



**HAL**  
open science

# Borrowing strength from adults: Transferability of AI algorithms for paediatric brain and tumour segmentation

Maxime Draï, Benoit Testud, Gilles Brun, Jean-François Hak, Didier Scavarda, Nadine Girard, Jan-Patrick Stellmann

## ► To cite this version:

Maxime Draï, Benoit Testud, Gilles Brun, Jean-François Hak, Didier Scavarda, et al.. Borrowing strength from adults: Transferability of AI algorithms for paediatric brain and tumour segmentation. European Journal of Radiology, 2022, 151, pp.110291. 10.1016/j.ejrad.2022.110291 . hal-03772222

**HAL Id: hal-03772222**

**<https://amu.hal.science/hal-03772222>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Full title page

**Title:** Borrowing strength from adults: Transferability of AI algorithms for paediatric brain and tumour segmentation

Maxime Drai, MD<sup>1\*</sup>, Benoit Testud, MD<sup>1,2,3\*</sup>, Gilles Brun, MD<sup>1</sup>, Jean-François Hak, MD<sup>1,4</sup>,  
Didier Scavarda, PhD<sup>5</sup>, Nadine Girard, PhD<sup>2,3</sup>, Jan-Patrick Stellmann, MD<sup>1,2,3</sup>

\*These authors contributed equally to this work.

1 APHM La Timone, Department of Neuroradiology, Marseille, France

2 Aix-Marseille Univ, CNRS, CRMBM, UMR 7339, Marseille, France

3 APHM La Timone, CEMEREM, Marseille, France

4 Aix-Marseille Univ, CERIMED, Marseille, France

5 APHM La Timone, Department of pediatric neurosurgery, Marseille, France

**Institution:** AP-HM La Timone, Department of Neuroradiology, 264 Rue Saint Pierre, 13385  
Marseille CEDEX 05, Marseille, France.

### Corresponding author:

Jan-Patrick STELLMANN

Mail: [jan-patrick.stellmann@univ-amu.fr](mailto:jan-patrick.stellmann@univ-amu.fr)

Street address: APHM La Timone, Department of diagnostic and interventional  
neuroradiology, 264 Saint Pierre Street, 13385 Marseille CEDEX 05, France

**Title:** Borrowing strength from adults: Transferability of AI algorithms for paediatric brain and tumour segmentation

**Manuscript Type:** Original Research

## **Abstract**

**Purpose:** AI brain tumour segmentation and brain extraction algorithms promise better diagnostic and follow-up of brain tumours in adults. The development of such tools for paediatric populations is restricted by limited training data but careful adaption of adult algorithms to paediatric population might be a solution. Here, we aim exploring the transferability of algorithms for brain (HD-BET) and tumour segmentation (HD-GLIOMA) in adults to paediatric imaging studies.

**Method:** In a retrospective cohort, we compared automated segmentation with expert masks. We used the dice coefficient for evaluating the similarity and multivariate regressions for the influence of covariates. We explored the feasibility of automatic tumor classification based on diffusion data.

**Results:** In 42 patients (mean age 7 years, 9 below 2 years, 26 males), segmentation was excellent for brain extraction (mean dice 0.99, range 0.85-1), moderate for segmentation of contrast-enhancing tumours (mean dice 0.67, range 0-1), and weak for non-enhancing T2-signal abnormalities (mean dice 0.41). Precision was better for enhancing tumour parts ( $p < 0.001$ ) and for malignant histology ( $p = 0.006$  and  $p = 0.012$ ) but independent from myelinisation as indicated by the age ( $p = 0.472$ ). Automated tumour grading based on mean diffusivity (MD) values from automated masks was good (AUC=0.86) but tended to be less accurate than MD values from expert masks (AUC=1,  $p = 0.208$ ).

**Conclusion:** HD-BET provides a reliable extraction of the paediatric brain. HD-GLIOMA works moderately for contrast-enhancing tumours parts. Without optimization, brain tumor AI algorithms trained on adults and used on paediatric patients may yield acceptable results depending on the clinical scenario.

**Key words:** Paediatric brain tumour; MRI; AI; Segmentation; Grading

## Introduction

Paediatric tumours of the central nervous system are the second leading cause of childhood cancer and the leading cause of death from solid tumours in those younger than 19 years[1]. The standard treatment strategy for paediatric tumours is often a mixed approach combining surgical resection possibly followed by radiotherapy and chemotherapy. Prognosis is associated with complete resection of the primary tumour[2]. Optimisation and standardisation of diagnosis, grading and follow-up are needed to improve and individualise treatments[3,4]. From a conceptual point of view, increasing the precision of imaging-based diagnosis and prognosis before biopsy or surgery would be very helpful in a complex and difficult care situation.

