GLIOCARE

THE ROAD TO GADOLINIUM FREE GLIOMA IMAGING

Ivar Johannes Hendrikus
Guillaume Wamelink

GLIOCARE

The Road to Gadolinium Free Glioma Imaging

Ivar Johannes Hendrikus Guillaume Wamelink

Colophon:

DOI: 10.5643/thesis.1632

Financial support for printing and publication of this thesis was kindly supported by stichting STOPhersentumoren and Cancer Center Amsterdam.

Cover art, design by Dilruba Tayfun
Printing by Ipskamp

VRIJE UNIVERSITEIT

GLIOCARE

The Road to Gadolinium Free Glioma Imaging

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
volgens besluit van de decaan
van de Faculteit der Geneeskunde
in het openbaar te verdedigen
op dinsdag 31 maart 2026 om 9.45 uur
in de universiteit

door

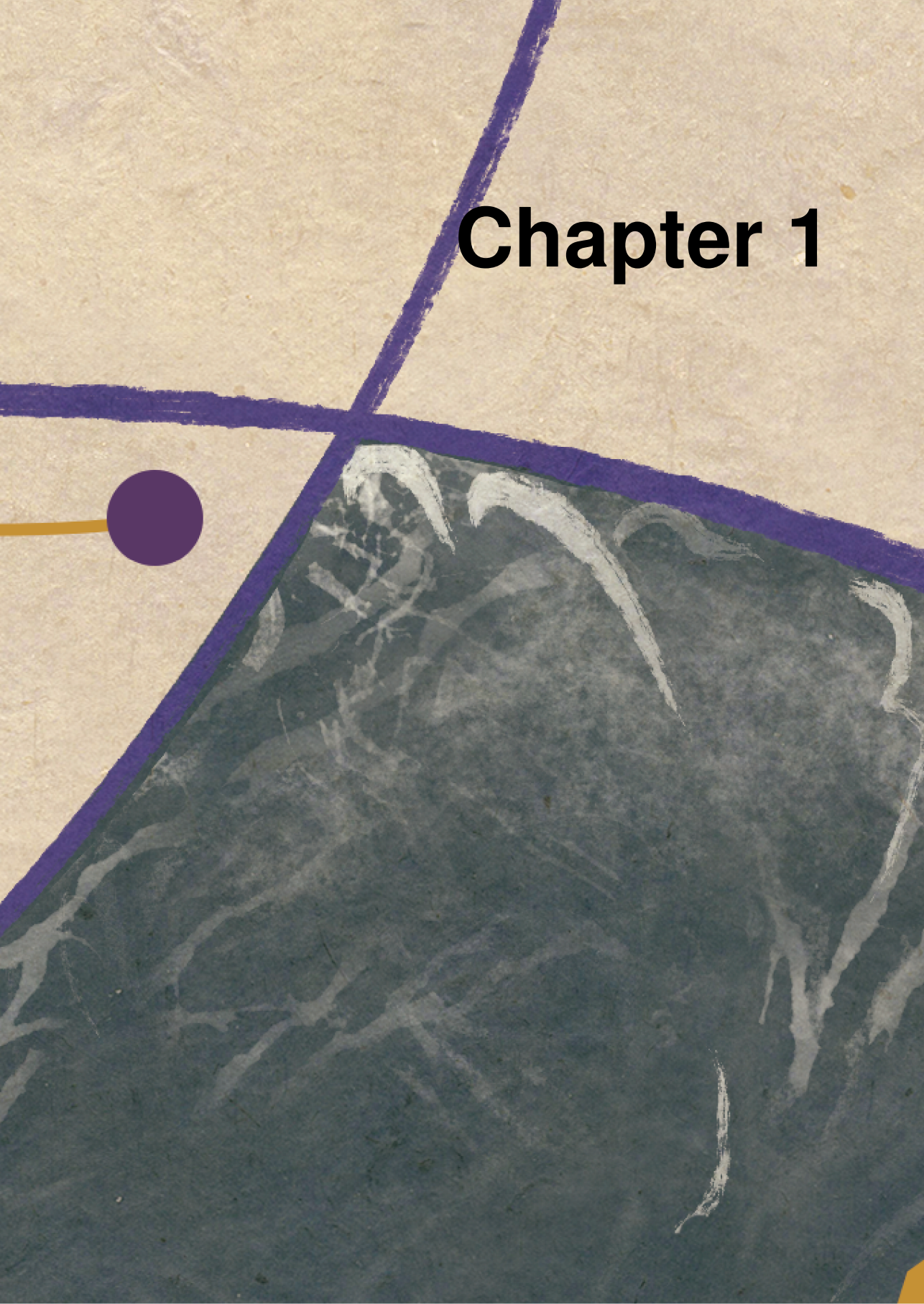
Ivar Johannes Hendrikus Guillaume Wamelink

geboren te 's-Hertogenbosch

promotor:	prof.dr. F. Barkhof
copromotoren:	dr. V.C.W Keil dr.ir. A.M. Wink
promotiecommissie:	prof.dr. D.P. Noske prof. N. Sollmann prof.dr.ir. A.J. Nederveen prof.dr. J.J.C. Verhoeff prof.dr. A. Van Der Hoorn prof.dr. S. Bisdas

Table of Contents

Chapter 1



Introduction

Glioma diagnosis and follow-up in a standard clinical setting

Adult-type diffuse gliomas are the most common type of primary brain cancer, with an incidence rate of ~ 6 per 100,000 globally.[1, 2] Gliomas originate from glial cells, which maintain neuronal homeostasis and signal transmission and outnumber neurons in a ratio of 10:1 in the central nervous system (CNS). Four types of glial cells exist: Oligodendrocytes, astrocytes, ependymal cells, and microglia. While each cell type can develop into a glioma, oligodendrogliomas, astrocytomas and glioblastomas (GBM) are the most common gliomas in adults. It is currently thought that glioblastoma cancer cells originate from astrocytes or neural stem cells.[3] Diagnosis is made by a neuropathologist after tumor tissue is obtained via biopsy or surgery. Pathological glioma subtyping is performed according to the 5th version of the WHO CNS criteria (CNS5 2021).[4] The CNS5 criteria distinguish three adult-type gliomas: oligodendroglioma (grades 2 and 3), astrocytoma (grades 2, 3, and 4), and glioblastoma (grade 4). The lower grade (2) gliomas are slow-growing, non-infiltrative tumors, whereas high-grade gliomas (grades 3 and 4) are fast-growing and infiltrative tumors with a dismal prognosis. According to the 2021 criteria, gliomas are graded using molecular markers complemented by phenotypic histological features, with the isocitrate dehydrogenase (IDH) gene playing a central role. IDH is an enzyme involved in the Krebs cycle and cellular homeostasis by catalyzing the oxidative decarboxylation of isocitrate.[5] IDH-mutated tumors tend to respond better to treatment and have a more favorable outcome. Oligodendrogliomas and astrocytomas are IDH-mutant tumors[6], whereas IDH wild-type (IDH-wt) gliomas are labeled as GBMs and are the most malignant form of glioma.[4] Another genetic marker is 1p/19q codeletion, where tumors with co-deletion are labeled as oligodendrogliomas and tumors without it are labeled as astrocytomas. 1p/19q-codeleted tumors have more favorable survival than non-codeleted counterparts. Extent of surgical resection, adjuvant therapy, and age are other prognostic factors.[7, 8]

Prognosis in glioma varies widely by histological grade and, increasingly, by molecular profile. Among lower-grade tumors, IDH-mutant oligodendrogliomas (1p/19q-codeleted) generally have the most favorable outcomes, with a 5-year survival of around 40%. IDH-wildtype tumors develop more aggressively, even during aggressive multimodal therapy; the 5-year survival rate is approximately 5%.[1, 9]

Symptoms of glioma patients are largely nonspecific. Patients with a high-grade glioma (HGG) more often have headaches, whereas patients with a low-grade glioma (LGG), particularly oligodendrogliomas, frequently suffer from seizures.[10] When a patient shows these symptoms, they typically undergo magnetic resonance imaging (MRI), which has become the central tool for establishing a suspected glioma diagnosis.[11] MRI is a highly versatile imaging modality, as it can be adjusted to show different soft-tissue contrasts.[12] This flexibility makes MRI particularly valuable for visualizing healthy brain tissue and pathological tissues such as gliomas. In neuro-oncology, structural MRI sequences remain fundamental for routine clinical assessment, offering essential information on tumor morphology, localization, and delineation.

MRI exploits the interaction between hydrogen nuclei and strong magnetic fields. By placing a patient in a magnetic field and applying a radiofrequency pulse tuned to the resonance frequency of water protons, MRI can use tissue-specific properties and translate these into images. Glioma MRI often follows a structured, standardized brain tumor imaging protocol to ensure consistency and generalizability. The routine protocol contains pre-, and postcontrast T1-weighted (T1w), T2-weighted (T2w), T2 fluid-attenuated inversion recovery (T2-FLAIR), and diffusion-weighted imaging (DWI). Dynamic susceptibility contrast (DSC) perfusion imaging MRI is often used to map brain perfusion.[13] Gliomas can present themselves either with or without contrast enhancement on contrast-enhanced T1w (CET1; postcontrast T1w) images and are generally hyperintense on FLAIR/T2w sequences (Figure 1).

Different MRI sequences yield distinct contrast profiles on the images. T2w images highlight tissues with high water content, causing fluids such as cerebrospinal fluid (CSF), edema, or cystic components to appear hyperintense, while fat-containing tissues appear hypointense. T2w imaging is therefore particularly useful for detecting edema, inflammatory changes, and cystic tumor regions. T2-FLAIR imaging extends the principles of T2-weighted imaging while removing signal from free fluids such as CSF. Free fluid removal enhances periventricular lesions, edema, and other fluid-associated abnormalities that might otherwise be obscured by the strong hyperintense signal of CSF in standard T2w images. As a result, T2-FLAIR is highly sensitive for detecting infiltrative tumor margins, small lesions adjacent to ventricles, and non-enhancing lesions.

On T1w images, fat- and calcium-rich tissues, among others, appear hyperintense, whereas water-containing tissues appear hypointense. T1w sequences become particularly informative after gadolinium-based contrast agent (GBCA) administration, producing CET1 images on which abnormal tissues often appear brighter. In gliomas, the enhancement pattern frequently correlates with biologic aggressiveness: high-grade tumors typically enhance, while many low-grade tumors do not. CET1 images are also used during surgery and radiotherapy planning, where they help to delineate the viable, contrast-enhancing tumor component for resection and irradiation. Clinically, GBCAs improve tumor detection, delineation of extent, and assessment of enhancement patterns, inform diagnosis, guide treatment planning, and facilitate response monitoring.

By extracting imaging phenotypes from different pulse sequences, radiologists make differential diagnoses that are discussed in multidisciplinary meetings to select surgery, radiotherapy, and/or chemotherapy, or active surveillance (watch-and-wait).[11] After treatment, patients undergo several follow-up MRI scans to assess treatment response and to guide subsequent care. Thus, throughout the course of the disease, patients undergo several MRI scans, each with contrast agents.

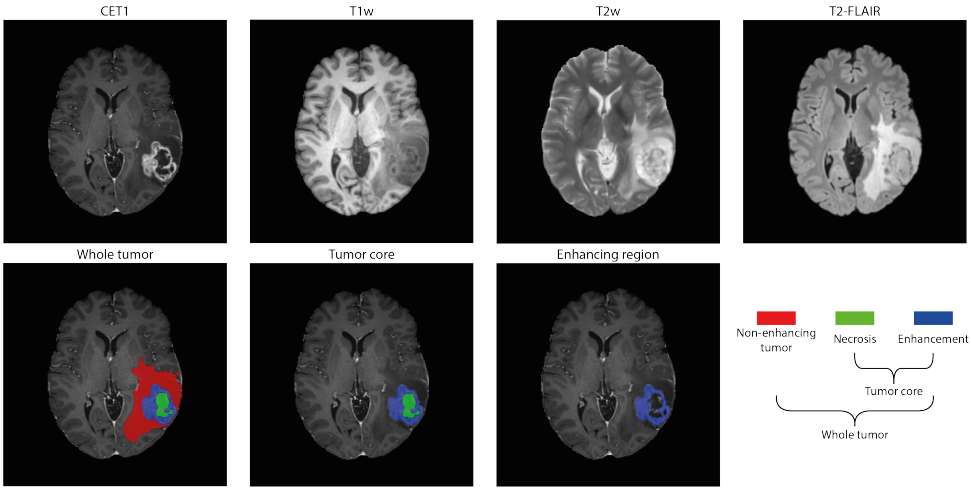


Figure 1: Overview of labeled tumor segmentation.

Gadolinium-Based Contrast Agents

Contrast-enhanced T1w imaging using GBCA represents the cornerstone of glioma diagnostics and monitoring.[11] GBCA was first introduced for clinical MRI in 1988 with gadopentetate dimeglumine (Magnevist®) and has since become widely adopted.[14] Since 1988, eight other gadolinium chelates have been developed and approved. When chelated, gadolinium can be safely administered intravenously, as the chelate prevents release of free gadolinium ions - a rare earth metal that is toxic to humans - and ensures almost complete renal clearance.[15]

GBCAs effectiveness stems from gadolinium's seven unpaired electrons, which provide strong paramagnetic properties. GBCAs are administered intravenously to enhance the visibility of abnormal tissues by shortening the T1 relaxation time (intrinsic relaxation properties of tissues that forms the basis for MRI contrast generation) of nearby water protons. T1 relaxation time shortening makes those areas appear brighter - referred to as 'enhancing' - on T1w images.

After injection, most GBCAs remain in the bloodstream and do not cross the intact blood–brain barrier (BBB). The BBB is a semi-permeable barrier formed by brain capillary endothelial cells and tight junctions between them. These tight junctions restrict the uncontrolled passage of large or charged molecules into the brain.[16] In many brain pathologies, including gliomas, the BBB becomes disrupted, and blood vessels may be abnormally leaky or highly vascular. Under these conditions, GBCAs can cross the broken BBB into the surrounding tissue, leading to increased T1 signal and contrast enhancement. This helps distinguish the viable components of the tumor (areas with disrupted BBB) from necrosis (areas with dead cells) and low-grade tumor areas. The RANO 2.0 criteria identify CET1 as the most sensitive and reliable MRI sequence for

assessing treatment response, particularly in high-grade gliomas.[17]

Surgical and radiotherapy planning has traditionally prioritized maximal resection or irradiation of contrast-enhancing tumors, as the non-enhancing part can only insufficiently be discriminated from brain edema. However, even after an apparent gross total resection and a radiation field that fully covers the enhancing lesion, most patients experience recurrence.[18, 19] Emerging evidence indicates that the enhancing margin underestimates the true extent of glioma, which infiltrate beyond areas of BBB disruption.[19, 20] Consequently, GBCA-enhanced MRI, while essential, does not capture the full "ground truth" tumor burden and should be interpreted alongside alternatives.

With 30 million GBCA administrations annually, downsides of GBCA administration seem unthinkable.[14] Yet, in recent years, various safety, environmental, patient discomfort, and economic concerns have emerged. From a safety perspective, adverse effects include nephrogenic systemic fibrosis (NSF), allergic reactions, and gadolinium deposition in the human body. NSF is a potentially fatal condition, predominantly affecting patients with impaired renal function exposed to linear-type GBCAs.[21, 22] Since 2014, gadolinium deposition in the brain (notably in the dentate nucleus) and in other organs has been documented, although its clinical consequences remain uncertain.[23, 24] As a result, regulatory bodies such as the European Medicines Agency (EMA) and the Federal Drugs Administration (FDA) have restricted the use of certain linear-type GBCAs, and clinical guidelines emphasize the need for clinical justification and careful use.[25, 26]

Environmental concerns have also gained prominence. GBCAs are excreted through the kidneys and enter the sewage system, ultimately ending up at wastewater treatment plants.[27] Current wastewater treatment plants are unable to remove GBCAs from wastewater, and processing may even promote conversion of chelate complexes to free, toxic Gd^{3+} ions.[28] This has led to rising levels of GBCAs in rivers, lakes, and even drinking water - also known as anthropogenic gadolinium - raising alarm about potential long-term effects on ecosystems and human health.[29]

On an individual level, patients may experience mild but unpleasant acute reactions such as coldness, warmth, and nausea, as well as pain at the injection site.[30] Particular caution is warranted in vulnerable groups, including prematures, pregnant and breastfeeding women, where the risk-benefit balance of GBCA administration is most delicate.[31, 32] Because these groups have reduced clearance, potential fetal/neonatal exposure, and limited long-term safety data, GBCA use should be restricted to clear indications, with preference for lower-risk (macrocyclic) agents and the minimum effective dose.[30]

From a societal perspective, GBCAs contribute substantially to healthcare costs. In the US alone, the GBCA contrast media market was valued at over half a billion USD in 2019.[33] While high-income countries can afford the costs of GBCA, low- and middle-income countries cannot, limiting their use in clinics.[34]

Despite these concerns, GBCA-enhanced MRI remains the clinical standard for glioma

imaging, as supported by consensus recommendations and the RANO criteria.[13, 17, 35] However, the evidence supporting this biomarker is surprisingly limited. Additional safety concerns have fueled growing interest in identifying GBCA-free alternatives for neuro-oncological imaging. This thesis focuses on two such approaches: (1) advanced MRI sequences, and (2) artificial intelligence–driven approaches.

Advanced MRI

Structural imaging (pre- and post-contrast T1w, T2w, T2-FLAIR) is limited in its ability to capture physiological and metabolic processes as it visualises 'static' tissue properties. Meanwhile, advanced MRI techniques focus on the extraction of (potential) 'active' biomarkers such as brain perfusion and/or permeability, microstructure, metabolism and chemical composition. Amide Proton Transfer-Chemical Exchange Saturation Transfer (APT-CEST) and Arterial Spin Labeling (ASL) are advanced sequences that can non-invasively characterize tumor properties, including protein content and perfusion dynamics.[36, 37] These methods provide complementary information beyond that of conventional structural MRI, aiding with diagnosis, prognosis, and treatment decisions.

APT imaging, a subset of CEST imaging, enables the detection of mobile amide protons in endogenous proteins and peptides.[36–38] These groups are abundant in gliomas, which often exhibit increased protein and peptide content. Rather than applying RF pulses at the water resonance frequency (as in conventional MRI), APT-CEST applies a frequency-selective saturation pulse at +3.5 ppm downfield from water, corresponding to the amide proton resonance. The saturated amide protons undergo chemical exchange with bulk water protons, and prolonged saturation reduces the observed water signal. However, saturation at +3.5 ppm also includes contributions from direct water saturation and magnetization transfer (MT) effects that are approximately symmetric about the water resonance. To account for these, a reference scan at –3.5 ppm (where no amide resonance is expected) is acquired. The difference between the signals at +3.5 ppm and –3.5 ppm (normalized by a reference image S_0) yields the magnetization transfer ratio asymmetry (MTR_{asym}), which indirectly reflects the concentration of mobile proteins and peptides (equation 1.1).[39]

$$APT_{CEST} = MTR_{asym}(3.5ppm) = \frac{S_{sat}(-3.5ppm) - S_{sat}(3.5ppm)}{S_0} \tag{1.1}$$

ASL MRI can noninvasively measure brain perfusion, more specifically, Cerebral Blood Flow (CBF). The technique measures the volume of arterial blood delivered to a unit mass of tissue per unit time (typically reported as mL blood/100g tissue/min). It does this by acquiring several pairs of control and label images to obtain sufficient signal. The control image is acquired before labeling and contains signals from blood and brain tissue. The label image is acquired with an identical MRI signal readout, but after labeling. During labeling, arterial blood water spins are magnetically inverted (or saturated), typically at the level of the cervical arteries to the brain. Once this labeled blood water flows

in the brain, it exchanges with tissue water. Because the blood magnetization has been inverted, when it mixes/exchanges with tissue water, the net longitudinal magnetization in that voxel decreases compared with the control condition. Thus, the label image contains the brain tissue signal and the attenuated signal from labeled (inverted) blood. By subtracting the label image from the control image, the signal of the brain tissue is effectively canceled, leaving the contribution of inflowing labeled blood water.[40]

Artificial Intelligence

Artificial intelligence (AI) broadly refers to computational methods that learn patterns from data to perform tasks such as prediction, classification, and segmentation. In practice, this learning is formulated as an optimization problem: given an algorithm with tunable parameters (hyperparameters) and a loss (objective) function that measures the discrepancy between the algorithm's outputs and desired targets, algorithms adjust the internal parameters to minimize this loss. This process (training) enables the algorithm to generalize from examples to unseen data.

When working with medical images (e.g., MRI), deep learning (DL) has become the dominant approach within AI. Deep learning algorithms are neural networks: compositions of layers that apply learned transformations to the data. In image-based DL, convolutional neural networks (CNNs) are especially common because they explicitly exploit spatial structure. A convolution applies a small, learnable filter (kernel) across the image, producing feature maps that highlight specific local patterns. These filters are not "the features" themselves; rather, they are parameters that the network learns so that the resulting activations become increasingly informative representations of the image. Early layers typically respond to simple image properties such as edges and corners; deeper layers combine these primitives into higher-level, more abstract concepts (e.g., anatomical structures like ventricles or tumor margins). CNN-based architectures often interleave convolutions with non-convolutional operations: nonlinear activations, normalization, and spatial down/up-sampling. Deep learning architectures learn from training data. Training data is composed of two data groups: 1) input data and 2) output data, where the goal of the algorithm is to predict the output based on the input. Depending on the optimization problem, architecture and input-output data pairs can differ.

Due to its adjustable nature, DL is an ideal method for reducing gadolinium usage. Two AI-based approaches have been pursued to reduce gadolinium usage: (1) segmentation algorithms that infer enhancing tumor regions from non-contrast inputs, effectively predicting where contrast uptake would appear, and (2) generative algorithms that synthesize contrast-enhanced T1-weighted (CET1) images from non-contrast sequences, aiming to approximate the diagnostic content of actual contrast administration.

Tumor Segmentation

In neuroradiology, delineating tumor borders is essential for surgical planning (e.g., resection margins) and for longitudinal assessment of tumor burden. In image processing,

the delineation of a volume is referred to as segmentation. In current practice, glioma segmentation typically relies on manual annotations by experts and is often reduced to 2D measurements in clinical workflows, despite tumors being inherently 3D. DL tumor segmentation algorithms address this by learning voxel-wise label maps from multi-modal MRI inputs (commonly T1w, CET1, T2w, and FLAIR). Outputs can be binary (e.g., whole tumor vs. background) or multi-class (e.g., edema/non-enhancing tumor, necrosis, and enhancing tumor). These supervised algorithms learn a mapping from input images to structured, per-voxel labels, but training is costly because each input-output pair requires detailed volumetric annotations drawn slice by slice for each patient. Encoder-decoder architectures with skip connections (e.g., U-Net variants) are widely used because they aggregate multi-scale context while preserving fine spatial detail (Figure 2, A). Loss functions are chosen to reflect the structured, imbalanced nature of segmentation. The Dice loss, which directly optimizes overlap, is common, often combined with cross-entropy or focal loss to improve calibration and class balance. Important considerations include class imbalance handling (e.g., region-weighted losses), 2D vs. 3D processing (3D CNNs better capture volumetric context but require more memory), patch-based training for large volumes, and post-processing (e.g., connected-component filtering).

Performance is typically reported with Dice, Hausdorff distance (95^{th} percentile), and volume error. Robustness across scanners and sites is addressed via intensity normalization, bias-field correction, resampling, data augmentation, and, when possible, domain adaptation.

Using contrast-enhanced sequences in the input, DL methods have achieved strong glioma segmentation performance (Figure 1).[41, 42] More recently, to reduce GBCA usage, researchers have explored algorithms that predict the enhancing region without using contrast-enhanced inputs.[43]

Contrast-Enhanced Image Synthesis

Image synthesis aims to map multi-modal inputs to a target imaging domain (output domain), for example, CET1 images. Formulating the task as conditional image-to-image translation encourages the algorithm to capture contrast uptake patterns consistent with anatomy. A common approach to generating synthetic CET1 images uses the CET1-T1w subtraction image as the target, focusing the algorithm on enhancement-specific signals rather than the entire CET1 image. Generative adversarial networks (GANs) are a popular choice for synthesis. They comprise a generator that produces synthetic images and a discriminator that distinguishes real from synthetic images, trained in an adversarial “game” (Figure 2, B). The generator typically optimizes a composite loss that combines a pixelwise loss term with an adversarial term. For the pixelwise loss, mean absolute error (MAE, L1) and structural similarity index measure (SSIM) are common; the discriminator commonly uses a cross-entropy loss.

At present, a complete shift to workflows that rely solely on segmentation (and omit CET1 acquisitions) is unlikely in the near term. In this context, CET1 synthesis algorithms may serve as an intermediary step toward gadolinium-based contrast agent (GBCA) reduc-

tion. Several proof-of-concept studies in neuro-oncology have demonstrated promising CET1 synthesis.[44–46] If visual similarity is high and clinical decision making is not affected, synthetic CET1 images could replace true CET1 images. By providing CET1-like information without additional contrast administration, such methods are likely to be more acceptable to clinicians and could facilitate a gradual, evidence-based shift in practice.[35]

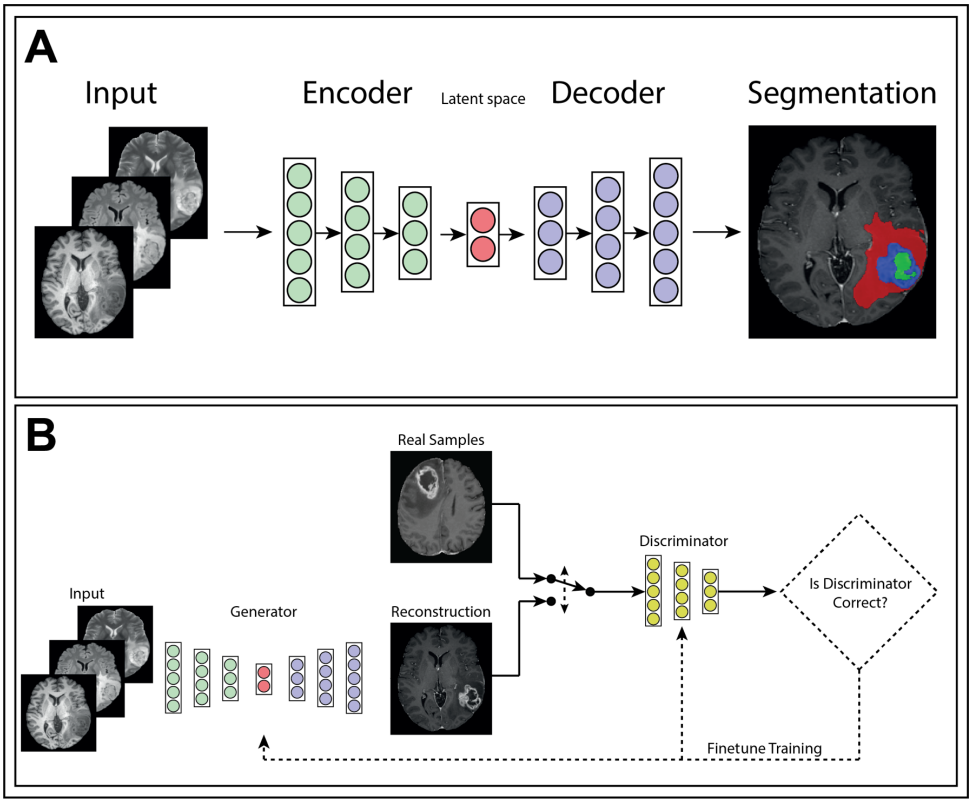


Figure 2: Encoder-Decoder UNet architecture (A) and a GAN architecture (B).

Aim and overview of this thesis

This thesis is the result of the GLIOCARE project (institutional review board VUmc_2021-0385) and was funded by Hanarth fonds call 2022. GLIOCARE: GLioma Imagin Omitting Contrast through Artificial Intelligence and Risk Evaluation aims to explore options to minimize GBCA usage in patients with glioma through the use of advanced MRI sequences and artificial intelligence. To this end, I scanned APT-CEST and ASL to investigate their diagnostic capabilities and use AI to develop synthetic post-contrast T1w images and GBCA-free segmentation tools.

I first assessed patient perspectives on GBCA use and reviewed existing literature on strategies to reduce GBCA exposure. In **Chapter 2** I performed a cross-sectional survey study to identify the patient experience in 100 patients with glioma. De novo patients and patients with long follow-up were included. In **Chapter 3** I did a nonsystematic literature review to identify current information regarding present and future opportunities and challenges of GBCA reduction in primary brain tumor imaging.

Next, I examined the value of APT-CEST and ASL sequences to reduce GBCA usage. **Chapter 4** focuses on APT-CEST reproducibility in healthy patients and served as a first pilot study in patients with glioma. The study investigated within-session, between-session, and between-days reproducibility in healthy volunteers. To limit patient burden, within-session reproducibility was assessed alone in this population. In **Chapter 5** tumor grade and genetic information are determined through preoperative APT-CEST and ASL imaging.

Before I applied AI approaches to minimize GBCA usage, I developed the Amsterdam UMC glioma imaging database IMAGO in **Chapter 6**. This database contains pre- and postoperative MRI data, along with clinical information, for 500 glioma patients. In **Chapter 7** I did a pilot study on 100 randomly selected GBM patients from the IMAGO database to compare segmentation performance between a radiologist assisted by AI and other radiologists working without AI. None of the parties, including the AI, had access to enhancing sequences. **Chapter 8** is a clinically oriented benchmark study that focuses on the technological readiness of algorithms that do not use CET1. Four large open-source datasets, including the IMAGO database, are used to train and test the identified algorithms. Rather than technological advancements, this study is clinically oriented and compares enhancement detection, prognostic performance, and response evaluation. In **Chapter 9** we developed an AI pipeline that can be adopted into the radiological glioma workflow. The pipeline focuses on triaging patients rather than reducing GBCA use overall. Patients can be labeled into one of three groups: 1) no GBCA is needed, 2) low-dose GBCA is needed, and 3) full-dose GBCA is warranted.

A general summary, discussion and future perspectives of the main findings can be found in **Chapter 10**.

References

1 Q. T. Ostrom et al. “CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2014-2018”. en. In: *Neuro. Oncol.* 23 (12 Suppl 2 Oct. 5, 2021), pp. iii1–iii105. doi: 10.1093/neuonc/noab200.

2 A. M. Molinaro et al. “Genetic and molecular epidemiology of adult diffuse glioma”. en. In: *Nat. Rev. Neurol.* 15 (7 July 2019), pp. 405–417. doi: 10.1038/s41582-019-0220-2.

3 H. J. Kim, J. W. Park, and J. H. Lee. “Genetic architectures and cell-of-origin in glioblastoma”. en. In: *Front. Oncol.* 10 (2020), p. 615400. doi: 10.3389/fonc.2020.615400.

4 D. N. Louis et al. “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary”. en. In: *Neuro. Oncol.* 23 (8 Aug. 2, 2021), pp. 1231–1251. doi: 10.1093/neuonc/noab106.

5 S. Han et al. “IDH mutation in glioma: molecular mechanisms and potential therapeutic targets”. en. In: *Br. J. Cancer* 122 (11 May 2020), pp. 1580–1589. doi: 10.1038/s41416-020-0814-x.

6 H. Yan et al. “IDH1 and IDH2 mutations in gliomas”. en. In: *N. Engl. J. Med.* 360 (8 Feb. 19, 2009), pp. 765–773. doi: 10.1056/NEJMoa0808710.

7 D. Li et al. “Management and survival trends for diffuse gliomas diagnosed at a single neurooncology center in China during 2000 to 2020”. In: *Sci Rep* 15 (1 Apr. 12, 2025). doi: 10.1038/s41598-025-95693-5.

8 P. Familiari et al. “Role of 1p/19q codeletion in diffuse low-grade glioma tumour prognosis”. en. In: *Anticancer Res.* 43 (6 June 2023), pp. 2659–2670. doi: 10.21873/anticanres.16432.

9 E. Miele et al. “Clinicopathological and molecular landscape of 5-year IDH-wild-type glioblastoma survivors: A multicentric retrospective study”. en. In: *Cancer Lett.* 588 (216711 Apr. 28, 2024), p. 216711. doi: 10.1016/j.canlet.2024.216711.

10 M. J. van den Bent et al. “Primary brain tumours in adults”. In: *Lancet* 402 (10412 Oct. 28, 2023). doi: 10.1016/S0140-6736(23)01054-1, pp. 1564–1579. doi: 10.1016/S0140-6736(23)01054-1.

11 M. Weller et al. “EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood”. en. In: *Nat. Rev. Clin. Oncol.* 18 (3 Dec. 8, 2020), pp. 170–186. doi: 10.1038/s41571-020-00447-z.

12 V. M. Runge and J. T. Heverhagen. *The physics of clinical MR taught through images*. en. 5th ed. Cham, Switzerland: Springer Nature, May 22, 2022. 373 pp. doi: 10.1007/978-3-030-85413-3.

13 B. M. Ellingson et al. “Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials”. en. In: *Neuro. Oncol.* 17 (9 Sept. 2015), pp. 1188–1198. doi: 10.1093/neuonc/nov095.

14 J. Lohrke et al. “25 Years of Contrast-Enhanced MRI: Developments, Current Challenges and Future Perspectives”. In: *Adv. Ther.* 33 (1 2016), pp. 1–28. doi: 10.1007/s12325-015-0275-4.

15 J. C. Weinreb et al. “Use of Intravenous Gadolinium-based Contrast Media in Patients with Kidney Disease: Consensus Statements from the American College of Radiology and the National Kidney Foundation”. In: *Radiology* 298 (1 Jan. 1, 2021), pp. 28–35. doi: 10.1148/radiol.2020202903.

16 D. Wu et al. “The blood-brain barrier: structure, regulation, and drug delivery”. en. In: *Signal Transduct. Target. Ther.* 8 (1 May 25, 2023), p. 217. doi: 10.1038/s41392-023-01481-w.

17 P. Y. Wen et al. “RANO 2.0: Update to the Response Assessment in Neuro-Oncology criteria for high- and low-grade gliomas in adults”. en. In: *J. Clin. Oncol.* 41 (33 Nov. 20, 2023), pp. 5187–5199. doi: 10.1200/JCO.23.01059.

18 L. S. Mirzayeva et al. “Pushing the boundaries of neurosurgical oncology: evaluating the superiority of supratotal resection over gross total resection in intraoperative MRI-guided glioma surgery”. en. In: *Neurosurg. Rev.* 48 (1 Feb. 6, 2025), p. 200. doi: 10.1007/s10143-025-03301-x.

19 A. F. Haddad et al. “FLAIRectomy: Resecting beyond the contrast margin for glioblastoma”. en. In: *Brain Sci.* 12 (5 Apr. 25, 2022), p. 544. doi: 10.3390/brainsci12050544.

20 F. Certo et al. “FLAIRectomy in supramarginal resection of glioblastoma correlates with clinical outcome and survival analysis: A prospective, single institution, case series”. en. In: *Oper. Neurosurg. (Hagerstown)* 20 (2 Jan. 13, 2021), pp. 151–163. doi: 10.1093/ons/opaa293.

21 Y. M. Shamam, M. F. Hashmi, and O. De Jesus. “Nephrogenic systemic fibrosis”. en. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, Jan. 2025.

22 T. Grobner. “Gadolinium–a specific trigger for the development of nephrogenic fibrosing dermopathy and nephrogenic systemic fibrosis?” In: *Nephrol. Dial. Transplant* 21 (4 2006), pp. 1104–1108. doi: 10.1093/ndt/gfk062.

23 T. Kanda et al. “High signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images: relationship with increasing cumulative dose of a gadolinium-based contrast material”. en. In: *Radiology* 270 (3 Mar. 2014), pp. 834–841. doi: 10.1148/radiol.13131669.

24 N. Iyad et al. “Gadolinium contrast agents- challenges and opportunities of a multidisciplinary approach: Literature review”. en. In: *Eur. J. Radiol. Open* 11 (100503 Dec. 2023), p. 100503. doi: 10.1016/j.ejro.2023.100503.

25 *Gadolinium-containing contrast agents - referral*. en. Mar. 18, 2016. (Visited on 05/16/2025).

26 *FDA Drug Safety Communication: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings*. May 16, 2018.

27 Y. Fujita et al. “Impacts of anthropogenic gadolinium on the activity of the ammonia oxidizing bacterium Nitrosomonas europaea”. en. In: *Chemosphere* 257 (127250 Oct. 2020), p. 127250. doi: 10.1016/j.chemosphere.2020.127250.

28 I. E. Oluwasola et al. “Gadolinium based contrast agents (GBCAs): Uniqueness, aquatic toxicity concerns, and prospective remediation”. en. In: *J. Contam. Hydrol.* 250 (104057 Oct. 2022), p. 104057. doi: 10.1016/j.jconhyd.2022.104057.

29 R. Brünjes and T. Hofmann. “Anthropogenic gadolinium in freshwater and drinking water systems”. en. In: *Water Res.* 182 (115966 Sept. 1, 2020), p. 115966. doi: 10.1016/j.watres.2020.115966.

30 ACR Committee on Drugs and Contrast Media. *ACR Manual on Contrast Media*. American College of Radiology, 2025.

31 E. Blumfield et al. “Survey of gadolinium-based contrast agent utilization among the members of the Society for Pediatric Radiology: a Quality and Safety Committee report”. en. In: *Pediatr. Radiol.* 47 (6 May 2017), pp. 665–673. doi: 10.1007/s00247-017-3807-z.

32 F. Proença et al. “Neuroimaging safety during pregnancy and lactation: a review”. en. In: *Neuroradiology* 63 (6 June 2021), pp. 837–845. doi: 10.1007/s00234-021-02675-1.

33 *U.S. Gadolinium-based Contrast Media Market (2020 - 2027)*.

34 U. C. Anazodo et al. “A framework for advancing sustainable magnetic resonance imaging access in Africa”. en. In: *NMR Biomed.* 36 (3 Mar. 2023), e4846. doi: 10.1002/nbm.4846.

35 F. Sanvito et al. “Standardized brain tumor imaging protocols for clinical trials: current recommendations and tips for integration”. en. In: *Front. Radiol.* 3 (Dec. 13, 2023), p. 1267615. doi: 10.3389/fradi.2023.1267615.

36 O. M. Henriksen et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 1: Perfusion and Diffusion Techniques”. en. In: *Front. Oncol.* 12 (Mar. 3, 2022), p. 810263. doi: 10.3389/fonc.2022.810263.

37 T. C. Booth et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 2: Spectroscopy, Chemical Exchange Saturation, Multiparametric Imaging, and Radiomics”. en. In: *Front. Oncol.* 11 (2021), p. 811425. doi: 10.3389/fonc.2021.811425.

38 T. C. Booth et al. “Machine learning and glioma imaging biomarkers”. In: *Clin. Radiol.* 75 (1 2020), pp. 20–32. doi: 10.1016/j.crad.2019.07.001.

39 P. C. M. van Zijl and N. N. Yadav. “Chemical exchange saturation transfer (CEST): what is in a name and what isn’t?” In: *Magn. Reson. Med.* 65 (4 Feb. 17, 2011), pp. 927–948. doi: 10.1002/mrm.22761.

40 P. Clement et al. “A beginner’s guide to arterial spin labeling (ASL) image processing”. en. In: *Front. Radiol.* 2 (June 14, 2022), p. 929533. doi: 10.3389/fradi.2022.929533.

41 H. G. Pemberton et al. “Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18911. doi: 10.1038/s41598-023-44794-0.

42 M. A. Naser and M. J. Deen. “Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images”. In: *Comput. Biol. Med.* 121 (2020), p. 103758. doi: 10.1016/j.compbiomed.2020.103758.

43 J. K. Ruffle et al. “Brain tumour segmentation with incomplete imaging data”. en. In: *Brain Commun.* 5 (2 Apr. 28, 2023), fcad118. doi: 10.1093/braincomms/fcad118.

44 C. Jayachandran Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *Lancet Digit Health* 3 (12 Dec. 2021), e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

45 E. Moya-Sáez, R. de Luis-García, and C. Alberola-López. “Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI”. In: *Frontiers in Neuroimaging* 2 (2023). doi: 10.3389/fnimg.2023.1055463.

46 L. Pasquini et al. “Synthetic Post-Contrast Imaging through Artificial Intelligence: Clinical Applications of Virtual and Augmented Contrast Media”. en. In: *Pharmaceutics* 14 (11 Nov. 4, 2022). doi: 10.3390/pharmaceutics14112378.

Chapter 2

The patients' experience of neuroimaging of primary brain tumors a cross-sectional survey study

Ivar J.H.G. Wamelink, Hugo L. Hempel, Elsmarieke van de Giessen, Mark H.M. Vries, Philip De Witt Hamer, Frederik Barkhof, Vera C. Keil

Journal of Neuro-Oncology

28-03-2023; 10.1007/s11060-023-04290

Abstract

Purpose: To gain insight into how patients with primary brain tumors experience MRI, follow-up protocols, and gadolinium-based contrast agent (GBCA) use.

