AJNR

Automated Diffusion Analysis for Non-Invasive Prediction of IDH Genotype in WHO Grade 2-3 Gliomas

Jiaming Wu, Stefanie C. Thust, Stephen J. Wastling, Gehad Abdalla, Massimo Benenati, John A. Maynard, Sebastian Brandner, Ferran Prados Carrasco and Frederik Barkhof

AJNR Am J Neuroradiol published online 10 June 2025
http://www.ajnr.org/content/early/2025/06/10/ajnr.A8776

Published June 10, 2025 as 10.3174/ajnr.A8776

ORIGINAL RESEARCH

Automated Diffusion Analysis for Non-Invasive Prediction of IDH Genotype in WHO Grade 2-3 Gliomas

Jiaming Wu, Stefanie C. Thust, Stephen J. Wastling, Gehad Abdalla, Massimo Benenati, John A. Maynard, Sebastian Brandner, Ferran Prados Carrasco, Frederik Barkhof

ABSTRACT

BACKGROUND AND PURPOSE: Glioma molecular characterization is essential for risk stratification and treatment planning. Non-invasive imaging biomarkers such as apparent diffusion coefficient (ADC) values have shown potential for predicting glioma genotypes. However, manual segmentation of gliomas is time-consuming and operator-dependent. To address this limitation, we aimed to establish a single-sequence-derived automatic ADC extraction pipeline using T2-weighted imaging to support glioma isocitrate dehydrogenase (IDH) genotyping.

MATERIALS AND METHODS: Glioma volumes from a hospital data set (University College London Hospitals; n=247) were manually segmented on T2-weighted MRI scans using ITK-Snap Toolbox and co-registered to ADC maps sequences using the FMRIB Linear Image Registration Tool in FSL, followed by ADC histogram extraction (Python). Separately, a nnUNet deep learning algorithm was trained to segment glioma volumes using T2w only from BraTS 2021 data (n=500, 80% training, 5% validation and 15% test split). nnUnet was then applied to the University College London Hospitals (UCLH) data for segmentation and ADC read-outs. Univariable logistic regression was used to test the performance manual and nnUNet derived ADC metrics for IDH status prediction. Statistical equivalence was tested (paired two-sided t-test).

RESULTS: nnUnet segmentation achieved a median Dice of 0.85 on BraTS data, and 0.83 on UCLH data. For the best performing metric (rADCmean) the area under the receiver operating characteristic curve (AUC) for differentiating IDH-mutant from IDH-wildtype gliomas was 0.82 (95% CI: 0.78-0.88), compared to the manual segmentation AUC 0.84 (95% CI: 0.77-0.89). For all ADC metrics, manually and nnUNet extracted ADC were statistically equivalent ($p<0.01$). nnUNet identified one area of glioma infiltration missed by human observers. In 0.8% gliomas, nnUnet missed glioma components. In 6% of cases, over-segmentation of brain remote from the tumor occurred (e.g. temporal poles).

CONCLUSIONS: The T2w trained nnUnet algorithm achieved ADC readouts for IDH genotyping with a performance statistically equivalent to human observers. This approach could support rapid ADC based identification of glioblastoma at an early disease stage, even with limited input data.

ABBREVIATIONS: AUC = Area under the receiver operating characteristic curve, BraTS = The brain tumor segmentation challenge held by MICCAI, Dice = Dice Similarity Coefficient, IDH = Isocitrate dehydrogenase, mGBM = Molecular glioblastoma , ADCmin = Fifth ADC histogram percentile, ADCmean = Mean ADC value, ADCNAWM = ADC in the contralateral centrum semiovale normal white matter, rADCmin = Normalized ADCmin , VOI rADCmean = Normalized ADCmean.

Received December, 2024; accepted after revision March 19, 2025.
From the Centre for Medical Image Computing (CMIC), Department of Medical Physics and Biomedical Engineering, University College London, London, UK (JW, FPC, FB), Neuroradiological Academic Unit, UCL Queen Square Institute of Neurology, University College London, London, UK (JW, ST, SW, JM, FB), Radiological Sciences, Mental Health and Neurosciences, School of Medicine, University of Nottingham, Nottingham, UK (ST), Sir Peter Mansfield Imaging Centre, School of Physics and Astronomy, University of Nottingham, Nottingham, UK (ST), NIHR Nottingham Biomedical Research Center, Nottingham, UK (ST), Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, London, UK (SW, GA, MB, JM, FB), Department of Neuroradiology, King's College Hospital NHS Foundation Trust, London, United Kingdom (GA), Department of Radiology, Mansoura University hospitals, Mansoura Faculty of Medicine, Egypt (GA), Department of Diagnostic Imaging and Radiotherapy, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy (MB), Department of Radiology, Guy's and St Thomas' NHS Foundation Trust, London, UK (JM), Department of Neurodegenerative Disease, UCL Institute of Neurology and Division of Neuropathology, National Hospital for Neurology and Neurosurgery, University College London NHS Foundation Trust, London (SB), Department of Neuroinflammation, Faculty of Brain Sciences, Queen Square MS Center, UCL Institute of Neurology, University College London, London, UK (FPC), e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain (FPC), Radiology & Nuclear Medicine, VU University Medical Center, Amsterdam, the Netherlands (FB).

Disclosure of potential conflicts of interest: ST discloses funding from the Nottingham NIHR Biomedical Research Centre and speaker fees from the British Society of Neuroradiologists. FPC is partially funded by University College London Hospital Biomedical Research Council. JW, SW, GA, MB, JM, SB and FB have nothing to declare.

Corresponding author: Dr. Stefanie Thust

Precision Imaging Beacon Hub

Room A39e, Medical School

Queens Medical Centre

Nottingham NG7 2RD, United Kingdom

Email: stefanie.thust@nottingham.ac.uk

Copyright 2025 by American Society of Neuroradiology.

INTRODUCTION

Diffuse gliomas are brain tumors with a variable prognosis that is determined by their genetic profile. Amongst the adult-type diffuse gliomas, isocitrate dehydrogenase-mutant (IDH^{mut}) tumors (astrocytomas and oligodendrogliomas) are associated with longer survival compared to IDH-wildtype (IDH^{wt}) (glioblastoma, IDH-wildtype). [1]. On microscopy, IDH^{mut} gliomas often resemble low to intermediate (WHO grade 2-3) tumors. In the 2021 World Health Organization (WHO) Classification of CNS Tumors, radiologically and histologically low grade appearing IDH^{wt} gliomas are classed as molecular glioblastoma (henceforth referred to as mGBM, CNS WHO grade 4) when exhibiting its typical genetic signature [2], i.e chromosome 7 gain, chromosome 10 loss, frequent TERT promoter mutation and EGFR amplification. Glioblastoma, IDH-wildtype is the most common primary malignant brain tumor in adults with a mean survival of 12-15 months, whereby mGBM is suspected to represent an early disease stage [3].