Automatic tumour segmentation would democratise the use of volumetric measurement to standardise the assessment of therapeutic response[5], and have many other uses such as pre-interventional feature extraction, multidimensional analysis, and classification for many applications[6]. Thanks to the continuous increase in computing power, the last few years have seen the emergence of deep learning, which seems to be the best approach for automatic segmentation and classification[7]. However, there are many methodological issues in creating and evaluating these algorithms. Several limitations can be attributed to one essential problem: lack of available image data for training and testing of AI algorithms. Large data sets with high-quality images and annotations are essential for supervised training, validation, and testing of AI algorithms, especially with external data set[8]. This is problematic for rare diseases like paediatric tumours of the central nervous system which shows an important imaging pattern variability, making it even more difficult to create sufficient databases to develop an AI algorithm. In addition, motion artifacts are frequent in the pediatric population, resulting in a higher exclusion rate for poor image quality

than in adults. Segmentation also presents a challenge in infants and newborns due to volume changes and myelination processes related to brain development[9].

Recently, easy to use automatic tumour segmentation and brain extraction algorithms, “HD-GLIOMA”[10] and “HD-BET”[11] ([github.com/neuroAI-HD](https://github.com/neuroAI-HD)) have been developed and validated. HD-GLIOMA allows a standardized follow-up of adult brain tumours and a better evaluation of the therapeutic response than the response assessment in neuro-oncology (RANO) criteria[12]. However, this algorithm was trained on an adult database and only on glial tumours. Given the problems for AI development in children, simply transferring an algorithm developed in adults seems conceptually challenging. However, given the lack of large data sets, even a use with some restrictions on performance could be an opportunity for rare diseases.

Here, we aimed exploring the value of the two segmentation algorithms on a heterogeneous paediatric tumour population. We hypothesized that segmentation performance might be less precise but sufficient for proofing feasibility of this alternative approach.

## **Methods**

### *Participants*

We retrospectively screened our clinical database of paediatric brain tumours from May 2015 to January 2020 that had a complete MRI pre-treatment examination with T1, T2, FLAIR and T1 post contrast (T1PC) sequences. The study was approved by the local review board (RGPD/AP-HM). The histological diagnosis was known for each case. Tumours were listed according to the 2016 World Health Organization (WHO) classification of central nervous system tumours[13].

### *Image acquisition*

MRI were performed on a 1.5T system (Aera, Siemens, Erlangen, Germany) using a 20-channel head coil and on a 3T system (Skyra, Siemens, Erlangen, Germany) using a 32-channel head coil. The MRI included standard sequences for brain tumours: 2D sagittal T1-weighted, axial T2-weighted, 2D coronal or 3D Fluid-Attenuated Inversion Recovery (FLAIR), axial gradient echo T2, diffusion weighted imaging and Gadolinium-enhanced three-dimensional MPRAGE (Magnetization Prepared - Rapid Gradient Echo) T1.

### *Image processing*

Images were processed with the functional imaging software library (FSL, version 5.0, [www.fmrib.ox.ac.uk](http://www.fmrib.ox.ac.uk)). Images were reoriented to standard MNI space and T1, T2 and FLAIR images were registered to the T1PC image. First, we used HD-BET to perform brain extraction (BET) on all four sequences. In addition, we computed brain masks from FSL BET[14] (all sequences) and ROBEX [15] (only on T1) for comparison. In a second step, we performed tumour segmentation with HD-GLIOMA which is based on the four classic anatomical images T1, T2, FLAIR and T1PC. We compared the result of the automated segmentation with manual expert masks for the brain and correspondingly for contrast-enhancing tumours (CE) and non-enhancing T2-signal abnormalities (NE) parts. All expert masks were generated on FSL by a radiology resident with 5 years of experience and checked by a neuroradiologist with 10 years of experience. Moreover, a post-AI quality check was performed to determine the efficiency of the HD-GLIOMA segmentation using a rating scale (good, moderate, poor, and completely wrong) and to check for lesions not seen by radiologists. In addition to the validation of the tumour masks and to explore the feasibility of this approach for automated feature extraction and classification, we extracted in a final step mean diffusivity (MD) values from the different expert and HD-GLIOMA segmentations. To follow this aim, we used non-linear registration between T1PC and the B0 image to transform all masks into the diffusion space.