Methods: Primary brain tumor patients answered a survey after their MRI exam. Questions were analyzed to determine trends in patients' experience regarding the scan itself, followup frequency, and the use of GBCAs. Subgroup analysis was performed on sex, lesion grade, age, and the number of scans. Subgroup comparison was made using the Pearson chi-square test and the Mann–Whitney U-test for categorical and ordinal questions, respectively.

Results: Of the 100 patients, 93 had a histopathologically confirmed diagnosis, and seven were considered to have a slowgrowing low-grade tumor after multidisciplinary assessment and follow-up. 61/100 patients were male, with a mean age±standard deviation of 44±14 years and 46±13 years for the females. Fifty-nine patients had low-grade tumors. Patients consistently underestimated the number of their previous scans. 92% of primary brain tumor patients did not experience the MRI as bothering and 78% would not change the number of follow-up MRIs. 63% of the patients would prefer GBCA free MRI scans if diagnostically equally accurate. Women found the MRI and receiving intravenous cannulas significantly more uncomfortable than men (p=0.003). Age, diagnosis, and the number of previous scans had no relevant impact on the patient experience.

Conclusion: Patients with primary brain tumors experienced current neuro-oncological MRI practice as positive. Especially women would, however, prefer GBCA-free imaging if diagnostically equally accurate. Patient knowledge of GBCAs was limited, indicating improvable patient information.

Keywords: *Primary brain neoplasms, Magnetic resonance imaging, Survey, Gadolinium*

Introduction

Patients with primary brain tumors, especially with gliomas, usually receive multiple MRI scans per year as standard care. Patients with a slow-growing low-grade glioma (LGG) may undergo dozens of MRI scans due to their chronic condition. Research endeavors towards faster and more informative MRI protocols are ongoing[1–4], but glioma MRI protocols remain lengthy and include gadolinium-based contrast agents (GBCA). The patient opinion on radiological care is largely unknown, despite the vulnerability of glioma patients and the relevant implications for patients and physicians.

MRI is a crucial pillar of therapy planning and response evaluation in neuro-oncology[5]. Brain tumor MRI protocols tend to adhere to consensus recommendations[6]. Most guidelines include an initial follow-up interval between three to six months after the completion of therapy, depending on the tumor histology. The scanning interval should be decreased to four to eight weeks in case of possible disease progression[7]. However, patients with brain tumors undergo particularly long and frequent MRI scans, while many low-grade brain tumors remain stable for long periods[8–10]. Furthermore, the benefit of fixed interval imaging remains unclear[11, 12].

Contrast-enhanced T1 weighted imaging (CET1w), often including contrast-enhanced dynamic susceptibility contrast perfusion imaging (DSC), is considered invaluable to the toolbox of neuroradiologists and is standard of care during the follow-up of brain tumors. However, research has shown long-term GBCA deposition, and current patient claims of GBCA-induced side effects are under investigation[13, 14]. Therefore, American and European pharmaco-safety agencies urge clinicians only to use GBCA when strictly necessary, but risk–benefit analyses for GBCA are awaited[15–17]. Many lesions never enhance or enhance without being high-grade brain tumors.

Against this backdrop, it becomes clear why neuroradiological research focuses on strategies to optimize imaging intervals. It also explores using advanced MRI sequences and artificial intelligence to shorten scan protocols and gain deeper insight into tumor biology[18–24]. This includes imaging without or with reduced GBCA, particularly for the low-grade tumor follow-up in the pediatric population[25–27].

The patient opinion on radiological care in brain tumor management is mainly unknown despite a general acknowledgement of the value of patient-centered research and shared decision-making[28, 29]. This includes patient opinions on GBCA use. There is a knowledge gap regarding the opinion of the patient on neuro-oncological MRI and research developments in particular, which also has a negative impact on the planning of future MRI research lines.

To gain more insight into the patient perspective on neuro-oncological MRI, its follow-up, and the use of GBCAs and to draw conclusions on the patient-perceived urgency of current research lines, we performed a cross-sectional survey on patients with primary brain tumors.

Methods

Study design and Participants

The local ethics committee approved the study. A questionnaire was designed in collaboration with our patient-reported experience measures department. One hundred patients were estimated as a sufficient sample size following the COSMIN study design checklist for patient-reported outcome measurement instruments[30].
Questionnaire targets were adult primary intra-axial brain tumor patients with at least 1 year of known diagnosis who had regular follow-up at our institution. The diagnosis had been histopathologically confirmed or was based on radiological phenotype and multidisciplinary consensus ("scan and wait"). Impairments due to tumor therapy were not an exclusion criterion, nor a selection criterion. All patients that had a neuro-oncological MRI scan and met all the inclusion-criteria were consecutively approached before their clinical MRI scan between 01-09-2021 and 04-08-2022. Patients needed to give informed consent before the clinically scheduled regular MRI scan and were interviewed directly after the scan. Patients with acute impairment, e.g. due to recent brain surgery (early postoperative MRI), were excluded from recruitment, as were patients under legal guardianship. The questionnaire was conducted on Dutch-speaking patients only. Participation was voluntary, and patients did not receive any compensation. Additional information on sex, age at study participation, tumor type and therapy course were added based on medical records. Grade 4 and 3 lesions were classified as high-grade gliomas (HGG), as most lesions were glioma-type. All others were in the LGG category. The WHO classification at the time of surgery defined the diagnosis. The number of previous MRI scans was registered from electronic hospital notes.

Questionnaire

Originally, the questionnaire contained ten questions plus one open comment space. It was extended by one additional question (question 11) during the course of the study in order to gain additional information on patient knowledge of GBCA.
First, the patient was asked to estimate their total number of tumor-related brain MRIs received until present. This was to evaluate if patients realistically assess the burden of MRI during the disease. The other questions required single-choice categorical and ordinal tick-box answers. Eight out of eleven questions allowed multiple answer options. Questions 2 and 10 were general questions on the burden of undergoing radiological follow-up as a patient with a primary brain tumor. Questions 3–6 covered the dimension of burden due to GBCA injection. Questions 7–9 covered the burden of the MRI scan procedure beyond GBCA injection. An open comment section ended the questionnaire. These comments were categorized into four groups: general burden of MRI, attitude towards GBCA injection, the burden due to follow-up/scan interval, and others.
The English version of the PENGUIN questionnaire can be found as Online Resource 1.

Analysis

Data was analyzed with descriptive statistics. Subgroup analyses were performed for sex, age, tumor grade, and the number of follow-up scans. Question 10 was transformed to a binary metric during the analysis, as patients were allowed to give multiple answers. If

patients left all the boxes empty, they were, after confirmation by the patient, categorized as not having any stress. The subgroup comparison involved parametric testing with the Pearson chi-square test and the Mann–Whitney U-test for categorical and ordinal questions, respectively. We also performed the Pearson correlation test. Bonferroni correction was performed for multiplicity. The significance level was set by dividing the significance threshold (0.05) by the number of subgroups (4) at $p < 0.01$.

Results

Patient Demographics

Of the one hundred filled-in questionnaires, 61 were from male patients (mean age ± standard deviation (SD): 44 ± 14 years) and 39 from female patients (mean age ± SD: 46 ± 13 years, Table 1). The age difference was insignificant between male and female patients ($p = 0.4$). Ninety-three patients had a histopathologically confirmed diagnosis. The other seven patients were considered to have a slow-growing low-grade lesion based on their radiophenotype and growth rates. In total, 41 patients had a HGG. Online Resource 2 shows the number of patients for each tumor entity.

Table 1: Patient subgroups split by sex and age for lesion subtypes and number of follow-up scans.

		Male (n = 61)			Female (n = 39)		
Age (in yrs.)[a]	**# of Patients**	**< 35**	**35–55**	**55+**	**< 35**	**35–55**	**55+**
High-grade	41	6	13	10	3	3	6
Low-grade	59	15	13	4	8	13	6
< 10 scans	33	8	8	3	4	5	5
10–20 scans	39	10	9	8	2	6	4
20–30 scans	18	2	7	3	3	3	0
> 30 scans	10	1	2	0	2	2	3

[a]Thirty-two patients were younger than 35 while 26 patients were older than 55. Forty-two patients were between the ages of 35 and 55.

General Burden of MRI

Patients systematically underestimated the number of scans they had undergone (49 underestimations vs 39 overestimations). An increase in the number of follow-up scans was associated with an increased underestimation of MRI burden (Fig. 1). If overestimation occurred, it was more marked than in cases of underestimation: 5 ± 6.57 scans overestimated compared to 3 ± 4 scans (median ± interquartile range) underestimated, respectively. There was no significant difference in over- or underestimation between men and women ($p = 0.83$).

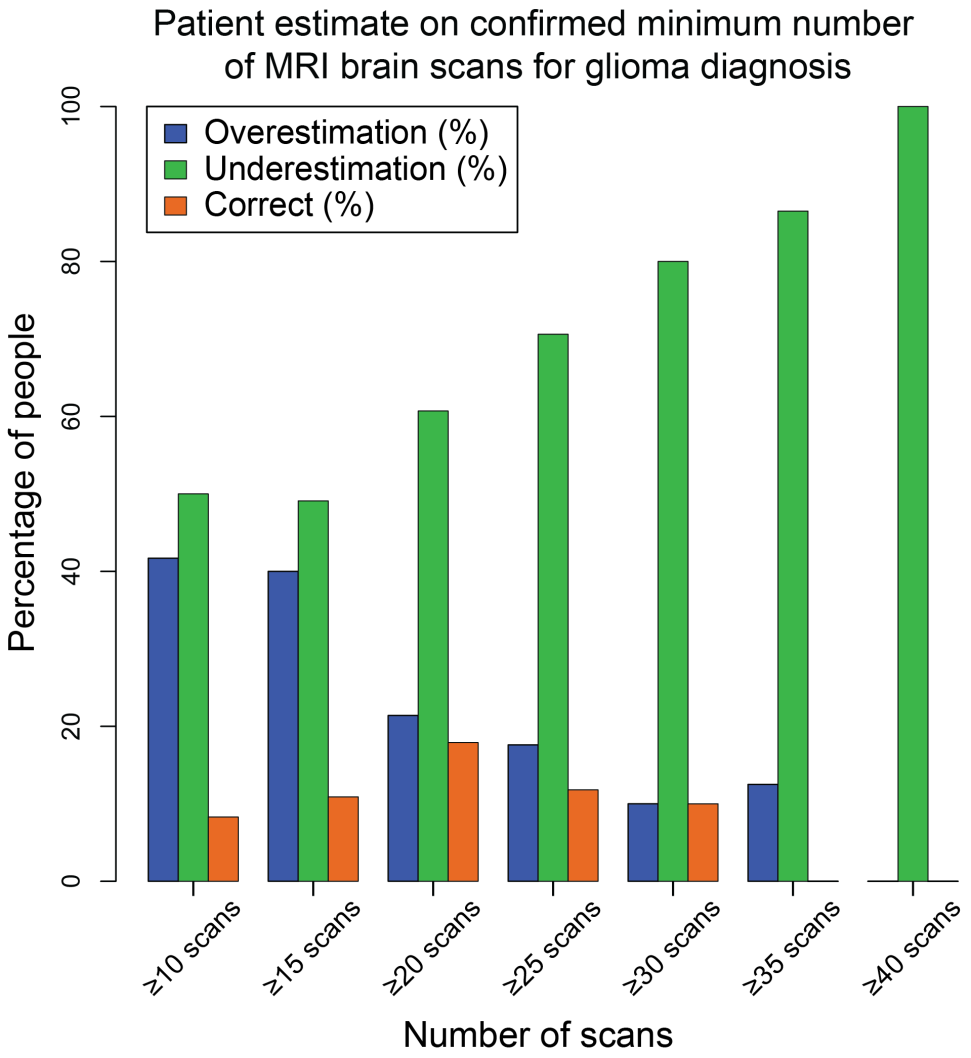


Figure 1: Percentage of patients overestimating, underestimating, or guessing correctly the number of MRI for glioma they had undergone until questionnaire session (y-axis) as sorted by the number of MRI they had truly undergone (x-axis). Underestimation of scan burden increases with the number of MRI undergone.

The general trend was that patients did not consider MRI scans burdensome, as shown in Fig. 3. It appeared that older patients experienced lying in the MRI as longer than younger patients, albeit not significantly (Q9, Online Resource 3). The most frequent stress factor was fear of outcome/bad news (Fig. 3). Five patients found the noise annoying or suggested extra hearing protection beyond the existing double layer.
Women reported more symptoms during or directly after the MRI (Q7) and found it more annoying than males (Q8). Women also experienced more stress from fearing bad news (Q10.1) and the travel times to the MRI unit (Q10.2) and experienced more

stress in general (Q10.6). Patients with less than ten previous MRIs were significantly more dissatisfied regarding the scan follow-up interval than patients with 30 or more scans. The same group (< 10 scans) showed significantly less fear towards receiving an intravenous (IV) cannula. Finally, LGG patients experienced more claustrophobia than HGG (Online Resource 3).

Attitude towards GBCA Injection

Most patients did not experience GBCA injections as burdensome (Fig. 2). However, 40% described at least some irritation when receiving IV administration. 63% preferred an MRI scan without GBCA if considered diagnostically equivalent. Fifty-eight patients answered the additional question about possible adverse effects from GBCA, but only three patients were aware of these. Nearly all patients found the wait between placing the cannula and taking the MRI perfect or short. Several patients wrote that they found cannulas annoying and painful and would prefer no cannula if diagnostic performance was maintained. One patient also mentioned that patients should receive more information regarding the use of GBCA agents.

Female patients found cannula placement significantly more unpleasant than males (Q4). There were no significant differences between males and females regarding preferences for MRI options without GBCA (Q5).

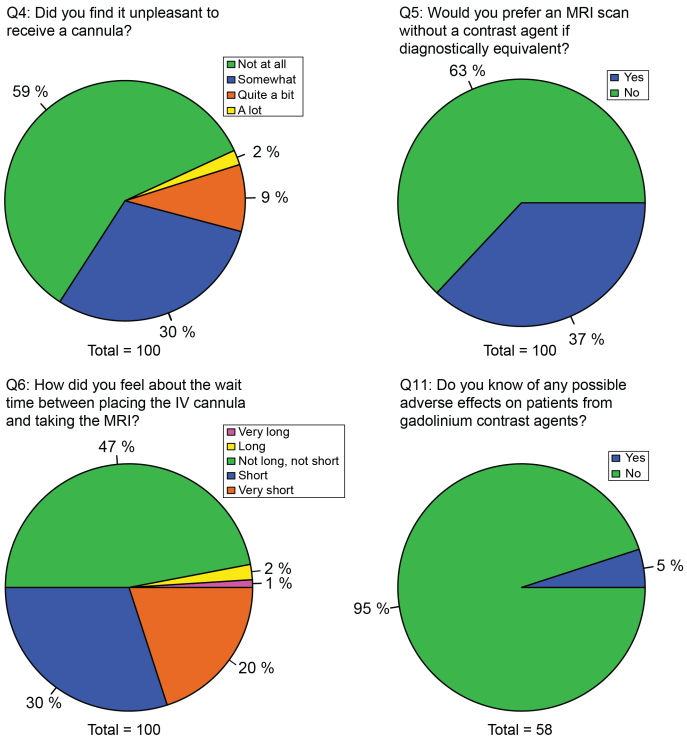


Figure 2: Questions 4–6 and 11 explore the dimensions of attitude towards GBCA injection.

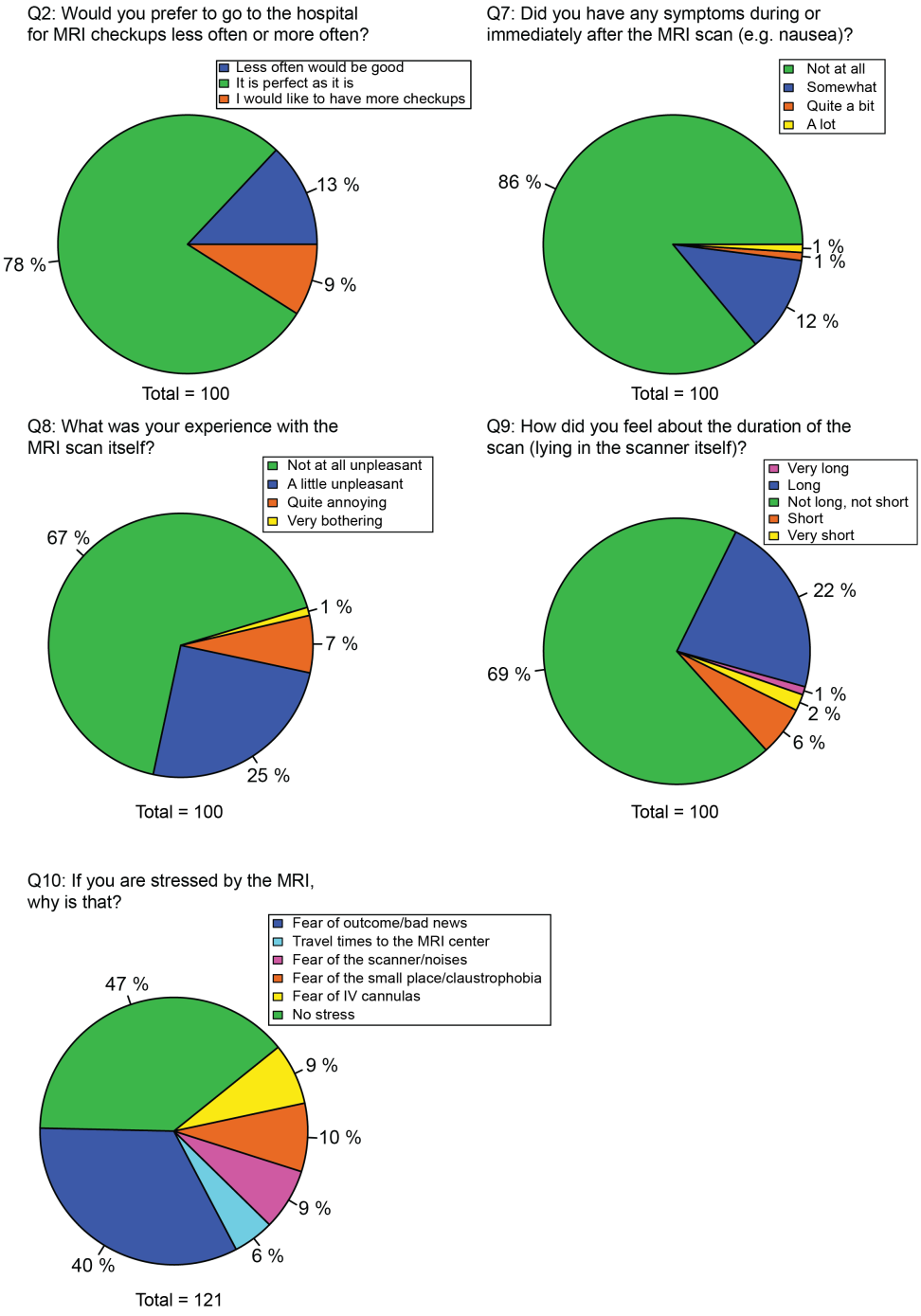


Figure 3: Questions 2 and 7–10 measured the dimensions of the general burden of the MRI. Patients were allowed to give multiple answers to question 10 if they experienced several types of stress.

Discussion

Patients with primary brain tumors expressed a generally positive attitude towards the current neuro-oncological MRI follow-up scheme in this monocentric survey at a tertiary academic center. However, GBCA-free MRI protocols would be preferred, provided their diagnostic non-inferiority. Importantly, patient knowledge about any potential adverse effects of GBCA was rare, and we identified women as less satisfied. At the same time, age, diagnosis and number of previous scans had no impact on satisfaction.

Patients underestimated the number of scans they had undergone, with underestimation being positively correlated with the number of previous scans. Scan burden underestimation can be explained by 'positivity bias in memory'—a phenomenon describing a person's inclination to remember pleasant events more vividly and favorably than unpleasant ones[31, 32]. Patients with 30 scans or more were significantly more satisfied with the number of scan follow-ups than patients with ten scans or fewer. We hypothesize that patients with more scans are more likely to think they are in a stable phase of their disease than patients who only recently got diagnosed. There is a tendency in the medical community to reduce both MRI frequency and protocol duration, scan time. Arguments are costs, waiting lists, and the assumption that patients find the MRI uncomfortable and have difficulty complying[33–35]. However, our results showed that most patients did not experience MRI as burdensome and that the follow-up intervals are perceived as appropriate—even by frequently scanned glioblastoma patients. The debate about whether scan frequency and protocol duration need to be reduced should include the patients, as they might oppose longer control intervals. On the other hand, data implies that most patients will tolerate moderate scan duration extension for imaging research.

However, certain patients experienced at least moderate discomfort during the MRI scan. It is worthwhile to study this group in more detail.

Our most remarkable and also most consistent finding is the role of sex in the perception of MRI. Overall, women found the MRI procedure more uncomfortable than men, which was characterized by experiencing the MRI procedure as more unpleasant, more often being afraid of bad news, and having a tendency to be more stressed about the travel times. These findings align with literature suggesting that women experience more stress and anxiety also when confronted with a brain tumor diagnosis[36, 37]. Women also found receiving a cannula more unpleasant. Research has reported that sex is a risk factor for difficult venous access and that catheter insertion in women is more difficult, explaining the difference in comfort[38]. We conclude that sex, and most likely gender, is not sufficiently reflected in the current MRI workflow of brain tumor patients despite indicators for relevant differences between male and female perception. According to our results, women will benefit from shorter MRI protocols–and should innovation permit it–even GBCA-free ones. The discussion between patient welfare and patient clinical needs should therefore be carefully balanced.

The age, number of previous scans, and diagnosis had surprisingly little impact on patients' MRI perception. Increasing age is a known stress factor for patients and MRI technicians[39]. In our study, we could only confirm a tendency below the significance threshold regarding age: older patients tended to experience the MRI scan as longer and

less comfortable than younger patients. This is relevant as brain tumor MRI protocols are particularly lengthy.
Patients with ten or fewer scans showed less fear of receiving a cannula than patients with 30 scans or more, at a significant level before and a possibly still relevant level after Bonferroni correction. While the probability of a negative experience with cannulas increases with the growing number of scans, this contrasts with what would be expected as dictated by exposure therapy. With exposure therapy, frequent engagement with anxiety-provoking stimuli, such as cannulas, can reduce and disconfirm a person's fearful projections towards the respective stimulus[40, 41].
Patients with a low-grade lesion experienced more claustrophobia than patients with a high-grade lesion. Patients in the low-grade group had a mean age of 41 years, while patients in the high-grade group had a mean age of 49, as expected. Even though the age differences between the two groups were normally distributed, the age difference could explain this finding as younger patients tend to be more stressed[42]. As younger patients with low-grade lesions will likely receive more follow-ups during their life time, any scientific innovation towards shorter protocols will be particularly in their favor. Clinically, the time between scans, the number of included sequences, and the decision of administering contrast is generally based on the lesion type. However, our results show that tumor type, a reflection of disease severity, does not play a relevant role in the perception of MRI. While our research did not focus on the patient-disease relationship, there may be a link between the patients' tolerance for number of follow-up scans and scan duration and the type of disease.
An estimated 40% of all MRI scans in neuroradiology are GBCA-dependent[43]. Therefore, patient opinion on gadolinium should be considered with the aim of shared decision-making[44]. Our results show that most patients would opt for an MRI scan without GBCAs if considered diagnostically non-inferior, supporting research in that direction. However, patients showed a profound lack of knowledge of GBCAs, including insufficient knowledge regarding possible adverse effects despite being a frequently prescribed diagnostic agent. Patients seemed to be poorly informed and could thus not make optimal decisions about their welfare. At this point, it must be understood that patients in the Netherlands will usually never meet with a radiologist, nor is written informed consent for MRI examinations with GBCA mandatory. Potential contraindications for MRI are ruled out by the clinician ordering the MRI scan. A detailed procedure description for the patient is usually not part of this conversation. Especially considering that pharmaco-safety agencies urge clinicians to reduce the use of gadolinium in the clinical workflow of glioma patients[16, 45], patients should be well-informed about the added value of GBCAs and their possible harm to the human body[13]. Beyond considerate use of GBCA and conciseness of scan protocols, there are other factors which may be relevant to increase patient comfort such as an acoustic optimization of sequences, as was confirmed by the comments of five of our participants[46].
There are several limitations to this study. First, this is a monocentric study with Dutch patients only, with consequences for data interpretation. The sample was, however, representative of the disease's general prevalence regarding diagnosis, age, and sex distribution[47]. Second, some of our patients have outdated diagnosis without an IDH classification which may result in high grade glioma patients being considered low grade

glioma. Patients for whom the questionnaire was too much of a burden were excluded, as were non-Dutch-speaking patients due to the language barrier. This biases the study, potentially underestimating the MRI burden by excluding the sickest patients and patients with a different cultural background. Further, the reference number of MRI scans derived from hospital records is a minimum estimate. Patients could also have been scanned elsewhere. However, patients in the Netherlands usually adhere to one clinic only for treatment and follow-up—making a relevant deviation in the correct total number of scans unlikely.

In summary, this study finds that patients with primary brain tumors generally have positive experiences with neuro-oncological MRI. Especially women, however, would support endeavors towards GBCA-free MRI diagnostics and shorter protocols. Approaches to reduce imaging frequency are neither a patient priority, nor preference. A lack of knowledge on GBCA indicates that shared decision-making remains an unreached goal in glioma imaging.

References

1 Y. Y. Cheung, E. M. Goodman, and T. O. Osunkoya. “No More Waits and Delays: Streamlining Workflow to Decrease Patient Time of Stay for Image-guided Musculoskeletal Procedures”. en. In: *Radiographics* 36 (3 May 2016), pp. 856–871. doi: 10.1148/rg.2016150174.

2 K. Beker et al. “Optimizing MRI Logistics: Prospective Analysis of Performance, Efficiency, and Patient Throughput”. en. In: *AJR Am. J. Roentgenol.* 209 (4 Oct. 2017), pp. 836–844. doi: 10.2214/AJR.16.17698.

3 M. P. Recht et al. “Optimization of MRI Turnaround Times Through the Use of Dockable Tables and Innovative Architectural Design Strategies”. en. In: *AJR Am. J. Roentgenol.* 212 (4 Apr. 2019), pp. 855–858. doi: 10.2214/AJR.18.20459.

4 R. Smith-Bindman et al. “Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016”. en. In: *JAMA* 322 (9 Sept. 3, 2019), pp. 843–856. doi: 10.1001/jama.2019.11456.

5 N. Tyldesley-Marshall et al. “The role of Magnetic Resonance Images (MRIs) in coping for patients with brain tumours and their parents: a qualitative study”. en. In: *BMC Cancer* 21 (1 Sept. 10, 2021), p. 1013. doi: 10.1186/s12885-021-08673-z.

6 B. M. Ellingson et al. “Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials”. en. In: *Neuro. Oncol.* 17 (9 Sept. 2015), pp. 1188–1198. doi: 10.1093/neuonc/nov095.

7 M. Weller et al. “EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood”. en. In: *Nat. Rev. Clin. Oncol.* 18 (3 Mar. 2021), pp. 170–186. doi: 10.1038/s41571-020-00447-z.

8 C. Gui et al. “Radiology reporting of low-grade glioma growth underestimates tumor expansion”. In: *Acta Neurochir.* 161 (3 2019), pp. 569–576. doi: 10.1007/s00701-018-03783-3.

9 R. Jooma, M. Waqas, and I. Khan. "Diffuse Low-Grade Glioma - Changing Concepts in Diagnosis and Management: A Review". en. In: *Asian J. Neurosurg.* 14 (2 Apr. 2019), pp. 356–363. doi: 10.4103/ajns.AJNS_24_18.

10 C. Gui et al. "Tumor growth dynamics in serially-imaged low-grade glioma patients". en. In: *J. Neurooncol.* 139 (1 Aug. 2018), pp. 167–175. doi: 10.1007/s11060-018-2857-x.

11 C. L. Monroe et al. "Does Surveillance-Detected Disease Progression Yield Superior Patient Outcomes in High-Grade Glioma?" en. In: *World Neurosurg.* 135 (Mar. 2020), e410–e417. doi: 10.1016/j.wneu.2019.12.001.

12 S. Y. Ji et al. "Radiological assessment schedule for high-grade glioma patients during the surveillance period using parametric modeling". en. In: *Neuro. Oncol.* 23 (5 May 5, 2021), pp. 837–847. doi: 10.1093/neuonc/noaa250.

13 C. A. Mallio et al. "Gadolinium Deposition Safety: Seeking the Patient's Perspective". en. In: *AJNR Am. J. Neuroradiol.* 41 (6 June 2020), pp. 944–946. doi: 10.3174/ajnr.A6586.

14 M. Parillo et al. "A Structured Survey on Adverse Events Occurring Within 24 Hours After Intravenous Exposure to Gadodiamide or Gadoterate Meglumine: A Controlled Prospective Comparison Study". In: *Invest. Radiol.* 54 (4 Apr. 2019), p. 191. doi: 10.1097/RLI.0000000000000528.

15 *Gadolinium-containing contrast agents*. Dec. 19, 2017.

16 *FDA Drug Safety Communication: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings*. May 16, 2018.

17 R. J. McDonald et al. "Gadolinium Retention: A Research Roadmap from the 2018 NIH/ACR/RSNA Workshop on Gadolinium Chelates". en. In: *Radiology* 289 (2 Nov. 2018), pp. 517–534. doi: 10.1148/radiol.2018181151.

18 P. Lehmann et al. "A comparative study of perfusion measurement in brain tumours at 3 Tesla MR: Arterial spin labeling versus dynamic susceptibility contrast-enhanced MRI". en. In: *Eur. Neurol.* 64 (1 June 19, 2010), pp. 21–26. doi: 10.1159/000311520.

19 C. Warmuth, M. Gunther, and C. Zimmer. "Quantification of blood flow in brain tumors: comparison of arterial spin labeling and dynamic susceptibility-weighted contrast-enhanced MR imaging". en. In: *Radiology* 228 (2 Aug. 2003), pp. 523–532. doi: 10.1148/radiol.2282020409.

20 H. Mehrabian et al. "Differentiation between Radiation Necrosis and Tumor Progression Using Chemical Exchange Saturation Transfer". en. In: *Clin. Cancer Res.* 23 (14 July 15, 2017), pp. 3667–3675. doi: 10.1158/1078-0432.CCR-16-2265.

21 J. Zhou et al. "Differentiation between glioma and radiation necrosis using molecular magnetic resonance imaging of endogenous proteins and peptides". en. In: *Nat. Med.* 17 (1 Jan. 2011), pp. 130–134. doi: 10.1038/nm.2268.

22 O. Togao et al. “Grading diffuse gliomas without intense contrast enhancement by amide proton transfer MR imaging: comparisons with diffusion- and perfusion-weighted imaging”. en. In: *Eur. Radiol.* 27 (2 Feb. 2017), pp. 578–588. doi: 10.1007/s00330-016-4328-0.

23 M. Patel et al. “Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma”. en. In: *Clin. Radiol.* 76 (8 Aug. 2021), 628.e17–628.e27. doi: 10.1016/j.crad.2021.03.019.

24 C. Jayachandran Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *Lancet Digit Health* 3 (12 Dec. 2021), e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

25 G. Morana et al. “Pediatric astrocytic tumor grading: comparison between arterial spin labeling and dynamic susceptibility contrast MRI perfusion”. en. In: *Neuroradiology* 60 (4 Apr. 2018), pp. 437–446. doi: 10.1007/s00234-018-1992-6.

26 J. Novak et al. “A comparison of pseudo-continuous arterial spin labelling and dynamic susceptibility contrast MRI with and without contrast agent leakage correction in paediatric brain tumours”. en. In: *Br. J. Radiol.* 92 (1094 Feb. 2019), p. 20170872. doi: 10.1259/bjr.20170872.

27 R. Vidyasagar et al. “Quantitative measurement of blood flow in paediatric brain tumours-a comparative study of dynamic susceptibility contrast and multi time-point arterial spin labelled MRI”. en. In: *Br. J. Radiol.* 89 (1062 June 2016), p. 20150624. doi: 10.1259/bjr.20150624.

28 A. Corell et al. “Shared decision-making in neurosurgery: a scoping review”. en. In: *Acta Neurochir.* 163 (9 Sept. 2021), pp. 2371–2382. doi: 10.1007/s00701-021-04867-3.

29 A. M. Robin et al. “Through the patient’s eyes: an emphasis on patient-centered values in operative decision making in the management of malignant glioma”. en. In: *J. Neurooncol.* 119 (3 Sept. 2014), pp. 473–479. doi: 10.1007/s11060-014-1492-4.

30 C. B. Terwee et al. “COSMIN methodology for assessing the content validity of Patient-Reported Outcome Measures (PROMs). User manual”. In: (2018).

31 C. Sedikides and J. D. Green. “Memory as a self-protective mechanism”. In: *Soc. Personal. Psychol. Compass* 3 (6 Dec. 2009), pp. 1055–1068. doi: 10.1111/j.1751-9004.2009.00220.x.

32 O. Adler and A. Pansky. “Chapter 7 - A “rosy view” of the past: Positive memory biases”. In: *Cognitive Biases in Health and Psychiatric Disorders*. Ed. by T. Aue and H. Okon-Singer. Academic Press, 2020, pp. 139–171. doi: 10.1016/B978-0-12-816660-4.00007-6.

33 E. Sartoretti et al. “Reduction of procedure times in routine clinical practice with Compressed SENSE magnetic resonance imaging technique”. en. In: *PLoS One* 14 (4 Apr. 12, 2019), e0214887. doi: 10.1371/journal.pone.0214887.

34 K. G. Hollingsworth. "Reducing acquisition time in clinical MRI by data undersampling and compressed sensing reconstruction". en. In: *Phys. Med. Biol.* 60 (21 Nov. 7, 2015), R297–322. doi: 10.1088/0031-9155/60/21/R297.

35 R. Mekle et al. "Combo acquisitions: balancing scan time reduction and image quality". en. In: *Magn. Reson. Med.* 55 (5 May 2006), pp. 1093–1105. doi: 10.1002/mrm.20882.

36 O. Remes et al. "A systematic review of reviews on the prevalence of anxiety disorders in adult populations". en. In: *Brain Behav.* 6 (7 July 2016), e00497. doi: 10.1002/brb3.497.

37 S. E. Braun et al. "Introducing FCR6-Brain: Measuring fear of cancer recurrence in brain tumor patients and their caregivers". en. In: *Neurooncol Pract* 9 (6 Dec. 2022), pp. 509–519. doi: 10.1093/nop/npac043.

38 A. F. Jacobson and E. H. Winslow. "Variables influencing intravenous catheter insertion difficulty and failure: an analysis of 339 intravenous catheter insertions". en. In: *Heart Lung* 34 (5 Sept. 2005), pp. 345–359. doi: 10.1016/j.hrtlng.2005.04.002.

39 A. Osmanovic-Thunström et al. "Do levels of perceived stress increase with increasing age after age 65? A population-based study". en. In: *Age Ageing* 44 (5 Sept. 2015), pp. 828–834. doi: 10.1093/ageing/afv078.

40 J. S. Abramowitz, B. J. Deacon, and S. P. H. Whiteside. *Exposure Therapy for Anxiety, Second Edition: Principles and Practice*. en. Guilford Publications, Apr. 9, 2019. 459 pp.

41 J. S. Joseph and M. J. Gray. "Exposure therapy for posttraumatic stress disorder". In: *The Journal of Behavior Analysis of Offender and Victim Treatment and Prevention* 1 (2008), pp. 69–79. doi: 10.1037/h0100457.

42 M. B. Donnellan and R. E. Lucas. "Age differences in the Big Five across the life span: evidence from two national samples". en. In: *Psychol. Aging* 23 (3 Sept. 2008), pp. 558–566. doi: 10.1037/a0012897.

43 "Guideline Safe Use of Contrast Media Part 3". In: *Radiological Society of The Netherlands.* (2022).

44 *Shared decision making*.

45 V. M. Runge. "Critical Questions Regarding Gadolinium Deposition in the Brain and Body After Injections of the Gadolinium-Based Contrast Agents, Safety, and Clinical Recommendations in Consideration of the EMA's Pharmacovigilance and Risk Assessment Committee Recommendation for Suspension of the Marketing Authorizations for 4 Linear Agents". en. In: *Invest. Radiol.* 52 (6 June 2017), pp. 317–323. doi: 10.1097/RLI.0000000000000374.

46 M. A. Oztek et al. "Practical Considerations for Radiologists in Implementing a Patient-friendly MRI Experience". en. In: *Top. Magn. Reson. Imaging* 29 (4 Aug. 2020), pp. 181–186. doi: 10.1097/RMR.0000000000000247.

47 Z. Lin et al. "Establishment of age group classification for risk stratification in glioma patients". en. In: *BMC Neurol.* 20 (1 Aug. 20, 2020), p. 310. doi: 10.1186/s12883-020-01888-w.

Supplementary Material

English version

Questionnaire PENGUIN study

1. How many MRI brain scans have you had: _____ times.
2. Would you prefer to go to the hospital for MRI checkups less often or more often?
    - Less often would be good
    - It is perfect as it is
    - I would like to have more checkups
3. Did you receive a contrast injection?
    - Yes
    - No
4. Did you find it unpleasant to receive a cannula?
    - Not at all
    - Somewhat
    - Quite a bit
    - A lot
5. Would you prefer to have an MRI scan without a contrast agent, if diagnostically equivalent?
    - Yes
    - No
6. How did you feel about the wait time between placing the IV cannula and taking the MRI?
    - Very long
    - Long
    - Not long, not short
    - Short
    - Very short
7. Did you have any symptoms during or immediately after the MRI scan (e.g., nausea)?
    - Not at all
    - Somewhat
    - Quite a bit

- A lot

8. What was your experience with the MRI scan itself?
    - Not at all unpleasant
    - A little unpleasant
    - Quite annoying
    - Very bothering
9. How did you feel about the duration of the MRI scan (lying in the scanner itself)?
    - Very long
    - Long
    - Not long, not short
    - Short
    - Very short
10. If you are stressed by the MRI, why is that?
    - Fear of outcome/bad news
    - Travel times to the MRI center
    - Fear of the scan/machine itself/noises
    - Fear of the small space/claustrophobia
    - Fear of IV cannulas
11. Do you know of any possible adverse effects on patients from gadolinium contrast agents?
    - ☐ Yes
    - ☐ No
12. Here is space for your own comments:

Dutch version

Vragenlijst PENGUIN studie

1. Vooraf: Hoeveel MRI hersenscans heeft u ruim gehad: _____ keer
2. Zou u liever minder vaak of vaker naar het ziekenhuis gaan voor MRI-controles?
    - Minder vaak zou goed zijn
    - Het is perfect zo
    - Ik zou graag meer controles willen hebben
3. Heeft u een infuus gekregen?
    - Ja
    - Nee
4. Vond u het vervelend een infuus te krijgen?
    - Helemaal niet
    - Een beetje
    - Nogal
    - Heel erg
5. Zou u liever een MRI scan zonder contrastmiddel hebben, indien het veilig is?
    - Ja
    - Nee
6. Wat vond u van de wachttijd tussen het plaatsen van de infuus en het maken van de MRI?
    - Erg lang
    - Lang
    - Niet lang, niet kort
    - Kort
    - Erg kort
7. Had u klachten tijdens of direct na de MRI scan (bijv. misselijkheid)?
    - Helemaal niet
    - Een beetje
    - Nogal
    - Heel erg
8. Wat was uw ervaring met de MRI scan zelf?

- Helemaal niet vervelend
- Een beetje vervelend
- Nogal vervelend
- Heel erg vervelend

9. Wat vond u van de duur van de MRI scan (het liggen in de scanner zelf)?
    - Erg lang
    - Lang
    - Niet lang, niet kort
    - Kort
    - Erg kort
10. Indien u stress heeft door de MRI, waarom is dat zo?
    - Angst voor uitslag/slecht nieuws
    - Reistijden naar het MRI-centrum
    - Angst voor het scannen/de machine zelf/geluiden
    - Angst voor de kleine ruimte/claustrofobie
    - Angst voor infuusprikken
11. Weet u van mogelijke nadelen voor patiënten door gadolinium contrastmiddel?
    - ☐ Ja
    - ☐ Nee
12. Hier is plek voor uw eigen commentaren:

    _____

    _____

    _____

Table S1: Tumor entities with the number of patients and latest confirmed WHO-based diagnosis/tumor grade.

