prognostic-significance-of-imaging-contrast>,  
540 doi:10.1215/15228517-2008-066.

Pedano, N., Flanders, A.E., Scarpace, L., Mikkelsen, T., Eschbacher, J.M., Hermes, B., Sisneros, V., Barnholtz-Sloan, J., Ostrom, Q., 2016. Radiology Data from The Cancer Genome Atlas Low Grade Glioma [TCGA-LGG] collection.  
545 The Cancer Imaging Archive URL: <https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG#{ba985fbb9bbf4d15987b8e98a6962471>,  
doi:10.7937/K9/TCIA.2016.L4LTD3TK.

Qi, S., Yu, L., Li, H., Ou, Y., Qiu, X., Ding, Y., Han, H., Zhang, X., 2014. Isocitrate dehydrogenase mutation is associated with tumor location and magnetic

550 resonance imaging characteristics in astrocytic neoplasms. *Oncology letters* 7,  
1895–1902. URL: <http://www.ncbi.nlm.nih.gov/pubmed/24932255><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4049752>,  
doi:10.3892/ol.2014.2013.

Reuss, D.E., Mamatjan, Y., Schrimpf, D., Capper, D., Hovestadt, V., Kratz,  
555 A., Sahm, F., Koelsche, C., Korshunov, A., Olar, A., Hartmann, C., Reijneveld, J.C., Wesseling, P., Unterberg, A., Platten, M., Wick, W., Herold-Mende, C., Aldape, K., von Deimling, A., 2015. IDH mutant diffuse and anaplastic astrocytomas have similar age at presentation and little difference in survival: a grading problem for WHO. *Acta Neuropathologica* 129, 867–873. URL: <http://link.springer.com/10.1007/s00401-015-1438-8>,  
560 doi:10.1007/s00401-015-1438-8.

Scarpace, L., Mikkelsen, T., Cha, S., Rao, S., Tekchandani, S., Gutman, D., Saltz, J., Erickson, B.J., Pedano, N., Flanders, A.E., Barnholtz-Sloan, J., Ostrom, Q., Barboriak, D., Pierce, L.J., 2016. Radiology Data from The Cancer  
565 Genome Atlas Glioblastoma Multiforme [TCGA-GBM] collection. The Cancer Imaging Archive doi:10.7937/K9/TCIA.2016.RNYFUYE9.

Skogen, K., Schulz, A., Dormagen, J.B., Ganeshan, B., Helseth, E., Server, A., 2016. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *European Journal of Radiology* 85, 824–829. URL: <https://www.sciencedirect.com/science/article/pii/S0720048X16300134>, doi:10.1016/J.EJRAD.2016.01.013.  
570

Sonoda, Y., Shibahara, I., Kawaguchi, T., Saito, R., Kanamori, M., Watanabe, M., Suzuki, H., Kumabe, T., Tominaga, T., 2015. Association between molecular alterations and tumor location and MRI characteristics in anaplastic gliomas. *Brain Tumor Pathology* 32, 99–104. URL: <http://link.springer.com/10.1007/s10014-014-0211-3>, doi:10.1007/s10014-014-0211-3.  
575

The Cancer Genome Atlas Research Network, 2015. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *New England Jour-*

nal of Medicine 372, 2481–2498. URL: <http://www.nejm.org/doi/10.1056/NEJMoa1402121>, doi:10.1056/NEJMoa1402121.

580

van der Voort, S.R., Incekara, F., Wijnenga, M.M., Kapas, G., Gardeniers, M., Schouten, J.W., Starmans, M.P., Nandoe Tewarie, R., Lycklama, G.J., French, P.J., Dubbink, H.J., van den Bent, M.J., Vincent, A.J., Niessen, W.J., Klein, S., Smits, M., 2019. Predicting the 1p/19q Codeletion Status of Presumed Low-Grade Glioma with an Externally Validated Machine Learning Algorithm. *Clinical Cancer Research* 25, 7455–7462. URL: <http://clincancerres.aacrjournals.org/http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-19-1127>, doi:10.1158/1078-0432.CCR-19-1127.