### *Statistics*

Besides descriptive statistics of the cohort, we computed dice coefficients between the different masks and the manual reference masks. The dice similarity coefficient is a standard measure to report the performance of a segmentation and measures the extent of spatial overlap between two binary segmentation masks[16]. The dice similarity coefficient is defined as twice the size of the intersection between the two masks normalised by the sum of their volumes. The dice coefficient can range between 0 (no overlap) and 1 (perfect agreement). We used multivariate linear models for comparison of dice coefficients and the role of covariates: age (0-2 vs >2 years as an indicator of MRI signal changes related to myelinisation), localisation (supra vs infratentorial), grading (high vs low) and field strength (1.5 vs 3T). The correlation between MD values was explored with Pearson's correlation coefficient  $r$ . To explore the accuracy of extracted MD values for tumour classification (low- vs high-grade) we applied ROC analyses and computed AUC values. ROC curves for HD-GLIOMA and expert masks were compared with bootstrapping. All analyses were performed with Statistics in R v4.0.3. P-values below 0.05 were considered as statistically significant.

## **Results**

### *Patients*

We included 42 paediatric patients with a mean age of 6.8 years  $\pm 5$  [standard deviation] (range 7 months to 17 years) and 26 males (60%). Demographic data is summarized in table 1. Most tumours were high grade tumours (e.g., glioblastoma and medulloblastoma) and they were more often located infratentorially. We defined the following sub-groups based on histology and biological appearance: paediatric-type glial/glioneuronal tumours (n=12), embryonal tumours (n=13), ependymal tumours (n=5), diffuse astrocytic and oligodendroglial tumours (n=13). Imaging patterns were highly variable including completely enhanced or completely non-enhanced tumors and diffuse to well delineated boundaries.



**Table 1: Patient Characteristics**

<b>Total number of patients</b>	N = 42
<b>Age</b> years mean (sd)	6.8 (4.9)
<b>Age groups</b> n (%)	
<b>0 - 2 years</b>	9 (21.4%)
<b>&gt;2 years</b>	33 (78.6%)
<b>Sex</b> n (%)	
Female	17 (40.4%)
Male	25 (59.5%)
<b>Localization</b> n (%)	
Infratentorial	25 (59.5%)
Supratentorial	17 (40.4%)
<b>Histology</b>	
<b>Diffuse astrocytic and oligodendroglial tumours</b>	<b>12 (28.6%)</b>
Unclassable malignant glioma	2
Glioblastoma	3
Pleiomorphic xanthoastrocytoma	1
Diffuse midline glioma	6
<b>Embryonal tumours</b>	<b>13 (31.0%)</b>
Pinealoblastoma	1
Medulloblastoma	8
Atypical teratoid rhabdoide tumor	2
Germinoma	1
Immature teratoma	1
<b>Ependymal tumours</b>	<b>5 (11.9%)</b>
Anaplastic ependymoma	5
<b>Paediatric-type glial / glioneuronal tumours</b>	<b>12 (28.6%)</b>
Pilocytic astrocytoma	8
Glioneuronal Tumour	2
DNET	1
Pilomyxoïd astrocytoma	1
<b>WHO Grade</b> n (%)	
I	10 (23.8 %)
II	1 (2.4 %)
III	6 (14.3 %)
IV	21 (50.0%)
unclassified	4 (9.5%)
<b>Field strength MRI</b> n (%)	
1.5 Tesla	21 (50.0%)
3.0 Tesla	21 (50.0%)

### *Brain extraction*

Table 2 and Figure 1 summarise the comparison of different BET algorithms with the expert masks. Over all sequences, HDBET performed best with a mean dice coefficient of 0.99 (range 0.85 – 1), followed by ROBEX (only T1, mean 0.96; range 0.91 - 0.98) and FSLBET (mean 0.86; range 0.45 – 0.98). The multivariate model revealed that for all BET algorithms performance was better for FLAIR images ( $\beta = 0.05$ ,  $p < 0.001$ ) if compared to the BET segmentations on T1PC. Performance was similar for T2 images ( $\beta = 0.01$ ,  $p = 0.623$ ) and worse on the precontrast 2D T1 ( $\beta = -0.04$ ;  $p < 0.001$ ). Tumour location ( $p = 0.385$ ), WHO grading ( $p = 0.165$ ), and MRI field strength ( $p = 0.793$ ) were not associated with the dice coefficient. Moreover, brain segmentation quality was independent from the age of the patient (age 0-2 vs >2 years:  $\beta = -0.02$ ;  $p=0.153$ ).