Tumor entity	Grade	Number of patients
Glioblastoma	4	18
Midline glioma, H3 K27-M mutant	4	1
Oligodendroglioma	3	13
Astrocytoma	3	7
Ependymoma, RELA fusion-positive	3	1
Pineal Parenchymal Tumor of Intermediate Differentiation	3	1
Diffuse astrocytoma	2	20
Oligodendroglioma	2	17
Pleomorphic xanthoastrocytoma	2	3
Pineocytoma	2	1
Ganglioglioma	1	4
Pilocytic astrocytoma	1	3
Dysembryoplastic neuroepithelial tumor	1	2
Gemistocytic astrocytoma	1	1
Astrocytoma	1	1
Unknown (radiological low-grade gliomas)	–	7

Table S2: Statistical tests between four different groups: male vs female, HGG vs LGG, having had less than 10 scans vs more than 30 scans, and being younger than 25 years vs being older than 55 years.

Q no.	Question	Test	Male vs female (n = 100; 61 vs 39)	HGG vs LGG (n = 100; 41 vs 59)	≤10 vs ≥30 scans (n = 43; 33 vs 10)	≤25 vs ≥55 years (n = 58; 32 vs 26)
2	Would you prefer to go to the hospital for MRI checkups less often or more often?	Chi-squared	0.285	0.342	0.046	0.945
4	Did you find it unpleasant to receive a cannula?	Mann–Whitney U test	0.003*	0.479	0.854	0.744
5	Would you prefer an MRI scan without a contrast agent, if diagnostically non-inferior?	Chi-squared	0.302	0.943	0.140	0.559
6	How did you feel about the wait time between placing the cannula and taking the MRI?	Mann–Whitney U test	0.247	0.821	0.788	0.086
7	Do you believe to have had any symptoms during or immediately after the MRI scan (e.g., nausea)?	Mann–Whitney U test	0.007*	0.103	0.724	0.703
8	What was your experience with the MRI scan itself?	Mann–Whitney U test	0.022	0.895	0.229	0.091
9	How did you feel about the duration of the MRI scan (lying in the scanner itself)?	Mann–Whitney U test	0.099	0.295	0.505	0.059
10.1	Did you fear the outcome/bad news?	Chi-squared	<0.001*	0.280	0.148	0.820

Continued on next page

Table S2 – *Continued from previous page*

Q no.	Question	Test	Male vs female (n = 100; 61 vs 39)	HGG vs LGG (n = 100; 41 vs 59)	≤10 vs ≥30 scans (n = 43; 33 vs 10)	≤25 vs ≥55 years (n = 58; 32 vs 26)
10.2	Did you experience stress due to the travel times to the MRI center?	Chi-squared	0.032	0.396	1	0.582
10.3	Did you experience fear of the scan/machine itself/noises?	Chi-squared	0.306	0.733	0.320	0.163
10.4	Did you fear the small space/experience claustrophobia?	Chi-squared	0.103	0.025	0.320	0.367
10.5	Did you fear getting a cannula?	Chi-squared	0.103	1	0.040	0.681
10.6	Patients did not experience any stress during the MRI.	Chi-squared	0.003*	0.481	0.481	0.475
11	Do you know of any possible adverse effects on patients from gadolinium contrast agents?	Chi-squared	1	0.593[a] (27 vs 31)	not computable[b]	1[a] (20 vs 14)

Note.—[a] This question was added later and only filled in by 58 patients. In brackets the distribution was denoted. [b] All 44 patients answered no. The Chi-squared test could therefore not be computed. * Significance level was set at 0.01. HGG = high-grade glioma; LGG = low-grade glioma; MRI = magnetic resonance imaging; IV = intravenous.

Chapter 3

Brain tumor imaging without gadolinium based contrast agents: feasible or fantasy?

Ivar J.H.G. Wamelink, Aynur Azizova, Thomas C. Booth, Henk J.M.M. Mutsaerts, Afolabi Ogunleye, Kshitij Mankad, Jan Petr, Frederik Barkhof, Vera C. Keil

Radiology

06-02-2024; 10.1148/radiol.230793

Abstract

Gadolinium-based contrast agents (GBCAs) form the cornerstone of current primary brain tumor MRI protocols at all stages of the patient journey. Though an imperfect measure of tumor grade, GBCAs are repeatedly used for diagnosis and monitoring. In practice, however, radiologists will encounter situations where GBCA injection is not needed or of doubtful benefit. Reducing GBCA administration could improve the patient burden of (repeated) imaging (especially in vulnerable patient groups, such as children), minimize risks of putative side effects, and benefit costs, logistics, and the environmental footprint. On the basis of the current literature, imaging strategies to reduce GBCA exposure for pediatric and adult patients with primary brain tumors will be reviewed. Early postoperative MRI and fixed-interval imaging of gliomas are examples of GBCA exposure with uncertain survival benefits. Half-dose GBCAs for gliomas and T2-weighted imaging alone for meningiomas are among options to reduce GBCA use. While most imaging guidelines recommend using GBCAs at all stages of diagnosis and treatment, non–contrast-enhanced sequences, such as the arterial spin labeling, have shown a great potential. Artificial intelligence methods to generate synthetic postcontrast images from decreased-dose or non-GBCA scans have shown promise to replace GBCA-dependent approaches. This review is focused on pediatric and adult gliomas and meningiomas. Special attention is paid to the quality and real-life applicability of the reviewed literature.

Introduction

An MRI protocol for primary brain tumor imaging without contrast-enhanced (CE) sequences is generally considered insufficient for diagnostic purposes.[1] With the administration of gadolinium-based contrast agents (GBCAs), T1-weighted CE images help delineate lesion borders, differentiate entities, and evaluate therapies. The most frequently used brain perfusion technique, dynamic susceptibility contrast (DSC) MRI, also relies on GBCA injection (Fig 1).

Health concerns caused by GBCA injection (objectifiable or subjective), patient comfort, and cost stimulate the search for GBCA-reduced and often shorter MRI protocols in neuro-oncologic patients.[2] Against the backdrop of prolonged survival and longer follow-up periods, particularly in children and adults with slow-growing lower-grade brain tumors, the introduction of reduced-GBCA imaging is clinically desired.

This review covers alternative image acquisition and evaluation methods for primary brain tumor subtyping, therapy planning, and follow-up to reduce GBCA exposure in neuro-oncology (Fig 2). The accuracy of radiologists and surgeons relying only on GBCA-free (or reduced) images is examined, as well as guideline recommendations on

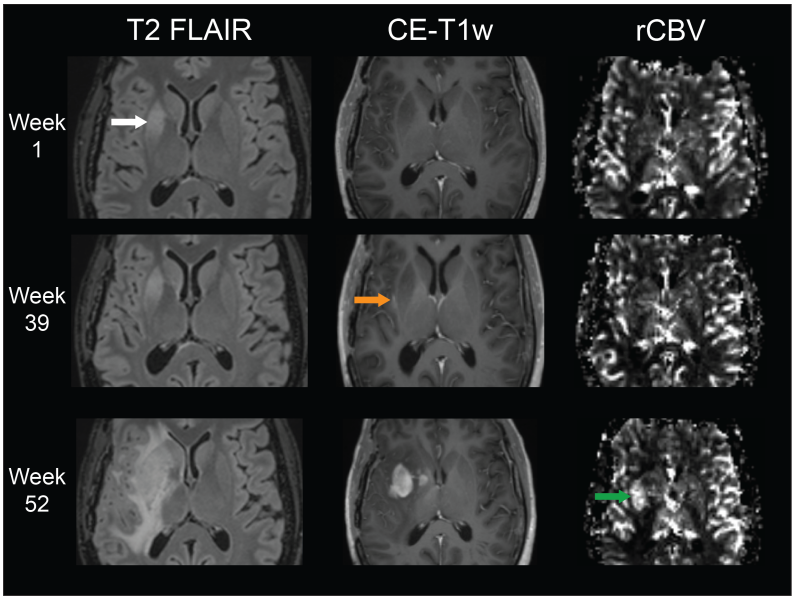


Figure 1: T2 fluid-attenuated inversion-recovery (FLAIR) images, contrast-enhanced T1-weighted (CE-T1w) images, and dynamic susceptibility contrast perfusion relative cerebral blood volume (rCBV) maps are pillars of primary brain tumor assessment. The example illustrates the independent value and behavior of all three sequences in a patient with isocitrate dehydrogenase wildtype glioblastoma after first-line therapy who presented for follow-up 3 months after the end of therapy (week 1). The right basal ganglia showed new hyperintensity (white arrow, week 1), and thus, 39 weeks before subtle contrast enhancement was observed (orange arrow, week 39) and 52 weeks before cerebral blood flow rose in the area (green arrow). Note that the brain areas affected differ between sequences.

GBCA use and imaging intervals. We also consider advanced MRI techniques probing tumor physiologic characteristics and metabolism and artificial intelligence applications, allowing diagnostic predictions from GBCA-free images and the creation of synthetic CE images. Outside the scope of this review are other imaging modalities, such as PET, and initial differential diagnostic considerations requiring GBCA administration, such as infectious or inflammatory conditions or brain metastasis. Gliomas and meningiomas are the primary focus of this review.

We aim to provide state-of-the-art information for clinicians regarding present and future opportunities and challenges of GBCA reduction in primary brain tumor imaging. For a general overview of GBCA reduction in neuroradiology, we refer to the article by Falk Delgado et al.[3]

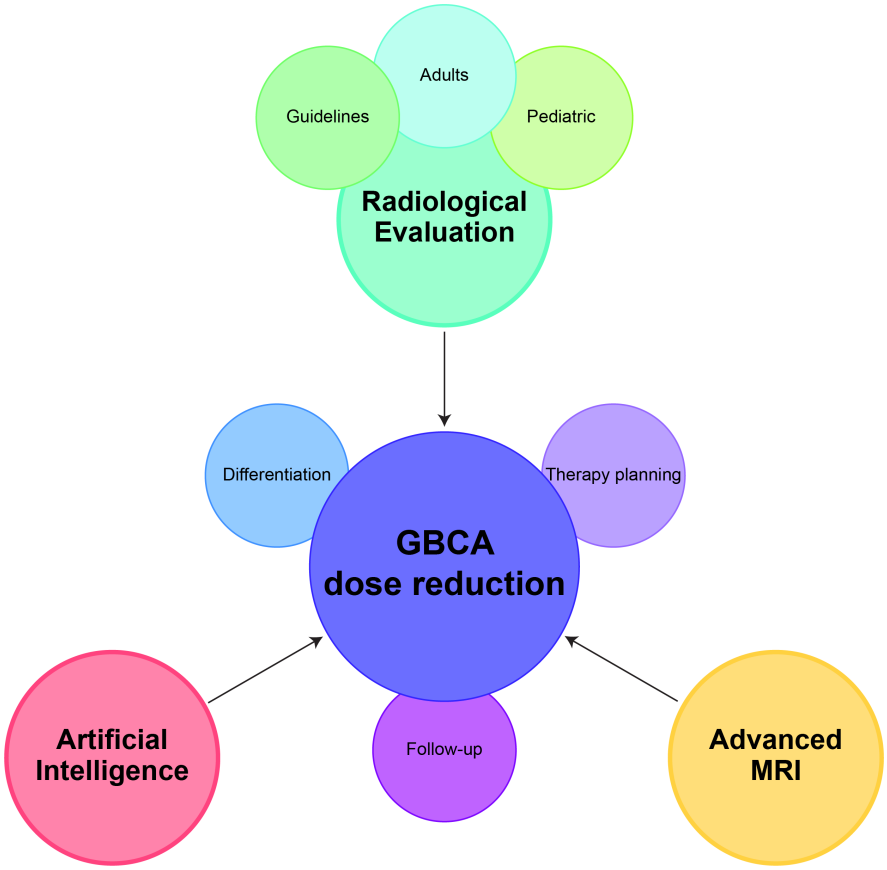


Figure 2: Diagram shows that radiologic evaluation practices (also expressed in guidelines), gadolinium-based contrast agent (GBCA)–free advanced MRI sequences, and artificial intelligence are the three pillars to help reduce GBCA use at all stages of neuro-oncologic MRI-based diagnostics.

Methods

Research Articles

This nonsystematic literature review includes studies published between January 2008 and March 2023 selected by teams of experts responsible for subtopics and who searched for original research articles and related meta-analyses and systematic reviews. Selected articles had to explicitly or implicitly present solutions to avoid or reduce GBCA use in primary brain tumor diagnostic MRI at various disease stages except for the differential diagnosis of the first MRI examination. Due to their publication date, most studies cannot adhere to the 2021 World Health Organization (WHO) classification of brain tumors.[4]

All authors were free in their search techniques, but PubMed was the recommended search tool. Potentially relevant articles were collected in a cloud online, read in full text, and evaluated for scope and inclusion criteria. Full-text reading also provided further search terms and potentially relevant articles. The most relevant articles were selected in consensus using the Standard for Reporting Diagnostic Accuracy 2015 and Checklist for Artificial Intelligence in Medical Imaging, where appropriate.

Appendix S1 elaborates on the method, with results depicted in Table S1 and Figure 3.

Guidelines

National and society guidelines on glioma and meningioma imaging were selected to allow global representation where possible, which involved inquiries for guideline use by contacting colleagues at several international organizations, such as the European Society for Magnetic Resonance in Medicine and Biology Consortium for Advancement of MRI Education and Research in Africa[5], or ESMRMB CAMERA working group, and the use of DeepL software[6] for automated translation where necessary.

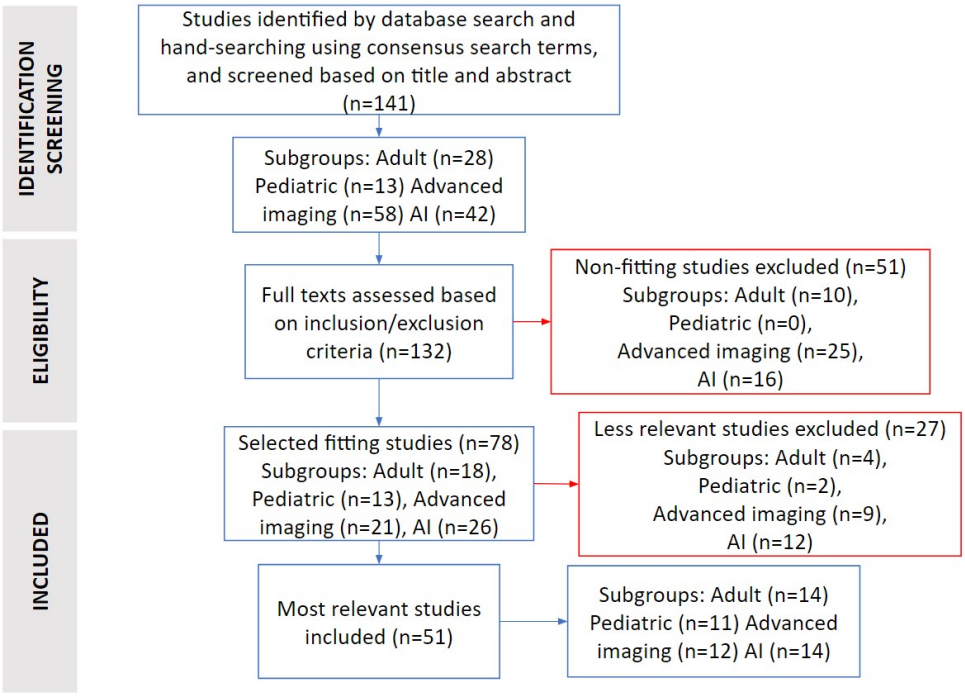


Figure 3: Flowchart visualizes the literature search conducted to shape the content of this review. Studies that did not contain a head-to-head comparison of a gadolinium-based contrast agent (GBCA)–dependent technique and a GBCA-free (or reduced) technique were considered less relevant. AI = artificial intelligence.

Results

Guidelines and Standard Radiological Evaluation

Review of Guidelines

We analyzed 14 national and eight international society guidelines regarding imaging practice in pediatric and adult patients with glioma and meningioma, aiming for representation of all global regions (Table 1). For several world regions and nations, imaging guidelines were, however, not traceable. In the analyzed guidelines, we found limited evidence supporting GBCA-free follow-up or less-frequent scanning.

Most guidelines recommend MRI with GBCA enhancement at all stages of diagnosis and treatment. Several glioma guidelines mention the lack of high-quality evidence for the optimal follow-up frequency (eg, National Institutes of Health and Care Excellence, or NICE, from the United Kingdom; guidelines of Spain; and European Association of Neuro-Oncology, or EANO). EANO guidelines suggest that longer follow-up intervals are appropriate for patients with stable low-grade glioma (LGG) (WHO grade I or II), with additional MRI examinations only for new symptoms.[1] NICE guidelines discuss possible disadvantages of frequent scanning follow-up, such as increased patient anxiety and costs. Danish guidelines suggest skipping early postoperative imaging (<48 hours) for nonenhancing gliomas due to the difficulty in evaluating nonenhancing residual tumors and instead suggest assessing the resection completeness only after 12 weeks.[7] Response Assessment in Pediatric Neuro-Oncology guidelines propose that GBCA-free follow-up imaging could be considered in nonenhancing pediatric LGG.[8]

Regarding meningiomas, EANO and Danish guidelines suggest a GBCA-free follow-up of small asymptomatic meningiomas relying on the measurements on T2-weighted images only.[9, 10] MRI intervals could also be extended to biennial in WHO grade I meningiomas after stable annual follow-up for 5 years.[9, 10]

Table 1: Guidelines Examined Regarding GBCAs

Guideline Type and No.	Country or Society Name	Reference No.	GBCA-free or GBCA-reduced Imaging or Longer Follow-up Intervals Mentioned
National guidelines			
1	Denmark	Glioma (7); meningioma (10)	Page 25 for glioma; page 3 for meningioma
2	United Kingdom (NICE)	Recommendations, brain tumors (primary) and brain metastases in over 16s (62)	None

Continued on next page

Table 1 – *Continued from previous page*

Guideline Type and No.	Country or Society Name	Reference No.	GBCA-free or GBCA-reduced Imaging or Longer Follow-up Intervals Mentioned
3	Spain	SEOM clinical guideline of diagnosis and management of low-grade glioma (63); SEOM clinical guidelines for anaplastic gliomas (2017) (64); SEOM clinical guidelines for diagnosis and treatment of glioblastoma (2017) (65)	None
4	Germany	Onkopedia: gliomas in adulthood (66)	None
5	France	ANOCEF glioblastoma repository (67)	None
6	Switzerland	A contemporary perspective on the diagnosis and treatment of diffuse gliomas in adults (68)	None
7	Australia	Optimal care pathway for people with high-grade glioma (69)	None
8	Sweden	National care program tumors in the brain and spinal cord (70)	None
9	China	Clinical practice guidelines for the management of adult diffuse gliomas (71)	None
10	India	ISNO consensus guidelines for practical adaptation of the WHO 2016 classification of adult diffuse gliomas (72)	None
11	Netherlands	Glioma guideline (73); Intracranial meningioma imaging guideline (74)	None
12	Japan	Japan Society of Clinical Oncology (75)	None
13	Russia	Association of Oncologists of Russia: Clinical recommendations for the diagnosis and treatment of patients with primary brain tumors (76)	None
14	Korea	The Overview of Practical Guidelines for Gliomas by KSNO, NCCN, and EANO (77)	None

Continued on next page

Table 1 – *Continued from previous page*

Guideline Type and No.	Country or Society Name	Reference No.	GBCA-free or GBCA-reduced Imaging or Longer Follow-up Intervals Mentioned
Society guidelines			
1	EANO	EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood (1); EANO guideline on the diagnosis and management of meningiomas (9)	Page 177 in reference 1; pages 1823 and 1825 in reference 9
2	ESNR	Glioma imaging in Europe: a survey of 220 centers and recommendations for best clinical practice (78) PMID: 29536240	None
3	RAPNO working group	Response assessment in paediatric low-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group (8); Response assessment in pediatric high-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group (79)	Page e310 in reference 8
4	ASCO and SNO	Therapy for diffuse astrocytic and oligodendroglial tumors in adults: ASCO-SNO guideline (80)	None
5	Congress of Neurological Surgeons and AANS	Guidelines in the management of CNS tumors (81)	None
6	ESMO	High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up (82)	None
7	Working Group, Royal College of Physicians	Good practice in the management of adults with malignant cerebral glioma: clinical guidelines (83)	None

Continued on next page

Table 1 – *Continued from previous page*

Guideline Type and No.	Country or Society Name	Reference No.	GBCA-free or GBCA-reduced Imaging or Longer Follow-up Intervals Mentioned
8	A joint meeting (January 30, 2014) among the FDA, NCI, clinical scientists, imaging experts, clinical trials cooperative groups, representatives from pharmaceutical and biotechnology companies, and patient advocate groups	Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials (84)	None

Note.—AANS = American Association of Neurological Surgeons, ASCO = American Society of Clinical Oncology, CNS = central nervous system, EANO = European Association of Neuro-Oncology, ESMO = European Society for Medical Oncology, ESNR = European Society of Neuroradiology, FDA = Food and Drug Administration, GBCA = gadolinium-based contrast agent, ISNO = Indian Society of Neurooncology, KSNO = Korean Society for Neuro-Oncology, NCCN = National Comprehensive Cancer Network, NCI = National Cancer Institute, NICE = National Institute for Health and Care Excellence, RAPNO = Response Assessment in Pediatric Neuro-Oncology, SEOM = Sociedad Española de Oncología Médica, SNO = Society for Neuro-Oncology, WHO = World Health Organization.

Tumor Subtype Differentiation in Adults

A GBCA dose reduction down to 50% and 75% of the suggested 0.1 mmol/kg standard dose was shown to not affect the diagnostic visibility of both gliomas and meningiomas in two prospective trials using several GBCA types (n = 141 and n = 352).[11, 12]

There need to be more studies focusing on glioma subtype differentiation without GBCAs. A 2010 study compared the diagnostic accuracy of GBCA-free imaging with standard imaging to differentiate LGG (n = 16) from high-grade glioma (HGG) (WHO grade III or IV; n = 32).[13] Areas under the receiver operating characteristic curve (AUCs) were 0.95 and 0.94, respectively. This study stressed the relevance of susceptibility-weighted imaging–based identification of microbleeds to discriminate LGG from HGG. The ratio of mean apparent diffusion coefficient of tumor to normal-appearing white matter at diffusion-weighted imaging was found to be the strongest single predictor for glioma isocitrate dehydrogenase status in another study (AUC, 0.83 compared with AUC of 0.65 for enhancement pattern; n = 290).[14] However, the enhancement pattern was valuable in a combined feature prediction model.[14] A small retrospective study (n = 29) on WHO 2016–graded mixed adult brainstem gliomas showed that apparent diffusion coefficient at first presentation was the only significant imaging marker in the prediction of survival.[15] Apparent diffusion coefficient had good sensitivity and specificity for glioma grade differentiation (81% [95% CI: 75, 86] and 87% [95% CI: 81, 91], respectively; n = 1172).[16] Results may, however, not be generalizable due to small sample sizes and class imbalance.

Meningioma WHO grading (n = 232) without GBCAs was noninferior compared with CE imaging (P = .10) in one retrospective study.[17] A meta-analysis (n = 1552) showed that apparent diffusion coefficient correlated inversely with the meningioma grade.[18]

Resection and Radiotherapy Planning in Adults

The standard of care for resection and radiation therapy planning in HGG is typically the outer CE margin, with a variably broad rim of surrounding tissue with T2 fluid-attenuated inversion-recovery (FLAIR) hyperintensity as a safety margin. While not the standard of care, there is a tendency to include nonenhancing tumor and/or edema areas at T2 FLAIR imaging, a so-called supramarginal resection. Several studies showed that FLAIR-based supramarginal resection improves survival without negatively impacting neurologic outcomes. For example, one single-center study (n = 1229) identified improved survival (20.7 vs 15.5 months; P < .001) in both treatment-naive and recurrent cases if resection involved the greater part of the surrounding FLAIR abnormalities in addition to the outer CE margins.[19] However, the retrospective design of most studies and the lack of differentiation between nonenhancing tumor and edema remain problematic.

FLAIR-guided radiation therapy planning in 174 patients with glioblastoma was shown to be feasible and safe and to lead to relatively long survival (median overall survival was 23 months in this single-arm study[20]) when compared with a previous study, which showed a median overall survival of 15 months.

Follow-up in Adults

We did not find studies comparing GBCA-free with standard CE imaging in glioma follow-up, nor studies regarding the effect of GBCA-free imaging on survival. A meta-analysis of 17 glioma studies showed that diffusion-weighted imaging could be used to distinguish recurrent tumors from therapy-related changes with a sensitivity of 82% (95% CI: 76, 87), specificity of 83% (95% CI: 76, 89), and AUC of 0.90 (95% CI: 0.87, 0.92), showing the potential of GBCA-free techniques for follow-up.[21]

There is a lack of high-quality studies investigating the patient benefits of fixed-interval imaging in glioma.[22] A retrospective study (n = 125) concluded that an early postoperative MRI examination did not impact survival.[23]

For meningioma follow-up, the noninferiority of using T2-weighted images alone was investigated in several studies. The most extensive study (n = 122) concluded that treated and untreated WHO grade I and II meningiomas can be followed up with T2-weighted imaging alone to identify tumor growth except for cavernous sinus lesions, with a preference for three-dimensional T2-weighted images to reduce measurement errors.[24]

Diagnosis and Response Assessment in Pediatric Tumors

Despite ongoing discussions on GBCAs in the diagnosis and follow-up of pediatric brain tumors, there is a paucity of high-quality studies. There are few alternative methods to improve the contrast conspicuity of lesions at diagnosis and during surveillance. Criteria for response assessment in children are not validated, and trials rely primarily on adult-based systems, such as the Response Assessment in Neuro-Oncology criteria.[25, 26]

It is the inherent nature of pediatric LGG to demonstrate waxing and waning contrast enhancement over time, with response assessment criteria including T2-weighted and FLAIR-weighted sequences.[8] In a cohort study of 88 patients with low-grade astrocytoma outside the context of neurofibromatosis type 1, Campion et al[27] showed that a change in CE characteristics was observed on only 2% of the scans. None of these resulted in a shift in management without other clinical symptoms. Dünger et al[28] studied a pediatric population of 7248 patients who had undergone a CE study for any indication (30% with suspected or known tumor). They showed that GBCAs provided additional information in 0.3% of individuals, questioning the benefit from standard GBCA use in the pediatric population.

GBCAs do, however, have a role in detecting leptomeningeal metastatic disease. CE MRI has a higher yield than cytologic evaluation of lumbar puncture in these cases (n = 17 and n = 18).[29, 30] To our knowledge, no dedicated pediatric studies regarding resection or radiation planning using GBCA-free images exist.

Follow-up in Pediatric Tumors

The highest potential for GBCA reduction is in the realm of surveillance imaging. Maloney et al[31] showed that response assessment of isolated optic pathway gliomas did not necessitate CE studies (n = 42). In a further study of 17 children with neurofibromatosis

type 1 and 21 without who were followed for about 8 years[32], the authors concluded that eight children had a change in management based on CE MRI findings, but the change in tumor size was also apparent on the other sequences. Malbari et al[33] studied 28 patients with chiasmatic-hypothalamic LGG with 683 surveillance scans. They found 67 progressions needing a management change, detectable on all GBCA-free sequences. Marsault et al[29] surveyed 17 similar patients, reporting a sensitivity of up to 88% and specificity of up to 100% regarding detection of tumor progression on unenhanced sequences. One may take these studies as indicators for the possibility of GBCA-free surveillance in indolent pediatric LGG, but small sample sizes urge caution. A 2016 study of 67 patients with pilocytic astrocytoma suggests stopping nonclinically motivated follow-up of WHO grade I pilocytic astrocytoma after gross total resection and 5 years of progression-free survival.[34]

Advanced Imaging Techniques

The past decade has brought major developments in MRI for characterizing tumor physiologic characteristics and molecular signature without GBCAs. While such techniques are often available on clinical MRI scanners, most methods await large-scale methodologic and clinical validation.[35]

ASL

Arterial spin labeling (ASL) noninvasively depicts tumor vascularization by measuring cerebral blood flow (CBF) (Fig 4) and is a potential replacement for DSC imaging. Equally high differentiation performance for HGG versus LGG (AUC, 0.9) was shown for ASL and DSC imaging (ASL relative CBF cutoff, 1.36; n = 44).[36] As in DSC imaging, more reliable results with ASL are achieved when normalizing tumor CBF to contralateral normal gray matter to reduce measurement and physiologic variability of CBF. However, optimal cutoff values differ across scanners and ASL implementations, and multicenter validation studies are needed. An alternative noninvasive tumor vascularization assessment is capillary microcirculation measurement through intravoxel incoherent motion modeling of diffusion-weighted imaging. It was a better predictor (AUC, 0.81) of isocitrate dehydrogenase status than the dynamic CE imaging–derived parameter volume transfer constant, or Ktrans (n = 30; AUC, 0.773.[37]

Direct comparisons of GBCA-based and GBCA-free methods in pediatric populations are scarce. However, ASL reached a similar performance (100% sensitivity, 95.5% specificity for a cutoff of 0.82) in differentiating high- from low-grade astrocytomas (n = 37) at DSC imaging.[38]

ASL is also a potential noninvasive alternative to DSC imaging for therapy monitoring. It is comparable in performance to DSC in differentiating glioma recurrence from radiation effects by lower CBF (n = 69; ASL: cutoff, 1.86 and accuracy, 79.7%; DSC accuracy, 82.6%)[39], separating recurrence and pseudoprogression (n = 116; ASL AUC, 0.72; DSC AUC, 0.87).[40] ASL may also help identify pseudoprogression at a lower field strength (ie, 1.5 T) (n = 26; ASL accuracy, 69%; MR spectroscopy accuracy, 74%; dynamic CE MRI accuracy, 69%; and DSC accuracy, 79%).[41]

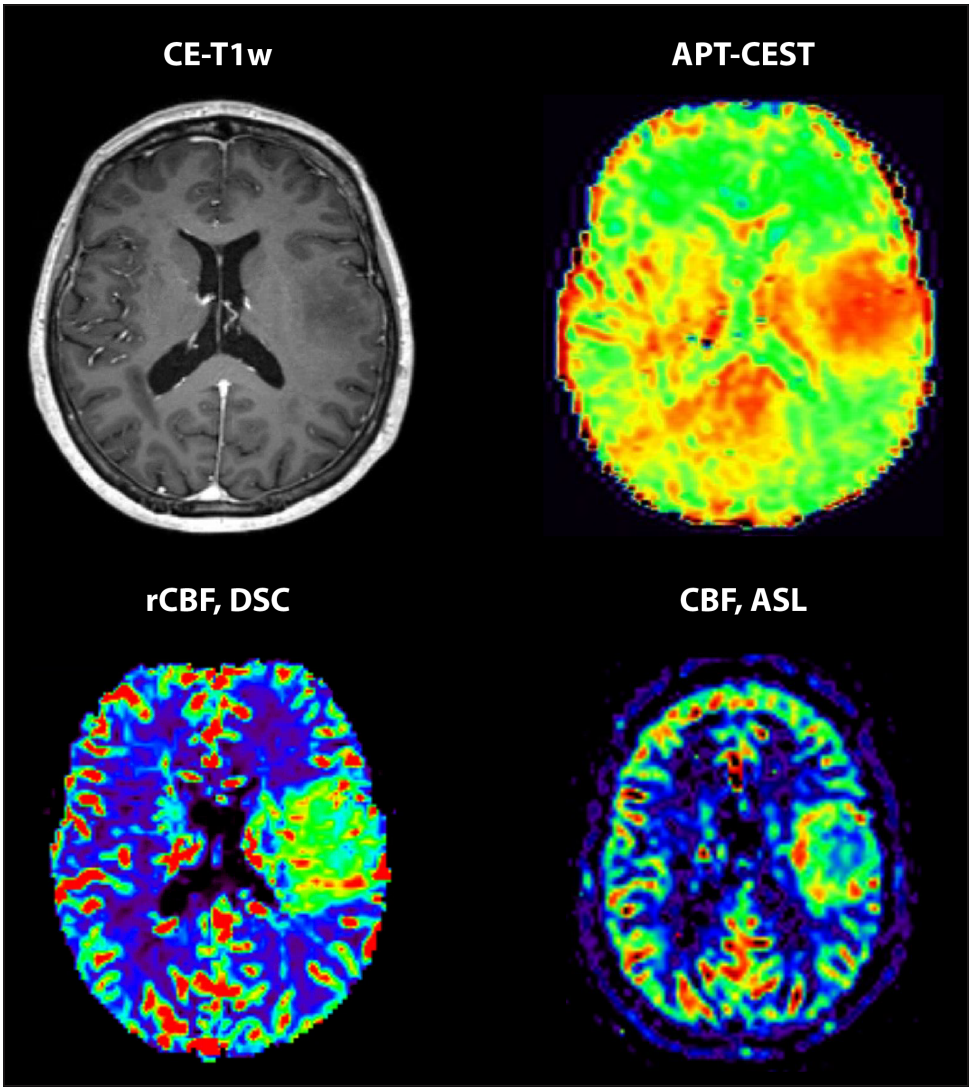


Figure 4: Contrast-enhanced T1-weighted (CE-T1w), amide-proton transfer chemical exchange saturation transfer (APT-CEST), dynamic susceptibility contrast (DSC), and arterial spin labeling (ASL) images in a nonenhancing left frontoinsular glioblastoma (isocitrate dehydrogenase wildtype). At the location of the tumor, there is an increase in amide content (red on the APT-CEST scan) even beyond the area of T1 hypointensity. There is good concordance between the DSC relative cerebral blood flow (rCBF) map and the ASL cerebral blood flow (CBF) map, both showing hyperperfusion (green and red at the location of the lesion) of the lesion as a marker of malignancy.

Molecular Imaging: Amide-Proton Transfer Chemical Exchange Saturation Transfer and MR Spectroscopy

Among the best-studied methods to image tumor molecular properties is amide-proton transfer chemical exchange saturation transfer (APT-CEST) (Fig 4, a novel sequence that is sensitive to proteins and peptides with exchangeable protons. There are far fewer studies using APT-CEST than ASL. Initial results suggest that APT-CEST (AUC, 0.911) can outperform ASL (AUC, 0.852) in identifying pseudoprogression (n = 48).[42] APT-CEST (cutoff, 1.53; AUC, 0.877) and DSC (AUC, 0.927) did not only perform comparably well in classifying HGG versus LGG in patients with glioma (n = 46)[43], but APT-CEST was able (cutoff, 2.56; AUC, 0.886) to help grade tumors among gliomas without intense contrast enhancement (n = 34).[44]

MR spectroscopy, which can be used to evaluate choline, N-acetylaspartate, and 2-hydroxyglutarate levels in tissue, is the most commonly studied advanced metabolic imaging modality in brain tumors. A meta-analysis of proton (1H) MR spectroscopy in 1228 patients showed that HGG can be differentiated from LGG with use of the choline-to–N-acetylaspartate ratio (AUC, 0.87).[45] While its performance is relatively stable, studies use different thresholds, metabolites, or metabolite ratios, complicating 1H MR spectroscopy harmonization for tumor imaging.[45] A meta-analysis of 460 patients with 2-hydroxyglutarate MR spectroscopy acquisition demonstrated excellent pooled sensitivity and specificity in tumor grading (95% and 91%, respectively) and in isocitrate dehydrogenase status identification (75% and 94%).[46]

Pilot studies have identified other promising GBCA-free techniques, such as vessel architecture imaging, diffusion kurtosis imaging, relaxometry and fingerprinting, and MR elastography, for preoperative and follow-up glioma imaging.[35, 47]

Artificial Intelligence Approaches

Quantitative image information (features) can be extracted to develop biomarkers invisible to the naked eye. Artificial intelligence, specifically machine learning, has advanced analysis capability by leveraging predesigned image features (radiomics) or by automating feature creation (deep learning). The latter allows advanced image segmentation and generation used in studies with GBCA reduction.

Synthetic CE Imaging

Synthetic GBCA-based images can be generated using deep learning with GBCA-free or GBCA-reduced image inputs. Currently, to our knowledge, no studies use advanced MRI sequences as inputs. Of the dozen proof-of-concept publications in this nascent area of neuro-oncology, we observe that generative adversarial networks (especially cycle generative adversarial networks with double the architectures) appear to have superior performance accuracy compared with U-Nets. Synthetic CE T1-weighted images are often the output, with a notable exception in one study where precontrast FLAIR yielded CE FLAIR images, which may be more sensitive to small lesions than T1-weighted images.[48] A large multicenter GBCA replacement study has demonstrated generalizability by location and scanner type, showing only a 7% CE volume mismatch

between image pairs. Thus, at the group level, treatment response assessment was feasible without a GBCA.[49] Some researchers used images with only 10%–25% of the standard GBCA dose as input for the deep learning models and were also successful in recovering full-quality CE images, as discussed in detail elsewhere.[50, 51] The different relaxivities of GBCAs are an unaddressed research topic in this context.

Radiological Biomarkers

Imaging biomarker studies performed head-to-head comparisons of machine learning accuracy by using features derived from GBCA-free imaging compared with features derived from CE imaging. However, reducing GBCA use was rarely the primary objective. Instead, an iterative discovery process is typically used to develop an optimal biomarker by means of training available data that may consist of a single sequence or a combination of sequences, with or without clinical information. The focus then shifts to validating the model on suitable testing data. In the pediatric setting, one study examined the added value of CE T1-weighted images in the differentiation of pediatric posterior fossa tumor types, including astrocytoma, by manually segmenting tumors, applying a range of radiomic features, and classifying the different groups with a support vector machine (n = 136).[52] The authors showed that CE T1-weighted images have no added value, as astrocytoma is optimally differentiated in combined models of GBCA-free sequences (AUC, 0.955 [95% CI: 0.810, 0.997]).

Some proof-of-concept studies evaluated the nonenhancing peritumoral and CE tumoral regions for a head-to-head analysis of features derived from CE and GBCA-free images. For example, one study applied PyRadiomics (open-source standardized radiomic features) to the segmented regions on images in 285 patients with glioma. It was noninferior (sensitivity and specificity >85% GBCA-free) in distinguishing LGG from HGG.[53] Another comparative study (n = 46) found a high correlation between ASL and DSC PyRadiomics features (Spearman rho or Pearson r > 0.7 for more than half of the 75 tested features). Both methods performed well in distinguishing LGG and HGG (AUC >0.89, P = .133 between methods).[54] A study using Qmazda features (another open-source radiomic repository) for the same glioma grading task found that performance accuracy using FLAIR images was noninferior to CE T1-weighted images (n = 181; AUCs >0.86).[55] In this and other similar diagnostic biomarker machine learning studies, combinations of features derived from different sequences performed better than those derived from single input sequences without GBCAs. Further analytical and clinical validation, especially in prospective multicenter studies, is needed before machine learning models can be considered clinic-ready tools.[56, 57]

While, to our knowledge, there are no therapy planning studies to reduce GBCA use with artificial intelligence, some studies allow a head-to-head subgroup comparative analysis. For example, a large study in patients with meningioma (n = 1728) to determine brain invasion, which influences operative treatment, showed similar discriminative performance (AUC >0.7) when T2-weighted and CE T1-weighted images were used separately for PyRadiomics feature extraction and support vector machine classification.[56] In a multicenter study (n = 496), deep learning (specifically residual convolutional neural networks) classified gliomas by isocitrate dehydrogenase status, with similar accuracy whether

contrast material had been used or not in T1-weighted images (AUCs, 0.92 and 0.86, respectively; no significance level reported).[58] Both examples show that features derived from combinations of sequences allow higher classification performance accuracy than when derived from a single GBCA-free sequence.

One machine learning study performed a head-to-head comparison of T1-weighted sequences with and without GBCAs to differentiate progression and pseudoprogression in HGG.[59] There was a significant performance loss without GBCAs, using PyRadiomics features and a generalized boosted regression model (n = 124; AUC, 0.82 vs 0.65). The authors concluded that GBCAs could not be omitted.[59] However, no comparison was drawn with other GBCA-free sequences, and it is known that T2-weighted sequences, diffusion-tensor imaging, ASL, and MR spectroscopy may be discriminative in this scenario.[60] Another study using deep learning to distinguish progression and pseudoprogression found that FLAIR and diffusion-weighted imaging combined outperformed the CE T1 model (n = 55; AUCs, 0.82 and 0.57, respectively).[61]

Conclusion

Despite the paucity of prospective high-quality studies, advocating for decreased gadolinium-based contrast agent (GBCA) use based on the current evidence could potentially be considered for (a) GBCA dose reduction per brain scan, (b) abandonment of early postoperative MRI in patients with nonenhancing tumor, (c) clinically driven instead of fixed-interval follow-up imaging, and (d) complete GBCA omission in convexity meningioma follow-up. Monitoring, but not initial diagnostics, of low-grade pediatric gliomas without GBCA-enhanced sequences can be considered in selected cases. For adult glioma, evidence is lacking that monitoring without GBCA is sufficient. High-quality and, in particular, prospective multicenter studies can address this need. Advanced imaging techniques and emerging artificial intelligence solutions will likely challenge the GBCA dependence of neuro-oncologic imaging in the near future. GBCA-free arterial spin labeling–based perfusion imaging is an acceptable alternative already today. In the end, contrast-enhanced sequences are an imperfect surrogate for tumor grading that can and should be challenged to benefit patients, reduce costs, enhance safety, and protect the environment.