The standard of care for glioma involves maximum safe resection, which provides tissue for molecular testing[4]. Securing the molecular diagnosis can encounter delays and be subject to geographical inequalities in access to specialized pathology services [5]. Imaging biomarkers, such as apparent diffusion coefficient (ADC) values derived from diffusion-weighted imaging (DWI), offer a non-invasive prediction of IDH mutation status preceding surgical planning, also in cases where a resection is not immediately feasible [6, 7]. Previous studies highlighted a role for ADC in the identification of WHO grade 2/3 glioma IDH status compared to other metrics [8–10]. Specifically, ADC values tend to be lower in IDH^{wt} tumors [8, 11]**.** We previously demonstrated that ADC values extracted from manually selected regions within the tumor can identify glioma molecular signatures of prognostic importance [12, 13]. However, quantitative volumetric analysis is too time-consuming for easy adoption, and radiologist measurements depend on operator experience. An automated extraction procedure of ADC parameters could facilitate clinical translation.

For diffuse gliomas, T2-weighted (T2w) imaging represents an standard of care in the diagnostic evaluation [14]. Low-grade gliomas may not exhibit significant contrast enhancement on post-contrast T1-weighted imaging [15]. More generally, non-contrast scans may be preferred for various reasons, including contraindications or risk factors precluding the administration of GBCAs [16, 17].

Deep learning models have revolutionized medical image segmentation, advancing from traditional convolutional neural networks (CNNs) to more sophisticated architectures like nnUNet [18, 19]. These models have been extensively applied to glioma segmentation using multimodal MRI data (e.g., T1, T2/FLAIR, and T1 post-Gadolinium sequences) [20]. However, missing sequences, for example due to contraindications for gadolinium-based contrast agents (GBCAs), pose technical barriers to AI-based segmentation [21, 22]. Sparse network approaches, such as those proposed by Roelant et al., attempt to resolve this issue but nevertheless rely on post contrast sequences[23].

This investigation focused on developing a deep learning model trained exclusively on non-contrast data for rapid tumor volume delineation and diffusion measurement. We identified nnUNet as an effective method for deep learning-based segmentation [22]and re-trained this algorithm using T2w single-sequence data.

Our hypothesis was that automated extraction of ADC parameters from T2w-based segmentations could reliably predict IDH genotype in WHO grade 2-3 gliomas. The objective of this study was to establish an automated pipeline for ADC extraction in support of glioma genotyping whilst demonstrating the feasibility of non-contrast-based glioma genotyping.

MATERIALS AND METHODS

University College London Hospitals (UCLH) dataset

This retrospective study represents a computational analysis of a published dataset of manually segmented gliomas (UCLH and Health Research Authority UK) [13]. The Checklist for Artifical Intelligence for Medical Imaging (CLAIM) used in this study is provided in **Supplementary Material 1**. Patients consecutively diagnosed at our specialist brain tumor referral center between July 2008 and January 2018 were included. Inclusion criteria encompassed histologically confirmed World Health Organization (WHO) grade 2 and 3 gliomas, documented IDH and 1p19q genetic testing results, and availability of pre-treatment MRI scans. Exclusion criteria were prior treatment for glioma; diagnoses other than WHO grade 2 and 3 gliomas; incomplete, uncertain or ambiguous molecular results (e.g., IDH wild-type/1p19q co-deleted); prolonged interval between MR imaging and surgery (≥1 year); incomplete imaging data (missing or severely degraded sequences), failure of image co-registration (ADC to T2) [13] or input failure for the automated segmentation.

Reference standard

All tissue samples underwent analysis at our neuropathology department according to the 2016 Classification of Tumors of the Central Nervous System, which corresponds to the timepoint of data enrolment [20] as described previously [11]. Multi-gene Sanger sequencing was performed for IDH R132H-negative tumors to identify rarer IDH mutations, and 1p/19q status was determined via quantitative polymerase chain reaction-based copy number assays.

MR Imaging Acquisition and Postprocessing

All MR imaging examinations included T2-weighted and DWI sequences (1.5-3T) as part of preoperative multisequence imaging. Because our institution is a quaternary center, the imaging originated from 23 different MRI machines with no individual scanner contributing >14% of any glioma subtype. ADC maps were calculated from 3-directional DWI acquired with 2 gradient values (b = 0 and b = 1000 s/mm^2) using proprietary software (Olea Sphere, Version 2.3; Olea Medical). Sufficient comparability of ADC among scanners was

assumed based on previous technical studies, and furthermore normalized ADC values were assessed as part of the previous and current study to mitigate technical differences [25, 26].

Manual volumetric ADC map analysis

Whole-tumor VOIs were segmented by a general radiologist (MB, 5 years' experience) using the ITK-Snap Toolbox, Version 3.6 (www.itksnap.org)[20] following training and under supervision of a neuroradiologist specialized in brain tumor imaging (SCT, 9 years' experience). Segmentations incorporated the entire T2-weighted signal abnormality. For multicentric gliomas, the total volume of signal abnormality was treated as one lesion. ADC maps were then co-registered to T2-weighted sequences using the FMRIB Linear Image Registration Tool (FLIRT; http://www.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT)[27]. Subsequently, ADC histogram data were obtained for each tumor ROI, using an in-house script written in Python 2.7. For each tumor, the fifth ADC histogram percentile was designated ADC_{min}. Additionally, VOI ADCmean, and the normalized values (divided by ADC in the contralateral centrum semiovale normal white matter, ADCNAWM) rADCmin and rADCmean were extracted.

BraTS Dataset

To form a training dataset for nnUNet, we randomly chose 500 glioma cases with a random mixture of WHO grades 2-4 from the BraTS 2021 challenge [28] using T2w sequences only. BraTS (The Brain Tumor Segmentation Challenge) is an open-source dataset from Medical Image Computing and Computer Assisted Intervention (MICCAI) widely used for benchmarking in segmentation of brain tumors in MR scans. The dataset includes pre-operative MRI scans that have been manually segmented by experts into different tumor regions, consisting of the enhancing tumor, non-enhancing tumor (including oedema/infiltration), and necrotic core. These cases were divided into 80% training, 5% validation and 15% test data, respectively **(Table 1)**. As a change from the BraTS challenges, only the whole tumor (WT) class contributed to the T2w segmentation. Subsequently, the nnUnet algorithm was applied to the UCLH data for T2w segmentation (without further training), see **Figure 1**.

Table 1. Breakdown of training, validation and test data for the study.