**Table 2: Comparison of brain extraction tools with expert masks**

Sequence	Method	Dice				
		mean	sd	median	min	max
T1	FSLBET	0.78	0.13	0.75	0.54	0.96
	HDBET	0.99	0.01	0.99	0.96	1
	ROBEX	0.96	0.02	0.97	0.91	0.98
T2	FSLBET	0.86	0.09	0.89	0.63	0.95
	HDBET	0.98	0.03	0.99	0.85	1
FLAIR	FSLBET	0.94	0.05	0.95	0.77	0.98
	HDBET	0.99	0.01	1	0.96	1
T1PC	FSLBET	0.85	0.14	0.90	0.44	0.98
	HDBET	0.99	0.01	1	0.98	10

### *Tumour segmentation*

Results for the tumour segmentation with HD-GLIOMA are summarised in Figure 2. We observed a moderate agreement between automated and expert masks, such as the mean

dice coefficient was 0.67 (median 0.78, range 0 - 1) for enhancing tumour parts (Figure 2A). HD-GLIOMA segmentation of T2 signal abnormalities, i.e., non-enhancing tissular components was less accurate (Figure 2B) compared to the enhancing tumour parts (beta = -0.26; p < 0.001). HD-GLIOMA segmented the T2 abnormalities with a mean dice coefficient of 0.42 (median 0.45, range 0 - 1). Moreover, supratentorial localisation reduced the dice coefficient (beta = -0.29; p < 0.001). In line with better abilities to detect enhancing tumour parts, the algorithm tended to perform better in grade III and IV tumours than in low grade tumours (beta = 0.32; p = 0.081). Moreover, the algorithm performed worse for embryonal (beta = -0.29; p = 0.006) and ependymal tumours (beta = -0.33; p = 0.012) if compared to the other two histology groups. Magnetic field strength seems to have no major influence on the accuracy of the algorithm (p = 0.176). Importantly, we observed no difference between the two age groups (0-2 vs >2 years; p = 0.472).

Visual inspection (Table 3) of the CE segmentation revealed a good quality in 30 patients (71.4%) while the algorithm completely failed in 4 patients (9.5 %). Results for non-enhancing T2 abnormalities were clearly worse with only 11 (26.2%) good segmentations, and completely wrong masks in 15 patients (35.7%). Figure 3 shows some examples of segmentation errors. The algorithm revealed no tumor parts not detected by the radiologist.

**Table 3: Quality of HD-GLIOMA Segmentation**

	<b>good</b>	<b>moderate</b>	<b>poor</b>	<b>completely wrong</b>
<b>CE</b>	30 (66.7%)	6 (14.3%)	2 (4.8%)	4 (9.3%)
<b>T2 abnormalities</b>	11 (26.2%)	8 (19.0%)	8 (19.0%)	15 (35.7%)

### *Feasibility*

Finally, we were interested to proof the feasibility of our approach by applying the automated pipeline to classify tumour grading (low vs high) based on MD. We extracted MD values from expert and HD-GLIOMA segmentations. While the correlation of values was very good for

CE segmentations ( $r = 0.79$ ,  $p < 0.001$ , Figure 4A), we observed no correlation for non-enhancing tumour parts ( $r = 0.07$ ,  $p = 0.68$ , Figure 4B). However, the ability to classify low- and high-grade tumours based on MD values from automated CE masks was good (AUC = 0.863) but seems to be less accurate than a classification based on expert masks (AUC = 1,  $p = 0.208$ ) (Figure 4C). The best of cut-off value for grading was a MD value of  $1.4 \cdot 10^{-3}$   $\text{mm}^2/\text{s}$  for both masks. For the classification of tumours based on non-enhancing tumour masks, extracted values from HD-GLIOMA segmentation (AUC = 0.773, cut-off =  $1.2 \cdot 10^{-3}$   $\text{mm}^2/\text{s}$ ) performed not significantly different than the expert masks (AUC = 0.644, cut-off =  $1.3 \cdot 10^{-3}$   $\text{mm}^2/\text{s}$ ,  $p = 0.47$ , Figure 4D).