References

1 M. Weller et al. “EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood”. en. In: *Nat. Rev. Clin. Oncol.* 18 (3 Mar. 2021), pp. 170–186. doi: 10.1038/s41571-020-00447-z.

2 W. A. Mehan Jr et al. “Optimal brain MRI protocol for new neurological complaint”. en. In: *PLoS One* 9 (10 Oct. 24, 2014), e110803. doi: 10.1371/journal.pone.0110803.

3 A. Falk Delgado et al. “Diagnostic value of alternative techniques to gadolinium-based contrast agents in MR neuroimaging-a comprehensive overview”. en. In: *Insights Imaging* 10 (1 Aug. 23, 2019), p. 84. doi: 10.1186/s13244-019-0771-1.

4 D. N. Louis et al. “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary”. en. In: *Neuro. Oncol.* 23 (8 Aug. 2, 2021), pp. 1231–1251. doi: 10.1093/neuonc/noab106.

5 *Consortium for Advancement of MRI Education and Research in Africa*.

6 *DeepL translator*.

7 *Gliomer hos voksne*. Version 3. July 25, 2022.

8 J. Fangusaro et al. “Response assessment in paediatric low-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group”. en. In: *Lancet Oncol.* 21 (6 June 2020), e305–e316. doi: 10.1016/S1470-2045(20)30064-4.

9 R. Goldbrunner et al. “EANO guideline on the diagnosis and management of meningiomas”. en. In: *Neuro. Oncol.* 23 (11 Nov. 2, 2021), pp. 1821–1834. doi: 10.1093/neuonc/noab150.

10 *Meningeomer*. Version 2.0. Dec. 7, 2021.

11 B. P. Liu et al. “Clinical Efficacy of Reduced-Dose Gadobutrol Versus Standard-Dose Gadoterate for Contrast-Enhanced MRI of the CNS: An International Multicenter Prospective Crossover Trial (LEADER-75)”. en. In: *AJR Am. J. Roentgenol.* 217 (5 Nov. 2021), pp. 1195–1205. doi: 10.2214/AJR.21.25924.

12 M. C. DeLano et al. “Dose-Lowering in Contrast-Enhanced MRI of the Central Nervous System: A Retrospective, Parallel-Group Comparison Using Gadobenate Dimeglumine”. en. In: *J. Magn. Reson. Imaging* 54 (5 Nov. 2021), pp. 1660–1675. doi: 10.1002/jmri.27731.

13 S. M. Park et al. “Combination of high-resolution susceptibility-weighted imaging and the apparent diffusion coefficient: added value to brain tumour imaging and clinical feasibility of non-contrast MRI at 3 T”. en. In: *Br. J. Radiol.* 83 (990 June 2010), pp. 466–475. doi: 10.1259/bjr/34304111.

14 J. Maynard et al. “World Health Organization Grade II/III Glioma Molecular Status: Prediction by MRI Morphologic Features and Apparent Diffusion Coefficient”. en. In: *Radiology* 296 (1 July 2020), pp. 111–121. doi: 10.1148/radiol.2020191832.

15 A. Leibetseder et al. “Prognostic factors in adult brainstem glioma: a tertiary care center analysis and review of the literature”. en. In: *J. Neurol.* 269 (3 Mar. 2022), pp. 1574–1590. doi: 10.1007/s00415-021-10725-0.

16 Q.-P. Wang et al. “Accuracy of ADC derived from DWI for differentiating high-grade from low-grade gliomas: Systematic review and meta-analysis”. en. In: *Medicine* 99 (8 Feb. 2020), e19254. doi: 10.1097/MD.0000000000019254.

17 Y. Yao et al. “Predicting the grade of meningiomas by clinical-radiological features: A comparison of precontrast and postcontrast MRI”. en. In: *Front. Oncol.* 12 (Dec. 1, 2022), p. 1053089. doi: 10.3389/fonc.2022.1053089.

18 Y.-T. Tsai et al. “Preoperative Apparent Diffusion Coefficient Values for Differentiation between Low and High Grade Meningiomas: An Updated Systematic Review and Meta-Analysis”. en. In: *Diagnostics (Basel)* 12 (3 Mar. 4, 2022). doi: 10.3390/diagnostics12030630.

19 Y. M. Li et al. “The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection?” en. In: *J. Neurosurg.* 124 (4 Apr. 2016), pp. 977–988. doi: 10.3171/2015.5.JNS142087.

20 C. M. Duma et al. “Upfront boost Gamma Knife "leading-edge" radiosurgery to FLAIR MRI-defined tumor migration pathways in 174 patients with glioblastoma multiforme: a 15-year assessment of a novel therapy”. en. In: *J. Neurosurg.* 125 (Suppl 1 Dec. 2016), pp. 40–49. doi: 10.3171/2016.7.GKS161460.

21 X. Du et al. “Diagnostic accuracy of diffusion-weighted imaging in differentiating glioma recurrence from posttreatment-related changes: a meta-analysis”. en. In: *Expert Rev. Anticancer Ther.* 22 (1 Jan. 2022), pp. 123–130. doi: 10.1080/14737140.2022.2000396.

22 G. Thompson et al. “Interval brain imaging for adults with cerebral glioma”. en. In: *Cochrane Database Syst. Rev.* 12 (Dec. 24, 2019), p. CD013137. doi: 10.1002/14651858.CD013137.pub2.

23 O. D. Mrowczynski et al. “Utility of Early Postoperative Magnetic Resonance Imaging After Glioblastoma Resection: Implications on Patient Survival”. en. In: *World Neurosurg.* 120 (Dec. 2018), e1171–e1174. doi: 10.1016/j.wneu.2018.09.027.

24 J. Boto et al. “Is Contrast Medium Really Needed for Follow-up MRI of Untreated Intracranial Meningiomas?” en. In: *AJNR Am. J. Neuroradiol.* 42 (8 Aug. 2021), pp. 1421–1428. doi: 10.3174/ajnr.A7170.

25 B. Tamrazi et al. “Current concepts and challenges in the radiologic assessment of brain tumors in children: part 2”. en. In: *Pediatr. Radiol.* 48 (13 Dec. 2018), pp. 1844–1860. doi: 10.1007/s00247-018-4232-7.

26 F. D’Arco et al. “Current concepts in radiologic assessment of pediatric brain tumors during treatment, part 1”. en. In: *Pediatr. Radiol.* 48 (13 Dec. 2018), pp. 1833–1843. doi: 10.1007/s00247-018-4194-9.

27 T. Campion et al. “Surveillance imaging of grade 1 astrocytomas in children: can duration and frequency of follow-up imaging and the use of contrast agents be reduced?” en. In: *Neuroradiology* 63 (6 June 2021), pp. 953–958. doi: 10.1007/s00234-020-02609-3.

28 D. Dünger et al. “Do we need gadolinium-based contrast medium for brain magnetic resonance imaging in children?” en. In: *Pediatr. Radiol.* 48 (6 June 2018), pp. 858–864. doi: 10.1007/s00247-017-3999-2.

29 P. Marsault et al. “Diagnostic performance of an unenhanced MRI exam for tumor follow-up of the optic pathway gliomas in children”. en. In: *Neuroradiology* 61 (6 June 2019), pp. 711–720. doi: 10.1007/s00234-019-02198-w.

30 L. Porto et al. “Leptomeningeal metastases in pediatrics: magnetic resonance image manifestations and correlation with cerebral spinal fluid cytology”. en. In: *Pediatr. Int.* 52 (4 Aug. 2010), pp. 541–546. doi: 10.1111/j.1442-200X.2010.03171.x.

31 E. Maloney et al. “Non-inferiority of a non-gadolinium-enhanced magnetic resonance imaging follow-up protocol for isolated optic pathway gliomas”. en. In: *Pediatr. Radiol.* 52 (3 Mar. 2022), pp. 539–548. doi: 10.1007/s00247-021-05226-1.

32 E. Maloney et al. “Surveillance magnetic resonance imaging for isolated optic pathway gliomas: is gadolinium necessary?” en. In: *Pediatr. Radiol.* 48 (10 Sept. 2018), pp. 1472–1484. doi: 10.1007/s00247-018-4154-4.

33 F. Malbari et al. “Gadolinium is not necessary for surveillance MR imaging in children with chiasmatic-hypothalamic low-grade glioma”. en. In: *Pediatr. Blood Cancer* 68 (10 Oct. 2021), e29178. doi: 10.1002/pbc.29178.

34 A. J. Dodgshun et al. “Low rates of recurrence and slow progression of pediatric pilocytic astrocytoma after gross-total resection: justification for reducing surveillance imaging”. en. In: *J. Neurosurg. Pediatr.* 17 (5 May 2016), pp. 569–572. doi: 10.3171/2015.9.PEDS15449.

35 L. Hirschler et al. “Advanced MR Techniques for Preoperative Glioma Characterization: Part 1”. en. In: *J. Magn. Reson. Imaging* (Mar. 3, 2023). doi: 10.1002/jmri.28662.

36 Y. Qu et al. “Perfusion measurement in brain gliomas using velocity-selective arterial spin labeling: comparison with pseudo-continuous arterial spin labeling and dynamic susceptibility contrast MRI”. en. In: *Eur. Radiol.* 32 (5 May 2022), pp. 2976–2987. doi: 10.1007/s00330-021-08406-7.

37 X. Wang et al. “Simplified perfusion fraction from diffusion-weighted imaging in preoperative prediction of IDH1 mutation in WHO grade II-III gliomas: comparison with dynamic contrast-enhanced and intravoxel incoherent motion MRI”. en. In: *Radiol. Oncol.* 54 (3 June 19, 2020), pp. 301–310. doi: 10.2478/raon-2020-0037.

38 G. Morana et al. “Pediatric astrocytic tumor grading: comparison between arterial spin labeling and dynamic susceptibility contrast MRI perfusion”. en. In: *Neuroradiology* 60 (4 Apr. 2018), pp. 437–446. doi: 10.1007/s00234-018-1992-6.

39 Y.-L. Wang et al. “Differentiation between radiation-induced brain injury and glioma recurrence using 3D pCASL and dynamic susceptibility contrast-enhanced perfusion-weighted imaging”. en. In: *Radiother. Oncol.* 129 (1 Oct. 2018), pp. 68–74. doi: 10.1016/j.radonc.2018.01.009.

40 A. Lavrova et al. “Diagnostic Accuracy of Arterial Spin Labeling in Comparison With Dynamic Susceptibility Contrast-Enhanced Perfusion for Brain Tumor Surveillance at 3T MRI”. en. In: *Front. Oncol.* 12 (May 20, 2022), p. 849657. doi: 10.3389/fonc.2022.849657.

41 A. Seeger et al. “Comparison of three different MR perfusion techniques and MR spectroscopy for multiparametric assessment in distinguishing recurrent high-grade gliomas from stable disease”. en. In: *Acad. Radiol.* 20 (12 Dec. 2013), pp. 1557–1565. doi: 10.1016/j.acra.2013.09.003.

42 H. Hou et al. “Differentiation of true progression from treatment response in high-grade glioma treated with chemoradiation: a comparison study of 3D-APTW and 3D-PcASL imaging and DWI”. en. In: *NMR Biomed.* 36 (1 Jan. 2023), e4821. doi: 10.1002/nbm.4821.

43 Y. S. Choi et al. “Amide proton transfer imaging to discriminate between low- and high-grade gliomas: added value to apparent diffusion coefficient and relative cerebral blood volume”. en. In: *Eur. Radiol.* 27 (8 Aug. 2017), pp. 3181–3189. doi: 10.1007/s00330-017-4732-0.

44 O. Togao et al. “Grading diffuse gliomas without intense contrast enhancement by amide proton transfer MR imaging: comparisons with diffusion- and perfusion-weighted imaging”. en. In: *Eur. Radiol.* 27 (2 Feb. 2017), pp. 578–588. doi: 10.1007/s00330-016-4328-0.

45 Q. Wang et al. “The diagnostic performance of magnetic resonance spectroscopy in differentiating high-from low-grade gliomas: A systematic review and meta-analysis”. en. In: *Eur. Radiol.* 26 (8 Aug. 2016), pp. 2670–2684. doi: 10.1007/s00330-015-4046-z.

46 C. H. Suh et al. “2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and meta-analysis using individual patient data”. en. In: *Neuro. Oncol.* 20 (12 Nov. 12, 2018), pp. 1573–1583. doi: 10.1093/neuonc/noy113.

47 O. M. Henriksen et al. “High-grade glioma treatment response monitoring biomarkers: A position statement on the evidence supporting the use of advanced MRI techniques in the clinic, and the latest bench-to-bedside developments. Part 1: Perfusion and diffusion techniques”. en. In: *Front. Oncol.* 12 (Mar. 3, 2022), p. 810263. doi: 10.3389/fonc.2022.810263.

48 Y. Wang et al. “Deep learning-based 3D MRI contrast-enhanced synthesis from a 2D noncontrast T2Flair sequence”. en. In: *Med. Phys.* 49 (7 July 2022), pp. 4478–4493. doi: 10.1002/mp.15636.

49 C. Jayachandran Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *Lancet Digit Health* 3 (12 Dec. 2021), e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

50 E. Moya-Sáez, R. de Luis-García, and C. Alberola-López. “Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI”. In: *Frontiers in Neuroimaging* 2 (2023). doi: 10.3389/fnimg.2023.1055463.

51 L. Pasquini et al. “Synthetic Post-Contrast Imaging through Artificial Intelligence: Clinical Applications of Virtual and Augmented Contrast Media”. en. In: *Pharmaceutics* 14 (11 Nov. 4, 2022). doi: 10.3390/pharmaceutics14112378.

52 J. Dong et al. “Differentiation of paediatric posterior fossa tumours by the multiregional and multiparametric MRI radiomics approach: a study on the selection of optimal multiple sequences and multiregions”. en. In: *Br. J. Radiol.* 95 (1129 Jan. 1, 2022), p. 20201302. doi: 10.1259/bjr.20201302.

53 J. Cheng et al. “Prediction of Glioma Grade Using Intratumoral and Peritumoral Radiomic Features From Multiparametric MRI Images”. en. In: *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19 (2 Apr. 1, 2022), pp. 1084–1095. doi: 10.1109/TCBB.2020.3033538.

54 T. Hashido, S. Saito, and T. Ishida. “A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas”. In: *Sci. Rep.* 10 (1 Apr. 9, 2020), p. 6121. doi: 10.1038/s41598-020-62658-9.

55 D. Alis et al. “The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas”. en. In: *Clin. Radiol.* 75 (5 May 2020), pp. 351–357. doi: 10.1016/j.crad.2019.12.008.

56 J. Zhang et al. “A radiomics model for preoperative prediction of brain invasion in meningioma non-invasively based on MRI: A multicentre study”. en. In: *EBioMedicine* 58 (Aug. 2020), p. 102933. doi: 10.1016/j.ebiom.2020.102933.

57 D. N. Cagney et al. “The FDA NIH Biomarkers, EndpointS, and other Tools (BEST) resource in neuro-oncology”. en. In: *Neuro. Oncol.* 20 (9 Aug. 2, 2018), pp. 1162–1172. doi: 10.1093/neuonc/nox242.

58 K. Chang et al. “Residual Convolutional Neural Network for the Determination of IDH Status in Low- and High-Grade Gliomas from MR Imaging”. en. In: *Clin. Cancer Res.* 24 (5 Mar. 1, 2018), pp. 1073–1081. doi: 10.1158/1078-0432.CCR-17-2236.

59 O. Mammadov et al. “Radiomics for pseudoprogression prediction in high grade gliomas: added value of MR contrast agent”. en. In: *Heliyon* 8 (8 Aug. 2022), e10023. doi: 10.1016/j.heliyon.2022.e10023.

60 T. C. Booth et al. “Imaging Biomarkers of Glioblastoma Treatment Response: A Systematic Review and Meta-Analysis of Recent Machine Learning Studies”. en. In: *Front. Oncol.* 12 (Jan. 31, 2022), p. 799662. doi: 10.3389/fonc.2022.799662.

61 S. Bacchi et al. “Deep learning in the detection of high-grade glioma recurrence using multiple MRI sequences: A pilot study”. en. In: *J. Clin. Neurosci.* 70 (Dec. 2019), pp. 11–13. doi: 10.1016/j.jocn.2019.10.003.

Supplementary Material

Literature scope

Importantly, this study was not set out as a systematic review.

This review comprises:

Original articles and related meta-analyses/systematic reviews that explicitly or implicitly ("de facto") aim to investigate solutions to avoid or reduce GBCA use in primary brain tumor diagnostic MRI at all disease stages except for the differential diagnosis of the first MRI.

Scope dimensions are:

1. Imaging interval extension
2. Intermitting GBCA-free MRI
3. Reduced dose of GBCA per scan
4. AI approaches to avoid GBCA injection or reduce the dose
5. Alternative techniques (like CEST or ASL) to replace GBCA-dependent techniques

Further, guidelines on brain tumor imaging are examined.

Inclusion and exclusion criteria

To keep the focus on more frequent primary brain tumors, and provide a literature selection with a higher level of evidence for the clinical reader, we restricted the number of entities portrayed in this review. This includes all articles differentiating lesion types except for molecular subtyping and WHO grading of single entities. Since metastases have a higher tendency to develop leptomeningeal metastases and very small lesions, we excluded secondary brain tumors as being GBCA-dependent entities.

Inclusion criteria:

- meta-analyses
- systematic reviews
- peer-reviewed and pubmed-listed
- English-language
- prospective and retrospective research
- brain MRI
- adults and children
- both sexes
- intra-axial primary brain tumor types:
  - glioma
  - meningioma (adults)
- article has a clinical research question covering at least 1 of these:
  - brain tumor subtype differentiation
  - therapy planning

  - survival prediction
  - therapy response evaluation/follow-up
- experimental set-up directly compares standard diagnostic imaging with GBCA with a technique resulting in less OR no GBCA (either by dosereduction/ omission OR imaging interval extension) **OR**
- experimental setup evaluates a technique resulting in less OR no GBCA (either by dose-reduction/omission OR imaging interval extension)

Exclusion criteria:

- research without a clearly described purpose of OR de facto result in the reduction/omission of GBCA use
- research presenting mainly or exclusively data from lesions other than the above-mentioned entities
- research using another invasive technique as the alternative (PET, USPIO, etc.) to GBCA injection
- non-MRI studies
- studies published before 2008 (15 years)
- animal or phantom research, including the use of synthetic source images
- case series, case reports, commentaries, editorials, position statements, pictorial essays

Search technique and final selection

All authors were free in their search techniques for appropriate literature sources. It was, however, advised to use pubmed.gov as a primary search tool. All topics were distributed among the authors to avoid redundant searches, and a cloud-based file with already applied search terms was maintained.
Potentially relevant articles (according to title and abstract) were collected in a cloud online, read in full text, and evaluated for scope and inclusion criteria. Full-text reading also provided further search terms and further potentially suitable articles.
Since the number of references for this review is limited, the most relevant articles were selected in consensus using STARD 2015 checklist (PMID: 26511519; `https://www.equator-network.org/reporting-guidelines/stard/`). For AI articles, quality criteria as presented by the Radiology Editorial Board article “Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers” were chosen to find potentially relevant articles (PMID: 31891322). Subsequently, we used the CLAIM checklist to select the most relevant articles from this subset (PMID: 33937821).

Table S1: Overview of the References

Guidelines					
Reference (PMID)	**Year**	**Tumor type(s)**	**Purpose**	**Sequences**	**Reference**
DNOG	2021/2022	meningioma/glioma	follow-up	National	7, 10
EANO (33293629, 34181733)	2021/2020	meningioma/glioma	follow-up	Society	1, 9
RAPNO (32502457)	2020	glioma	follow-up	Society	8
Adults					
Reference (PMID)	**Year**	**Tumor type(s)**	**Purpose**	**Sequences**	**Reference**
Liu et al (34133205)*	2021	Glioma, meningioma, lymphoma, metastasis, cerebellopontine angle tumors, choroid plexus tumors	visualization	T2w, T1w, FLAIR, and CE-T1w	11
DeLano et al (34018290)	2021	Intra-axial (such as meningioma) and extra-axial tumors (such as metastasis)	visualization	T2w, T1w, FLAIR, and CE-T1w	12
Park et al (19690076)*	2010	glioma, neurocytoma, lymphoma, metastasis	differentiation	T2w, T1w, T2*W, DWI, HR-SWI, DSC MRI, and CE-T1w	13
Yao et al (36530973)	2022	meningioma	differentiation	T2w, T1w, FLAIR, DWI, and CE-T1w	17
Li et al (26495941)	2016	glioblastoma	therapy planning	FLAIR and CE-T1w	19

Continued on next page

Table S1 – *Continued from previous page*

Reference (PMID)	Year	Tumor type(s)	Purpose	Sequences	Reference
Duma et al (27903197)	2016	glioblastoma	therapy planning	FLAIR and MRI-SPECT (in some cases)	20
Mrowczynski et al (30218799)	2018	glioblastoma	follow-up	not indicated	23
Boto et al (34117017)	2021	meningioma	follow-up	T2w and CE-T1w	24
		Pediatrics			
Reference (PMID)	**Year**	**Tumor type(s)**	**Purpose**	**Sequences**	**Reference**
Campion et al (33241451)	2021	Grade 1 astrocytoma	follow-up	T2w, T1w, FLAIR, DWI, SWI, and CE-T1w	27
Dünger et al (29623352)	2018	not specified	diagnosis	T2w, T1w, FLAIR, DWI, and CE-T1w	28
Marsault et al (30904949)	2019	Optic pathway glioma	follow-up	T2w, T1w, FLAIR, and CE-T1w	29
Maloney et al (34751813)	2022	Optic pathway glioma	follow-up	at least T1w and CE-T1w	31
Maloney et al (29789890)	2018	Optic pathway glioma	follow-up	T2w, FLAIR, and CE-T1w	32
Malbari et al (34133064)	2021	Chiasmatic-hypothalamic low-grade glioma	follow-up	T2w, T1w, FLAIR, and CE-T1w	33
		Advanced MRI			
Reference (PMID)	**Year**	**Tumor type(s)**	**Purpose**	**Sequences**	**Reference**
Qu et al (35066634)	2022	Glioma	Differentiation	PCASL, VSASL, and DSC-PWI	36

Continued on next page

Table S1 – *Continued from previous page*

Reference (PMID)	Year	Tumor type(s)	Purpose	Sequences	Reference
Wang et al (32559177)*	2020	Glioma	IDH status prediction	IVIM-PSF (DWI), and DCE	37
Choi et al (28116517)*	2017	Glioma	Differentiation	APT-CEST, DWI, and DSC	43
Togao et al (27003139)	2016	Glioma	Differentiation	APT-CEST, DWI, and DSC	44
Morana et al (29453753)	2018	Pediatric astrocytic tumor	Differentiation	ASL and DSC	38
Wang et al (29398151)	2018	Glioma	Follow-up	PCASL and DSC	39
Lavrova et al (35669426)	2022	Glioma, metastases, lymphoma	Follow-up	PCASL and DSC	40
Seeger et al (24200483)	2013	Glioma	Follow-up	DSC, DCE, MRS, and ASL	41
Artificial Intelligence					
Reference (PMID)	**Year**	**Tumor type(s)**	**Purpose**	**Sequences**	**Reference**
Wang et al (35396712)	2022	Not specified	Synthetic imaging	FLAIR and CE-FLAIR	48
Preetha et al (34688602)	2021	Glioblastoma	Synthetic imaging	T1w, FLAIR, T2, and CE-T1w	49
Dong et al (34767476)	2022	Pediatric posterior fossa tumors including astrocytoma	Differentiation	T1w, FLAIR, T2, DWI, and CE-T1w	52
Cheng et al (33104503)	2022	Glioma	Differentiation	T1w, FLAIR, T2, and CE-T1w	53
Hashido et al (32273523)	2020	Glioma	Differentiation	ASL and DSC	54

Continued on next page

Table S1 – *Continued from previous page*

Reference (PMID)	Year	Tumor type(s)	Purpose	Sequences	Reference
Alis et al (31973941)	2020	Glioma	Differentiation	FLAIR and CE-T1w	55
Zhang et al (32739863)	2020	Meningioma	Therapy planning	T2w and CE-T1w	56
Chang et al (29167275)	2017	Glioma	IDH status prediction	T1w, FLAIR, T2, and CE-T1w	58
Mammadov et al (35965975)	2022	Glioma	Follow-up	T1w and CE-T1w	59
Bacchi et al (31648967)	2019	Glioma	Follow-up	DWI, ADC, FLAIR, and CE-T1w	61

Note.—DWI = diffusion-weighted imaging, ADC = apparent diffusion coefficient, FLAIR = fluid-attenuated inversion recovery, CE-T1w = contrast-enhanced T1-weighted, DSC = dynamic susceptibility contrast, DCE = dynamic contrast enhanced, MRS = MR spectroscopy, ASL = arterial spin labeling, PCASL = pseudo continuous arterial spin labeling, APT-CEST = amide-proton transfer chemical exchange saturation transfer, VSASL = velocity selective arterial spin labeling, IVIM-SPF = intravoxel incoherent motion-derived perfusion fraction, T2w = T2-weighted, T1w = T1-weighted, T2*w = T2 star weighted, HR-SWI = high-resolution susceptibility-weighted imaging, PMID = PubMed identifier. * Prospective study design

Chapter 4

Reproducibility of 3T APT-CEST in Healthy Volunteers and Patients With Brain Glioma

Ivar J.H.G. Wamelink, Joost P.A. Kuijer, Beatriz E. Padrela, Yi Zhang, Frederik Barkhof, Henk J.M.M. Mutsaerts, Jan Petr, Eslmarieke van de Giessen, Vera C. Keil

Journal of Magnetic Resonance Imaging

28-03-2022; 10.1002/jmri.28239

Abstract

Background: Amide proton transfer (APT) imaging is a chemical exchange saturation transfer (CEST) technique offering potential clinical applications such as diagnosis, characterization, and treatment planning and monitoring in glioma patients. While APT-CEST has demonstrated high potential, reproducibility remains underexplored.

Purpose: To investigate whether cerebral APT-CEST with clinically feasible scan time is reproducible in healthy tissue and glioma for clinical use at 3T.

Study type: Prospective, longitudinal.

Subjects: 21 healthy volunteers (11 females; mean age±standard deviation: 39±11 years) and 6 glioma patients (3 females; 50±17 years: four glioblastomas, one oligodendroglioma, one radiologically suspected low-grade glioma). Field strength/sequence: 3T, TSE SPACE-CEST.

Assessment: APT-CEST measurement reproducibility was assessed within-session (glioma patients, scan session 1; healthy volunteers scan sessions 1, 2, and 3), between-sessions (healthy volunteers scan sessions 1 and 2), and between-days (healthy volunteers, scan sessions 1 and 3). The mean APT_{CEST} values and standard deviation of the within-subject difference (SD_{diff}) was calculated in whole tumor enclosed by regions of interest (ROIs), and eight ROIs - whole-brain, cortical gray matter, putamen, thalamus, orbitofrontal gyri, occipital lobe, central brain – and compared.

Statistical tests: Brown-Forsythe tests and variance component analyses (VCA) were used to assess the reproducibility of ROIs for the three time intervals. Significance was set at $P < 0.003$ after Bonferroni correction.

Results: Intratumoral mean APT_{CEST} was significantly higher than APT_{CEST} in healthy-appearing tissue in patients (0.5±0.46%). The average within-session, between-sessions, and between days SD_{diff} of healthy control brains was 0.2% and did not differ significantly with each other ($0.76>P>0.22$). The within-session SD_{diff} of whole-brain was 0.2% in both healthy volunteers and patients, and 0.21% in the segmented tumor. VCA showed that the within-session factors was the most important factor (60%) for scanning variance.

Data conclusion: Cerebral APT-CEST imaging may show good scan-rescan reproducibility in healthy tissue and tumors with clinically feasible scan times at 3T. Short-term measurement effects may be the dominant components for reproducibility.

Keywords: reproducibility, glioma, APT, CEST, brain

Introduction

Novel imaging biomarkers in glioma patients may help diagnosis, prognosis, and treatment decisions by improving tumor characterization and delineation.[1–3] One such imaging biomarker is amide proton transfer (APT) imaging. APT is a subset of chemical exchange saturation transfer (CEST) imaging that allows for quantitative amide and peptide detection.[1, 4, 5] Specifically, APT-CEST may be especially useful in glioma imaging as these primary brain tumors express increased amounts of protein.[6–8] This magnetic resonance imaging (MRI) technique measures the intensity change in bulk water after the magnetization in amide protons is selectively saturated and transferred by chemical exchange to water protons.[6] This intensity change indirectly reflects protein concentration in the tissue, computed by the magnetization transfer ratio asymmetry (MTR_{asym}).[4, 5, 9] Multiple clinical applications have been proposed in recent years for neuro-oncological APT-CEST imaging and show promising results regarding diagnosis, characterization, and treatment planning and monitoring in glioma.[8, 10–14] Also in other clinical contexts, APT-CEST has demonstrated its potential clinical usability, e.g., in the detection of stroke for highlighting ischemic areas, paving its way to a near-future standard application.[15–17] However, particularly for routine clinical application, high reproducibility in healthy and tumorous tissue is crucial. Quantitative assessment of the reproducibility of differences between healthy and tumorous tissue is mandatory for the development of imaging biomarkers with established cut-off values to non-invasively differentiate between tissue conditions.

Previous preliminary findings in reproducibility studies were promising for clinical 3 T APT-CEST applications.[18–20] One study showed consistently higher within-session than between-sessions reproducibility for 3 T APT-CEST with partial brain coverage and relatively long scanning time in healthy volunteers and glioma patients.[18] Another study found good reproducibility in brain tumors.[19] Spatially homogenous radiofrequency (RF) shimming and B_0-inhomogeneity correction could contribute to this reproducibility, but the authors of this study exclusively investigated mean tumor APT-CEST values in a single slice. A third study found good reproducibility in only two healthy volunteers.[20]

The purpose of this study was to investigate the short- and long-term reproducibility of 3 T APT-CEST measurements in a comparatively large group of healthy volunteers and a pilot level group of glioma patients, with a clinically applicable scan protocol with full brain coverage and a short scan duration.

Methods

This prospective study was approved by the institutional review board (VUmc_2021-0038) and written informed consent was obtained from all participants.

Study Design

Healthy volunteers were recruited non-commercially through external advertisements. Inclusion criteria were 1) age > 18 years and 2) neither clinical history nor MRI evidence of brain pathology. Prospective recruitment also included patients with suspected recurrent or de-novo glioma on external MRI and referred for tumor resection between May 2021

and July 2021. Patients were included if 1) age > 18 years, 2) they had a radiological diagnosis of de-novo glioma or recurrent already confirmed glioma, 3) received clinically indicated MRI, and 4) had no other concurrent brain pathology at the time of diagnosis. Participants with severe motion artifacts in T1-weighted (T1w), contrast-enhanced T1-weighted (ceT1w), or APT-CEST scans were excluded. No participants needed to be excluded due to compromised image quality.

Participants

Twenty-one healthy volunteers (mean age ± SD: 39 ± 11 years, 11 women) and 6 glioma patients (mean age ± SD: 50 ± 17 years, 3 women) were scanned. Five patients subsequently received a histopathological confirmation according to the 2021 World Health Organization (WHO) criteria: four glioblastomas (GBMs; WHO grade 4) and one oligodendroglioma (WHO grade 2)[21]. One patient opted against surgery due to the high radiological likelihood of a slow-growing low-grade glioma. Power analysis to determine group size to detect change in APT-CEST for GBM and LGG is given in Table S1 in the Supplemental Material.

Protocol of Scan-Rescan Test

Healthy volunteers were subjected to three APT-CEST scan sessions (Fig. 1). The first two sessions were separated by participant repositioning (between-sessions reproducibility) and the last session occurred 4-14 days later (between-days reproducibility). Patients underwent one scan session (Fig. 1). Each session, in both volunteers and patients, contained two consecutive APT-CEST scans (within-session reproducibility) with a B_0-map in between. All participants received a structural three-dimensional (3D) T1-weighted acquisition (T1w) while the patients additionally received a complete neuro-oncological MRI protocol according to the European Organisation for Research and Treatment of Cancer standard including 3D T1-weighted post-contrast (ceT1w) and fluid-attenuated inversion recovery.[22] APT-CEST scanning was performed before the neuro-oncological MRI protocol.

APT-CEST Image Acquisition and Post-Processing

All participants were scanned on a 3-T MRI (Vida, Siemens Healthineers, Erlangen, Germany) using a 20-channel head coil, including whole-brain 3D MPRAGE T1w (repetition time (TR) 2300 msec, echo time (TE) 2.32 msec, inversion time (TI) 900 msec, refocusing flip angle 8°, field of view 240 mm, slice thickness 0.9 mm) and TSE SPACE APT-CEST (TR 3000 msec, TE 17 msec, turbo factor 140, refocusing flip angle 120°, 2.8 x 2.8 x 2.8, mm^3 resolution, 7 frequencies with saturation pulses [10 Gaussian pulses of 100 msec at 2.0 μT] at ± 3.0 ppm, ± 3.5 ppm, ± 4.0 ppm, and 1560 ppm off-resonance, scan duration: 4:36 minutes) plus a dual-echo gradient recalled echo (GRE) B_0 map for correction of spatial inhomogeneity of the resonance frequency.[23]
All off-resonance scans were registered to the 1560 ppm off-resonance scan to correct for head motion (MCFLIRT, FSL 5.0.9).[24] APT-CEST maps were then calculated as MTR_{asym} at 3.5 ppm off-resonance from motion-corrected images:

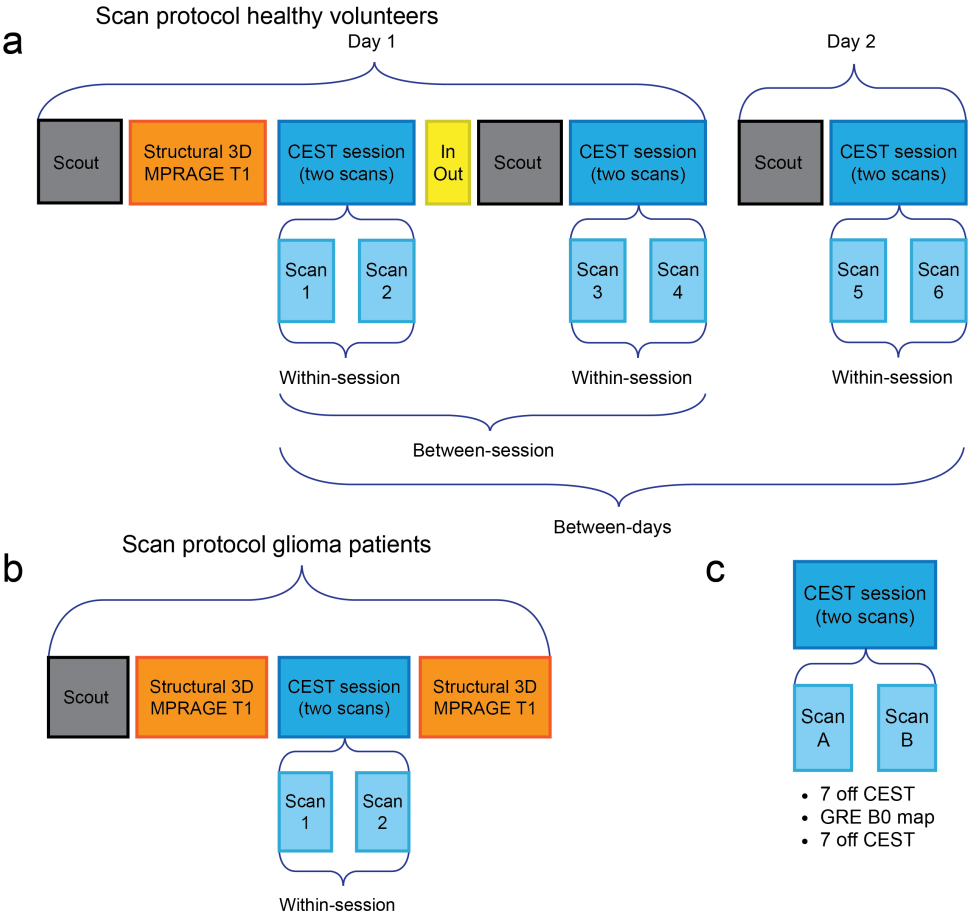


Figure 1: (a) Scan protocol for the healthy volunteers. (b) Scan protocol for the patients. (c) Scan session. A session contains two consecutive APT-CEST scans and one B_0 map. (a) Scans: 1 = baseline, 2 = scan directly after the baseline—i.e., 1–2 = within-session reproducibility—scan 4 is a scan 30 minutes after the first—i.e., 1–4 = between-sessions reproducibility—scan 6 = another day—i.e., 1–6 = between-day reproducibility. (b) Glioma patients underwent one session.

$$APT_{CEST} = MTR_{asym}(3.5ppm) = \frac{S_{sat}(-3.5ppm) - S_{sat}(3.5ppm)}{S_0} \quad (4.1)$$

The S_{sat} and S_0 represent the imaging signal intensities measured with RF saturation at ± 3.5 ppm and at 1560 ppm, respectively. The B_0 map was expressed in ppm after unwrapping by FSL (Version 5.0.9) Prelude.[24] The measured points on the Z-spectrum (off-resonance saturation spectrum) were interpolated to the actual frequency shift as measured by the B_0 map. We extrapolated the measured points on the Z-spectrum if the frequency shift was between ± 0.5 ppm and ± 1 ppm. Voxels with a shift larger than 1 ppm were masked (set to 0).

Next, the 1560 ppm off-resonance APT-CEST scan was registered to the non-contrast 3D T1w image with co-registration of the APT-CEST map using FLIRT (FSL 5.0.9) for healthy volunteers and patients.[24] The 3D T1w scan was spatially normalized to Montreal Neurological Institute (MNI) space using CAT12.7 (build 1615)[25] and again with co-registration of the APT-CEST map, which was then smoothed by a 3.5 x 3.5 x 3.5, mm^3 full width at half maximum (FWHM) Gaussian filter. No skull stripping was performed.

Image Analysis

Five anatomical regions of interest (ROIs) were available in MNI space using the Harvard-Oxford atlas[26]: the whole brain including cerebellum and internal cerebral spinal fluid spaces, cortical gray matter, deep white matter, putamen, and thalami. Three additional anatomical ROIs were manually delineated near the orbitofrontal gyri, the occipital lobes, and the central brain by (I.J.H.G. Wamelink) with 1 year of experience under supervision of a neuroradiologist with 10 years of experience (V.C. Keil) (Fig. 2). The orbitofrontal gyri represent a region that is susceptible to B_0-inhomogeneities.[27] All three ROIs contained both white and gray matter.

In patients with glioma, the ROIs enclosing presumably vital glioma tissue were manually delineated, together with similar-sized contralateral tissue ROIs, on the ceT1w for enhancing tumors and based on FLAIR for the non-enhancing tumors by a neuroradiologist with 10 years of experience (V.C. Keil).

By multiplying the binary ROI atlases with the APT-CEST scans the mean APT_{CEST} values were calculated for all scans. Next, the SD across all measurements was taken for the eight ROIs. The SD across subjects of the within-subject difference (SD_{diff}) between the mean APT_{CEST} values was calculated for the eight ROIs and each of the three reproducibility time intervals—within-session, between-sessions, and between-days. Additionally, voxel-wise SD_{diff} maps were computed from non-averaged APT_{CEST} values for the three reproducibility time intervals to visually identify locations of low reproducibility.