	BraTS	*UCLH*	*Total*
BraTS Training	400 (80%)	0	400
BraTS Validation	25 (5%)	0	25
BraTS Test	75 (15%)	0	75
External test set	0	247	247
Total	500	247	747

Results are listed as absolute number (percentage of cohort). UCLH=University College London Hospitals

Pre-processing

Using the Advanced Normalization Tools (ANTs) method [23], ADC images were rigidly registered to the corresponding T2w images. Subsequently, all scans were registered to the SRI24 atlas [24] (https://www.nitrc.org/projects/sri24/) and resampled to 1mm isotropic voxels. N4 bias field correction was then applied, followed by skull stripping with the "HD-bet" algorithm [25] (https://github.com/MIC-DKFZ/HD-BET).

Deep Learning Segmentation Models

We selected the nnUnet architecture as the network [32]. This method combines 2D, 3D, and cascaded U-nets for automated segmentation [27]. With this strategy, three U-net models are trained simultaneously, whereby the best-trained model is automatically selected. This architecture was chosen due to high performance in brain tumor segmentation challenges [28]. The cascaded 3D-2D-3D U-Net design of this architecture allows it to fully leverage 3D contextual information and 2D high-resolution details, thereby enabling precise segmentation.

Model training and testing

Models were trained to perform a binary class segmentation of WT and background (unaffected brain parenchyma) for each tumor. After randomly splitting the BraTS cases (80% training, 5% validation and 15% test), the test data from the BraTS dataset were used to assess

the performance of each nnUNet model. The UCLH clinical data were only used for testing (excluded during training) and are henceforth referred to as the external test set. This process served to interpret the generalizability of each model and to determine whether there was likely to be a future need for further retraining of algorithms on local hospital data. In addition, we applied our previously multimodal segmentation model [22], which includes post contrast T1-weighted MRI for comparison in the UCLH data.

Model performance assessment: Segmentation performance metrics

Tumor segmentations for the T2w-only algorithm versus the ground truth labels were assessed using median dice similarity coefficient (Dice) and Hausdorff distance (HD) for all experiments.

Dice is defined as:

$$DICE = \frac{2|S_m \cap S_a|}{|S_m| + |S_a|} = \frac{2TP}{2TP + FP + FN}$$

where S_m are the manually segmented tumor voxels, S_a the automatically segmented tumor voxels, true positive voxels (TP), false positive voxels (FP) and false negative voxels (FN). HD calculates the greatest distance between boundaries of two images and is therefore sensitive to assessment direction and outliers. The undirected 95th percentile HD was used:

$$d_{u95H} = \max\left\{95percentile \min_{p_m \epsilon P_m} d(p_a, p_m), 95percentile \min_{p_a \epsilon P_a} d(p_a, p_m)\right\}$$

where P_m is the collection of vertices defining the manual segmentation boundaries, P_a as the collection of vertices defining the automatic segmentation boundaries, and $d(p_a, p_m)$ the distance between two vertices.

Performance metrics were generated using methods described by Taha and Hanbury [34] and associated software – https://github.com/Visceral-Project/EvaluateSegmentation. Additionally, each test case underwent a visual inspection of the nnUNet mask overlap with tumor.

ADC readout evaluation metrics

ADC metrics were extracted and underwent binary logistic regression in SPSS to predict glioma IDH status (mutant vs. wild-type). For the comparison of manual and nnUNet ADC parameters, we used the equivalence test, which is the paired two one-sided t-tests (TOST) for testing equivalence.

Specifically, the hypotheses tested in TOST are:

$$H0: |\mu 1 - \mu 2| \geq \Delta \ (mean\ difference\ exceeds\ equivalence\ bounds)$$

$$H1: |\mu 1 - \mu 2| < \Delta \ (mean\ difference\ within\ equivalence\ bounds)$$

Where μ1 and μ2 are the means of the two groups, and Δ is the pre-specified equivalence margin.

In our study, we defined the equivalence bounds as low_eqbound = -0.01 and high_eqbound = 0.01. This means that if the confidence interval of the mean difference between the manual VOI and nnUNet VOI methods falls entirely within (-0.01, 0.01), the two methods can be considered clinically equivalent. We extracted the p-values and confidence interval bounds for the paired TOST using the Python "statsmodels" library (version 0.14.2). For the manual method, Youden's index was used to define an ADC threshold for IDH^{wt} identification [35].