## Discussion

Borrowing strength from large datasets is an interesting but controversial approach to benefit from the advantages of AI for rare diseases in clinical practice. Here, we challenged an automatic MRI brain extraction algorithm (HD-BET) and an automated tumour segmentation (HD-GLIOMA) trained on an adult population in a heterogeneous real-life cohort of paediatric brain tumours. While we observed an excellent performance for brain extraction, tumour segmentation showed less impressive results with lower performance mainly for non-enhancing parts of the tumours. However, the feasibility of the segmentation and its putative clinical value could be illustrated by distinguishing between low- and high-grade tumours based on the diffusion sequence. This approach seems interesting to provide future support to radiologists in clinical practice but without optimization, algorithms trained for brain tumor detection on an adult population are not easily transferable to pediatric patients.

Brain extraction with HD-BET was fast, reliable and robust for different weighted images in 2D and 3D planes and in the presence of pathology and in line with previous studies[11,17]. Tumour segmentation with HD-GLIOMA showed lower performance in our paediatric cohort than in the original adult cohort [10] with consistent results, both on a visual rating scale and the dice coefficients. The results are not unexpected and conceptually bound to the challenging design of our study: For example, a hydrocephalus leading to periventricular hyperintensities on FLAIR and T2 weighted images mislead the algorithm for non-enhancing tumour parts. Hydrocephalus is more frequent in paediatric brain tumours than in adults[18] because of the preferential location of lesions in the posterior fossa. In addition, we observed a lack of distinction between peri-lesional oedema and the NE of the tumour in some cases. Glial peri-lesional oedema cannot be differentiated from tumour infiltration based on morphological imaging data alone[19] While this differentiation is less important in high grade glioma, it becomes more relevant in other histology[20] which were common in our cohort. In line with the original training set, we observed the best performance of the algorithm in the

segmentation of high-grade tumours indicating probably an easy use of the algorithm for surveillance of high-grade tumours in children. We observed more limitations for NE segmentation, but results were still sufficiently precise to perform a tumour grading based on the diffusion data, which is a rather old but reliable technique that outstands other approaches like perfusion imaging in the grading of paediatric brain tumour[21]. Interestingly, we did not observe a relevant impact of age on segmentation results. On a conceptual level, one might consider that algorithms trained in adult populations perform worse in young children below the age of two as MRI signal is related to myelinisation and thus very different in comparison to adults. Here, nearly one out of four patients were in this age group, but performance of the algorithms was not worse.

Our study has several limitations, notably the small sample size which contrasts the high heterogeneity of the cohort. However, the sample size was large enough for indicating feasibility of the transfer concept. Moreover, the sample size allowed excluding large effects of conceptually important covariates including age related myelinisation and field strength. Moreover, the data set included only baseline MRI where a better and more precise tumour diagnosis might improve the counselling of patients and relatives. However, at diagnosis tumour histology will always be the most important information for treatment decisions and prognosis. Thus, even an optimized and high performing AI supported MRI diagnosis has probably minor impact. The importance of MRI is much higher in the post-surgical follow-up, where early detection of tumour recurrence and exclusion of pseudo-progression or radionecrosis could be improved by specialized algorithms. Based on our results we believe that evaluating longitudinal data as a next step seems reasonable. Lastly, in the absence of a universally accepted standard, we have used DICE coefficient to evaluate the algorithms but other parameters may also be relevant[22].

Two approaches are available to improve the precision of algorithms like HD-GLIOMA for clinical use. A natural way to benefit from the existing algorithm would be a re-training of the

data including adult and paediatric cases, which might increase precision without affecting the network architecture[23].

In contrast, developing new algorithms in sufficiently large paediatric cohorts might be an opportunity but remains challenging with low prevalence rates. Very recently, Pen and al [24] proposed a convolutional neural network trained segmentation of brain tumours in a paediatric population to calculate tumour volumes with promising results (CE : dice 0.72 ; NE: dice : 0.72). However, their training population was more restrictive, consisting of posterior fossa tumours as medulloblastomas and high-grade gliomas, unlike our cohort, which is more representative of daily practice. Thus, even if the algorithm was trained in a more specific populations, expected limitations in clinical practice might be very similar if compared to our approach. In addition, optimized tumour segmentations are probably only an intermediate step. The goal of automated tumour surveillance and grading will be the integration with clinical decisions based on the evidence that such automated pipelines improve surveillance and, to that end, relevant outcomes such as disability or survival. Here, an extension towards automated radiomic feature extraction is an interesting opportunity[25]. Overall, our study design made use of a heterogenous but representative cohort of paediatric central nervous system tumours[1]. Our approach seems to be a cost-effective method to easily explore the feasibility of existing algorithms which requires fewer resources than developing new algorithms.