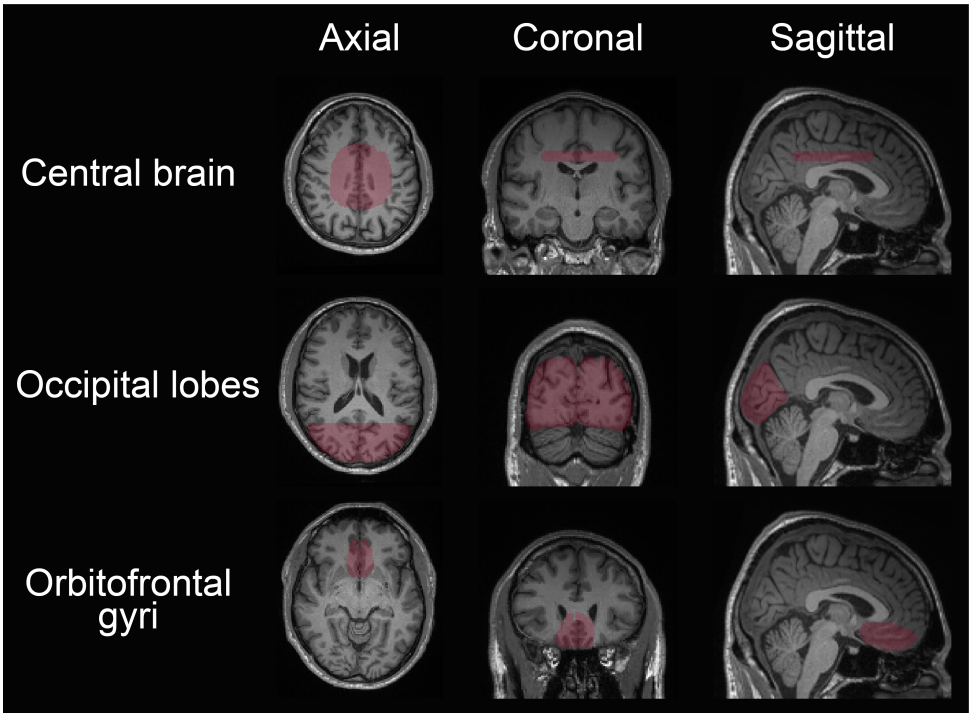


Figure 2: Axial, coronal, and sagittal slices of the three manually delineated regions of interest (ROIs).

Statistics

The Brown–Forsythe test (Python, Python Software Foundation, version 3.6.9, package: SciPy, version 1.5.2) was used to test the variance of the SD_{diff} among the three reproducibility time intervals of the eight anatomical ROIs. Bonferroni correction was performed by dividing the significance threshold (0.05) by the number of separate tests (8 ROIs x 2 tests = 16). Variance component analysis (VCA) (RStudio Version: 1.4.1106; VCA Version: 1.4.3) was performed on the mean APT_{CEST} value of the eight anatomical ROIs to find the contributions of the different effects. The Brown–Forsythe test and VCA were only performed on healthy volunteers.

The agreement of the APT_{CEST} values in tumor patients was assessed by the intraclass correlation coefficient (ICC) (Python, Python Software Foundation, version 3.6.9, package: Pingouin, version 0.3.12). The within-session ICC was computed for the tumor and contralateral ROI. The ICCs were calculated with a 95% confidence interval (CI) using an average random raters model, and given the following classifications: 0.00–0.39, poor; 0.40–0.59, fair; 0.60–0.74, good; and 0.75–1.00 excellent.[28]

Bland–Altman plots were created to show differences between the three reproducibility time intervals in healthy volunteers. The Mann–Whitney U test (Python, Python Software Foundation, version 3.6.9, package: SciPy, version 1.5.2) was performed to compare the mean APT_{CEST} values of the different structures in healthy volunteers with the mean

tumor values in patients. Statistical significance was determined at P < 0.006 with Bonferroni correction (0.05/8).

The effect size (E) was only computed for patients:

$$Effectsize(E) = \frac{tumor - to - normal - tissuedifference}{SD_{\text{diff}}} \tag{4.2}$$

Results

APT_{CEST} Descriptive Statistics in Volunteers and Patients

The mean and SD for the anatomical and tumor ROIs are shown in Table 1. The mean APT_{CEST} difference between ROIs in GBMs and contralateral tissue was 1.11% (N = 4, Table 1, Fig. 3). Intratumoral mean APT_{CEST} (1.59 ± 0.67%, N = 6) was significantly higher than contralateral similar-sized ROI APT_{CEST} in healthy-appearing tissue in patients (0.50 ± 0.46%), and also significantly higher than mean APT_{CEST} in anatomical ROIs in healthy volunteers (between 0.42 and 1.02, N = 21).

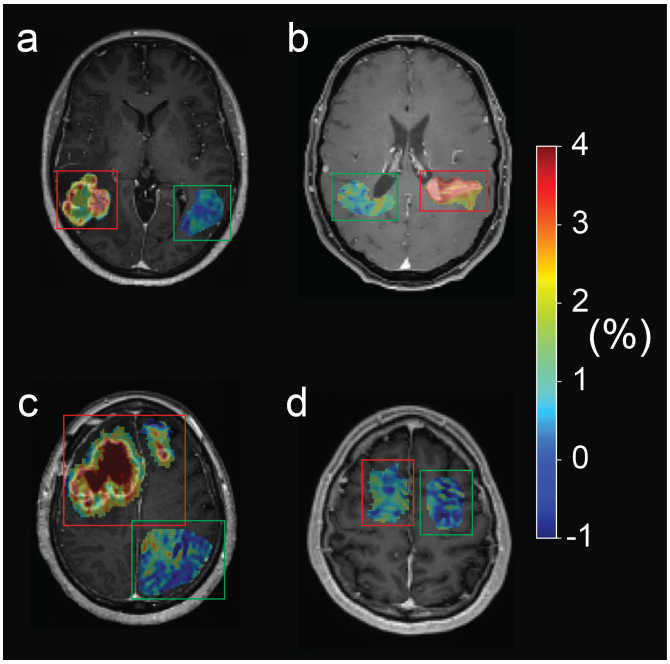


Figure 3: Single transversal APT-CEST slices of four different glioblastoma patients. APTCEST value in color scale projected on the post-contrast T1-weighted image, within the tumor ROI only. Red and green boxes are positioned around the tumor and contralateral ROIs, respectively. Note the heterogeneous hyperintensity in the tumorous regions that show hyperintensity on the contrast-enhanced T1-weighted image when compared to the contralateral region (a–c). (d) The APT-CEST value of a non-enhancing glioblastoma. APT = amide proton transfer; CEST = chemical exchange saturation transfer; ROI = region of interest.

Table 1: Mean regional APTCEST values in healthy volunteers and glioma patients.

Dataset	Whole Brain	Cortical Gray Matter	Deep White Matter	Putamen	Thalami	Orbitofrontal Gyri	Occipital Lobes	Central Brain	Tumor ROI / Contralateral ROI
Healthy volunteers (N = 21)	0.88 ± 0.93	0.82 ± 0.85	0.42 ± 0.53	0.95 ± 0.45	1.39 ± 0.46	0.26 ± 1.81	1.02 ± 0.65	0.87 ± 0.54	–
Patients (N = 6)	0.93 ± 0.95	–	–	–	–	–	–	–	1.59 ± 0.67 / 0.50 ± 0.46
GBM, WHO 4 (N = 4)	1.02 ± 0.17	–	–	–	–	–	–	–	1.93 ± 0.51 / 0.83 ± 0.34
Oligodendroglior WHO 2 (N = 1)	0.66 ± 0.04	–	–	–	–	–	–	–	0.57 ± 0.29 / 0.14 ± 0.05
LGG, radiologically WHO 2 (N = 1)	0.93 ± 0.06	–	–	–	–	–	–	–	1.24 ± 0.04 / 0.14 ± 0.02

Note.—APTCEST = amide proton transfer chemical exchange saturation transfer; GBM = glioblastoma; LGG = low grade glioma; ROI = region of interest.

Table 2: The SD_{diff} of the eight ROIs and each of the three reproducibility time intervals.

Time Interval	Whole Brain	Cortical Gray Matter	Deep White Matter	Putamen	Thalami	Orbitofrontal Gyri	Occipital Lobes	Central Brain
Within-session (%)	0.20	0.20	0.19	0.22	0.26	0.61	0.22	0.20
Between-sessions (%)	0.24 P = 0.35	0.25 P = 0.28	0.26 P = 0.39	0.28 P = 0.22	0.28 P = 0.31	0.67 P = 0.76	0.26 P = 0.33	0.27 P = 0.33
Between-days (%)	0.22 P = 0.17	0.22 P = 0.20	0.24 P = 0.11	0.33 P = 0.01	0.33 P = 0.05	0.48 P = 0.58	0.24 P = 0.14	0.26 P = 0.03

Note.—The P-value is the comparison between the within-session difference and either the between-sessions or between-days difference for the brain structures in healthy volunteers. P-values less than 0.003 were considered significant. SD_{diff} = standard deviation across subjects of the within-subject difference; ROI = region of interest.

Reproducibility in Healthy Volunteers

Figure 4 shows high within-session reproducibility and only slightly higher between-sessions and between-days voxel-wise SD_{diff}. The highest within-session voxel-wise SD_{diff} variance was found in the orbitofrontal gyri, as also shown by the higher ROI SD_{diff} (Table 2). The SD_{diff} was consistent within each of the three time intervals and only showed a small increase (maximum of 0.11%) between them.

The between-days SD_{diff} was consistently slightly lower than the between-sessions SD_{diff}. However, the Brown–Forsythe test showed a non-significant difference between the between-sessions and between-days reproducibility (P = 0.75, N = 21). The SD_{diff} increase between sessions was small and the reproducibility did not significantly differ between the within-session and between-sessions and the within-session and between-days (Table 2). No statistically significant differences were found after Bonferroni correction. The within-session SD_{diff} of the whole brain for all patients was 0.2% (N = 6).

The VCA showed that total variance (0.03–0.07) was relatively low compared to the difference between GBMs and contralateral tissue (1.11%). The variance appears to be predominated by the factor within-session rather than any other factor (Table 3). This is consistent with the SD_{diff} maps (Fig. 4). Figures 5 and 6 show Bland–Altman plots of the mean APT_{CEST} value against the APT_{CEST} value session differences of the brain ROIs.

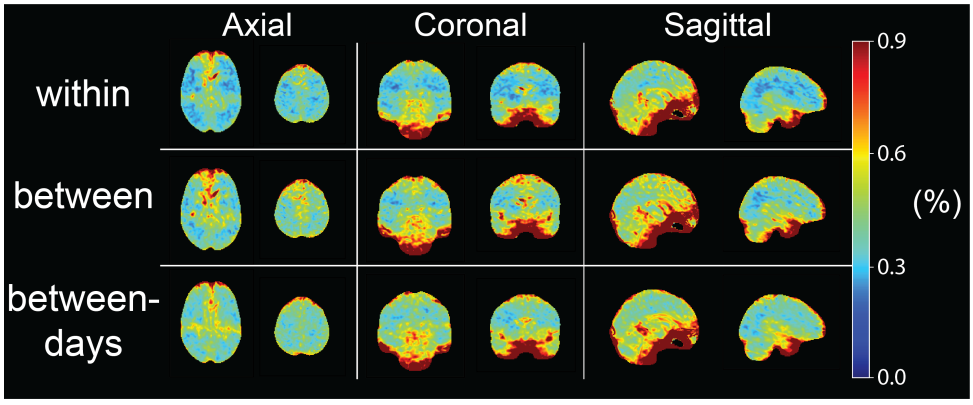


Figure 4: Voxel-wise SD maps of the within-subject difference between scan sessions 1–2, 1–4, and 1–6 (rows a, b, and c, respectively), for all healthy volunteers (N = 21). Color scale represents the SD of the difference in APT_{CEST} value (%) between sessions (SD_{diff}), where higher SD_{diff} means lower reproducibility.

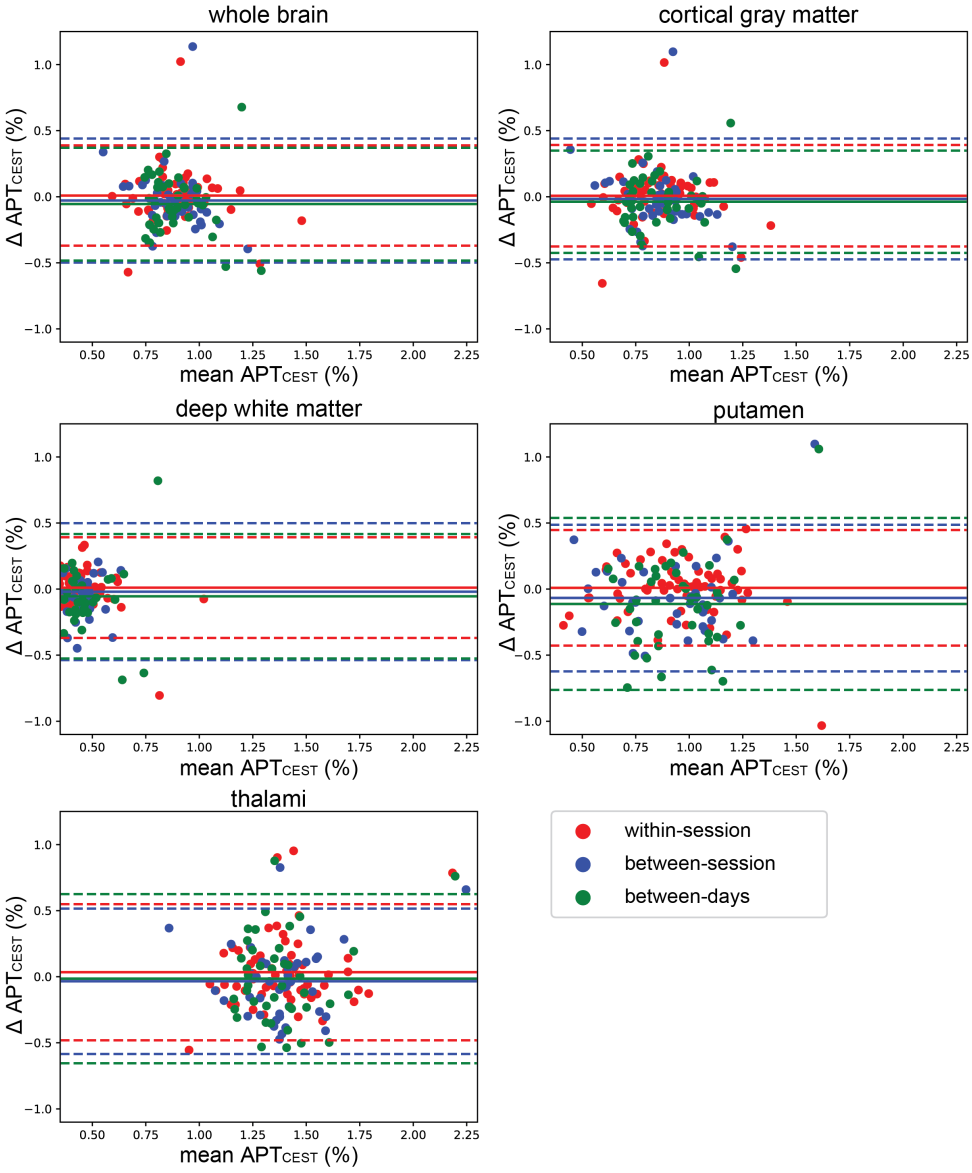


Figure 5: Bland–Altman plots. Amide proton transfer chemical exchange saturation transfer (APT_{CEST}) differences of the five different brain structures are plotted against their mean APT_{CEST} value. Continuous and broken lines indicate mean difference and 95% limits of agreement (mean difference ± 1.96 SD of the paired difference) respectively.

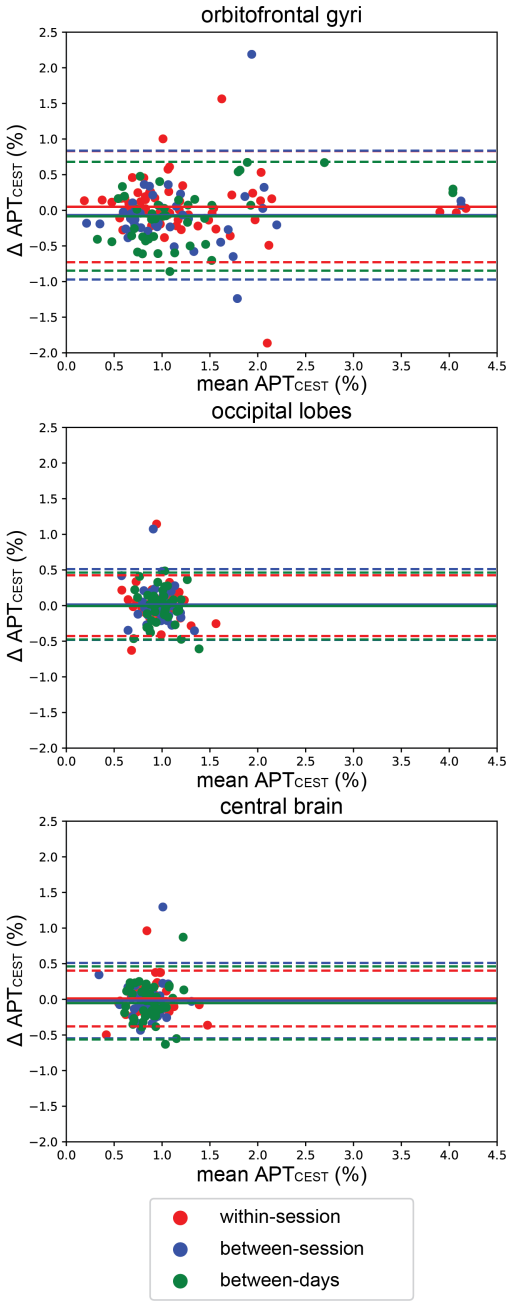


Figure 6: Bland–Altman plots. Showing differences of reproducibility in the orbitofrontal gyri, central brain, and occipital lobes. Note the larger limits of agreement in the orbitofrontal gyri and the narrower limits of agreement in the occipital lobes and central brain regions.

Table 3: Percentage of Total Variance Computed by the Variance Component Analysis.

	Whole Brain	Cortical Gray Matter	Deep White Matter	Putamen	Thalami	Orbitofrontal Gyri	Occipital Lobes	Central Brain
Total variance (APTCEST %)	0.033	0.039	0.034	0.069	0.058	0.205	0.041	0.043
Factor subject	16%	28%	11%	**34%**	13%	**42%**	12%	22%
Factor day	19%	16%	**21%**	16%	4%	14%	**20%**	8%
Factor session	5%	2%	13%	14%	22%	1%	8%	**24%**
Factor within-session	**60%**	**54%**	55%	36%	**61%**	43%	60%	46%

Note.—Largest contributor to total variance is marked in bold. APTCEST = amide proton transfer chemical exchange saturation transfer.

Reproducibility in Patients

The within-session reproducibility of the tumors was slightly lower (SD_{diff}, 0.21%, N = 6) than in the contralateral ROIs (SD_{diff}, 0.11%), but without a statistical significance (P = 0.73). The mean whole-brain APT_{CEST} values did not significantly differ between healthy volunteers and patients (0.88 ± 0.93%, 0.93 ± 0.95% respectively; P = 0.13). Mean tumor values did however differ significantly with the eight ROIs in healthy volunteers. Within-session scan agreement was the highest in contralateral ROIs (ICC = 0.99; 95% CI, 0.92–1.00; N = 21) followed by glioma (ICC = 0.97; 95% CI, 0.82–1.00; N = 6). The effect size of the APT_{CEST} in GBM was 5 when considering the ROI-wise within-session SD_{diff} of the putamen.

Discussion

We found high reproducibility across the three reproducibility time intervals in healthy volunteers. Within-session measurement agreement in glioma, contralateral-to-glioma, and healthy brain ROIs was excellent. Furthermore, within-session reproducibility in whole-brain APT-CEST scans of glioma patients was equal to the reproducibility in healthy volunteers and even better for tumor ROIs. The variance thus seems to be predominated by general error contributors such as thermal noise, subject motion, B_0 inhomogeneity, and stress on scanner hardware, rather than long-term instrumental errors, physiological variance, or between-subject differences. These results may suggest that 3 T APT-CEST provides sufficient reproducibility for clinical application, especially given the larger APT_{CEST} value changes between glioma and normally appearing tissue.

While the between-sessions reproducibility was not evaluated previously, the within-session and between-days reproducibility of APT-CEST scans of this study are in line with previous studies in healthy volunteers[18, 20] and brain tumor patients.[18–20] The within-session, between-sessions, and between-days reproducibility presented in the current study showed a trend of decrease—although not statistically significant—for the SD_{diff} and the VCA. Between-days reproducibility showed higher consistency than between-sessions reproducibility, albeit not to a statistically significant degree. The first two studies used scan protocols and acquisition parameters similar to ours with an exclusive focus on mean APT_{CEST} values.[18, 19] Our current study pursues a different perspective by also performing a voxel-wise reproducibility analysis in healthy tissue of patients and healthy participants to focus on the reproducibility of relative APT_{CEST} values. Both studies found similar within-session scan agreement for glioma.[18, 19]

Our results show that the different imaging sessions had no statistically significant effect on APT_{CEST} measurement values, which is in agreement with a previous similar analysis of variance in healthy volunteers and glioma patients.[18] Our findings are comparable to another study with two healthy volunteers.[20] Furthermore, our difference between glioma measurements and healthy appearing contralateral tissue at all time points was in agreement with previous studies.[13, 18, 29] While APT-CEST reproducibility in the brain has not yet been investigated extensively, our reproducibility was similar to that of studies in other body parts such as breast (7 T) and prostate (3 T) tissue and different CEST sequences such as glutamate-CEST.[30–33]

We found the lowest reproducibility in the orbitofrontal gyri, which can be explained by the susceptibility of these regions to B_0-inhomogeneity that impacts the Z-spectrum.[13] Moreover, regions with large veins, such as areas close to the superior sagittal sinus, also showed high variance consistent with the literature.[29] We corrected for the B_0-inhomogeneity by creating a B_0-map between both APT-CEST scans.[34] The B_0-map was scanned once for each session in order to limit the session duration. We thus implicitly assumed that the measurement error of the B_0-map itself does not contribute to APT_{CEST} variability. However, the within-session variability does include subject motion between CEST and B_0 scans. Other potential causes are subject movement in general, and the pre- and post-processing steps, which is why we applied the same sequence parameters during the entire study. Finally, low reproducibility could be caused by physiological variation of amide concentrations.

With good reproducibility, APT-CEST imaging shows potential for clinical implementation. The SD_{diff} was consistent within each of the three time intervals and only slightly increased between the different time intervals. Moreover, the SD_{diff} was small compared to the tumor signal change, denoting a detectable difference between healthy and tumorous tissue, in particular for GBMs. This opens the door for potential clinical biomarker applications such as percentage differences in APT-CEST values that can be used to differentiate tumors from viable tissue. A potential benefit of APT-CEST over conventional tumor imaging, more specifically ceT1w imaging, is that APT-CEST might be able to identify non-contrast-enhancing parts of high-grade gliomas.[29] Previous studies have also shown that APT imaging is able to differentiate between solitary brain metastases and GBMs or radiation necrosis and viable tumor tissue.[35–37] These findings are particularly important for quick and non-invasive disease staging and tumor classification, which may avoid unnecessary biopsies, but also open potential alternatives to gadolinium contrast administration.[38]

Limitations

We only investigated intra-vendor reproducibility in this study. It is thus unclear whether the results can be generalized to different scanners and vendors. Inter-vendor reproducibility is important to study for clinical implementation as scanning a patient on the same scanner for follow-up assessment is not always clinically feasible. Another limitation is that our sample size for glioma patients remains on a pilot study level. Finally, we could have scanned more offset frequencies to improve correction for off-resonance effects but this would have prolonged the scan duration to clinically unfeasible times.

Conclusion

The findings of this study indicate that whole-brain APT-CEST imaging with clinical scan time has a sufficiently high short- and long-term reproducibility in tumors and healthy tissue at clinically feasible scan times at 3 T. The tumor–normal tissue contrast was larger than the APT-CEST scan–rescan errors. Instrumental noise seems to be the largest component, while long-term physiological variance and between-subject variability were smaller. Reproducibility might be lower near regions with B_0-inhomogeneity.

References

1 T. C. Booth et al. “Machine learning and glioma imaging biomarkers”. In: *Clin. Radiol.* 75 (1 2020), pp. 20–32. doi: 10.1016/j.crad.2019.07.001.

2 O. M. Henriksen et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 1: Perfusion and Diffusion Techniques”. en. In: *Front. Oncol.* 12 (Mar. 3, 2022), p. 810263. doi: 10.3389/fonc.2022.810263.

3 T. C. Booth et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 2: Spectroscopy, Chemical Exchange Saturation, Multiparametric Imaging, and Radiomics”. en. In: *Front. Oncol.* 11 (2021), p. 811425. doi: 10.3389/fonc.2021.811425.

4 J. Zhou et al. “Amide proton transfer (APT) contrast for imaging of brain tumors”. en. In: *Magn. Reson. Med.* 50 (6 Dec. 2003), pp. 1120–1126. doi: 10.1002/mrm.10651.

5 J. Zhou et al. “Using the amide proton signals of intracellular proteins and peptides to detect pH effects in MRI”. en. In: *Nat. Med.* 9 (8 Aug. 2003), pp. 1085–1090. doi: 10.1038/nm907.

6 H. Schmidt et al. “Systematic Evaluation of Amide Proton Chemical Exchange Saturation Transfer at 3 T: Effects of Protein Concentration, pH, and Acquisition Parameters”. In: *Invest. Radiol.* 51 (10 2016).

7 R. Scheidegger, E. T. Wong, and D. C. Alsop. “Contributors to contrast between glioma and brain tissue in chemical exchange saturation transfer sensitive imaging at 3 Tesla”. en. In: *Neuroimage* 99 (Oct. 1, 2014), pp. 256–268. doi: 10.1016/j.neuroimage.2014.05.036.

8 Z. Wen et al. “MR imaging of high-grade brain tumors using endogenous protein and peptide-based contrast”. en. In: *Neuroimage* 51 (2 June 2010), pp. 616–622. doi: 10.1016/j.neuroimage.2010.02.050.

9 P. C. M. van Zijl and N. N. Yadav. “Chemical exchange saturation transfer (CEST): what is in a name and what isn’t?” en. In: *Magn. Reson. Med.* 65 (4 Feb. 17, 2011), pp. 927–948. doi: 10.1002/mrm.22761.

10 T. Zou et al. “Differentiating the histologic grades of gliomas preoperatively using amide proton transfer-weighted (APTW) and intravoxel incoherent motion MRI”. en. In: *NMR Biomed.* 31 (1 Nov. 3, 2017). doi: 10.1002/nbm.3850.

11 S. Jiang et al. “Amide proton transfer-weighted magnetic resonance image-guided stereotactic biopsy in patients with newly diagnosed gliomas”. en. In: *Eur. J. Cancer* 83 (July 10, 2017), pp. 9–18. doi: 10.1016/j.ejca.2017.06.009.

12 W. Dou et al. “Chemical exchange saturation transfer magnetic resonance imaging and its main and potential applications in pre-clinical and clinical studies”. en. In: *Quant. Imaging Med. Surg.* 9 (10 Oct. 2019), pp. 1747–1766. doi: 10.21037/qims.2019.10.03.

13 C. K. Jones et al. “Amide proton transfer imaging of human brain tumors at 3T”. en. In: *Magn. Reson. Med.* 56 (3 Sept. 2006), pp. 585–592. doi: 10.1002/mrm.20989.

14 J. Zhou et al. “Practical data acquisition method for human brain tumor amide proton transfer (APT) imaging”. en. In: *Magn. Reson. Med.* 60 (4 Oct. 2008), pp. 842–849. doi: 10.1002/mrm.21712.

15 P. Z. Sun. “Consistent depiction of the acidic ischemic lesion with APT MRI-Dual RF power evaluation of pH-sensitive image in acute stroke”. en. In: *Magn. Reson. Med.* (Sept. 30, 2021). doi: 10.1002/mrm.29029.

16 L. S. Foo et al. “Clinical translation of amide proton transfer (APT) MRI for ischemic stroke: a systematic review (2003-2020)”. en. In: *Quant. Imaging Med. Surg.* 11 (8 Aug. 2021), pp. 3797–3811. doi: 10.21037/qims-20-1339.

17 L. S. Foo et al. “Study of common quantification methods of amide proton transfer magnetic resonance imaging for ischemic stroke detection”. en. In: *Magn. Reson. Med.* 85 (4 Apr. 2021), pp. 2188–2200. doi: 10.1002/mrm.28565.

18 J. B. Lee et al. “Repeatability of amide proton transfer–weighted signals in the brain according to clinical condition and anatomical location”. In: *Eur. Radiol.* 30 (1 Jan. 1, 2020), pp. 346–356. doi: 10.1007/s00330-019-06285-7.

19 O. Togao et al. “Scan-rescan reproducibility of parallel transmission based amide proton transfer imaging of brain tumors”. en. In: *J. Magn. Reson. Imaging* 42 (5 Nov. 2015), pp. 1346–1353. doi: 10.1002/jmri.24895.

20 S. Goerke et al. “Clinical routine acquisition protocol for 3D relaxation-compensated APT and rNOE CEST-MRI of the human brain at 3T”. en. In: *Magn. Reson. Med.* 86 (1 July 2021), pp. 393–404. doi: 10.1002/mrm.28699.

21 D. N. Louis et al. “The 2021 WHO Classification of Tumors of the Central Nervous System: a summary”. en. In: *Neuro. Oncol.* 23 (8 Aug. 2, 2021), pp. 1231–1251. doi: 10.1093/neuonc/noab106.

22 B. M. Ellingson et al. “Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials”. en. In: *Neuro. Oncol.* 17 (9 Sept. 2015), pp. 1188–1198. doi: 10.1093/neuonc/nov095.

23 Y. Zhang et al. “Whole-brain chemical exchange saturation transfer imaging with optimized turbo spin echo readout”. In: *Magn. Reson. Med.* 84 (3 2020), pp. 1161–1172. doi: 10.1002/mrm.28184.

24 M. Jenkinson et al. “Improved optimization for the robust and accurate linear registration and motion correction of brain images”. en. In: *Neuroimage* 17 (2 Oct. 2002), pp. 825–841. doi: 10.1016/s1053-8119(02)91132-8.

25 C. Gaser and R. Dahnke. “CAT-A Computational Anatomy Toolbox for the Analysis of Structural MRI Data”. In: 2016.

26 N. Makris et al. “Decreased volume of left and total anterior insular lobule in schizophrenia”. In: *Schizophr. Res.* 83 (2 2006), pp. 155–171. doi: 10.1016/j.schres.2005.11.020.

27 S. Yoo et al. “Feasibility of head-tilted brain scan to reduce susceptibility-induced signal loss in the prefrontal cortex in gradient echo-based imaging”. In: *Neuroimage* 223 (Dec. 1, 2020), p. 117265. doi: 10.1016/j.neuroimage.2020.117265.

28 K. A. Hallgren. “Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial”. en. In: *Tutor. Quant. Methods Psychol.* 8 (1 2012), pp. 23–34. doi: 10.20982/tqmp.08.1.p023.

29 J. Zhou et al. “Three-dimensional amide proton transfer MR imaging of gliomas: Initial experience and comparison with gadolinium enhancement”. en. In: *J. Magn. Reson. Imaging* 38 (5 Nov. 2013), pp. 1119–1128. doi: 10.1002/jmri.24067.

30 A. N. Dula et al. “Optimization of 7-T chemical exchange saturation transfer parameters for validation of glycosaminoglycan and amide proton transfer of fibroglandular breast tissue”. en. In: *Radiology* 275 (1 Apr. 2015), pp. 255–261. doi: 10.1148/radiol.14140762.

31 V. S. Evans et al. “Optimization and repeatability of multipool chemical exchange saturation transfer MRI of the prostate at 3.0 T”. en. In: *J. Magn. Reson. Imaging* 50 (4 Oct. 2019), pp. 1238–1250. doi: 10.1002/jmri.26690.

32 R. P. R. Nanga et al. “Reproducibility of 2D GluCEST in healthy human volunteers at 7 T”. en. In: *Magn. Reson. Med.* 80 (5 Nov. 2018), pp. 2033–2039. doi: 10.1002/mrm.27362.

33 P. Bagga et al. “In vivo GluCEST MRI: Reproducibility, background contribution and source of glutamate changes in the MPTP model of Parkinson’s disease”. en. In: *Sci. Rep.* 8 (1 Feb. 13, 2018), p. 2883. doi: 10.1038/s41598-018-21035-3.

34 P. Schuenke et al. “Simultaneous mapping of water shift and B1 (WASABI)-Application to field-Inhomogeneity correction of CEST MRI data”. en. In: *Magn. Reson. Med.* 77 (2 Feb. 2017), pp. 571–580. doi: 10.1002/mrm.26133.

35 H. Yu et al. “Applying protein-based amide proton transfer MR imaging to distinguish solitary brain metastases from glioblastoma”. en. In: *Eur. Radiol.* 27 (11 Nov. 2017), pp. 4516–4524. doi: 10.1007/s00330-017-4867-z.

36 H. Mehrabian et al. “Differentiation between Radiation Necrosis and Tumor Progression Using Chemical Exchange Saturation Transfer”. en. In: *Clin. Cancer Res.* 23 (14 July 15, 2017), pp. 3667–3675. doi: 10.1158/1078-0432.CCR-16-2265.

37 J. Zhou et al. “Differentiation between glioma and radiation necrosis using molecular magnetic resonance imaging of endogenous proteins and peptides”. en. In: *Nat. Med.* 17 (1 Jan. 2011), pp. 130–134. doi: 10.1038/nm.2268.

38 A. Falk Delgado et al. “Diagnostic value of alternative techniques to gadolinium-based contrast agents in MR neuroimaging-a comprehensive overview”. en. In: *Insights Imaging* 10 (1 Aug. 23, 2019), p. 84. doi: 10.1186/s13244-019-0771-1.

Supplementary Material

Table S1: Power analysis to determine group size for statistical tests.

Time interval and region	Sigma	Normal brain vs GBM	Normal brain vs LGG
Within-session thalamus	0.18	1	3
Between-sessions thalamus	0.20	1	4
Within-session orbitofrontal gyri	0.43	2	17
Between-sessions orbitofrontal gyri	0.47	2	20

Note.—Alpha=0.05, power=0.8, and sigma=SD_{diff}/sqrt(2). SD_{diff} = standard deviation across subjects of the within-subject difference; GBM = glioblastoma; LGG = low grade glioma.

Chapter 5

Amide Proton Transfer and Arterial Spin Labeling for Non-Invasive Molecular Stratification of Glioma: A Multi-Dataset Imaging Biomarker Study

Rajeev A. Essed, Ivar J.H.G. Wamelink, Jan Petr, Joost P.A. Kuijer, Alle Meije Wink, Shuncong Wang; Frederik Barkhof, Vera C. Keil

As submitted to Neuroradiology

Abstract

Background: Arterial spin labeling (ASL) and amide proton transfer chemical exchange saturation transfer (APT-CEST) imaging are noninvasive methods to probe brain tumor perfusion and protein content, potentially reducing the need for invasive biopsies. However, protocol variability and lack of large multi-center datasets limit their clinical translation.

Purpose: To evaluate the ability of APT-CEST magnetization transfer ratio asymmetry (MTR_{asym}) and ASL cerebral blood flow (CBF) features to predict isocitrate dehydrogenase (IDH) mutation status, 1p/19q co-deletion status, and tumor grade using univariate tests on our internal dataset. We assessed model generalizability via pooling and cross-dataset/cross-vendor validation on two external datasets, and examined if combining MTR_{asym} and CBF features improves predictions.

Materials and Methods: This multi-center study includes 219 grade 2–4 glioma patients from three datasets (Netherlands/D1: n=48, Siemens; Russia/D2: n=42, Philips; China/D3: n=129, Siemens). Descriptive and first-order radiomic features were extracted from MTR_{asym} (n=219) and CBF maps (n=90). Univariate diagnostic performance was assessed using ROC curves, and random forest radiomics classifiers were evaluated using 5-fold cross-validation internally and models were validated externally. Combined MTR_{asym}+CBF models were assessed on D1+D2 (n=90).

Results: Despite significant inter-site differences in absolute values, contralateral ROI-normalized ratios were comparable across sites. In univariate analysis, MTR_{asym} achieved AUCs of 0.94 (1p/19q), 0.87 (grade), and 0.76 (IDH); contralateral ROI-normalized CBF showed comparable performance for IDH (AUC=0.77) and 1p/19q (AUC=0.75). Radiomics models achieved AUCs of 0.91 (1p/19q), 0.81 (IDH), and 0.89 (grade) for MTR_{asym}; combined MTR_{asym}+CBF models significantly improved AUCs of 0.85 (IDH, p=0.007) and 0.91 (grade, p=0.001). External validation revealed complementary vendor-specific generalizability: MTR_{asym} performed well on D3 (Siemens), while normalized CBF generalized better to D2 (Philips). Pooled multi-center analysis achieved AUCs of 0.86 (IDH/grade) and 0.76 (1p/19q).

Conclusion: MTR_{asym} and normalized CBF univariate analyses and radiomics features demonstrate potential for internal molecular stratification of glioma. Combining MTR_{asym} and CBF may improve predictive performance. However, inter-dataset variability limits cross-site generalization, underscoring the need for standardized acquisition protocols before clinical translation.

Introduction

Gliomas are the most common primary malignant brain tumors and remain a critical neuro-oncological challenge, with glioblastoma (WHO grade 4) carrying a dismal prognosis despite aggressive therapy.[1] Tumor characterization and molecular profiling are vital for prognosis and treatment planning under the current WHO classification[2, 3] and rely on invasive surgical procedures to obtain tissue for analysis. Such procedures carry risks like infection, hemorrhage, and neurological deficit, while not always capturing the molecular heterogeneity of the tumor.[4, 5] Conventional MRI, including gadolinium-based contrast agent (GBCA)- enhanced sequences, is indispensable for initial diagnosis and surgical planning.[6] Still, it exposes patients to the potential risks of repeated contrast agent administration.[7, 8]

Advanced MRI techniques like arterial spin labeling (ASL) and amide proton transfer (APT) imaging aim to reduce reliance on more invasive methods requiring GBCA or PET-tracer injection.[9] ASL, a perfusion-weighted imaging (PWI) technique, quantifies cerebral blood flow (CBF) using blood protons as an endogenous tracer.[10–13] APT, on the other hand, is based on chemical exchange saturation transfer (CEST), detecting amide protons resonating at 3.5 ppm downfield from water, thereby reflecting intracellular protein concentrations.[14–16] APT-CEST-derived magnetization transfer ratio asymmetry (MTR_{asym}) around 3.5 parts per million (ppm) and ASL-derived CBF maps have previously demonstrated promise in predicting tumor grade, isocitrate dehydrogenase (IDH) mutation status, and 1p/19q codeletion status and have potential as non-GBCA dependent clinical assessment tools with comparable efficiency.[17–22] However, most APT-CEST and ASL literature lacks external validation datasets and uses varying pulse sequence parameters and post-processing methods, hindering external generalizability.[15, 23, 24]

Radiomics offers a framework to overcome these limitations by extracting quantitative features from MRI scans, allowing for the identification of tumor heterogeneity patterns that may correlate with tumor biology.[25–27] While there is criticism of using histogram radiomics features for biomarker identification due to reproducibility constraints and processing complexity[28], these features still contain rich information about voxel relationships and histogram shapes, which may be key to overcoming vendor- or sequence-based variability by describing the distribution of voxel intensities rather than only their raw values if correct normalization and/or harmonization approaches are applied.

We hypothesize that combining perfusion (CBF) and proteomic (MTR_{asym}) features within a comprehensive machine learning framework that relies on a larger combined multi-vendor dataset will synergistically improve the non-invasive prediction of histopathological characteristics (IDH, 1p/19q codeletion and WHO grade), as tumor angiogenesis and protein metabolism reflect distinct yet complementary biological processes.[15, 29–31] TAs literature on multimodal MTR_{asym}+CBF radiomics feature modeling for glioma mutation prediction is scarce, this study has two interrelated aims: first, to demonstrate that histogram-based radiomics can yield reproducible and clinically meaningful results when coupled with appropriate cross-site validation and harmonization strategies; and

second, to evaluate whether the combination of MTR_{asym} and CBF features improves classification of IDH mutation status, 1p/19q codeletion, and WHO grade beyond either modality alone.

Methods

Overview

This study comprised five main analytical components: (i) characterization of MTR_{asym} and CBF signal distributions across three international datasets, including assessment of inter-site and inter-vendor variability; (ii) univariate group comparisons and diagnostic performance evaluation of MTR_{asym} and CBF histogram features for classifying IDH mutation status, 1p/19q codeletion, and WHO grade using ROC analyses and logistic regressions; (iii) development of Random Forest classifiers using first-order radiomic features from single and combined modalities (MTR_{asym}+CBF), validated internally with 5-fold cross-validation over 10 repeats; (iv) assessment of cross-site and cross-vendor generalizability through external validation on two independent datasets (Russia, n=42, Philips; China, n=129, Siemens); and (v) evaluation of classification performance when pooling multi-center data.