585

Weller, M., van den Bent, M., Tonn, J.C., Stupp, R., Preusser, M., Cohen-Jonathan-Moyal, E., Henriksson, R., Rhun, E.L., Balana, C., Chinot, O., Bendszus, M., Reijneveld, J.C., Dhermain, F., French, P., Marosi, C., Watts, C., Oberg, I., Pilkington, G., Baumert, B.G., Taphoorn, M.J.B., Hegi, M., Westphal, M., Reifenberger, G., Soffietti, R., Wick, W., 2017. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *The Lancet Oncology* 18, e315–e329. URL: <https://www.sciencedirect.com/science/article/pii/S1470204517301948?via=ihub>, doi:10.1016/S1470-2045(17)30194-8.

590

595

Wijnenga, M.M.J., Mattni, T., French, P.J., Rutten, G.J., Leenstra, S., Kloet, F., Taphoorn, M.J.B., van den Bent, M.J., Dirven, C.M.F., van Veelen, M.L., Vincent, A.J.P.E., 2017. Does early resection of presumed low-grade glioma improve survival? A clinical perspective. *Journal of neuro-oncology* 133, 137–146. URL: <http://www.ncbi.nlm.nih.gov/pubmed/28401374><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5495869>, doi:10.1007/s11060-017-2418-8.

600

605

Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos,

I., Batinic-Haberle, I., Jones, S., Riggins, G.J., Friedman, H., Friedman, A.,  
Reardon, D., Herndon, J., Kinzler, K.W., Velculescu, V.E., Vogelstein, B.,  
610 Bigner, D.D., . IDH1 and IDH2 mutations in gliomas. *New England Journal*  
*of Medicine* , 765–773doi:10.1056/NEJMoa0808710.

Yang, Y., Yan, L.F., Zhang, X., Han, Y., Nan, H.Y., Hu, Y.C., Hu, B.,  
Yan, S.L., Zhang, J., Cheng, D.L., Ge, X.W., Cui, G.B., Zhao, D.,  
Wang, W., 2018. Glioma Grading on Conventional MR Images: A  
615 Deep Learning Study With Transfer Learning. *Frontiers in neuroscience*  
12, 804. URL: <http://www.ncbi.nlm.nih.gov/pubmed/30498429><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6250094>,  
doi:10.3389/fnins.2018.00804.

Yu, J., Shi, Z., Lian, Y., Li, Z., Liu, T., Gao, Y., Wang, Y., Chen, L., Mao,  
620 Y., 2017. Noninvasive IDH1 mutation estimation based on a quantitative ra-  
diomics approach for grade II glioma. *European Radiology* 27, 3509–3522.  
URL: <http://link.springer.com/10.1007/s00330-016-4653-3>, doi:10.  
1007/s00330-016-4653-3.

Zhang, B., Chang, K., Ramkissoon, S., Tanguturi, S., Bi, W.L., Reardon,  
625 D.A., Ligon, K.L., Alexander, B.M., Wen, P.Y., Huang, R.Y., 2017. Multi-  
modal MRI features predict isocitrate dehydrogenase genotype in high-grade  
gliomas. *Neuro-Oncology* 19, 109–117. URL: [https://academic.oup.com/  
neuro-oncology/article-lookup/doi/10.1093/neuonc/now121](https://academic.oup.com/neuro-oncology/article-lookup/doi/10.1093/neuonc/now121), doi:10.  
1093/neuonc/now121.

630 Zhou, H., Chang, K., Bai, H.X., Xiao, B., Su, C., Bi, W.L., Zhang, P.J.,  
Senders, J.T., Vallières, M., Kavouridis, V.K., Boaro, A., Arnaout, O.,  
Yang, L., Huang, R.Y., 2019. Machine learning reveals multimodal  
MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status  
in diffuse low- and high-grade gliomas. *Journal of neuro-oncology* 142,  
635 299–307. URL: <http://www.ncbi.nlm.nih.gov/pubmed/30661193><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6711193>.

[//www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6510979](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6510979),  
doi:10.1007/s11060-019-03096-0.

Journal Pre-proof