In conclusion, brain extraction with HD-BET looks promising to paediatric populations (even in infants) while tumour segmentation with HD-GLIOMA needs further developed to increase precision. Our results indicate that, even without fine-tuning or out-of-distribution detection, the use of AI algorithms on rare diseases, e.g., trained on adults and used on pediatric patients, may yield acceptable results depending on the clinical scenario.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Data statement

The datasets analysed during the current study are available from the corresponding author on reasonable request.

## References

- 1 Bauchet L, Rigau V, Mathieu-Daudé H, *et al.* Clinical epidemiology for childhood primary central nervous system tumors. *J Neurooncol* 2009;**92**:87–98. doi:10.1007/s11060-008-9740-0
- 2 Albright AL, Wisoff JH, Zeltzer PM, *et al.* Effects of medulloblastoma resections on outcome in children: a report from the Children’s Cancer Group. *Neurosurgery* 1996;**38**:265–71. doi:10.1097/00006123-199602000-00007
- 3 Erker C, Tamrazi B, Poussaint TY, *et al.* Response assessment in paediatric high-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. *The Lancet Oncology* 2020;**21**:e317–29. doi:10.1016/S1470-2045(20)30173-X
- 4 Huang R. Response assessment in high-grade glioma: tumor volume as endpoint. *Neuro Oncol* 2017;**19**:744–5. doi:10.1093/neuonc/nox035
- 5 Chang K, Beers AL, Bai HX, *et al.* Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement. *Neuro-Oncology* 2019;**21**:1412–22. doi:10.1093/neuonc/noz106
- 6 Quon JL, Bala W, Chen LC, *et al.* Deep Learning for Pediatric Posterior Fossa Tumor Detection and Classification: A Multi-Institutional Study. *AJNR Am J Neuroradiol* 2020;:ajnr;ajnr.A6704v1. doi:10.3174/ajnr.A6704



- 7 Havaei M, Davy A, Warde-Farley D, *et al.* Brain tumor segmentation with Deep Neural Networks. *Med Image Anal* 2017;**35**:18–31. doi:10.1016/j.media.2016.05.004
- 8 Willemink MJ, Koszek WA, Hardell C, *et al.* Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020;**295**:4–15. doi:10.1148/radiol.2020192224
- 9 Phan TV, Smeets D, Talcott JB, *et al.* Processing of structural neuroimaging data in young children: Bridging the gap between current practice and state-of-the-art methods. *Developmental Cognitive Neuroscience* 2018;**33**:206–23. doi:10.1016/j.dcn.2017.08.009
- 10 Kickingereeder P, Isensee F, Tursunova I, *et al.* Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *The Lancet Oncology* 2019;**20**:728–40. doi:10.1016/S1470-2045(19)30098-1
- 11 Isensee F, Schell M, Pflueger I, *et al.* Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping* 2019;**40**:4952–64. doi:<https://doi.org/10.1002/hbm.24750>
- 12 Yang D. Standardized MRI assessment of high-grade glioma response: a review of the essential elements and pitfalls of the RANO criteria. *Neuro-Oncology Practice* 2016;**3**:59–67. doi:10.1093/nop/npv023
- 13 Louis DN, Perry A, Reifenberger G, *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 2016;**131**:803–20. doi:10.1007/s00401-016-1545-1
- 14 Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002;**17**:143–55. doi:10.1002/hbm.10062
- 15 Iglesias JE, Liu C-Y, Thompson PM, *et al.* Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Trans Med Imaging* 2011;**30**:1617–34. doi:10.1109/TMI.2011.2138152
- 16 Menze BH, Jakab A, Bauer S, *et al.* The Multimodal Brain Tumor Image

Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging* 2015;**34**:1993–2024.

doi:10.1109/TMI.2014.2377694

17 Zhao X, Zhao X-M. Deep learning of brain magnetic resonance images: A brief review. *Methods* 2021;**192**:131–40. doi:10.1016/j.ymeth.2020.09.007

18 Wong T-T, Liang M-L, Chen H-H, *et al.* Hydrocephalus with brain tumors in children. *Childs Nerv Syst* 2011;**27**:1723–34. doi:10.1007/s00381-011-1523-9