Patient selection and Inclusion

Data acquisition in the Netherlands was approved by the institutional review board (VUmc_2021-0385), and all patients provided written informed consent. The external validation datasets were approved by their respective institutions, with written informed consent obtained from all patients, and are publicly available.[32–34]

Participants for the internal dataset were prospectively included between November 2022 and February 2025 meeting the following criteria: (i) pre-operative brain tumor imaging protocol MRI plus APT-CEST, and/or ASL sequences present; (ii) histomolecular verification of adult diffuse glioma diagnosis (according to the WHO CNS criteria (CNS5 2021)[3]) present; (iii) availability of molecular/histopathological data, including IDH mutation status, 1p/19q codeletion status, and tumor grade.[35]

Exclusion criteria were: (i) age <18 years; (ii) diagnosis not clearly fitting adult-type diffuse glioma; (iii) corrupted MTR_{asym} or ASL CBF maps.

MRI acquisition and pre-processing

Internal dataset (Netherlands, D1)

Scanning was performed at 3T (Vida, Siemens Healthineers, Erlangen, Germany) with a 20-channel head coil. The protocol included 3D T1-weighted MPRAGE (TR/TE/TI = 2300/2.32/900 ms, flip angle = 8°, resolution = 0.9 mm isotropic, FoV = 240 mm, 192 sagittal slices, GRAPPA = 2), T2-weighted TSE (TR/TE = 4100/74 ms, flip angle = 150°, resolution = 0.4×0.4×5.0 mm^3, FoV = 220 mm, 28 transversal slices), 3D FLAIR (TR/TE/TI = 5000/388/1650 ms, T2 prep = 125 ms, resolution = 0.5×0.5×0.9 mm^3, FoV = 230 mm, 176 sagittal slices, GRAPPA = 2), post-contrast 3D T1-weighted MPRAGE (identical parameters to pre-contrast), multi-post labeling delay (PLD) pseudocontinuous

ASL (pCASL) sequences, and a 3D TSE SPACE-CEST sequence. Detailed ASL and APT-CEST sequence parameters can be found in Tables S1 and S2. APT-CEST quantitative maps were calculated using the magnetization transfer ratio asymmetry (MTR_{asym}) at 3.5 ppm, defined as:

$$\text{APTCEST} = \text{MTR}_{\text{asym}}(3.5\text{ppm}) = \frac{S(-3.5\text{ppm}) - S(+3.5\text{ppm})}{S_0}$$

where S represents the signal intensity at the specified offset frequency, and S_0 is the reference signal with a specific saturation. CBF maps were quantified using ExploreASL[36]

External dataset 1 (Russia, D2)
Imaging was conducted on a 3T MRI system (Ingenia, Philips Healthcare, Best, The Netherlands) utilizing a 16-channel head coil.[32, 33] The acquisition protocol comprised 3D T1-weighted TFE (TR/TE = 6.63/3.01 ms, flip angle = 8°, resolution = 1 mm isotropic, matrix = 240×240×192, SENSE = 2.2), 3D T2-weighted TSE (TR/TE = 3500/300 ms, flip angle = 90°, resolution = 1 mm isotropic, matrix = 252×252×180, SENSE = 2.4×2.1, fat saturation), 3D FLAIR (TR/TE/TI = 4800/340/1650 ms, flip angle = 90°, resolution = 1.2 mm isotropic, matrix = 228×228×140, SENSE = 3×1.7, fat saturation), T1-weighted post-contrast (identical parameters to pre-contrast T1), SWI, DWI, ASL, and APT-CEST sequences. Detailed ASL and APT-CEST sequence parameters can be found in Tables S1 and S2. MTR_{asym} maps were calculated by the MRI scanner. Participants were adults histologically diagnosed with diffuse brain gliomas according to the WHO CNS criteria (CNS5 2021), and collected at Federal Neurosurgical Center, Novosibirsk, Russia, in 2023.

External dataset 2 (China, D3)
Imaging of the D3[34] was conducted on a 3T MRI system (Prisma, Siemens Healthcare, Erlangen, Germany), which included T1-weighted, T2-weighted, FLAIR, and T1-weighted post-contrast and APT-CEST imaging. Detailed APT-CEST sequence parameters are provided in Table S1, and quantitative maps were calculated in the same manner as for D1. Participants were adults with diagnosed diffuse brain gliomas according to the WHO CNS criteria (CNS5 2021), and collected at Tiantan Hospital, Beijing, China, from December 2020 to August 2022.

Registration, segmentation and ROI selection

Tumors were segmented automatically using the picture nnunet package[37, 38] on the structural MRI sequences and subsequently manually adjusted if necessary by two researchers (I.W. and R.E., with 3 and 2 years of experience in neuroradiology, respectively) under the supervision of a neuroradiologist (V.K., 15 years of experience) using ITK-SNAP (version 3.8.0). All researchers were blinded, and disagreements were resolved through consensus discussions. To avoid uncertainty regarding APT-based regions-of-interest (ROI), three distinct tumor compartments were delineated based on 3D structural MRI sequences: non-enhancing tumor/edema (label 1, based on FLAIR hyperintensity), necrotic tumor core (label 2, based on T1 hypointensity inside T1-post

contrast tumor borders), and enhancing tumor (label 3, based on T1-post contrast enhancing lesion parts).

The tumor ROI was defined by the enhancing tumor region (label 3) in case of an enhancing tumor. In the absence of tumor enhancement, the non-enhancing tumor components (label 1), were selected as the tumor ROI. Necrotic regions (label 2) were removed from the segmentation due to lack of fluid suppression in the APT-CEST scans in D1 and D2. A visual representation of the ROI selection is shown in Figure 1.

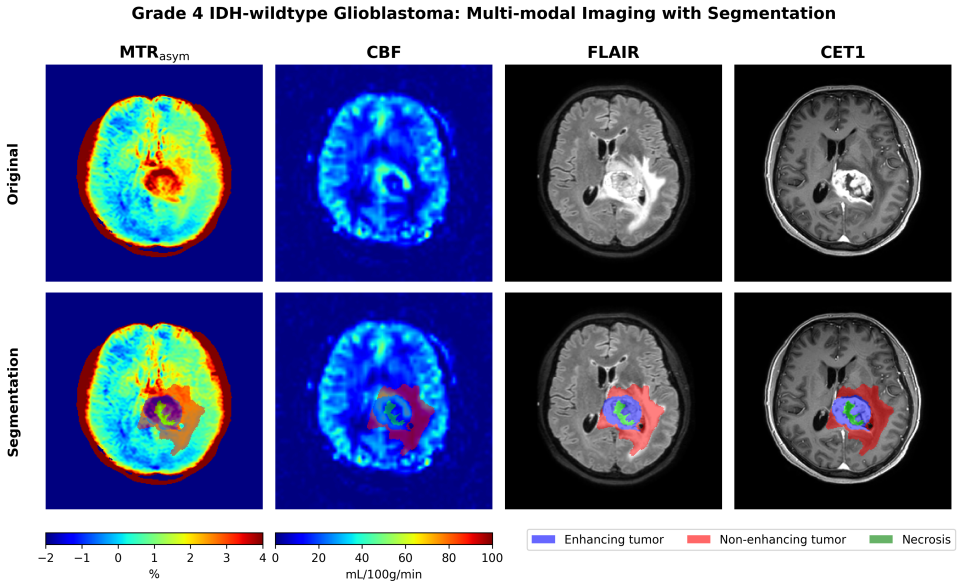


Figure 1: Magnetization transfer ratio asymmetry (MTR_{asym}), cerebral blood flow (CBF), Fluid-Attenuated Inversion Recovery (FLAIR), and post-contrast T1 (CET1) scans of a grade 4 IDH-wildtype glioma case with segmentation labels for enhancing tumor (blue), non-enhancing tumor (red) and necrosis (green).

For D1, APT-CEST and ASL sequences were co-registered to the corresponding 3D T1-weighted images. In contrast, in D2, structural and ASL-sequences were co-registered to APT-CEST images to minimize interpolation artifacts resulting from slice thickness variations. Registrations were performed using a rigid body transformation in FLIRT (FSL version 5.0.9). Image post-processing for D3 is described by Wu et al.[34].

Feature extraction

Mean MTR_{asym} and CBF values were extracted from 5 brain regions, being the whole brain, white matter (WM), grey matter (GM), the tumor ROI, and a similar-sized contralateral ROI. Additionally, for D1 and D2, first-order histogram features were extracted to characterize the distribution of MTR_{asym} and CBF values within the tumor ROI to focus on fundamental intensity-based differences across datasets. Using the open-source PyRadiomics library (version 3.0[39]), we extracted 18 first-order features which can be

found in the supplementary material. We restricted our analysis to first-order features as these are generally more reproducible across acquisition protocols and scanner platforms than higher-order texture features, which are highly sensitive to voxel size, resolution, and reconstruction methods.[40] All features were computed separately for the raw MTR_{asym} and CBF maps. Features from D3 had already been computed, and the source data could not be obtained.[34]

Statistical analysis

Datasets and Available Modalities D1 comprised 48 patients from the Netherlands with MTR_{asym} and CBF maps available alongside contralateral ROI reference tissue. D2 included 42 patients with MTR_{asym}, CBF, and contralateral ROI available. D3 contained 129 patients from China with MTR_{asym} features. For D1 and D2, both absolute tumor values and contralateral ROI-normalized tumor values (calculated as the tumor-to-contralateral ROI ratio) were computed for MTR_{asym} and CBF.

Mean MTR_{asym} and CBF signal intensities were compared between datasets and vendors using pairwise Mann-Whitney U tests to identify differences. The comparisons included: mean absolute tumor value, mean contralateral ROI value, and contralateral ROI-normalized tumor value for MTR_{asym} and CBF for dataset where available. Additionally, mean absolute tumor values were compared between enhancing versus non-enhancing tumors within D1 and D2.

Univariate Diagnostic Performance

Using the D1 dataset as an internal validation dataset, univariate diagnostic performance was assessed for three binary classification tasks: IDH status (mutant versus wildtype), 1p/19q codeletion status (codeleted versus non-codeleted), and WHO grade (low versus high grade). Three histogram features were selected for analysis: mean signal intensity, 90^{th} percentile signal intensity, and maximum signal intensity. These metrics were chosen to capture central tendency and the upper tail of the signal distribution within tumor regions, as elevated signals in the most metabolically active tumor subregions may provide complementary diagnostic information. Absolute tumor values and contralateral ROI-normalized values were analyzed for MTR_{asym} and CBF.

Group comparisons between molecular subgroups were performed using Mann-Whitney U tests for two-group comparisons and Kruskal-Wallis tests for three-group comparisons. Multiple-comparison correction was applied using the Benjamini-Hochberg false discovery rate (FDR) method across all histogram features and classification tasks. These comparisons are used to justify the selection of features (e.g. contralateral ROI normalized or not) for the diagnostic performance assessments.

Diagnostic performance was evaluated using receiver operating characteristic (ROC) curve analysis. Area under the ROC curve (AUC) with 95% confidence intervals was calculated using 2000 bootstrap iterations. To assess whether combining imaging modalities improved classification, logistic regression was used to integrate MTR_{asym} and CBF features, and DeLong tests were employed to compare AUCs between single-modality

and combined models.

Machine Learning Classification

Machine learning classification was performed to evaluate the predictive value of radiomic features for the same three binary classification tasks (IDH status, 1p/19q codeletion status, and WHO grade). A Random Forest classifier was used to model non-linear relationships between radiomic features and molecular status. The classifier was implemented using scikit-learn[41] with the following parameters: 100 estimators, maximum depth of 5, minimum samples per split of 10, and minimum samples per leaf of 5. These conservative hyperparameters were chosen to limit model complexity and reduce the risk of overfitting, given the relatively small dataset sizes. The feature set consisted of all 18 first-order radiomic statistics extracted from tumor regions.

Model performance was first established on the D1 dataset using 10 times repeated 5-fold stratified cross-validation, yielding 50 AUC values per model configuration. Three feature configurations were evaluated: MTR_{asym}-only features, CBF-only features (tested with both raw and Contralateral ROI-normalized versions), and combined MTR_{asym} and CBF features. The mean AUC and 95% confidence intervals were calculated from the distribution of the 50 AUC values using the t-distribution. Given that dataset-specific univariate and/or radiomics results have been previously published for D2 and D3, our study focused on combined and cross-site analyses.

To assess generalizability across institutions and vendors, models trained on D1 were subsequently applied to D2 and D3 as external validation datasets. Performance was evaluated following Z-score harmonization, in which external dataset features were transformed to match the D1 reference distribution. DeLong tests were used to compare performance between harmonized and non-harmonized conditions. Given the absence of CBF data in D3, this dataset was used only to validate MTR_{asym}'s generalizability.

Finally, to evaluate whether pooling data from multiple centers could improve classification performance, datasets were combined and evaluated using 5-fold stratified cross-validation. Pooling with Z-score harmonization to the D1 reference distribution was used. Four dataset combinations were tested to account for differences in data availability: MTR_{asym} features across all three datasets (n=217), MTR_{asym} features across D1 and D2 only (n=84), CBF features across D1 and D2, and combined MTR_{asym} and CBF features across D1 and D2.

To assess whether combined MTR_{asym} and CBF features provided synergistic improvement over single-modality models, paired t-tests were performed on the AUC values obtained from repeated cross-validation on the D1 internal diagnostic performance evaluation. Since both models were evaluated on identical fold partitions across all 10 repetitions, the resulting 50 AUC values were paired, allowing for a direct comparison that accounts for fold-to-fold variability. Statistical significance was set at $p<0.05$. For external and pooled multi-dataset evaluations, all AUC comparisons between the combined and unimodal models were performed using the DeLong test. Statistical significance was defined as $p<0.05$.

Results

Descriptive statistics

Patient demographics are provided in Table 1. Table 2 presents mean values for MTR_{asym} and CBF across the whole brain, contralateral ROI, and tumor ROI, stratified by tumor type and grade in D1. Supplementary Table S3 provides absolute and normalized histogram features (mean, 90th percentile, and maximum values) for different molecular subtypes based on IDH and 1p/19q codeletion status in D1. Supplementary Figure S1 visualizes the comparison of mean MTR_{asym} and CBF signal intensity between tumor types, including tumor-to-contralateral ROI ratios and within-type tumor versus contralateral ROI comparisons in D1.

Significant inter-dataset/vendor differences were observed in baseline signal intensities: mean tumor MTR_{asym} was significantly lower in D1 (1.94%±0.75%) compared to both D2 (2.61%±0.81%; p<0.001; r=0.44) and D3 (2.42%±1.25%; p=0.007; r=0.26). Mean contralateral ROI MTR_{asym} was also lower in D1 versus D2 (0.97%±0.58% vs 1.38%±0.56%; p<0.001; r=0.45). Mean CBF values showed the opposite pattern, with D1 demonstrating significantly higher mean tumor CBF (41.97±20.91 vs 30.09±12.55 ml/100g/min; p=0.002; r=-0.38) and mean contralateral ROI CBF (31.35±14.41 vs 21.43±8.44 ml/100g/min; p<0.001; r=-0.67) compared to D2. However contralateral ROI-normalized ratios were comparable (p>0.05) between datasets (MTR_{asym} normalized: D1 2.25±3.03, D2 2.13±1.42; CBF normalized: D1 1.36±0.63, D2 1.47±0.60). Detailed descriptive statistical comparisons across multiple datasets are shown in Figure 2.

Table 1: Demographics per dataset.

	Internal (NL), D1	External 1 (RU), D2	External 2 (CN), D3
Number of patients	48	42	129
Age (years)	56.6 ± 15.0	53.5 ± 14.3	54.64 ± 12.86
Sex (F:M)	21:27	23:19	60:69
IDH state (wt:mutated)	27:18	15:27	88:41
1p/19q state (codeleted:non-codeleted)	7:41	8:34	29:100
WHO grade distribution (2/3/4)	13/5/30	5/13/24	77/52*

* Only binary data available: grade 4 versus grade 2/3 *Abbreviations:* 1p/19q = Chromosome 1p and chromosome 19q, CN = China, D1 = Dataset 1, D2 = Dataset 2, D3 = Dataset 3, F = Female, IDH = Isocitrate dehydrogenase, M = Male, NL = Netherlands, RU = Russia, WHO = World Health Organization, wt = Wild-type

Group comparisons revealed that MTR_{asym} features were significantly associated with molecular subtypes, whereas CBF features showed more limited associations. For 1p/19q codeletion prediction, only MTR_{asym} maximum was significantly higher in non-codeleted than codeleted tumour patients after correction for multiple comparisons (p<0.001). For IDH mutation status, MTR_{asym} 90th percentile (p=0.017), mean (p=0.029), and normalized CBF 90th percentile (p=0.017) and mean (p=0.029) were significantly

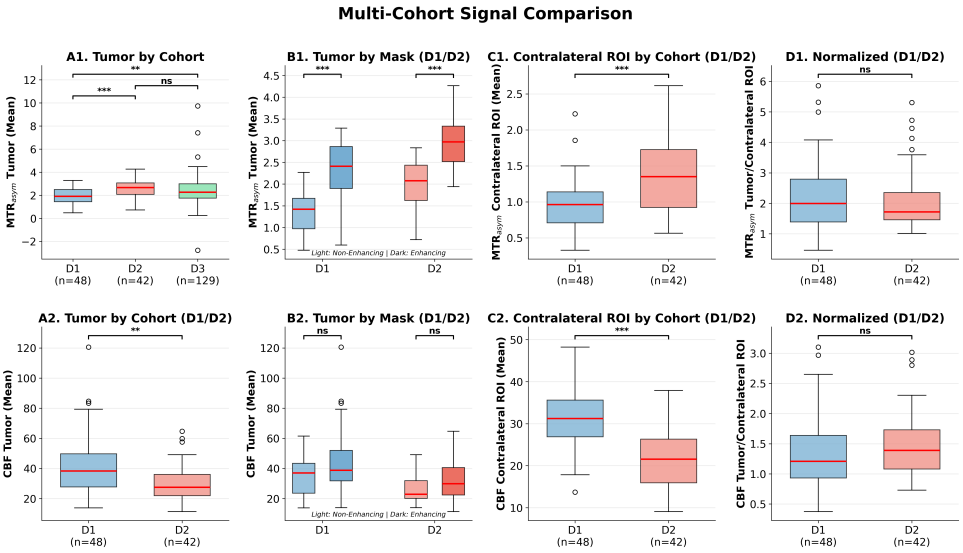


Figure 2: Box plots comparing magnetization transfer ratio asymmetry (MTR_{asym}) at 3.5 ppm (top row, A1–D1) and cerebral blood flow (CBF) (bottom row, A2–D2) across datasets and tumor regions. (A1) MTR_{asym} tumor signal across the Netherlands (n=48), Russia (n=42), and China (n=129) datasets. (B1) MTR_{asym} tumor signal in non-enhancing (n=33) versus enhancing (n=57) tumor regions. (C1) MTR_{asym} in normal-appearing white matter (NAWM) between the Netherlands and Russia datasets. (D1) Normalized MTR_{asym} (tumor/NAWM ratio) between datasets. (A2) CBF tumor signal between the Netherlands and Russia datasets. (B2) CBF tumor signal in non-enhancing versus enhancing tumor regions. (C2) CBF in NAWM between datasets. (D2) Normalized CBF (tumor/NAWM ratio) between datasets. ODG = oligodendroglioma; GBM = glioblastoma. Statistical significance: *p < 0.05, **p < 0.01, ***p < 0.001; ns = not significant.

different between IDH-wildtype and IDH-mutant tumors (see Figure 3 for illustration). Pairwise grade comparisons showed the biggest differences between grade 2 and grade 4 tumors, with MTR_{asym} 90th percentile (p=0.002), mean (p=0.002), and maximum (p=0.012) all significantly higher in grade 4; normalized CBF 90th percentile (p=0.008) and mean (p=0.023) also differed significantly. Grade 3 versus 4 comparisons showed only MTR_{asym} 90th percentile reaching significance (p=0.046), while no features distinguished grade 2 from grade 3 tumors.

Discriminative diagnostic performance mirrored these findings. For grade prediction, MTR_{asym} 90th percentile showed strong performance (AUC=0.87; 95% CI: 0.75–0.96), while normalized CBF 90th percentile achieved moderate discrimination (AUC=0.72; 95% CI: 0.56–0.87). For IDH prediction, MTR_{asym} and normalized CBF 90th percentile showed comparable moderate performance (AUC=0.76; 95% CI: 0.61–0.89 and AUC=0.77; 95% CI: 0.62–0.90, respectively). For 1p/19q codeletion, normalized CBF 90th percentile (AUC=0.75; 95% CI: 0.59–0.89) slightly outperformed MTR_{asym} (AUC=0.73; 95% CI: 0.57–0.87). Combining MTR_{asym} and CBF 90th percentile features improved performance for grade (AUC=0.88; 95% CI: 0.77–0.97), IDH (AUC=0.80; 95% CI: 0.65–0.92), and 1p/19q prediction (AUC=0.79; 95% CI: 0.63–0.92), however, DeLong tests were not significant for any comparison. Comparisons of signal intensities between groups and

Table 2: Mean MTR_{asym} and CBF values for the different tumor types and tumor grades in D1.

Group	N	Modality	Tumor ROI	Contralateral ROI	WM	GM	Whole Brain
IDH-mut 1p/19q-codel ODG	7	MTR_{asym}	1.53 ± 0.42	0.89 ± 0.64	0.62 ± 1.45	0.86 ± 1.05	0.80 ± 1.15
IDH-mut 1p/19q-codel ODG	7	CBF	30.66 ± 19.29	37.20 ± 33.07	24.46 ± 14.89	50.64 ± 14.57	38.80 ± 26.08
IDH-mut Astrocytoma	11	MTR_{asym}	1.52 ± 0.58	0.83 ± 0.61	0.48 ± 1.50	0.85 ± 1.05	0.79 ± 1.14
IDH-mut Astrocytoma	11	CBF	32.91 ± 24.29	34.50 ± 28.49	24.77 ± 14.42	50.51 ± 14.79	40.36 ± 23.29
IDH-wt GBM	30	MTR_{asym}	2.15 ± 0.84	1.03 ± 0.72	0.56 ± 1.48	0.87 ± 1.10	0.80 ± 1.20
IDH-wt GBM	30	CBF	43.85 ± 27.14	38.34 ± 26.33	26.07 ± 17.76	51.67 ± 14.81	37.92 ± 22.58
Grade 2	13	MTR_{asym}	1.35 ± 0.51	0.96 ± 0.65	0.57 ± 1.49	0.89 ± 1.05	0.82 ± 1.15
Grade 2	13	CBF	29.20 ± 22.21	33.39 ± 26.25	25.21 ± 15.24	49.71 ± 14.39	39.25 ± 24.47
Grade 3	5	MTR_{asym}	1.34 ± 0.44	0.88 ± 0.50	0.41 ± 1.40	0.81 ± 1.02	0.72 ± 1.12
Grade 3	5	CBF	47.03 ± 22.90	47.40 ± 33.87	30.16 ± 17.91	54.41 ± 15.21	41.02 ± 25.76
Grade 4	30	MTR_{asym}	2.26 ± 0.85	0.99 ± 0.73	0.56 ± 1.49	0.87 ± 1.10	0.80 ± 1.20
Grade 4	30	CBF	42.58 ± 27.18	37.30 ± 27.50	24.91 ± 17.01	51.40 ± 14.86	37.93 ± 22.30

* Only binary data available: grade 4 versus grade 2/3 *Abbreviations:* 1p/19q = Chromosome 1p and chromosome 19q, APT = Amide Proton Transfer, ASL = Arterial Spin Labeling, CBF = Cerebral Blood Flow, D1 = Dataset 1 from the Netherlands, GBM = Glioblastoma, IDH = Isocitrate dehydrogenase, MTR_{asym} = Magnetization Transfer Ratio asymmetry, N = Number of patients, NAWM = Normal-Appearing White Matter, ROI = Region of Interest, SD = Standard Deviation.

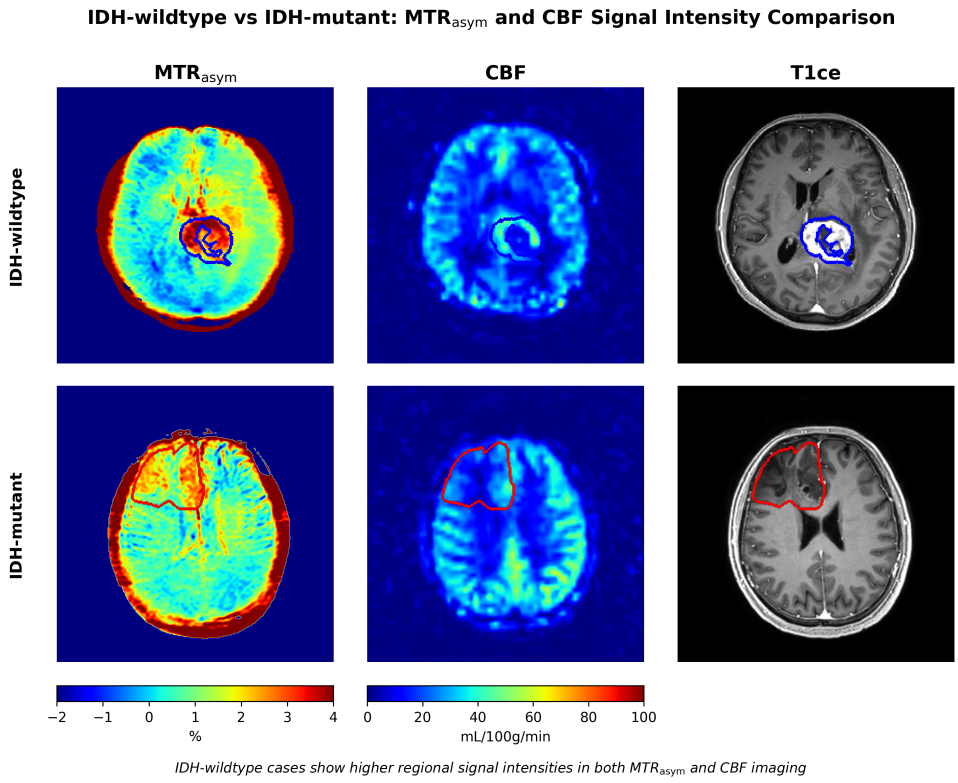


Figure 3: Example of enhancing IDH-wildtype (top row) versus non-enhancing IDH-mutant cases (bottom row) with raw signal intensity in magnetization transfer ratio asymmetry (MTR_{asym}) maps and cerebral blood flow (CBF) maps showing higher regional signal intensities for both MTR_{asym} and CBF in the IDH-wildtype case.

diagnostic performances can be found in Figures 4 and 5.

Radiomics

Dutch dataset evaluation

The predictive performance of MTR_{asym}, CBF, and the combined model for molecular and histological classification in the D1 (n=48) can be found in Figure 6. Multimodal modeling yielded significant improvements in AUC in IDH (p=0.007) and grade (p=0.001) classification compared to the best single modality. Diagnostic performance using the contralateral ROI-normalized variants of the MTR_{asym} and CBF maps can be found in Figure S5. Table S4 provides a side-by-side comparison between diagnostic performance of the best performing univariate feature compared to radiomics models.

External validation

IDH and 1p/19q prediction using MTR_{asym} with the previously trained model demonstrated good performance in D3, in contrast to the poor diagnostic performance of D2 (n=42) (see Figure 7). The contralateral ROI-normalized CBF models demonstrated

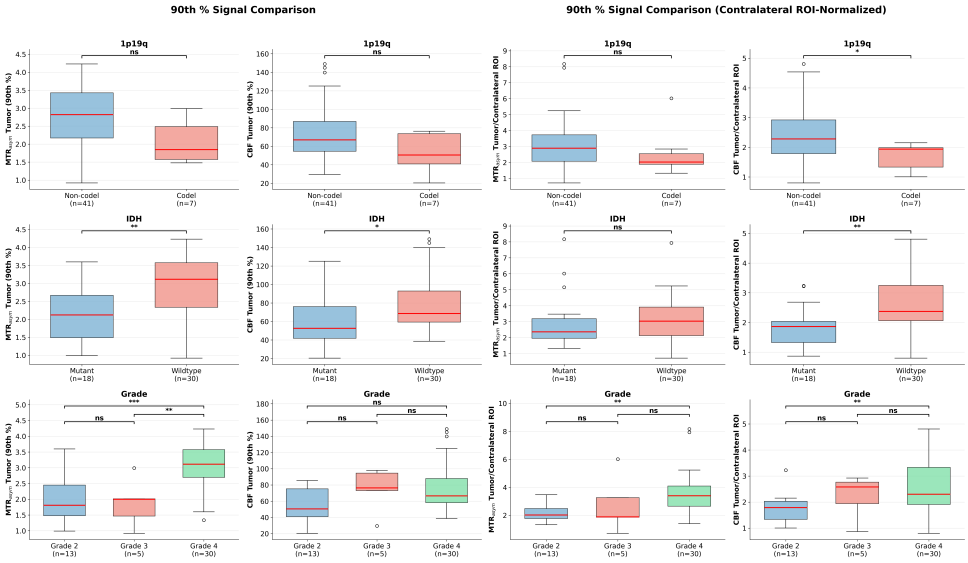


Figure 4: Boxplots comparing raw (left 2x3 plots) and normal appearing white matter (NAWM)-normalized (right 2x3 plots) 90th percentile MTR_{asym} and cerebral blood flow (CBF) tumor values between groups for 1p/19q co-deletion status (row 1), IDH status (row 2) and grade (row 3) in dataset D1).

more promising diagnostic utility when cross-validated on D2. Raw CBF results are shown in Figure S4.

Combined dataset

When combining all three datasets (n=219), MTR_{asym} achieved comparable classification performance for grade and IDH mutation compared to the D1 only radiomics analysis. In the two-site analysis combining the D1 and D2 datasets (n=90), the combined MTR_{asym} + contralateral ROI-normalized CBF model demonstrated the highest performance for grade and IDH prediction; however, the improvement was not significant over the best performing single modality ($p>0.05$). Figure 8 and S4 show contralateral ROI-normalized model results and raw model results, respectively.

Discussion

This multi-center study demonstrates that combining APT-CEST and ASL-derived features improves non-invasive prediction of IDH mutation status and WHO grade compared to single-modality approaches, supporting the integration of metabolic and perfusion imaging for their complementary tumor information. Furthermore, it shows that histogram-based radiomics can yield reproducible results for IDH and grade classification across datasets when coupled with appropriate normalization and harmonization strategies, however, cross-site performance did not fully match internal benchmarks.

Our research observed significant differences in baseline tumor and/or contralateral ROI signal intensities across datasets, reflecting the well-documented sensitivity of both

APT-CEST and ASL to acquisition parameters and scanner hardware.[15] Consequently, tumor-to-contralateral ratios showed no significant differences between datasets suggesting that internal tissue normalization is a pragmatic approach for mitigating intra-site variability in MTR_{asym} and CBF based features, similarly to amino-acid positron emission tomography (PET) imaging.[42]

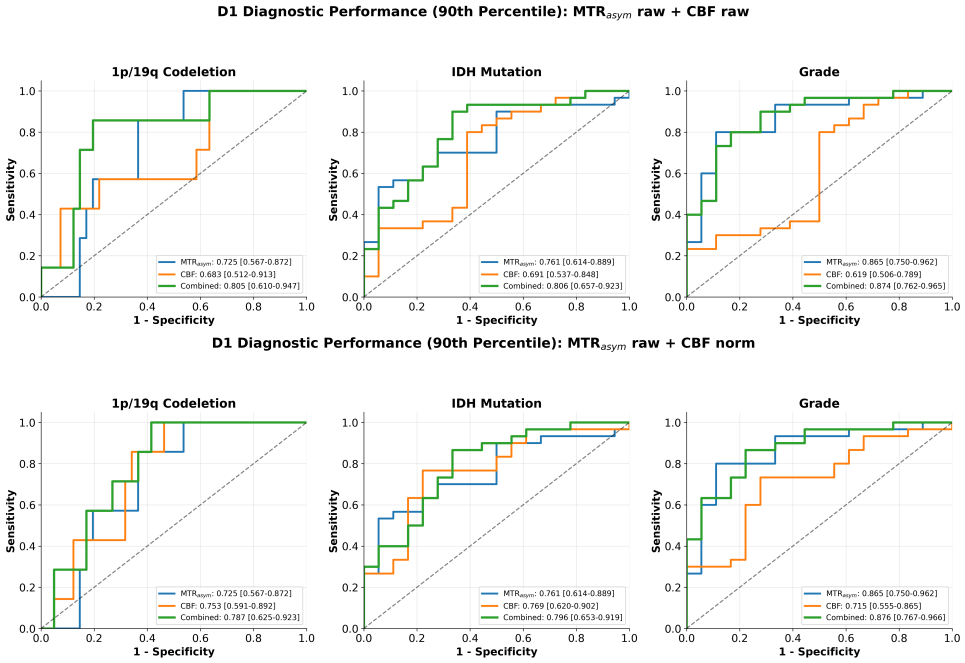


Figure 5: Area under the receiver operator characteristics curves (AUROC) for cerebral blood flow (CBF), magnetization transfer ratio asymmetry (MTR_{asym}) and combined (MTR_{asym}+CBF) in predicting IDH status, 1p/19q codeletion status, and grade (high versus low) using raw (top row) and normalized by normal appearing white matter (NAWM) (bottom row) 90th percentile tumor values in dataset D1.

Group comparisons revealed that MTR_{asym} and normalized CBF features demonstrated consistent and significant associations with glioma molecular subtypes. These findings are biologically plausible. The MTR_{asym} signal at 3.5 ppm predominantly reflects mobile protein and peptide concentrations, which are elevated in more proliferative and metabolically active tumors.[14–16] IDH-wildtype gliomas and higher-grade tumors typically exhibit greater cellularity, increased protein synthesis, and altered pH—all factors that contribute to elevated MTR_{asym} signal.[17, 19, 43] The strong association with 1p/19q codeletion status may reflect the distinct metabolic phenotype of oligodendrogliomas. However, this finding should be interpreted cautiously, given the small number of codeleted tumors in all datasets. The significant CBF associations with IDH status and grade align with the known hypervascular phenotype of IDH-wildtype and high-grade tumors, particularly glioblastomas.[22, 44, 45] While IDH-wildtype glioblastomas are often hypervascular, considerable overlap exists with IDH-mutant astrocytoma grade 3 and 4 tumors. Importantly, normalized CBF showed comparable univariate diagnostic performance to MTR_{asym} for

IDH mutation and 1p/19q codeletion prediction, suggesting that perfusion imaging provides complementary molecular information when appropriately normalized to account for inter-subject variability.

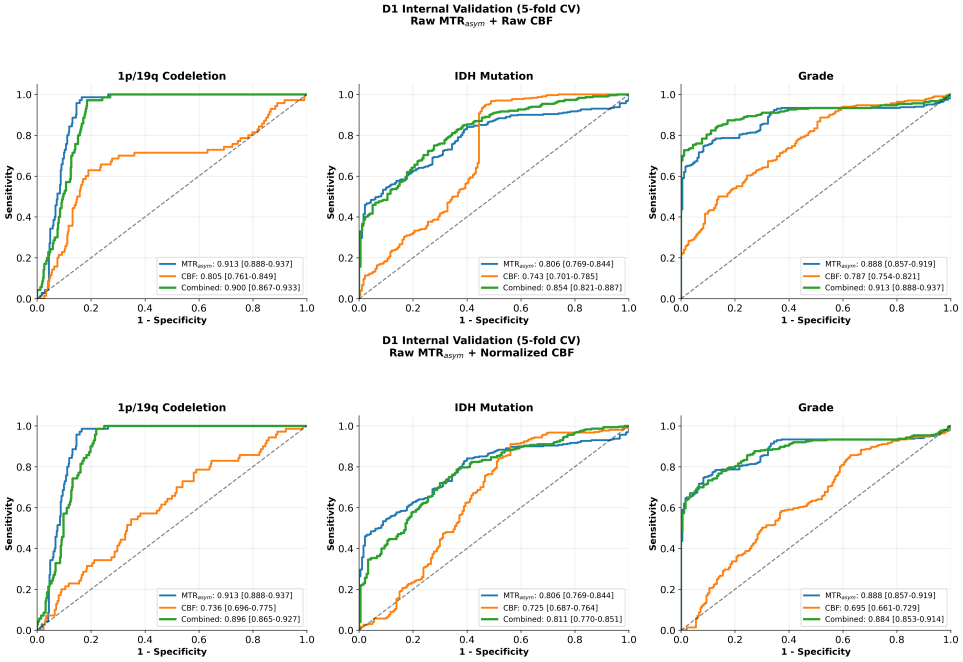


Figure 6: Area under the receiver operator characteristics curves (AUROC) for cerebral blood flow (CBF), magnetization transfer ratio asymmetry (MTR_{asym}) and combined (MTR_{asym}+CBF) in predicting IDH status, 1p/19q codeletion status, and grade (high versus low) using random forest radiomics models in dataset D1.

Beyond univariate associations, our radiomics models generally outperformed univariate features, indicating that the full radiomic feature set captures additional discriminative information beyond simple intensity statistics. The complementary contributions of MTR_{asym}+CBF to model performance may reflect distinct but related aspects of tissue pathophysiology. For example, tumors may be highly cellular yet relatively hypoperfused, or moderately cellular yet hypervascular. The non-redundant nature of these signals means multiparametric integration can provide a more complete portrait of tumor biology. Prior clinical studies similarly showed added value when combining MTR_{asym} with perfusion measures for glioma grading[17, 43] and prior research also observed this synergistic performance in treatment-response assessment.[46]

Cross-dataset validation revealed an intriguing pattern of complementary generalizability between modalities. MTR_{asym} models derived from D1 showed good generalization to D3 (Siemens scanner, matching our internal dataset) but poor performance on D2 (Philips scanner), while normalized CBF demonstrated the opposite pattern, outperforming MTR_{asym} on D2 for all classification tasks. This modality-specific vendor sensitivity may reflect differences in how CEST and ASL acquisitions are affected by scanner-dependent factors such as B0/B1 inhomogeneities, labeling efficiency, and se-

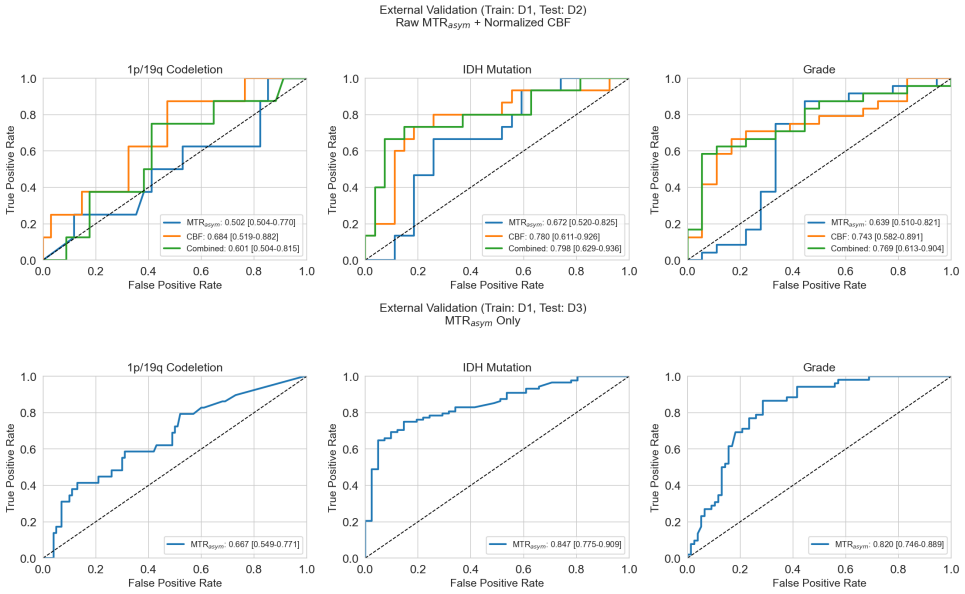


Figure 7: Area under the receiver operator characteristics curves (AUROC) for contralateral region of interest (ROI)-normalized cerebral blood flow (CBF), magnetization transfer ratio asymmetry (MTR_{asym}) and/or combined (MTR_{asym}+CBF) in predicting IDH status, 1p/19q codeletion status, and grade (high versus low) using random forest radiomics models on the D2 (top row) and D3 (bottom row) datasets.

quence implementation. Similarly to our internal radiomics model, multimodal classifiers demonstrated higher performance for both IDH mutation and grade classification tasks, however, these improvements did not reach statistical significance, likely due to the limited statistical power of the DeLong tests used in small samples (n=42). Additionally, the observation that contralateral ROI-normalized ratios were comparable across sites despite large differences in absolute intensities suggests that the cross-site performance drop for MTR_{asym} in D2 may be primarily driven by systematic intensity shifts rather than more complex technical confounders.