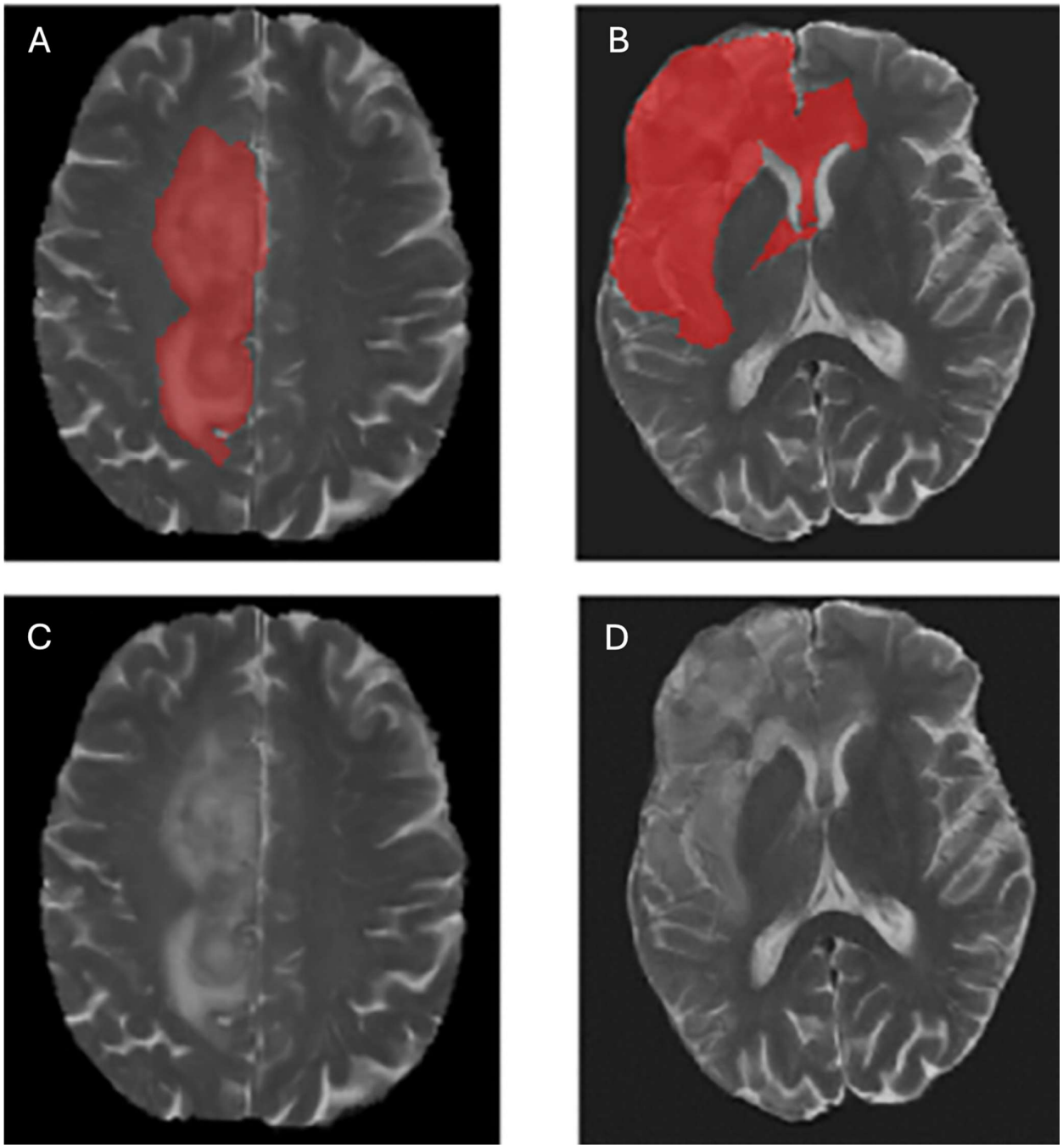

Figure 1. The left column shows a case from the BraTS 2021 dataset (whole tumor segmentation (WT) in red (A, B)). The WT class includes the full extent of the lesion including the tumor core (solid tumor and, where applicable, necrosis) and infiltration-oedema, indicated by hyperintensity. The right column shows a glioma from the UCLH glioma dataset (whole tumor segmentation (WT)) in red (B, D).

RESULTS

nnUNet segmentation performance BraTS

The deep learning brain tumor segmentation model achieved a median Dice Similarity Coefficient (Dice) of 0.85 on the BraTS test sets for delineating T2w glioma volumes. Furthermore, the model exhibited sensitivity and specificity values of 0.88 and 0.92, respectively, on the BraTS test set. The observed Hausdorff distance, a measure of the maximum distance between the predicted and ground truth segmentation boundaries, was 7.85 mm on the BraTS data sets **(Table 2).**

Table 2. Performance of whole tumor T2w-based segmentation model on internal and external tests sets.

Test set	Dice (IQR)	HD (IQR)	Sensitivity (IQR)	Specificity (IQR)
Internal (BRATS)	0.85 (0.09)	7.85 (8.58)	0.88 (0.08)	0.92 (0.11)
External (UCLH)	0.83 (0.13)	12.61 (10.11)	0.85 (0.11)	0.90 (0.13)

Median Dice Similarity Coefficient (Dice) and Hausdorff Distance (HD) on the BraTS test set (n=75) and UCLH test set (n=247) of the nnUNet segmentation model. The interquartile range (IQR) is also provided in comparing the performance of models in terms of sensitivity and specificity.

nnUNet segmentation validation in real-world data

On the UCLH dataset, the model maintained a consistent level of performance with a median Dice of 0.83 for whole tumor segmentation. The sensitivity and specificity values were 0.85 and 0.90, respectively. The median Hausdorff distance observed on the UCLH dataset was 12.61 mm **(Table 2)**. DeLong's test showed that there was no statistically significant difference (P =0.58) between the ROC profile of the nnUNet model (AUC = 0.82) and the manual labeling method (AUC = 0.84). When compared to the multimodal model, T2w-only yield comparable results on whole tumor label (paired t-test p<0.05).

Prediction of IDH Status

The performance of the nnUNet segmentation and ADC extraction pipeline for predicting glioma IDH status was evaluated on the external dataset (UCLH n=247, 70 IDHwt and 177 IDHmut). **Figure 2** shows the diagnostic performance of manually and automatically derived ADC metrics for IDH genotyping. The receiver operating characteristic area under the curve (AUC) values provide a quantitative measure of the diagnostic accuracy, with values closer to 1 indicating better performance.

Using the $rADC_{mean}$ extracted from the nnUNet segmentation masks, the AUC for differentiating IDH-mutant from IDH-wildtype gliomas was 0.82 (95% CI: 0.78-0.88), compared to the manual segmentation mask AUC 0.84 (95% CI: 0.77-0.89). The ADC_{mean} metric demonstrated an AUC of 0.77 (95% CI: 0.74-0.86) for the same task, compared to 0.79 (95% CI: 0.75-0.86) using manual masks. The ADC_{min} extracted from nnUNet masks yielded an AUC of 0.64 (95% CI: 0.60-0.74), while manual masks provided an AUC of 0.67 (95% CI: 0.61-0.75). Finally, the $rADC_{min}$ metric from nnUNet segmentations achieved an AUC of 0.70 (95% CI: 0.66-0.80), compared to 0.73 (95% CI: 0.67-0.81) using manual segmentations.

The equivalency between nnUNet segmentations and manual ADC readouts was assessed using a paired two one-sided t-test (TOST) on the University College London Hospitals test set comprising 247 cases. the TOST results for different VOI ADC metrics between nnUNet and manual ADC readout shows below: for ADC_{min},$rADC_{min}$, ADC_{mean}, and $rADC_{mean}$ were 0.0040, 0.0026, 0.0095, and 0.0018, respectively. These p-values indicate statistical equivalence of the manual and nnUNet segmentation derived ADC readouts.

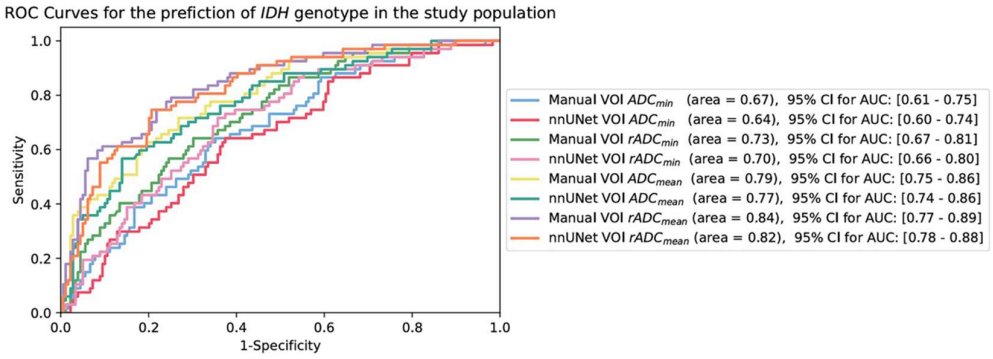

Figure 2. ROC curves of volumetric (VOI) ADC parameters for IDH genotype prediction.