19 Eidel O, Burth S, Neumann J-O, *et al.* Tumor Infiltration in Enhancing and Non-Enhancing Parts of Glioblastoma: A Correlation with Histopathology. *PLoS ONE* 2017;**12**:e0169292. doi:10.1371/journal.pone.0169292

20 Warren KE, Vezina G, Poussaint TY, *et al.* Response assessment in medulloblastoma and leptomeningeal seeding tumors: recommendations from the Response Assessment in Pediatric Neuro-Oncology committee. *Neuro Oncol* 2018;**20**:13–23.

doi:10.1093/neuonc/nox087

21 Testud B, Brun G, Varoquaux A, *et al.* Perfusion-weighted techniques in MRI grading of pediatric cerebral tumors: efficiency of dynamic susceptibility contrast and arterial spin labeling. *Neuroradiology* Published Online First: 27 January 2021. doi:10.1007/s00234-021-02640-y

22 Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 2015;**15**. doi:10.1186/s12880-015-0068-x

23 Ding Y, Acosta R, Enguix V, *et al.* Using Deep Convolutional Neural Networks for Neonatal Brain Image Segmentation. *Frontiers in Neuroscience* 2020;**14**:207.

doi:10.3389/fnins.2020.00207

24 Peng J, Kim DD, Patel JB, *et al.* Deep Learning-Based Automatic Tumor Burden Assessment of Pediatric High-Grade Gliomas, Medulloblastomas, and Other Leptomeningeal Seeding Tumors. *Neuro-Oncology* Published Online First: 26 June 2021.

doi:10.1093/neuonc/noab151

25 Pei L, Vidyaratne L, Rahman MM, *et al.* Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images. *Sci Rep* 2020;**10**:19726. doi:10.1038/s41598-020-74419-9

## FIGURE LEGENDS

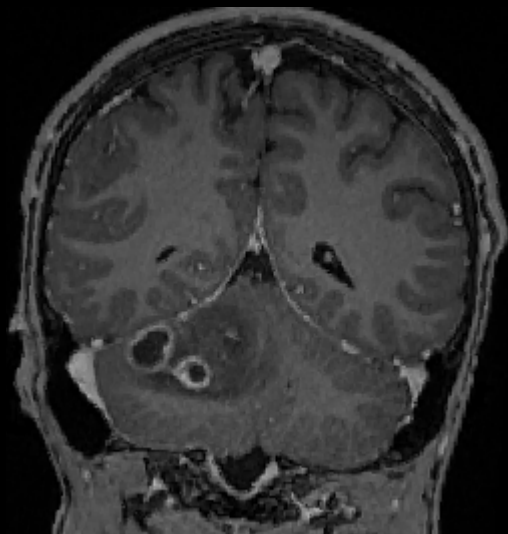
**Figure 1:** Histograms and boxplots of Dice coefficients of the three BET algorithms compared to an expert reference mask showing an excellent performance of HDBET. ROBEX was only applied on T1 images.

**Figure 2:** Histograms and boxplots of Dice coefficients for the two segmentations from HDGLIOMA compared to expert reference masks show moderate performance for CE (A) and poor performance for T2-abnormalities (B). CE = contrast-enhancing tumour part