A limitation of our current analysis is the difference in tumor distribution across sites, which may have contributed to some of the observed variability in reported population values. Another methodological consideration is that we did not have full control over the quantification process: we obtained pre-quantified data from site D2 while performing the quantification ourselves for D1; this inconsistency could further reduce reproducibility. Radiomic studies frequently suffer from spurious associations arising from high-dimensional data and oversampling - as well as poor reproducibility due to inconsistent ROI segmentation and lack of external reproducibility/validation.[40, 47, 48] This is exacerbated for ASL and APT-CEST, where sparse literature, technical heterogeneity, inconsistent ROI selection standards, and small datasets limit clinical translation.[15] We correct this by additionally evaluating the diagnostic performance on independent datasets. Given the strong univariate performance of simple intensity metrics, future work should investigate whether radiomic models offer true added value over parsimonious

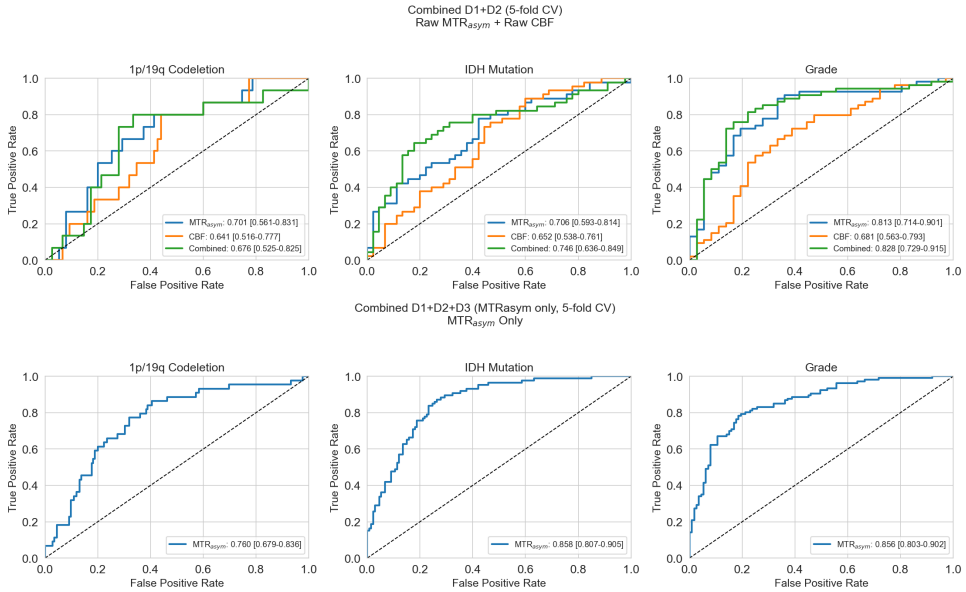


Figure 8: Area under the receiver operator characteristics curves (AUROC) for cerebral blood flow (CBF), magnetization transfer ratio asymmetry (MTR_{asym}) and combined ($MTR_{asym}MTR_{asym}$+CBF) in predicting IDH status, 1p/19q codeletion status, and grade (4 vs 2/3) using random forest radiomics models in pooled datasets (D1+D2 top row; D1+D2+D3 bottom row) using 5-fold cross validation.

approaches using only a handful of features (e.g., mean, maximum, standard deviation), and whether appropriate reference tissue normalization could sufficiently address cross-vendor variability without requiring more complex harmonization strategies.

In conclusion, both MTR_{asym} and CBF-based histogram features show promise for non-invasive prediction of glioma molecular characteristics, with contralateral ROI normalization effectively mitigating inter-site variability. Cross-vendor validation, however, reveals that harmonization benefits are context-dependent. Combined MTR_{asym}+CBF radiomics may offer modest advantages over single-modality approaches when both are available.

References

1 N. F. Brown et al. “Survival outcomes and prognostic factors in glioblastoma”. en. In: *Cancers (Basel)* 14 (13 June 28, 2022), p. 3161. doi: 10.3390/cancers14133161.

2 C. Horbinski et al. “NCCN guidelines® insights: Central Nervous System cancers, version 2.2022”. en. In: *J. Natl. Compr. Canc. Netw.* 21 (1 Jan. 2023), pp. 12–20. doi: 10.6004/jnccn.2023.0002.

3 D. N. Louis et al. “The 2021 WHO classification of tumors of the Central Nervous System: A summary”. en. In: *Neuro. Oncol.* 23 (8 Aug. 2, 2021), pp. 1231–1251. doi: 10.1093/neuonc/noab106.

4 C. Balana et al. “Preoperative diagnosis and molecular characterization of gliomas with liquid biopsy and radiogenomics”. en. In: *Front. Neurol.* 13 (May 26, 2022), p. 865171. doi: 10.3389/fneur.2022.865171.

5 M. Riche et al. “Complications after frame-based stereotactic brain biopsy: a systematic review”. en. In: *Neurosurg. Rev.* 44 (1 Feb. 2021), pp. 301–307. doi: 10.1007/s10143-019-01234-w.

6 F. M. Doniselli et al. “Recommendations on the use of gadolinium-based contrast agents in the diagnosis and monitoring of common adult intracranial tumours”. en. In: *Eur. Radiol.* 35 (12 Dec. 2025), pp. 7774–7785. doi: 10.1007/s00330-025-11646-6.

7 E. J. Bernstein, C. Schmidt-Lauber, and J. Kay. “Nephrogenic systemic fibrosis: a systemic fibrosing disease resulting from gadolinium exposure”. en. In: *Best Pract. Res. Clin. Rheumatol.* 26 (4 Aug. 2012), pp. 489–503. doi: 10.1016/j.berh.2012.07.008.

8 Y. Li et al. “Noninvasive determination of the IDH status of gliomas using MRI and MRI-based radiomics: Impact on diagnosis and prognosis”. en. In: *Curr. Oncol.* 29 (10 Sept. 23, 2022), pp. 6893–6907. doi: 10.3390/curroncol29100542.

9 N. Soni et al. “Amino acid tracer PET MRI in glioma management: What a neuroradiologist needs to know”. en. In: *AJNR Am. J. Neuroradiol.* 44 (3 Mar. 2023), pp. 236–246. doi: 10.3174/ajnr.A7762.

10 M. Taso and D. C. Alsop. “Arterial spin labeling perfusion imaging”. en. In: *Magn. Reson. Imaging Clin. N. Am.* 32 (1 Feb. 2024), pp. 63–72. doi: 10.1016/j.mric.2023.08.005.

11 J. G. Woods et al. “Recommendations for quantitative cerebral perfusion MRI using multi-timepoint arterial spin labeling: Acquisition, quantification, and clinical applications”. en. In: *Magn. Reson. Med.* 92 (2 Aug. 2024), pp. 469–495. doi: 10.1002/mrm.30091.

12 D. C. Alsop et al. “Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia”. en. In: *Magn. Reson. Med.* 73 (1 Jan. 2015), pp. 102–116. doi: 10.1002/mrm.25197.

13 T. Iutaka et al. “Arterial spin labeling: Techniques, clinical applications, and interpretation”. en. In: *Radiographics* 43 (1 Jan. 2023), e220088. doi: 10.1148/rg.220088.

14 J. Zhou et al. “Amide proton transfer (APT) contrast for imaging of brain tumors”. en. In: *Magn. Reson. Med.* 50 (6 Dec. 2003), pp. 1120–1126. doi: 10.1002/mrm.10651.

15 J. Zhou et al. “Review and consensus recommendations on clinical APT-weighted imaging approaches at 3T: Application to brain tumors”. en. In: *Magn. Reson. Med.* 88 (2 Aug. 2022), pp. 546–574. doi: 10.1002/mrm.29241.

16 H. Koike et al. “Amide proton transfer-chemical exchange saturation transfer imaging of intracranial brain tumors and tumor-like lesions: Our experience and a review”. en. In: *Diagnostics (Basel)* 13 (5 Feb. 28, 2023). doi: 10.3390/diagnostics13050914.

17 O. Togao et al. “Amide proton transfer imaging of adult diffuse gliomas: correlation with histopathological grades”. en. In: *Neuro. Oncol.* 16 (3 Mar. 2014), pp. 441–448. doi: 10.1093/neuonc/not158.

18 Y. Han et al. “Amide proton transfer imaging in predicting isocitrate dehydrogenase 1 mutation status of grade II/III gliomas based on support vector machine”. en. In: *Front. Neurosci.* 14 (Feb. 21, 2020), p. 144. doi: 10.3389/fnins.2020.00144.

19 S. Jiang et al. “Predicting IDH mutation status in grade II gliomas using amide proton transfer-weighted (APTw) MRI”. en. In: *Magn. Reson. Med.* 78 (3 Sept. 2017), pp. 1100–1109. doi: 10.1002/mrm.26820.

20 S. Jiang et al. “Applications of chemical exchange saturation transfer magnetic resonance imaging in identifying genetic markers in gliomas”. en. In: *NMR Biomed.* 36 (6 June 2023), e4731. doi: 10.1002/nbm.4731.

21 S. Okuchi et al. “Endogenous chemical exchange saturation transfer MRI for the diagnosis and therapy response assessment of brain tumors: A systematic review”. en. In: *Radiol. Imaging Cancer* 2 (1 Jan. 2020), e190036. doi: 10.1148/rycan.2020190036.

22 Y. Prysiazhniuk et al. “Diffuse glioma molecular profiling with arterial spin labeling and dynamic susceptibility contrast perfusion MRI: A comparative study”. en. In: *Neurooncol. Adv.* 6 (1 Jan. 2024), vdae113. doi: 10.1093/noajnl/vdae113.

23 A. Falk Delgado et al. “Arterial spin labeling MR imaging for differentiation between high- and low-grade glioma-a meta-analysis”. en. In: *Neuro. Oncol.* 20 (11 Oct. 9, 2018), pp. 1450–1461. doi: 10.1093/neuonc/noy095.

24 A. Alsaedi et al. “The value of arterial spin labelling in adults glioma grading: systematic review and meta-analysis”. en. In: *Oncotarget* 10 (16 Feb. 22, 2019), pp. 1589–1601. doi: 10.18632/oncotarget.26674.

25 S. A. Mali et al. “Making radiomics more reproducible across scanner and imaging protocol variations: A review of harmonization methods”. en. In: *J. Pers. Med.* 11 (9 Aug. 27, 2021), p. 842. doi: 10.3390/jpm11090842.

26 S. Sanduleanu et al. “Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score”. en. In: *Radiother. Oncol.* 127 (3 June 2018), pp. 349–360. doi: 10.1016/j.radonc.2018.03.033.

27 R. J. Gillies, P. E. Kinahan, and H. Hricak. “Radiomics: Images are more than pictures, they are data”. en. In: *Radiology* 278 (2 Feb. 2016), pp. 563–577. doi: 10.1148/radiol.2015151169.

28 A. Stanzione. “Feasible does not mean useful: Do we always need radiomics?” en. In: *Eur. J. Radiol.* 156 (110545 Nov. 2022), p. 110545. doi: 10.1016/j.ejrad.2022.110545.

29 A. Rajendran et al. “Amide proton transfer imaging-arterial spin labeling mismatch: a new imaging biomarker for pilocytic astrocytoma”. en. In: *Sci. Rep.* 13 (1 Sept. 29, 2023), p. 16377. doi: 10.1038/s41598-023-43235-2.

30 V. Dangouloff-Ros et al. “Arterial spin labeling to predict brain tumor grading in children: Correlations between histopathologic vascular density and perfusion MR imaging”. en. In: *Radiology* 281 (2 Nov. 2016), pp. 553–566. doi: 10.1148/radiol.2016152228.

31 H. Pang et al. “3D-ASL perfusion correlates with VEGF expression and overall survival in glioma patients: Comparison of quantitative perfusion and pathology on accurate spatial location-matched basis”. en. In: *J. Magn. Reson. Imaging* 50 (1 July 2019), pp. 209–220. doi: 10.1002/jmri.26562.

32 E. Filimonova et al. “Utilizing the amide proton transfer technique to characterize diffuse gliomas based on the WHO 2021 classification of CNS tumors”. en. In: *Neuroradiol. J.* 37 (4 Aug. 2024), pp. 490–499. doi: 10.1177/19714009241242658.

33 E. Filimonova et al. *Utilizing amide proton transfer technique to characterise diffuse gliomas based on WHO 2021 classification of CNS tumors*. Openneuro, 2023. doi: 10.18112/OPENNEURO.DS004717.V1.0.0.

34 M. Wu et al. “Amide proton transfer-weighted imaging and derived radiomics in the classification of adult-type diffuse gliomas”. en. In: *Eur. Radiol.* 34 (5 May 2024), pp. 2986–2996. doi: 10.1007/s00330-023-10343-6.

35 B. M. Ellingson et al. “Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials”. en. In: *Neuro. Oncol.* 17 (9 Sept. 2015), pp. 1188–1198. doi: 10.1093/neuonc/nov095.

36 H. J. M. M. Mutsaerts et al. “ExploreASL: An image processing pipeline for multi-center ASL perfusion MRI studies”. en. In: *Neuroimage* 219 (117031 Oct. 1, 2020), p. 117031. doi: 10.1016/j.neuroimage.2020.117031.

37 H. G. Pemberton et al. “Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18911. doi: 10.1038/s41598-023-44794-0.

38 F. Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. en. In: *Nat. Methods* 18 (2 Feb. 2021), pp. 203–211. doi: 10.1038/s41592-020-01008-z.

39 J. J. M. van Griethuysen et al. “Computational radiomics system to decode the radiographic phenotype”. en. In: *Cancer Res.* 77 (21 Nov. 1, 2017), e104–e107. doi: 10.1158/0008-5472.CAN-17-0339.

40 A. Traverso et al. “Repeatability and reproducibility of radiomic features: A systematic review”. en. In: *Int. J. Radiat. Oncol. Biol. Phys.* 102 (4 Nov. 15, 2018), pp. 1143–1158. doi: 10.1016/j.ijrobp.2018.05.053.

41 F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *arXiv [cs.LG]* (2012). doi: 10.48550/ARXIV.1201.0490.

42 M. Unterrainer et al. “Towards standardization of 18F-FET PET imaging: do we need a consistent method of background activity assessment?” en. In: *EJNMMI Res.* 7 (1 Dec. 2017), p. 48. doi: 10.1186/s13550-017-0295-y.

43 T. Wada et al. “Grading of gliomas using 3D CEST imaging with compressed sensing and sensitivity encoding”. en. In: *Eur. J. Radiol.* 158 (110654 Jan. 2023), p. 110654. doi: 10.1016/j.ejrad.2022.110654.

44 N. Wang et al. “Arterial spin labeling for glioma grade discrimination: Correlations with IDH1 genotype and 1p/19q status”. en. In: *Transl. Oncol.* 12 (5 May 2019), pp. 749–756. doi: 10.1016/j.tranon.2019.02.013.

45 A. M. Ahmadzadeh et al. “Predicting 1p/19q codeletion status in glioma using MRI-derived radiomics: A systematic review and meta-analysis of diagnostic accuracy”. en. In: *AJNR Am. J. Neuroradiol.* 46 (10 Aug. 14, 2025), pp. 2098–2106. doi: 10.3174/ajnr.A8771.

46 R. A. Essed et al. “Performance of amide proton transfer imaging to differentiate true progression from therapy-related changes in gliomas and metastases”. en. In: *Eur. Radiol.* 35 (2 Feb. 2025), pp. 580–591. doi: 10.1007/s00330-024-11004-y.

47 S. S. F. Yip and H. J. W. L. Aerts. “Applications and limitations of radiomics”. en. In: *Phys. Med. Biol.* 61 (13 July 7, 2016), R150–66. doi: 10.1088/0031-9155/61/13/R150.

48 J. E. Park et al. “Reproducibility and generalizability in radiomics modeling: Possible strategies in radiologic and statistical perspectives”. en. In: *Korean J. Radiol.* 20 (7 July 2019), pp. 1124–1137. doi: 10.3348/kjr.2018.0070.

Supplementary Material

List of all extracted first-order features:

1. Energy - The energy of voxel intensities
2. Total Energy - The total energy of voxel intensities
3. Entropy - Measure of randomness/uncertainty in intensity distribution
4. Minimum - Minimum gray level intensity in the ROI
5. 10th Percentile - The 10th percentile of voxel intensities
6. 90th Percentile - The 90th percentile of voxel intensities
7. Maximum - Maximum gray level intensity in the ROI
8. Mean - Average gray level intensity
9. Median - Middle value of the intensity distribution
10. Interquartile Range - Difference between 75th and 25th percentiles
11. Range - Difference between maximum and minimum values
12. Mean Absolute Deviation (MAD) - Average absolute deviation from the mean
13. Robust Mean Absolute Deviation (rMAD) - MAD calculated using the 10th-90th percentile range
14. Root Mean Squared (RMS) - Square root of the mean of squared intensities
15. Standard Deviation - Measure of intensity variability
16. Skewness - Measure of asymmetry of the distribution
17. Kurtosis - Measure of the "tailedness" of the distribution (note: PyRadiomics calculates standard kurtosis, which is +3 compared to IBSI's excess kurtosis)
18. Variance - Square of the standard deviation
19. Uniformity - Measure of homogeneity of intensity distribution

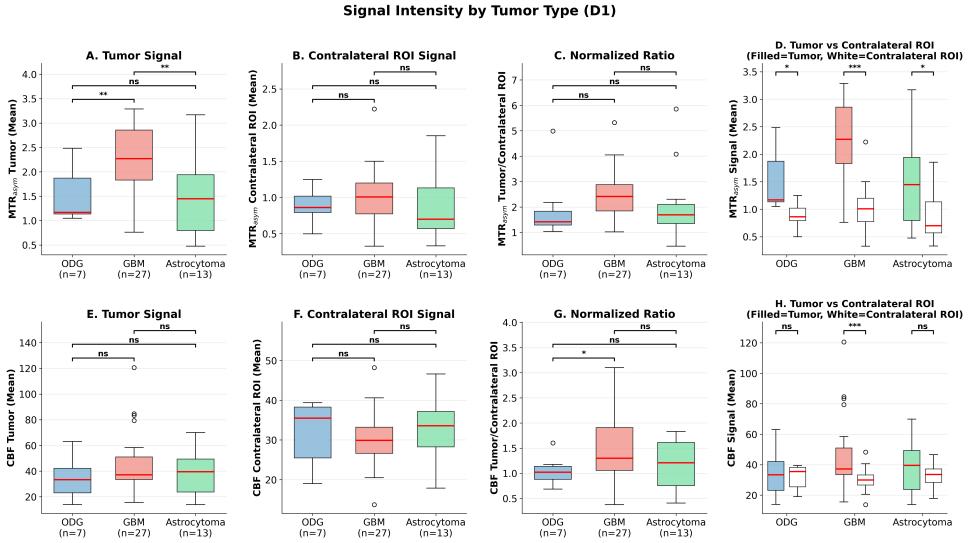


Figure S1: MTR_{asym}(3.5ppm) and CBF signal intensity comparison across glioma subtypes in the D1 dataset. (A) Mean tumor MTR_{asym}(3.5ppm) signal values stratified by histological subtype. (B) Mean contralateral region of interest (ROI) MTR_{asym}(3.5ppm) signal showed no significant differences across tumor types. (C) Tumor-to-contralateral ROI MTR_{asym}(3.5ppm) signal ratio did not differ significantly between groups. (D) Within-group comparison of tumor versus contralateral ROI MTR_{asym}(3.5ppm) signal intensity. (E) Mean tumor CBF values stratified by histological subtype showed no significant differences. (F) Mean Contralateral ROI CBF signal showed no significant differences across tumor types. (G) Tumor-to-contralateral ROI CBF ratio did not differ significantly between groups. (H) Within-group comparison of tumor versus contralateral ROI CBF signal intensity.

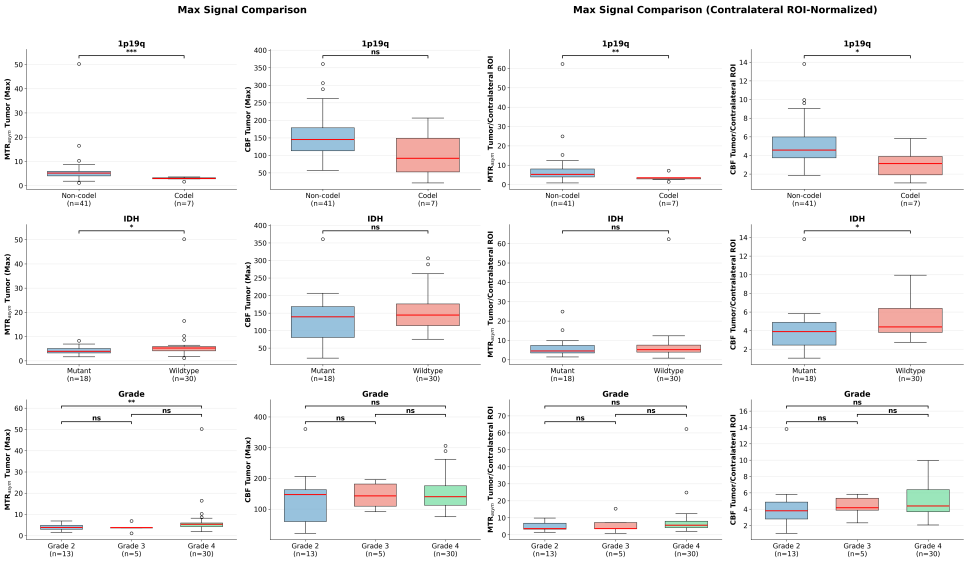


Figure S2: Boxplots comparing raw (A) and normal contralateral region of interest (ROI)-normalized maximum MTR_{asym} and cerebral blood flow (CBF) tumor values between groups for 1p/19q co-deletion status (row 1), IDH status (row 2) and grade (row 3) in dataset D1.

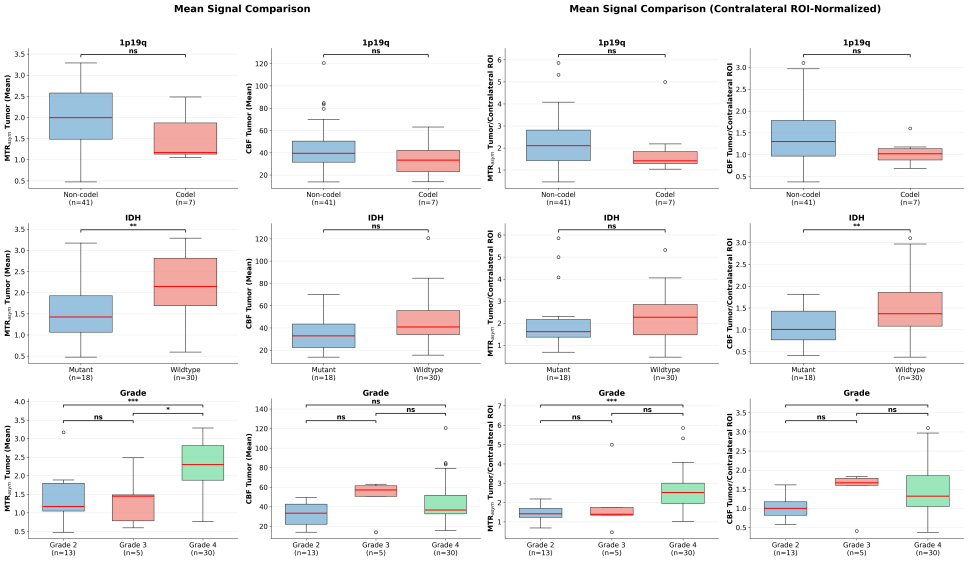


Figure S3: Boxplots comparing raw (A) and normal contralateral region of interest (ROI)-normalized mean MTR_{asym} and cerebral blood flow (CBF) tumor values between groups for 1p/19q co-deletion status (row 1), IDH status (row 2) and grade (row 3) in dataset D1.

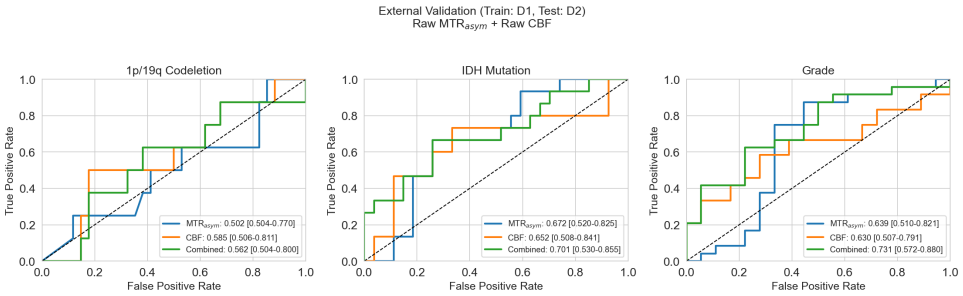


Figure S4: Area under the receiver operator characteristics curves (AUROC) for cerebral blood flow (CBF), magnetization transfer ratio asymmetry (MTR_{asym}) and combined (MTR_{asym}+CBF) in predicting IDH status, 1p/19q codeletion status, and grade (high versus low) using random forest radiomics models trained on the D1 dataset and tested on the D2 dataset using Z-score feature normalisation/harmonization

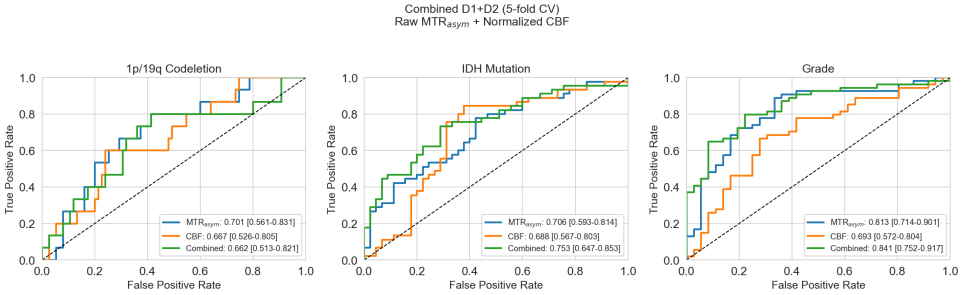


Figure S5: Area under the receiver operator characteristics curves (AUROC) for cerebral blood flow (CBF), magnetization transfer ratio asymmetry (MTR_{asym}) and combined (MTR_{asym}+CBF) in predicting IDH status, 1p/19q codeletion status, and grade (4 versus 2/3) using random forest radiomics models in pooled datasets using raw MTR_{asym} and CBF with z-score normalisation on combined D1 and D2 datasets.

Table S1: CEST imaging parameters

Dataset	Hardware	Saturation parameters	Offset frequencies	Readout	B0 correction	Voxel size
Netherlands	Vida Siemens 3T	T_{sat}=1.1s, t_p=100ms, t_d=10ms, $B1_{rms}$=2μT, DC_{sat}=91%	±3ppm, ±3.5ppm, ±4ppm, -1560ppm	3D TSE SPACE-CEST	dual-echo GRE	1.4×1.4×2.8mm^3
Russia (32,33)	Philips 3T	unknown	unknown	3D TSE	unknown	2×2×6mm^3
§ China (34)	Siemens Prisma 3T	T_{sat}=1s, t_p=1s, $B1_{rms}$=2.5μT, DC_{sat}=100%	±3ppm, ±3.5ppm, ±4ppm, -1560ppm	3D TSE SPACE-CEST	dual-echo GRE	2.8×2.8×2.8mm^3

Table S2: ASL imaging parameters

Dataset	Hardware	Sequence	Background suppression	TR/TE (ms)	Labeling duration (ms)	PLD times	Readout	Voxel size (mm^3)
Netherlands	Vida Siemens 3T	pCASL	Yes	5000/22.1	1800	1600, 2400, 3200, 3600, 3800, 4000, 4200, 4400	3D GRASE	1.8×1.8×4
Russia	Philips 3T	pCASL	Unknown	4300/11.6	1800	Unknown	3D	3.75×3.75×6

Table S3: Absolute and normalized histogram feature values for different tumor molecular subtypes and grades.

Modality	Group	N	Mean (abs)	90th (abs)	Max (abs)	Mean (norm)	90th (norm)	Max (norm)
APT	Non-codeleted	41	2.01 ± 0.75	2.74 ± 0.87	6.48 ± 7.32	2.90 ± 2.91	1.81 ± 0.82	0.85 ± 0.59
APT	1p/19q codel	7	1.53 ± 0.51	2.06 ± 0.55	2.88 ± 0.58	3.18 ± 3.47	1.45 ± 0.83	0.42 ± 0.24
CBF	Non-codeleted	41	43.27 ± 21.21	74.79 ± 29.98	154.82 ± 63.07	1.27 ± 0.65	1.10 ± 0.45	1.02 ± 0.49
CBF	1p/19q codel	7	34.39 ± 15.33	53.75 ± 19.84	103.26 ± 61.86	0.96 ± 0.31	0.80 ± 0.38	0.48 ± 0.42
APT	IDH mutant	18	1.57 ± 0.69	2.16 ± 0.71	4.32 ± 1.66	3.24 ± 3.28	1.65 ± 0.95	0.80 ± 0.77
APT	IDH wildtype	30	2.17 ± 0.67	2.93 ± 0.82	6.93 ± 8.46	2.75 ± 2.80	1.83 ± 0.73	0.78 ± 0.41
CBF	IDH mutant	18	35.11 ± 15.55	59.42 ± 24.73	133.96 ± 75.77	1.00 ± 0.36	0.86 ± 0.31	0.85 ± 0.63
CBF	IDH wildtype	30	46.09 ± 22.25	79.10 ± 29.94	155.31 ± 56.96	1.35 ± 0.70	1.17 ± 0.48	0.99 ± 0.42
APT	Grade 2	13	1.40 ± 0.65	1.98 ± 0.67	3.97 ± 1.37	1.91 ± 0.92	1.25 ± 0.32	0.56 ± 0.32
APT	Grade 3	5	1.36 ± 0.66	1.88 ± 0.69	3.86 ± 1.83	3.42 ± 4.03	1.58 ± 1.01	0.95 ± 0.56
APT	Grade 4	30	2.27 ± 0.57	3.06 ± 0.69	7.16 ± 8.40	3.30 ± 3.27	2.01 ± 0.83	0.86 ± 0.63
CBF	Grade 2	13	32.67 ± 11.66	56.56 ± 20.01	137.35 ± 83.80	1.03 ± 0.40	0.90 ± 0.33	0.90 ± 0.68
CBF	Grade 3	5	49.18 ± 18.22	74.27 ± 24.45	144.26 ± 40.04	1.02 ± 0.29	0.88 ± 0.27	0.99 ± 0.50
CBF	Grade 4	30	44.80 ± 22.74	77.86 ± 31.57	152.12 ± 59.01	1.34 ± 0.70	1.15 ± 0.49	0.95 ± 0.42

Abbreviations.—1p/19q = Chromosome 1p and 19q; abs = absolute; APT = Amide Proton Transfer; CBF = Cerebral Blood Flow; IDH = Isocitrate dehydrogenase; N = Number of patients; norm = normalized by contralateral normal appearing white matter.

Table S4: Univariate features versus radiomics model performance comparison.

Classification task	Modality	Best univariate feature	Univariate AUC	Random forest AUC
1p/19q Codeletion	MTR_{asym}	Max (raw)	0.944 (0.865–1.000)	0.913 (0.888–0.937)
1p/19q Codeletion	CBF	90^{th} percentile (norm)	0.753 (0.591–0.892)	0.805 (0.761–0.849)
1p/19q Codeletion	Combined	Max (raw_norm)	0.847 (0.665–0.978)	0.900 (0.867–0.933)
IDH Status	MTR_{asym}	90^{th} percentile (raw)	0.761 (0.614–0.889)	0.806 (0.769–0.844)
IDH Status	CBF	90^{th} percentile (norm)	0.769 (0.620–0.902)	0.743 (0.701–0.785)
IDH Status	Combined	90^{th} percentile (with CBF raw)	0.806 (0.657–0.923)	0.854 (0.821–0.887)
Grade (High vs Low)	MTR_{asym}	90^{th} percentile (raw)	0.865 (0.750–0.962)	0.888 (0.857–0.919)
Grade (High vs Low)	CBF	90^{th} percentile (norm)	0.715 (0.555–0.865)	0.787 (0.754–0.821)
Grade (High vs Low)	Combined	90^{th} percentile (with CBF norm)	0.876 (0.767–0.966)	0.913 (0.888–0.937)

Abbreviations.—1p/19q = Chromosome 1p and chromosome 19q, AUC = Area Under the Curve, CBF = Cerebral Blood Flow, IDH = Isocitrate dehydrogenase, Max = Maximum, MTR_{asym} = Magnetization Transfer Ratio asymmetry, vs = versus.

Chapter 6

Amsterdam Imaging and Clinical Glioma Database; IMAGO

Ivar J.H.G. Wamelink, Alle Meije Wink, Niels Verburg, Roelant S. Eijgelaar, Emmanouil Koltsakis, Marcus Cakmak, Henk J.M.M. Mutsaerts, Philip de Witt Hamer, Mathilde Kouwenhoven, Pieter Wesseling, Frederik Barkhof, Vera C. Keil

As submitted to Nature Scientific data

Abstract

The Amsterdam University Medical Centers IMAging in Glioma (IMAGO) database contains 1,700 adult-type diffuse glioma patients with structural MRI, histological, molecular, and survival information. Its first release includes 500 patients with preoperative structural MRI (T1w pre-contrast, T1w post-contrast, T2w, and T2-FLAIR). Postoperative MRI with the same sequences as the preoperative MRI are included for patients scanned within 14 days of surgery (381 patients). Pre- and postoperative apparent diffusion coefficient (ADC; derived from diffusion-weighted MRI; 497/500 and 377/381 respectively) and relative cerebral blood volume (rCBV; derived from dynamic susceptibility contrast MRI; 269/500 and 24/381 respectively) images are provided. All sequences are registered to MNI standard space and skull stripped. Histological and molecular data include IDH (500/500) and 1p/19q codeletion status (170/500), and WHO CNS5 (2021) tumor grade (500/500). Preoperative and postoperative nnUnet-based tumor region segmentations with human quality check are available for a subgroup of cases. The dataset is usable to create synthetic images, tumor segmentations, and to predict histological and molecular features in adult-type diffuse glioma.

Background & Summary

Research involving patients with adult-type diffuse glioma has expanded significantly, driven by a convergence of recent technological and methodological advancements. One key contributor has been the evolution of magnetic resonance imaging (MRI), which has enabled a more detailed and multi-parametric characterization of brain tumors, including structural, functional, and metabolic information.[1–4] The increasing availability of high-resolution imaging data has, in turn, facilitated the application of artificial intelligence (AI) and machine learning approaches to neuro-oncology.[4, 5] AI-driven techniques have shown considerable promise across a wide range of neuro-oncological applications, including automated tumor segmentation, which reduces observer variability and could enhance workflow efficiency.[6–9] Furthermore, predictive modeling using radiomic and deep learning features extracted from MRI has enabled more accurate survival prognostication and treatment response assessment.[9] Recent advances also include the development of synthetic imaging methods, such as generating contrast-enhanced T1w imaging.[10, 11] Collectively, these innovations are reshaping the landscape of diffuse glioma research, facilitating more personalized, data-driven approaches to diagnosis, prognosis, and therapeutic planning.

A critical factor enabling these advancements is the increasing availability of accessible external MRI datasets. Notable examples include the Multimodal Brain Tumor Segmentation (BraTS)[12–14] challenge and The Cancer Genome Atlas Glioblastoma Multiforme (TCGA-GBM)[15] dataset. More recently, these resources have been complemented by additional datasets such as the Rotterdam Erasmus Glioma Database (EGD)[16] and the University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM)[17] collections.

Despite the growth of open source imaging data, these datasets typically include a

limited set of clinical variables and are often restricted to conventional structural MRI sequences, such as T1-weighted (T1w, pre- and post-contrast), T2-weighted (T2w), and T2-FLAIR imaging. This constraint highlights the ongoing need for more comprehensive and multimodal data resources to support continued progress in both imaging research and AI-based approaches in glioma studies.

The Imaging in Glioma (IMAGO) database[18] contains MRI and clinical data of 1,700 subjects with cerebral adult-type diffuse gliomas scanned between 2010 and 2024. This first release of the IMAGO database[18] consists of 500 patients and contains four different data types: (1) preoperative and postoperative structural MRI and perfusion and diffusion MRI where available; (2) histological and molecular labels specifying the WHO CNS5 (2021) tumor types; (3) clinical parameters; (4) nnU-Net-based segmentations of edema/non-enhancing lesion, necrosis, and enhancing tissue.

The inclusion of perfusion and diffusion MRI is important as these sequences provide critical information on tumor vascularity and cellularity, respectively. Perfusion MRI helps assess tumor angiogenesis and blood flow, while diffusion MRI informs on tumor cellular density and potential treatment response.[1] Together, they enhance the characterization of glioma biology and are therefore a valuable addition to the dataset. The inclusion of follow-up imaging, an increased number of pre-operative scans, and more extensive longitudinal clinical data has the potential to substantially enhance the performance of predictive AI models and to support clinical decision-making in patients with adult-type diffuse cerebral gliomas. Furthermore, the availability of such data can contribute to improved patient management by enabling more accurate modeling of tumor progression and prediction of growth dynamics over time. In addition, incorporating data from multiple centers can increase the robustness and generalizability of these segmentation and predictive models, making them more trustworthy and easier to implement across diverse clinical settings.[19, 20] Finally, the use of external validation datasets is one of the criteria for the METRICS and CLEAR tools.[21, 22]

Methods

Data from patients with an adult-type diffuse glioma treated at the Amsterdam University Medical Center location VUmc, The Netherlands, between 2010 and 2024 were retrospectively collected. Patients were included if they were 16 years or older, if isocitrate dehydrogenase (IDH) status was known, and in the case of IDH-mutated patients if 1p/19q codeletion status was known. This study was approved by the institutional review board (VUmc_2021-0385) and all patients gave written informed consent. Additionally, all four structural sequences—pre-contrast T1w, post-contrast T1w (CET1), T2w, and T2-FLAIR; preferably 3D—needed to be available for the preoperative MRI. Subsequently, the surgery date was used to identify the first postoperative MRI. Postoperative MRI scans were only included if the patient was scanned within 14 days of the surgery to avoid confounders from, e.g., early progression. All four structural sequences also had to be available for the postoperative MRI.

MRI

Structural preoperative MRI were retrospectively collected from the hospital picture archiving and communication system (PACS) (pre-contrast T1w, CET1, T2w, and T2-FLAIR) for all patients. The MRI sequences were automatically labeled using an in-house Digital Imaging and Communications In Medicine (DICOM) labeling tool CINDERELLA,[23] which uses a minimal number of sequence parameters for classification. T2w images were scanned as 2D multislice. Apparent diffusion coefficient (ADC) maps are provided when available. Relative cerebral blood volume (rCBV) maps are provided when available and are superseded by motion-corrected rCBV maps if available. All rCBV maps were obtained using a 50% Dotarem contrast agent preload followed by the 50% dose dynamic scanning with a 4 mL per second injection rate using the autoinjector to the antecubital vein. The ADC and rCBV maps were post-processed with scanner-integrated software and are, therefore, not post-processed with the same software. Units may differ between images. rCBV maps from before 2022 were leakage corrected, while the presence of leakage correction before 2022 was inconsistent. An overview of all available images can be found in Figure 1.

Images were converted from DICOM to Neuroimaging Informatics Technology Initiative (NIfTI) using dcm2niix version v1.0.20230411. All sequences were registered to the CET1 images. ADC maps were registered to the CET1 images via the T2w images. The post-contrast T1w scan was registered to the Montreal Neurological Institute (MNI) 152 atlas, version ICBM 2009a nonlinear, which has a 1 mm isotropic resolution and a size of 197 x 233 x 189 voxels, using ANTs version 0.4.2 in Python version 3.9.18.[24] After registration to the MNI152 atlas, the CET1 images were skull stripped using HD-BET version 1.0. The computed CET1-to-MNI152 atlas transform and brain mask were subsequently applied to all other sequences.

Technical details on scanner and sequence settings, such as MRI vendor, scanner model, repetition time, and inversion time (among others) are provided per sequence.