Segmentation error rate

In most cases, nnUNet achieved a visually perfect match (**Figure 3**). In one case, an additional site of multifocal glioma infiltration was identified by the deep learning algorithm, which had been missed by the human operator. In the UCLH test set (n=247), n=2 (0.8%) gliomas were missed by the automated segmentation (false negative, **Figure 4**). In 15 (6%) of cases, normal temporal poles were erroneously segmented (false positive, **Figure 5**), remote from tumor boundaries.

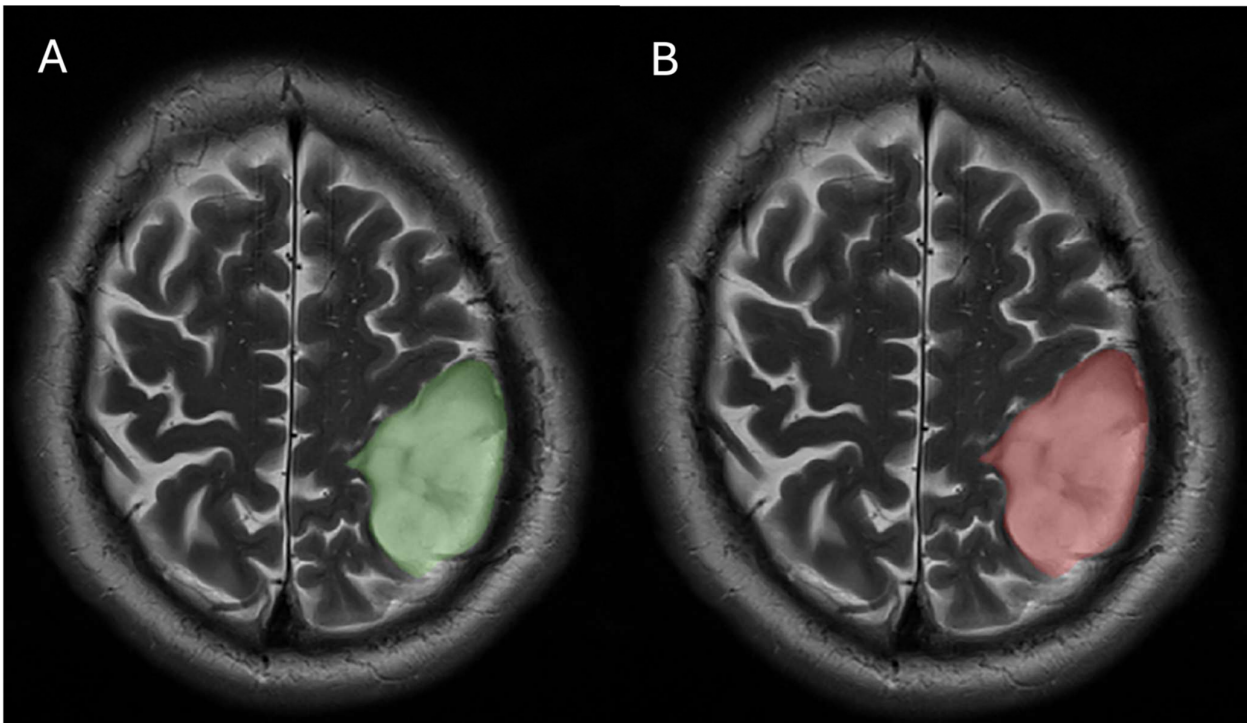

Figure 3. An example of the nnUnet model's segmentation output: manual (green mask, A) versus nnUNet (red mask, B) segmentation. This was perceived as a perfect visual match between the two methods.

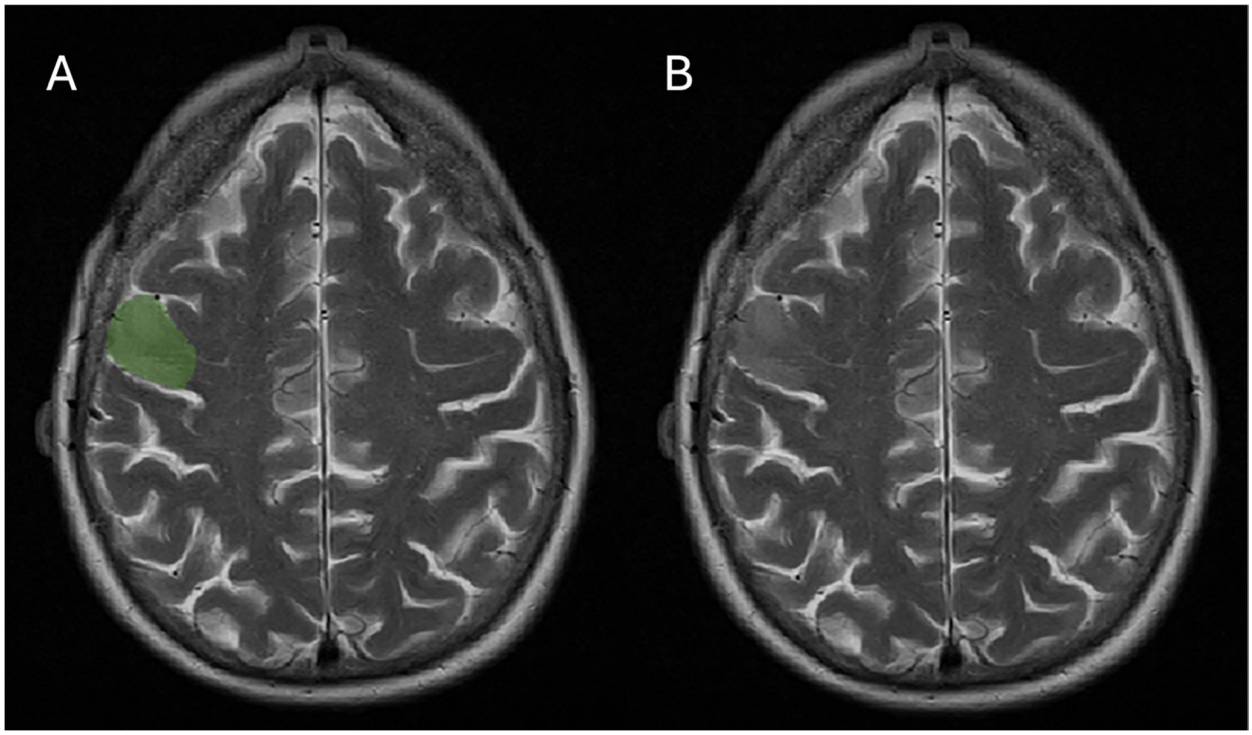

Figure 4. An example of a false-negative (2/247, 0.8%) nnUNet segmentation: Glioma infiltration was identified by the manual segmentation (green mask, A) but missed by the automated segmentation (missing red mask, B).