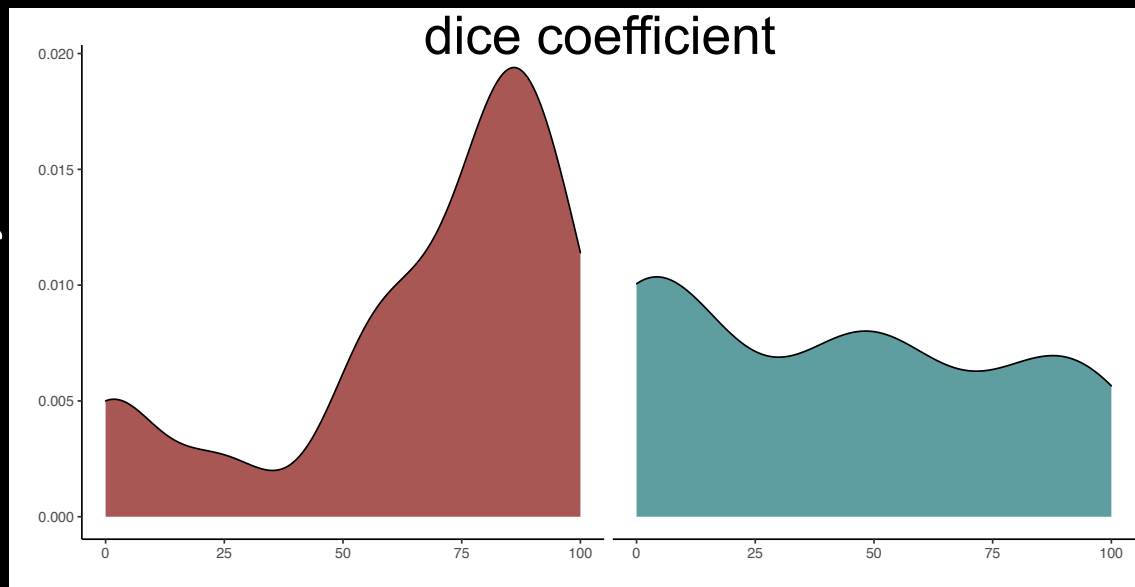
**Figure 3:** Examples of tumour segmentation with HDGLIOMA. (A) Segmentation of good quality of posterior fossa tumour (glioblastoma) displayed on 3D T1 post contrast (T1PC); (B) Segmentation error with misclassification of transependymal resorption besides peritumoral oedema (medulloblastoma) illustrated on coronal FLAIR sequence; (C) False segmentation of peri-lesional oedema (atypical teratoid/rhabdoid tumour) displayed on T1PC; (D) Segmentation error due to dental metal artifact (glioblastoma) displayed on T1PC. Red = Contrast enhancing; Blue= non-enhancing signal abnormalities on T2

**Figure 4:** Dot plots (A and B) with linear regression estimate illustrate the correlation between mean diffusivity values extracted from expert segmentations and HDGLIOMA masks, showing a very good correlation CE and no correlation with T2-abnormalities. Boxplots (C and D) compare mean diffusivity values extracted from the HDGLIOMA segmentations in low- and high-grade tumours showing a good ability for automated CE mask to classify low- and high-grade. CE = contrast-enhancing tumour part

# CT1 - Glioblastoma



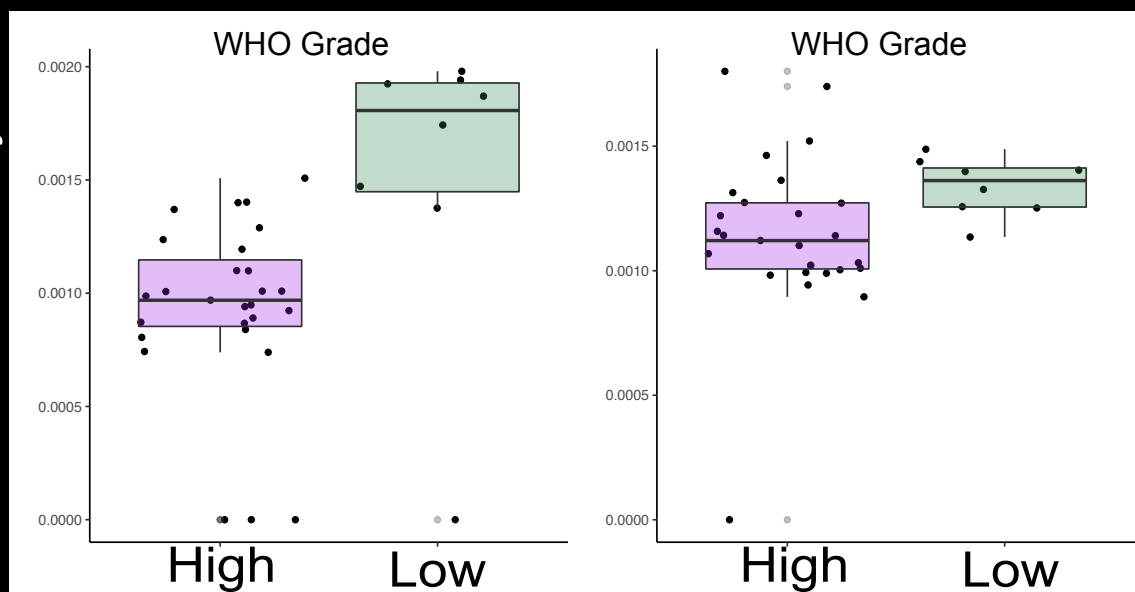
density



## CE tumor parts

## T2 abnormalities

Mean diffusivity



## AI segmentation