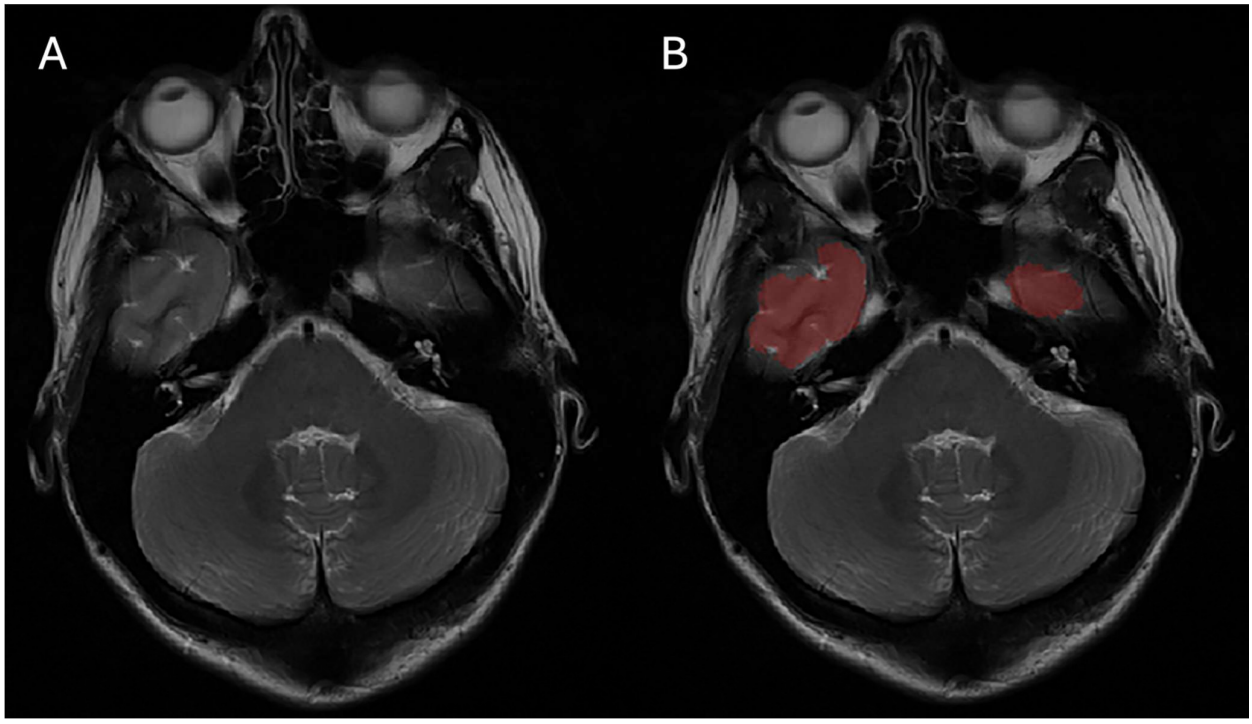

Figure 5. An example of a false-positive (n=15/247, 6%) nnUNet segmentation: No glioma infiltration is visible at the temporal poles (A), but these were outlined by the automated segmentation (red mask, B).

DISCUSSION

In this study, a nnUNet deep learning algorithm trained exclusively on T2w sequences achieved glioma segmentation similar to (Dice score 0.83-0.85) to manual expert annotations. There was no significant difference between the AUC of the nnUNet and manual ADC extraction methods (AUC= 0.82 versus 0.84, p=0.58). Interestingly, the algorithm operated accurately in histological CNS WHO grades 2-3 after being trained on a dataset (BraTS 2021) rich in WHO grade 4 glioblastoma. Glioblastoma is the commonest type of glioma, and thus overrepresented in several of the currently published datasets, whereas solid mGBM stages are less prevalent [30]. Despite the potential bias, this data imbalance did not appear detrimental for the segmentation task.

The decision to employ only T2w sequences for training was based on their capacity to provide valuable information about tumor boundaries whilst being universally performed in glioma imaging [37] without the need for a contrast injection. Our targeted training approach served to optimize the segmentation performance for rapid workflows, or circumstances where only T2w images might be available. Because our center is a regional referral institution, the preoperative MRI test sets in this research are 'real-world' data originating from a variety of referral institutions [12, 38]. Our findings highlight the potential of non-contrast genotyping, although we acknowledge the role of contrast-enhanced imaging in current clinical practice.

In comparison with deep learning segmentations using multimodal annotations [22, 39], the T2w-only approach was similarly comparable with human 'ground truth' for whole lesion delineation (paired t-test $p<0.05$). For IDH genotyping, nnUNet-based segmentations produced normalized glioma ADC values statistically equivalent to measurements derived from manual segmentations. This reflects the close overlap of manual and automated lesion masks but may also be influenced by tumor properties. In a comparison of whole lesion and regional measurements [13], we previously observed minimal differences in glioma diffusivity.

Identical to the manually derived results, the best performance for IDH genotyping was achieved using rADC values. By the process of normalization, ADC values become independent from acquisition influences and vendor differences [40]. The reason for ADC minimum values performing less well remains unclear; a possible explanation could be the presence of a relatively more cellular rim in IDH mutant glioma compared to their core [41], which would reduce ADC differences between genotypes, partial volume influences from normal brain bordering glioma segmentations, or registration imperfections.

The results from this research indicate potential to expedite clinical workflows whilst improving diagnostic accuracy. Qualitative ADC map inspection for signs of diffusion restriction is common radiologist practice [31], but this is not predictive of IDH genotype [36]. While our algorithm showed a strong performance (Dice 0.83, AUC >0.82), it may not overcome all variations in tumor morphology and imaging conditions. In a limited proportion of cases (6%), false positive segmentation of temporal pole parenchyma occurred. In a post hoc analysis, these areas were perceived to exhibit slightly greater T2w parenchymal signal (unquantified). Such segmentation errors, which occurred in a symmetrical fashion remote from tumor, are easily identifiable and could be remedied by manual erasure in a hospital workflow. To further address this issue, one could consider implementing additional post-processing steps, such as morphological filtering or threshold-based refinement, to reduce false positives in future iterations of the algorithm. In 0.8% cases, small tumors were missed by the algorithm. However, in one case nnUNet discovered an area of glioma infiltration which had been missed by the human observer in multifocal disease. For the task of direct comparison to human-derived tumor ADC values, the comparison values in contralateral normal appearing white matter (NAWM) were kept identical; it is possible that rADC results slightly differ for automatically derived NAWM ADC values.

A potential limitation of this study is the molecular testing prior to the latest WHO 2021 classification. The UCLH IDH^{wt} cohort underwent state-of-the art tissue analysis at the time of enrolment, however refinements were made to the mGBM genetic definition subsequently [3].

The model's performance may be further improved by analyzing larger sets of data, and/or by training on a greater number of histological WHO grade 2-3 gliomas to increase target representation. A broader program of validation of this method would be necessary to ensure its applicability in diverse clinical settings.

CONCLUSIONS

We present a novel approach to automatically quantify ADC in adult-type diffuse gliomas (WHO grade 2-3 IDH-mutant astrocytomas, oligodendrogliomas, and molecular glioblastoma with no histological high-grade features, i.e. lacking microvascular proliferation and/or necrosis (mGBM)), based on a single T2w sequence-based segmentation. This approach could support rapid ADC based identification of glioblastoma at an early disease stage, even with limited input data. By leveraging deep learning techniques, we succeeded in automating the laborious process of manual segmentation to produce equivalent results.

REFERENCES

1. Eckel-Passow JE, Lachance DH, Molinaro AM, et al (2015) Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. N Engl J Med 372:2499–2508. https://doi.org/10.1056/NEJMOA1407279
2. Louis DN, Perry A, Wesseling P, et al (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol 23:1231–1251. https://doi.org/10.1093/NEUONC/NOAB106
3. Brat DJ, Aldape K, Colman H, et al (2018) cIMPACT-NOW update 3: recommended diagnostic criteria for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV." Acta Neuropathol 136:805–810. https://doi.org/10.1007/S00401-018-1913-0
4. Alnahhas I (2024) Molecular Testing in Gliomas: What is Necessary in Routine Clinical Practice? Curr Oncol Rep 26:1277.

https://doi.org/10.1007/S11912-024-01602-W
5. BNOS Guideline for Tissue Sampling of Brain Tumours - BNOS. https://www.bnos.org.uk/bnos-guideline-for-tissue-sampling-of-brain-tumours/. Accessed 14 May 2024
6. Zhang J, Peng H, Wang YL, et al (2021) Predictive Role of the Apparent Diffusion Coefficient and MRI Morphologic Features on IDH Status in Patients With Diffuse Glioma: A Retrospective Cross-Sectional Study. Front Oncol 11:640738. https://doi.org/10.3389/FONC.2021.640738/FULL
7. Verduin M, Compter I, Steijvers D, et al (2018) Noninvasive Glioblastoma Testing: Multimodal Approach to Monitoring and Predicting Treatment Response. Dis Markers 2018:2908609. https://doi.org/10.1155/2018/2908609
8. Thust SC, Hassanein S, Bisdas S, et al (2018) Apparent diffusion coefficient for molecular subtyping of non-gadolinium-enhancing WHO grade II/III glioma: volumetric segmentation versus two-dimensional region of interest analysis. Eur Radiol 28:3779–3788. https://doi.org/10.1007/S00330-018-5351-0
9. Leu K, Ott GA, Lai A, et al (2017) Perfusion and diffusion MRI signatures in histologic and genetic subtypes of WHO grade II-III diffuse gliomas. J Neurooncol 134:177–188. https://doi.org/10.1007/S11060-017-2506-9
10. Cho NS, Sanvito F, Le VL, et al (2024) Diffusion MRI is superior to quantitative T2-FLAIR mismatch in predicting molecular subtypes of human non-enhancing gliomas. Neuroradiology 66:. https://doi.org/10.1007/S00234-024-03475-Z
11. Chen H, Hu W, He H, et al (2019) Noninvasive assessment of H3 K27M mutational status in diffuse midline gliomas by using apparent diffusion coefficient measurements. Eur J Radiol 114:152–159. https://doi.org/10.1016/J.EJRAD.2019.03.006
12. Maynard J, Okuchi S, Wastling S, et al (2020) World Health Organization Grade II/III Glioma Molecular Status: Prediction by MRI Morphologic Features and Apparent Diffusion Coefficient. Radiology 296:111–121. https://doi.org/10.1148/RADIOL.2020191832
13. Thust SC, Maynard JA, Benenati M, et al (2021) Regional and Volumetric Parameters for Diffusion-Weighted WHO Grade II and III Glioma Genotyping: A Method Comparison. American Journal of Neuroradiology 42:441–447. https://doi.org/10.3174/AJNR.A6965
14. Stall B, Zach L, Ning H, et al (2010) Comparison of T2 and FLAIR imaging for target delineation in high grade gliomas. Radiation Oncology 5:1–7. https://doi.org/10.1186/1748-717X-5-5/FIGURES/4
15. Lasocki A, Gaillard F, Lasocki XA, Gaillard XF (2019) Non-Contrast-Enhancing Tumor: A New Frontier in Glioblastoma Research. American Journal of Neuroradiology 40:758–765. https://doi.org/10.3174/AJNR.A6025
16. Maier-Hein KH, Isensee F, Petersen J, et al (2021) Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. Lancet Digit Health 3:e784–e794. https://doi.org/10.1016/S2589-7500(21)00205-3
17. Conte GM, Weston AD, Vogelsang DC, et al (2021) Generative Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a Multisequence Brain Tumor Segmentation Model. Radiology 299:313–323. https://doi.org/10.1148/RADIOL.2021203786
18. Ghadimi DJ, Vahdani AM, Karimi H, et al (2024) Deep Learning-Based Techniques in Glioma Brain Tumor Segmentation Using Multi-Parametric MRI: A Review on Clinical Applications and Future Outlooks. J Magn Reson Imaging. https://doi.org/10.1002/JMRI.29543
19. Isensee F, Kickingereder P, Wick W, et al (2019) No new-net. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11384 LNCS:234–244. https://doi.org/10.1007/978-3-030-11726-9_21/TABLES/3
20. Decuyper M, Bonte S, Deblaere K, Van Holen R (2021) Automated MRI based pipeline for segmentation and prediction of grade, IDH mutation and 1p19q co-deletion in glioma. Comput Med Imaging Graph 88:. https://doi.org/10.1016/J.COMPMEDIMAG.2020.101831
21. Ruffle JK, Mohinta S, Gray R, et al (2023) Brain tumour segmentation with incomplete imaging data. Brain Commun 5:fcad118. https://doi.org/10.1093/BRAINCOMMS/FCAD118
22. Pemberton HG, Wu J, Kommers I, et al (2023) Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms. Scientific Reports 2023 13:1 13:1–15. https://doi.org/10.1038/s41598-023-44794-0
23. Eijgelaar RS, Visser M, Müller DMJ, et al (2020) Robust deep learning–based segmentation of glioblastoma on routine clinical MRI scans using sparsified training. Radiol Artif Intell 2:1–9. https://doi.org/10.1148/RYAI.2020190103/ASSET/IMAGES/LARGE/RYAI.2020190103.FIG6.JPEG
24. Louis DN, Perry A, Reifenberger G, et al (2016) The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol 131:803–820. https://doi.org/10.1007/S00401-016-1545-1
25. Grech-Sollars M, Hales PW, Miyazaki K, et al (2015) Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. NMR Biomed 28:468–485. https://doi.org/10.1002/NBM.3269
26. Thust SC, Maynard JA, Benenati M, et al (2021) Regional and Volumetric Parameters for Diffusion-Weighted WHO Grade II and III Glioma Genotyping: A Method Comparison. AJNR Am J Neuroradiol 42:441–447. https://doi.org/10.3174/AJNR.A6965
27. Jenkinson M, Beckmann CF, Behrens TEJ, et al (2012) FSL. Neuroimage 62:782–790. https://doi.org/10.1016/J.NEUROIMAGE.2011.09.015
28. Baid U, Ghodasara S, Mohan S, et al (2021) The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification
29. Avants BB, Tustison N, Johnson H (2014) Advanced Normalization Tools (ANTS) Release 2.x
30. Rohlfing T, Zahr NM, Sullivan E V., Pfefferbaum A (2010) The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp 31:798–819. https://doi.org/10.1002/HBM.20906
31. Isensee Msc F, Schell M, Tursunova I, et al Title: Automated brain extraction of multi-sequence MRI using artificial neural networks Short title: ANN-based brain extraction with HD-BET
32. Isensee F, Jaeger PF, Kohl SAA, et al (2020) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 2020 18:2 18:203–211. https://doi.org/10.1038/s41592-020-01008-z
33. Kharaji M, Abbasi H, Orouskhani Y, et al (2024) Brain tumor segmentation with advanced nnU-Net: Pediatrics and adults tumors. Neuroscience Informatics 4:100156. https://doi.org/10.1016/J.NEURI.2024.100156
34. Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Med Imaging. https://doi.org/10.1186/s12880-015-0068-x
35. Thust SC, Maynard JA, Benenati M, et al Regional and Volumetric Parameters for Diffusion-Weighted WHO Grade II and III Glioma Genotyping: A Method Comparison. https://doi.org/10.3174/ajnr.A6965
36. Ostrom QT, Price M, Neff C, et al (2022) CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. Neuro Oncol 24:v1–v95. https://doi.org/10.1093/NEUONC/NOAC202
37. Thust SC, Heiland S, Falini A, et al (2018) Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice. Eur Radiol 28:3306–3317. https://doi.org/10.1007/S00330-018-5314-5
38. Doig D, Thorne L, Rees J, et al (2023) Clinical, Imaging and Neurogenetic Features of Patients with Gliomatosis Cerebri Referred to a Tertiary Neuro-Oncology Centre. J Pers Med 13:. https://doi.org/10.3390/JPM13020222
39. Pati S, Baid U, Edwards B, et al (2022) Federated learning enables big data for rare cancer boundary detection. Nat Commun 13:.

https://doi.org/10.1038/S41467-022-33407-5
40. Pipe J (2009) Pulse sequences for diffusion-weighted MRI. Diffusion MRI 11–35. https://doi.org/10.1016/B978-0-12-374709-9.00002-X
41. Fujita Y, Nagashima H, Tanaka K, et al (2021) The Histopathologic and Radiologic Features of T2-FLAIR Mismatch Sign in IDH-Mutant 1p/19q Non-codeleted Astrocytomas. World Neurosurg 149:e253–e260. https://doi.org/10.1016/J.WNEU.2021.02.042
42. Hyare H, Rice L, Thust S, et al (2019) Modelling MR and clinical features in grade II/III astrocytomas to predict IDH mutation status. Eur J Radiol 114:120–127. https://doi.org/10.1016/J.EJRAD.2019.03.003

SUPPLEMENTAL FILES

CLAIM: Checklist for Artificial Intelligence in Medical Imaging

Section / Topic	No.	Item	
TITLE / ABSTRACT			
	1	Identification as a study of AI methodology, specifying the category of technology used (e.g., deep learning)	√
	2	Structured summary of study design, methods, results, and conclusions	√
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the AI approach	√
	4	Study objectives and hypotheses	√
METHODS			
Study Design	**5**	Prospective or retrospective study	√
	6	Study goal, such as model creation, exploratory study, feasibility study, non-inferiority trial	√
Data	**7**	Data sources	√
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (e.g., symptoms, results from previous tests, inclusion in registry, patient-care setting, location, dates)	√
	9	Data pre-processing steps	√
	10	Selection of data subsets, if applicable	√
	11	Definitions of data elements, with references to Common Data Elements	√
	12	De-identification methods	√
	13	How missing data were handled	√
Ground Truth	**14**	Definition of ground truth reference standard, in sufficient detail to allow replication	√
	15	Rationale for choosing the reference standard (if alternatives exist)	√
	16	Source of ground-truth annotations; qualifications and preparation of annotators	√
	17	Annotation tools	√
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies	√
Data Partitions	**19**	Intended sample size and how it was determined	√
	20	How data were assigned to partitions; specify proportions	√
	21	Level at which partitions are disjoint (e.g., image, study, patient, institution)	√
Model	**22**	Detailed description of model, including inputs, outputs, all intermediate layers and connections	√
	23	Software libraries, frameworks, and packages	√
	24	Initialization of model parameters (e.g., randomization, transfer learning)	√
Training	**25**	Details of training approach, including data augmentation, hyperparameters, number of models trained	√
	26	Method of selecting the final model	√
	27	Ensembling techniques, if applicable	√
Evaluation	**28**	Metrics of model performance	√
	29	Statistical measures of significance and uncertainty (e.g., confidence intervals)	√
	30	Robustness or sensitivity analysis	√
	31	Methods for explainability or interpretability (e.g., saliency maps), and how they were validated	√
	32	Validation or testing on external data	√
RESULTS			
Data	**33**	Flow of participants or cases, using a diagram to indicate inclusion and exclusion	√
	34	Demographic and clinical characteristics of cases in each partition	√
Model performance	**35**	Performance metrics for optimal model(s) on all data partitions	√

	36	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	✓
	37	Failure analysis of incorrectly classified cases	✓
DISCUSSION			
	38	Study limitations, including potential bias, statistical uncertainty, and generalizability	✓
	39	Implications for practice, including the intended use and/or clinical role	✓
OTHER INFORMATION			
	40	Registration number and name of registry	✓
	41	Where the full study protocol can be accessed	✓
	42	Sources of funding and other support; role of funders	✓