Clinical Data

Clinical information was retrospectively obtained from the electronic patient system. The postoperative MRI scan is only provided if the MRI was performed within 14 days of surgery (arbitrary cut-off to avoid the inclusion of cases showing potentially early progression and to ensure comparability among patients).The number of days the early-postoperative MRI were scanned are additionally marked to identify cases performed within the 72 h window. Tumor enhancement was obtained from the preoperative radiologic reports. All patients had undergone surgery (biopsy or resection) in the period 2010–2021 as part of their routine clinical care and were diagnosed with an adult-type diffuse glioma. Between 2014 and 2016, the genetic characteristics of the tumors were mostly assessed using immunohistochemistry for the IDH R132H mutant protein, in a subset of cases (especially those histologically showing oligodendroglial characteristics) supplemented with loss of heterozygosity (LOH) analysis for the chromosome 1p/19q codeletion status. During this phase, certain IDH-mutant gliomas may not have been detected as such due to limitations inherent to this diagnostic approach. However, approximately 90% of the IDH-mutant gliomas carry the IDH1 R132H mutation and can

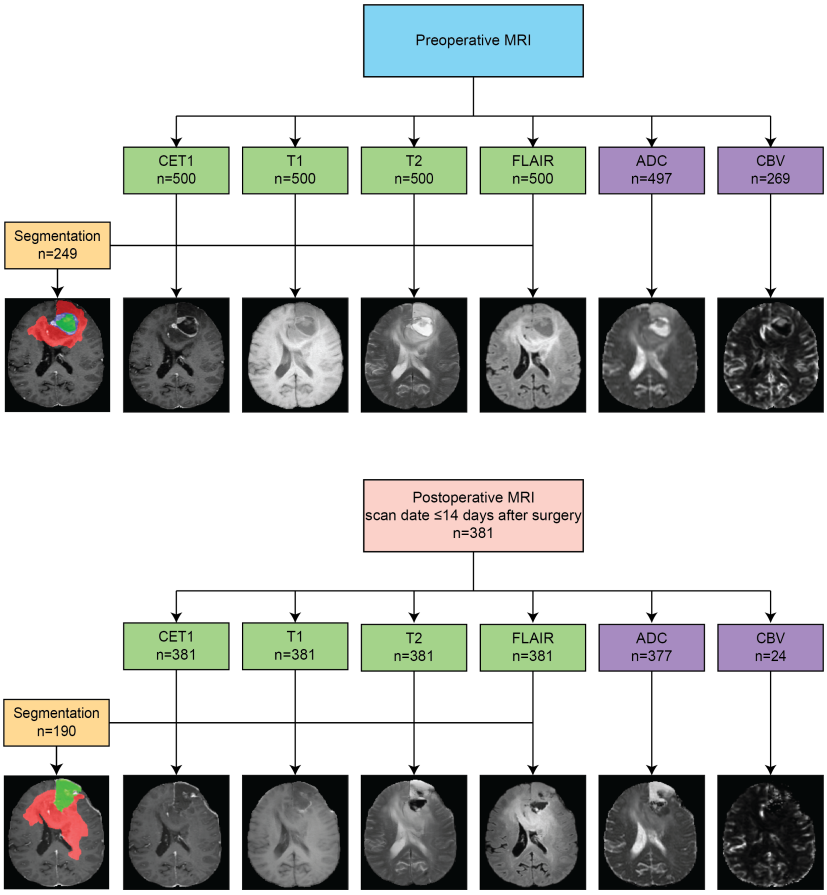


Figure 1: An overview of all available MRI sequences. Postoperative MRI scans are included onlyif performed within 14 days after surgery.

thus be identified as such by immunohistochemistry. Therefore, the vast majority of cases diagnosed between 2014 and 2016 can be expected to be accurately classified.[25] Between 2016 and 2019, a dedicated glioma DNA next-generation sequencing (NGS) panel was mostly used to screen for IDH1 and IDH2 hotspot mutations as well as for 1p/19q codeletion. Since 2019, methylome profiling has been applied to analyze these markers.[26] For patients who underwent surgery between 2010 and 2014, the IDH mutation and 1p/19q codeletion status of their gliomas were determined from tissue obtained at recurrence and analyzed at a later time. For other patients, retrospective assessment was conducted using immunohistochemistry for the IDH1 mutant protein, and in some cases, this was supplemented with LOH analysis for 1p/19q.

Tumor Segmentations

All images were automatically segmented using the PICTURE pre- and postoperative segmentation models.[6, 27–29] The preoperative model provides the following labels: 0) background, 1) edema/non-enhancing, 2) necrosis, and 3) enhancing tissue; the postoperative model solely provides the enhancing tumor residue. The postoperative model was only applied if the postoperative MRI scan occurred within three days of the surgery, as instructed by the software developers. Both models used all four structural MRI sequences. The segmentations were visually checked for acceptance by a neuroradiologist (6 years of experience).

Data Record

IMAGO has been made accessible through EUCAIM on the Health-RI XNAT platform and can be downloaded free of charge from the Health-RI website by academic researchers.[18] Access is provided through the EUCAIM platform. By default, the dataset may not be used for industry developments.

Patients can be downloaded separately, by session (i.e., pre- and postoperative), by scan type, and by data format (NIfTI and JSON). The NIfTI files have the study patient ID, scan session and scan type in the name, which allows the downloader to re-create a complete BIDS-compliant directory tree.

MRI

For all 500 patients, the images are provided as NIfTI files with the following names: D_T1.nii.gz (T1w precontrast MRI), `D_CET1.nii.gz` (T1w postcontrast MRI), `D_T2.nii.gz` (T2w MRI), `D_FLAIR.nii.gz` (T2-FLAIR MRI), `ADC.nii.gz` (DWI MRI), and `CBV.nii.gz` (DSC MRI), with `D` being the scanning dimension (2D or 3D). The scans were acquired on 23 different scanners from 4 vendors: General Electric (GE), Boston, Massachusetts, USA (DISCOVERY MR750 (3T), Signa HDxt (1.5T), and OPTIMA MR360 (1.5T)), Philips Healthcare, Best, The Netherlands (Achieva (1.5T and 3T), Ingenia (1.5T and 3T), Ingenia Elition X (3T), Ingenia Elition S (3T), Achieva Stream (1.5T and 3T), and Intera (1.5T)), Siemens Healthineers, Erlangen, Germany (Avanto (1.5T), Vida (3T), Sonata (1.5T), Avanto fit (1.5T), Aera (1.5T), MAGNETOM Sola (1.5T), Skyra (3T), Espree (1.5T), Verio (3T), ESSENZA (1.5T), Symphony (1.5T)), and Toshiba Medical Systems Corporation, Otawara, Japan (Titan3T (3T)).

Postoperative MRI scans are available for 381/500 patients. Axial 2D ADC and rCBV maps are available for 497/500 and 269/500 patients preoperatively and 377/381 and 24/381 postoperatively. Technical details on scanner and sequence settings, such as MRI vendor, scanner model, repetition time, and inversion time (among others) are provided in accompanying `.json` files per sequence.

Clinical data

Per patient, the following clinical information is provided: sex (male/female), date of preoperative MRI (in days after birth), year of diagnosis, age at preoperative MRI (age at

radiology detection), surgery type (biopsy (B), partial resection (P), gross total resection (G); obtained from the first postoperative radiological report), tumor type (glioblastoma IDH-wildtype, astrocytoma IDH-mutant, and oligodendroglioma IDH-mutant and 1p/19q-codeleted), CNS WHO (2021) grade (2, 3, 4), surgery date (days after preoperative MRI), the number of days the postoperative MRI was obtained after surgery, preoperative Karnofsky score (KPS) (250/500 patients), date of death (in days after birth or -1 if patient was still alive on 18-10-2024), enhancement (1: enhancement; 0: no enhancement), IDH state (wt-IDH and IDH-mutation; 500/500) and 1p/19q state (codeleted, not codeleted, and unknown) (170/500 patients). Clinical data are provided in a CSV file entitled Clinical_data.csv. Dataset demographics can be found in Table 1.

Table 1: Dataset demographics grouped by tumor type.

Tumor type	IDH-mutant 1p/19q codeleted oligoden-droglioma	IDH-mutant astrocytoma	IDH wild-type glioblastoma
Total	60	72	368
Sex			
Male	38	39	226
Female	22	33	142
Age at diagnosis (years)	46 (± 14.2)	39 (± 11.7)	59 (± 12.3)
Surgery type			
Biopsy	4	7	95
Partial	45	54	232
Gross total	11	11	41
WHO grade			
Grade 2	39	41	0
Grade 3	21	19	0
Grade 4	–	12	368
Number of days preoperative MRI was before surgery	24.2 (± 32.9)	14.3 (± 20.6)	8.2 (± 11.8)
Number of days postoperative MRI was after surgery	1.6 (± 0.9)	2.1 (± 2.15)	1.8 (± 1.6)
IDH state	60	72	368
1p/19q codeletion state	60	72	23

Tumor segmentations

Segmentations are provided for 249/500 preoperative cases and 190/500 postoperative cases.

Technical Validation

All patients were diagnosed with adult-type diffuse gliomas according to the WHO CNS5 (2021) criteria. Only patients with availability of all four structural MRI sequences (T1w precontrast, CET1, T2w, and T2-FLAIR) were included. MR images were skull stripped for anonymization. Cases were assigned a random number; to ensure anonymity, no key-file was stored. All automatically generated segmentations were verified by an experienced neuroradiologist and not included if incorrect. The data have been directly downloaded from the Health-RI website and assessed to ensure all images were in the same atlas and skull stripped. Bar plots (Figure 2) and patient survival curves (Figure 3) were generated, demonstrating survival outcomes that were comparable to or exceeded those reported in previous studies,[30, 31] and showed distributions as reported in other papers.

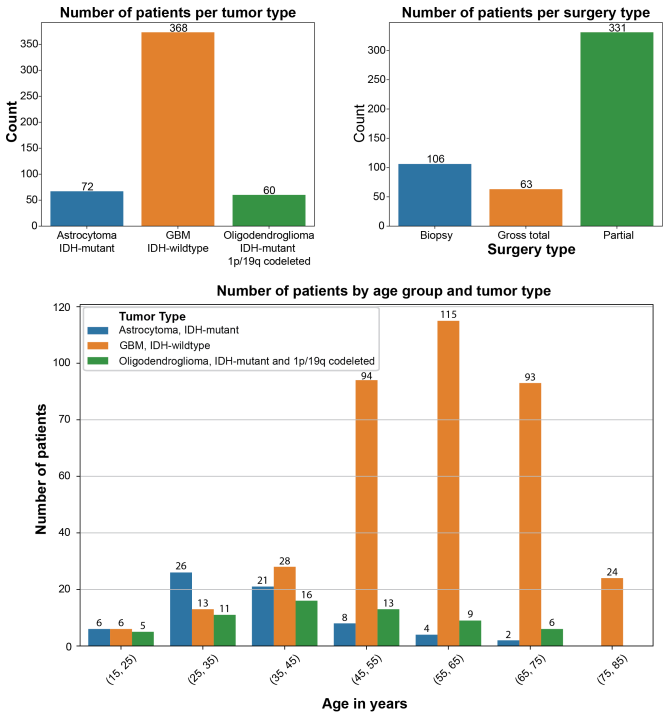
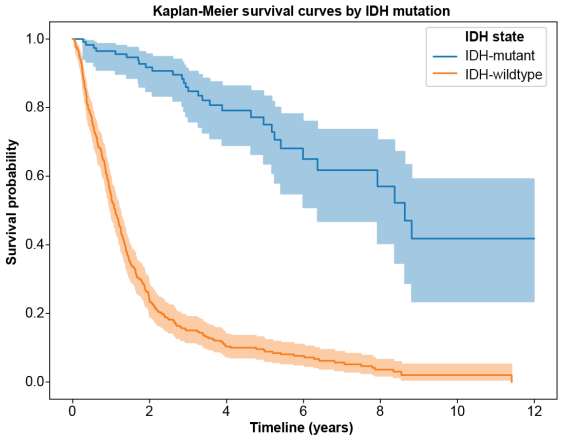
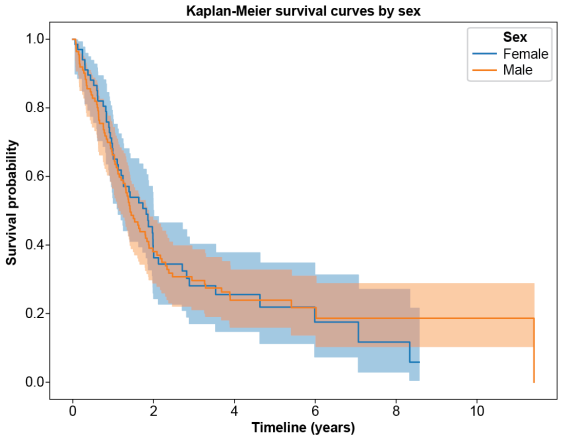


Figure 2: Bar graph of number of patients per surgery type and tumor type. Pane 2c shows thenumber of patients by age group and tumor type (bins of 10 years).





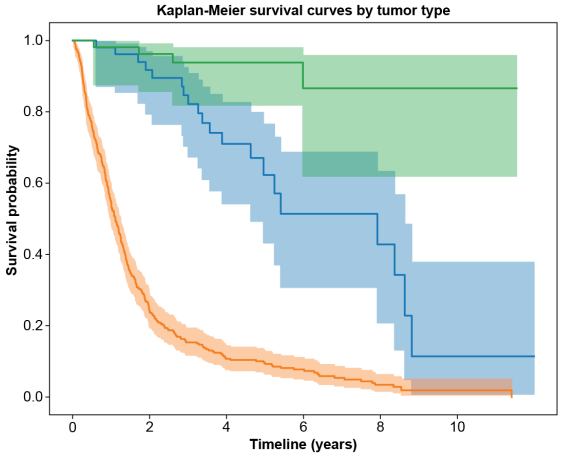


Figure 3: Kaplan-Meier survival curves byIDH mutation, sex, and tumor type.

Usage notes

Release 1.0 of the dataset is available through the EUCAIM platform and can be obtained after completing the registration steps. Users first register with EUCAIM (EUCAIM Catalogue) and request access to the IMAGO database (EUCAIM Catalogue). This can be done by (1) adding the database; (2) requesting the database; (3) sending the request to the negotiator. Then, the user can request an account for the data host by filling out the Health-RI request form on the EUCAIM helpdesk page and download the data.[18] MRI data can be visualized by platforms that support NIfTI format. Clinical data can be opened with any CSV viewer (e.g., Excel).

Code Availability

DICOMs were converted to NIfTI using dcm2niix version v1.0.20230411.[32] Image registration was done with the Advanced Normalization Tools (ANTs) version 0.4.2.[24] For skull stripping HD-BET version 1.0 was used.[33]

References

1 O. M. Henriksen et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 1: Perfusion and Diffusion Techniques”. en. In: *Front. Oncol.* 12 (Mar. 3, 2022), p. 810263. doi: 10.3389/fonc.2022.810263.

2 T. C. Booth et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 2: Spectroscopy, Chemical Exchange Saturation, Multiparametric Imaging, and Radiomics”. en. In: *Front. Oncol.* 11 (2021), p. 811425. doi: 10.3389/fonc.2021.811425.

3 J. Petr et al. “A systematic review on the use of quantitative imaging to detect cancer therapy adverse effects in normal-appearing brain tissue”. en. In: *MAGMA* 35 (1 Feb. 2022), pp. 163–186. doi: 10.1007/s10334-021-00985-2.

4 A. Chelliah et al. “Glioblastoma and radiotherapy: A multicenter AI study for Survival Predictions from MRI (GRASP study)”. en. In: *Neuro. Oncol.* 26 (6 June 3, 2024), pp. 1138–1151. doi: 10.1093/neuonc/noae017.

5 S. Khalighi et al. “Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment”. en. In: *NPJ Precis. Oncol.* 8 (1 Mar. 29, 2024), p. 80. doi: 10.1038/s41698-024-00575-0.

6 H. G. Pemberton et al. “Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18911. doi: 10.1038/s41598-023-44794-0.

7 P. Vollmuth et al. “Artificial intelligence (AI)-based decision support improves reproducibility of tumor response assessment in neuro-oncology: An international multi-reader study”. en. In: *Neuro. Oncol.* 25 (3 Mar. 14, 2023), pp. 533–543. doi: 10.1093/neuonc/noac189.

8 I. Sidibe et al. “Pseudoprogression in glioblastoma: Role of metabolic and functional MRI-systematic review”. en. In: *Biomedicines* 10 (2 Jan. 26, 2022), p. 285. doi: 10.3390/biomedicines10020285.

9 D. J. Ghadimi et al. “Deep learning-based techniques in glioma brain tumor segmentation using multi-parametric MRI: A review on clinical applications and future outlooks”. en. In: *J. Magn. Reson. Imaging* 61 (3 Mar. 2025), pp. 1094–1109. doi: 10.1002/jmri.29543.

10 C. Jayachandran Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *Lancet Digit Health* 3 (12 Dec. 2021), e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

11 E. Moya-Sáez, R. de Luis-García, and C. Alberola-López. “Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI”. In: *Frontiers in Neuroimaging* 2 (2023). doi: 10.3389/fnimg.2023.1055463.

12 U. Baid et al. “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification”. In: *arXiv [cs.CV]* (July 5, 2021).

13 B. H. Menze et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. en. In: *IEEE Trans. Med. Imaging* 34 (10 Oct. 2015), pp. 1993–2024. doi: 10.1109/TMI.2014.2377694.

14 S. Bakas et al. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”. en. In: *Sci Data* 4 (Sept. 5, 2017), p. 170117. doi: 10.1038/sdata.2017.117.

15 L. Scarpace et al. “Radiology Data from The Cancer Genome Atlas Glioblastoma Multiforme [TCGA-GBM] collection [Data set]”. In: *The Cancer Imaging Archive* (2016). doi: 10.7937/K9/TCIA.2016.RNYFUYE9.

16 S. R. van der Voort et al. “The Erasmus Glioma Database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma”. en. In: *Data Brief* 37 (107191 Aug. 2, 2021), p. 107191. doi: 10.1016/j.dib.2021.107191.

17 E. Calabrese et al. “The university of California San Francisco preoperative diffuse glioma MRI dataset”. en. In: *Radiol. Artif. Intell.* 4 (6 Nov. 2022), e220058. doi: 10.1148/ryai.220058.

18 Ivar J.H.G Wamelink, Alle Meije Wink, Niels Verburg, Roeland S Eijgelaar, Emmanouil Koltsakis, Marcus Cakmak, Maarten Balder, Henk J.M.M Mutsaerts, Philip Witt Hamer Frederik Barkhof, Vera C. Keil. *IMAGO*. July 24, 2025. doi: 10.5281/zenodo.17582072.

19 M. Paschali et al. “Foundation models in radiology: What, how, why, and why not”. en. In: *Radiology* 314 (2 Feb. 2025), e240597. doi: 10.1148/radiol.240597.

20 J. E. Park et al. “A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features”. en. In: *BMC Cancer* 20 (1 Jan. 10, 2020), p. 29. doi: 10.1186/s12885-019-6504-5.

21 B. Kocak et al. “METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII”. en. In: *Insights Imaging* 15 (1 Jan. 17, 2024), p. 8. doi: 10.1186/s13244-023-01572-w.

22 B. Kocak et al. “CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII”. en. In: *Insights Imaging* 14 (1 May 4, 2023), p. 75. doi: 10.1186/s13244-023-01415-8.

23 Ivar Wamelink, Joost Kuijer, Vera Keil, Frederik Barkhof, Alle Meije Wink. “CINDERELLA: Computerised INference and DERivation of MRI sequence Linguistic Labelling Algorithm”. In: *OHBM*. OHBM (Seoul; South Korea, 2024). https://archive.aievolution.com/2024/hbm2401/Abstracts/viewAbs?abs=3452.

24 B. Avants, N. Tustison, and G. Song. *Advanced Normalization Tools: V1.0*. 2009. doi: 10.54294/uvnhin.

25 R. Gupta et al. “Expanding the spectrum of IDH1 mutations in gliomas”. en. In: *Mod. Pathol.* 26 (5 May 2013), pp. 619–625. doi: 10.1038/modpathol.2012.210.

26 F. Sahm et al. “Molecular diagnostic tools for the World Health Organization (WHO) 2021 classification of gliomas, glioneuronal and neuronal tumors; an EANO guideline”. en. In: *Neuro. Oncol.* 25 (10 Oct. 3, 2023), pp. 1731–1749. doi: 10.1093/neuonc/noad100.

27 P. Kickingereder et al. “Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study”. en. In: *Lancet Oncol.* 20 (5 May 2019), pp. 728–740. doi: 10.1016/S1470-2045(19)30098-1.

28 R. H. Helland et al. “Segmentation of glioblastomas in early post-operative multimodal MRI with deep neural networks”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18897. doi: 10.1038/s41598-023-45456-x.

29 F. Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. en. In: *Nat. Methods* 18 (2 Feb. 2021), pp. 203–211. doi: 10.1038/s41592-020-01008-z.

30 M. Salvati et al. “Glioblastoma: Molecular profile and immunophenotypic analysis as prognostic tools for tailored therapy and decision making in a recent surgical series”. en. In: *Interdiscip. Neurosurg.* 20 (100697 June 2020), p. 100697. doi: 10.1016/j.inat.2020.100697.

31 A. Olar et al. “IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas”. en. In: *Acta Neuropathol.* 129 (4 Apr. 2015), pp. 585–596. doi: 10.1007/s00401-015-1398-z.

32 X. Li et al. “The first step for neuroimaging data analysis: DICOM to NIfTI conversion”. en. In: *J. Neurosci. Methods* 264 (May 1, 2016), pp. 47–56. doi: 10.1016/j.jneumeth.2016.03.001.

33 F. Isensee et al. “Automated brain extraction of multisequence MRI using artificial neural networks”. en. In: *Hum. Brain Mapp.* 40 (17 Dec. 1, 2019), pp. 4952–4964. doi: 10.1002/hbm.24750.

Chapter 7

Contrast-Enhancing Tumor Margin Detection in Gliomas using Non-Contrast MRI: From Human-Only to Human-AI Assisted Assessment

Ivar J.H.G. Wamelink, Aynur. Azizova, Elif Kaya, E. Koltsakis, Joao N. Ramos, Aziz A. Tan, Normann Kornemann, Frederik Barkhof, Alle Meije Wink , Vera C. Keil

As submitted to Journal of Mangetic Resonance Imaging

Abstract

Background: Contrast-enhanced MRI remains a cornerstone of neuro-oncological imaging, but growing concerns regarding safety, sustainability, and costs prompted interest in the development and evaluation of contrast-free alternatives.

Purpose/hypothesis: o investigate the ability of human raters to delineate the contrast-enhancing tumor region in glioblastoma using gadolinium-based contrast agent (GBCA)-free MRI, and whether a deep learning-based algorithm can improve this delineation.

Study Type: Retrospective

Subjects: One hundred glioblastoma, IDH-wildtype, patients from the IMAGO database. A segmentation model was trained using BraTS (n=989) and UCSF-PDGM (n=501) datasets.

Sequence: Pre- and postcontrast T1-weighted (T1w), T2-weighted (T2w), T2-FLAIR, DWI-B0 and B-1000, ADC.

Assessment: Ground truth segmentations of the enhancing tumor regions were generated using an established algorithm and refined by an independent neuroradiologist. All cases were then independently re-segmented by three radiologists and one non-clinician using only GBCA-free sequences. Another radiologist (R-AI) improved the segmentations produced by an AI algorithm without post-contrast sequence input. For each case, raters had to fill in a questionnaire.

Statistical tests: Intraclass Correlation Coefficient (ICC), Generalized Conformity Index (GCI), Dice (DC), Hausdorff distance (HD), Mann-Whitney U test, Kruskal-Wallis test, Pearson correlation coefficient.

Results: All raters showed a moderate-to-strong correlation with the ground truth ($r=0.79$; $P<0.001$). AI assistance increased Dice similarity by 0.09 (0.82 vs. 0.72; $P<0.001$), reduced Hausdorff distance by 6.2 mm (9.97 vs. 16.16 mm; $P<0.001$), and decreased absolute volume error by 10 ml (6.75 vs. 17.07 ml; $P<0.001$). Rater confidence increased with larger tumor volume, despite nonsignificant volumetric error metrics after correction (all $P \geq 0.24$)

Conclusion: Human raters predicted enhancing tumor areas from GBCA-free sequences with moderate accuracy, showing considerable inter-rater variability in glioblastoma. AI-based GBCA-free segmentation assistance significantly improved human segmentations.

Keywords: Glioblastoma, Segmentation, Artificial Intelligence, Gadolinium Based Contrast Agents, Glioma

Introduction

Contrast-enhanced magnetic resonance imaging (CE-MRI) is central in neuro-oncological practice as it highlights areas of blood-brain barrier disruption, increases lesion visibility, and provides critical information for lesion characterization.[1] Enhancement morphology, intensity, and extent serve as important imaging biomarkers that support histomolecular glioma classification and grading. In glioblastoma (GBM), the enhancing tumor core defines the surgical resection margin and guides radiotherapy planning, underscoring its essential role in treatment stratification.[2, 3]

Safety, sustainability, and cost issues of gadolinium-based contrast agents (GBCAs) have prompted growing surveillance and research into the feasibility of GBCA-free alternatives.[4–8] GBCA-free sequences, such as FLAIR, can guide surgical decision-making and, in some scenarios, superiorly reflect therapeutic outcomes.[9, 10] Deep learning (DL) analyses of GBCA-free imaging are a more recent approach to limit contrast agent exposure.[11–14]

Despite these developments, any transition to GBCA-free imaging must ensure that the enhancing tumor core can still be reliably identified, because treatment margins fundamentally depend on accurate segmentation of this region. Glioma segmentation typically distinguishes three classes: 1) non-enhancing tumor/edema, 2) necrosis, and 3) enhancing tumor volumes. Manual delineation of these subregions is time-consuming and impractical for routine workflows. Consequently, automated DL-based segmentation approaches using standard MRI sequences have been extensively investigated as efficient alternatives. However, these models rely on CE-MRI as the reference signal, limiting their applicability to GBCA-free imaging.[15, 16]

Recent work explored whether human raters can infer enhancement characteristics from GBCA-free imaging. Their proposed decision tree model for predicting enhancement intensity and morphology demonstrated sufficient diagnostic accuracy.[17] However, the proposed method did not provide spatial information about enhancing areas, which precludes its use for treatment decision-making. Second, knowledge of human benchmark performance and the impact of added AI assistance are crucial to develop the field of GBCA avoidance through AI in neuroradiology further in its entirety.

To address this gap, this study aims to 1) evaluate the ability of human raters to identify the enhancing tumor regions, i.e. the enhancing volume including central necrosis, in glioblastoma patients using standard GBCA-free MR sequences compared with CE-MRI, and 2) assess if DL-assisted segmentations based on GBCA-free imaging improve human segmentation accuracy.

Materials and Methods

The study was approved by our Institutional Review Board, and informed consent was waived due to the retrospective study design.

Datasets

Subjects included 100 glioblastoma, IDH-wildtype patients, WHO grade 4, classified according to the WHO CNS5 (2021) criteria. Cases were randomly drawn from the IMAging in GliOma (IMAGO) database.[18] All patients underwent treatment at Amsterdam University Medical Centre between November 2014 and June 2021. Clinical information included IDH mutation status, overall survival.

MRI data included pre- and post-contrast T1-weighted (T1w and CET1), T2-weighted (T2w), T2-FLAIR, diffusion-weighted imaging (DWI B0 and B1000), and apparent diffusion coefficient (ADC) images. Ground truth (GT) segmentations from the IMAGO database included annotations for non-enhancing tumor/edema (1), necrosis (2), and enhancing tumor (3), generated using the PICTURE segmentation algorithm on CET1, T1w, T2w, and T2-FLAIR images, converted into binary masks including the tumor core (enhancing tumor and necrosis).[19] GT segmentations were independently reviewed and refined by a board-certified neuroradiologist with 5 years of experience (V.C.K). Delineation of the enhancing tumor margin was performed on axial slices and subsequently validated across coronal and sagittal planes.

The GBCA-free segmentation algorithm was trained on the BraTS 2021 (n=1,251)[15, 16, 20] and UCSF-PDGM (n=501)[21] datasets. Due to overlap with the UCSF-PDGM dataset, 262 BraTS cases were excluded, leaving 1,490 cases for training. The IMAGO dataset has demographic distributions similar to those of UCSF-PDGM (male: 59% vs. 60%; mean age: 59±10 vs. 57±15 years). Clinical data for BraTS were unavailable. Only 3 of the IMAGO cases were non-enhancing.

AI-algorithm

The GBCA-free segmentation algorithm was developed using the nn-UNet architecture.[22] No additional model fine-tuning was performed on the IMAGO data. For details on model training, see.[22] Three GBCA-free structural MRI sequences (pre-contrast T1w, T2w, and T2-FLAIR) were used as inputs in an 80:20 training-to-validation split, and the output was a three-level mask similar to the one produced in the PICTURE segmentations. Since the human rater segmentations only captured the tumor core, the AI-generated three-label predictions were also converted into binary masks (enhancing + necrotic regions). Training was performed on one NVIDIA A100-SXM4 (80GB) GPU, using 10 CPU cores and 50GB of RAM.

Human rater segmentation

Five raters (E with 3 months of neuroradiology training and R-AI, B, C, D with 5, 5, 7, and 7 years of neuroradiology experience, respectively) performed GBCA-free manual segmentations using 3D Slicer software. Raters received training on 3D Slicer usage (by I.W.) and GBCA-free prediction of the resectable tumor margin, defined as the outer enhancing margin (by A.A.). The imaging features identifying the enhancing tumor margin were: necrosis, diffusion restriction, and T2 heterogeneity based on the findings of our previous study.[17] Segmentations were performed in five rounds, each with 20 cases. Four raters separately segmented patients in the same order. After each set,

raters were automatically directed to review post-contrast images without modifying their segmentations. This iterative process allowed them to self-train by identifying errors and improving their segmentation accuracy. During each round of 20 cases, raters completed a per-case questionnaire assessing (1) which sequence guided the segmentation (T2w, T1w + T2w, DWI, T2w + DWI), (2) how confident they were with their segmentation (5-point scale), (3) how difficult the case was to segment (5-point scale). Segmentation time was automatically tracked by the segmentation tool and was used for a time-quality tradeoff analysis (4). During the post-contrast review of each round, each rater was asked (per case) whether they identified the errors in their segmentation and whether they identified additional imaging clues beyond necrosis, diffusion restriction, and T2 heterogeneity for subsequent decision-making. One Rater (E) was excluded from answering all questions and segmented all 100 cases successively.

Additionally, an independent rater, R-AI, reviewed and refined the GBCA-free AI-based segmentation in 3D Slicer. This rater followed the same structured five-round approach as described earlier. A flowchart of the segmentation workflow is shown in Figure 1.

Analysis

All analyses were performed in Python (Version 3.11) using the SciPy package (Version 1.14.1). A P-value <.05 adjusted with Bonferroni correction for multiple comparisons was considered statistically significant.

Geometric Accuracy

To evaluate the ability of human raters to define the tumor core on GBCA-free MRI, we performed quantitative analyses compared to the GT. Segmentation similarity was assessed using the: (1) Dice Similarity Coefficient (DC) which measures voxel-level spatial overlap between the segmentation and the GT, (2) Intraclass correlation coefficient (ICC) that quantifies volumetric agreement across raters and with the GT (single random raters; poor <0.5, moderate 0.5–0.75, good 0.75–0.9, excellent >0.9)[23], (3) Hausdorff distance (HD) that measures boundary agreement between segmentations, and (4) generalized conformity index (GCI) which quantifies the spatial agreement among multiple segmentations. It generalizes overlap measures such as the dice coefficient to more than two raters.

Questionnaire Analyses

We performed the following exploratory analyses: (1) Which MRI sequences most strongly guided human rater segmentations, and whether reliance on specific sequences was associated with errors or time spent per case, (2) analysis of segmentation confidence, comparing self-reported rater confidence against actual segmentation performance and time, (3) analysis of case difficulty, examining whether cases rated as more “difficult” correlated with lower segmentation accuracy or longer segmentation times, and (4) analysis of time-quality tradeoff to examine relationships between segmentation time and accuracy (DC, volume error). For the segmentation confidence and case difficulty analyses, we formed case-level groups using the majority rating across raters. This yielded eight confidence groups and eight difficulty groups. For each construct

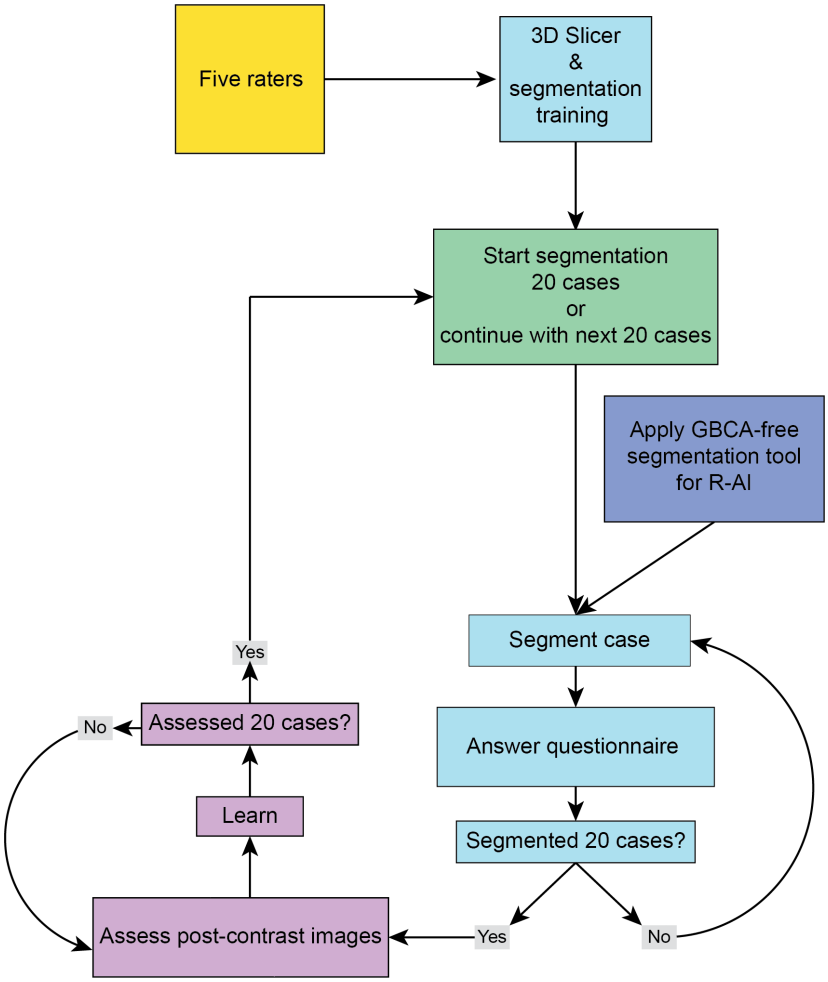


Figure 1: Workflow segmentation analysis.

(confidence and difficulty), five groups corresponded to the ordinal majority ratings from 1 to 5 (e.g., from “easy” (majority scored 1) to “difficult” (majority scored 5), and from “not confident” (majority scored 1) to “highly confident” (majority scored 5)). The remaining three groups captured different agreement patterns: (i) total agreement (all raters gave the same rating), (ii) total disagreement (no clear majority), and (iii) a “two rater pairs agree” pattern (two raters sharing one rating and the other two sharing another rating). Unless stated otherwise, segmentation performance was assessed for both human raters and R-AI. Inter-rater comparisons were evaluated with the Kruskal–Wallis test and post-hoc pairwise Mann-Whitney tests. Cliff’s δ was employed to quantify the effect size of differences in Dice Coefficient (DC) across confidence and difficulty groups. Bowker’s test was applied to evaluate whether raters exhibited systematic asymmetry in rating

cases as easy/confident or difficult/not confident.

We additionally performed a qualitative analysis of non-enhancing cases.

Results

Quantitative Analysis

Segmentation performance of the five raters is summarized in Table 1. The four manual raters achieved moderate DC scores (0.71–0.76 ± 0.19-0.25), whereas the AI model and R-AI had higher DC scores (0.83±0.18 and 0.82±0.19, respectively; Figure 2). R-AI achieved significantly higher DC than each manual rater (mean difference 0.08–0.12; all P<0.001 after Bonferroni correction) and a higher DC than the average of all manual raters (mean difference 0.09; P<0.001). A similar pattern was observed for Hausdorff Distance, where R-AI showed smaller Hausdorff Distances with the ground truth (mean reduction of 5-7 mm; P<0.001) than every manual rater and than the average among manual raters. Volume analysis using absolute error relative to GT demonstrated that the R-AI also produced more accurate volumetric estimates, with a reduction in mean absolute error of approximately 10 mL compared to the mean absolute error of the manual raters (P<0.001). All P-values for pairwise and mean comparisons were derived from one-sided paired t-tests and Wilcoxon signed-rank tests, with Bonferroni correction applied for multiple pairwise comparisons. These findings were confirmed in mixed-effects models with patients as a random effect, in which AI assistance was associated with higher DC and lower Hausdorff distance and volume error (all P-values <0.001). All human raters showed a tendency to oversegment, with one exception (Rater D, Table 1 and Figure 3). Overall accuracy analysis showed a mean absolute error (MAE) of 13.31 mL, a root mean square error (RMSE) of 24.05 mL, and a correlation coefficient of r=0.789 (P<0.001) with the GT, despite substantial variability between raters.

Inter-Rater Analysis

Inter-rater agreement analysis showed moderate agreement depending on the group of raters considered (Table 2). Excluding the inexperienced rater (E) improved the agreement from 0.72 to 0.77. Figure 4 shows pairwise correlations for DC scores and predicted volumes between raters and highlights patterns of variable but generally moderate-to-strong agreement.

Segmentation Justification Questionnaire

Segmentation accuracy differed significantly across sequence-decision subsets, reflecting the varying contribution of MRI sequences to delineating the tumor core margin (Kruskal–Wallis H = 18.0, P<0.001). The highest performance (mean DC 0.84, 95% CI 0.78–0.89) was achieved when T1w and T2w images were used for segmentation. Lower accuracy was observed when T2w alone (mean DC 0.78, 95% CI 0.75–0.81) and T2w combined with DWI (mean DC 0.76, 95% CI 0.74–0.78) were the main guidance and even lower when DWI was the primary driver (mean DC 0.58, 95% CI 0.44–0.72). Pairwise comparisons confirmed significant differences between T2w vs T2w+DWI (P<0.01, effect size = 0.20) and T2w vs DWI (P=0.02, effect size = 0.34), whereas other contrasts

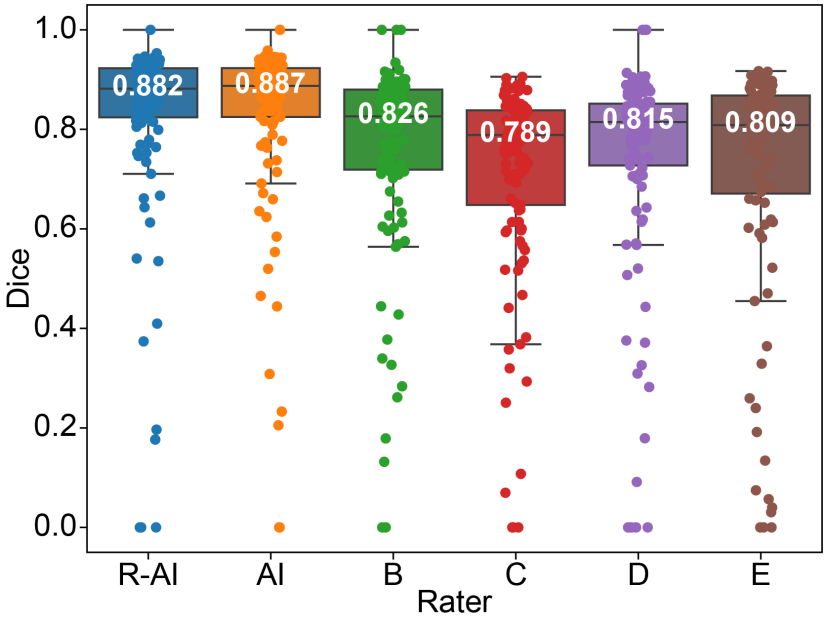


Figure 2: Dice similarity coefficients for AI and human raters. Boxplots show the distribution of Dice scores for the AI model, the human rater refining the AI output(rater+AI), and individual manual raters. Median values are indicated on each box. These results confirm that AI-based approaches outperformed manual raters.

were not significant after correction (Table 3).

Segmentation Confidence Questionnaire

Rater agreement on confidence scores was significantly associated with segmentation performance (DC), but not with time or volumetric error. Global testing revealed differences in DC across confidence groups (Kruskal–Wallis H = 15.99, p = 0.01). DC was highest when raters had perfect agreement (median 0.88) or majority high confidence (0.83–0.84) and lowest when confidence was split or very low (median was 0.23 when not confident, though n=2). Time (H=4.74, P=0.58) showed no effect with segmentation confidence. Larger tumors tended to fall into the high-agreement groups, but volumetric error metrics (all Pgeq0.24) did not remain significant after correction.

Case Difficulty Questionnaire

Global testing across case difficulty groups revealed no significant differences in segmentation time or volumetric error (all P>0.05). DC was different between difficulty groups (Kruskal–Wallis H=26.34, P<0.001), but the only pairwise contrast that remained significant after correction was between moderately-easy and difficult, with a strong separation (Cliff's δ=0.979). Other contrasts showed trends toward differences, but no significant ones after adjustment for multiple comparisons, likely due to very small subgroup sizes (n$\leq$3 for easy and difficult).

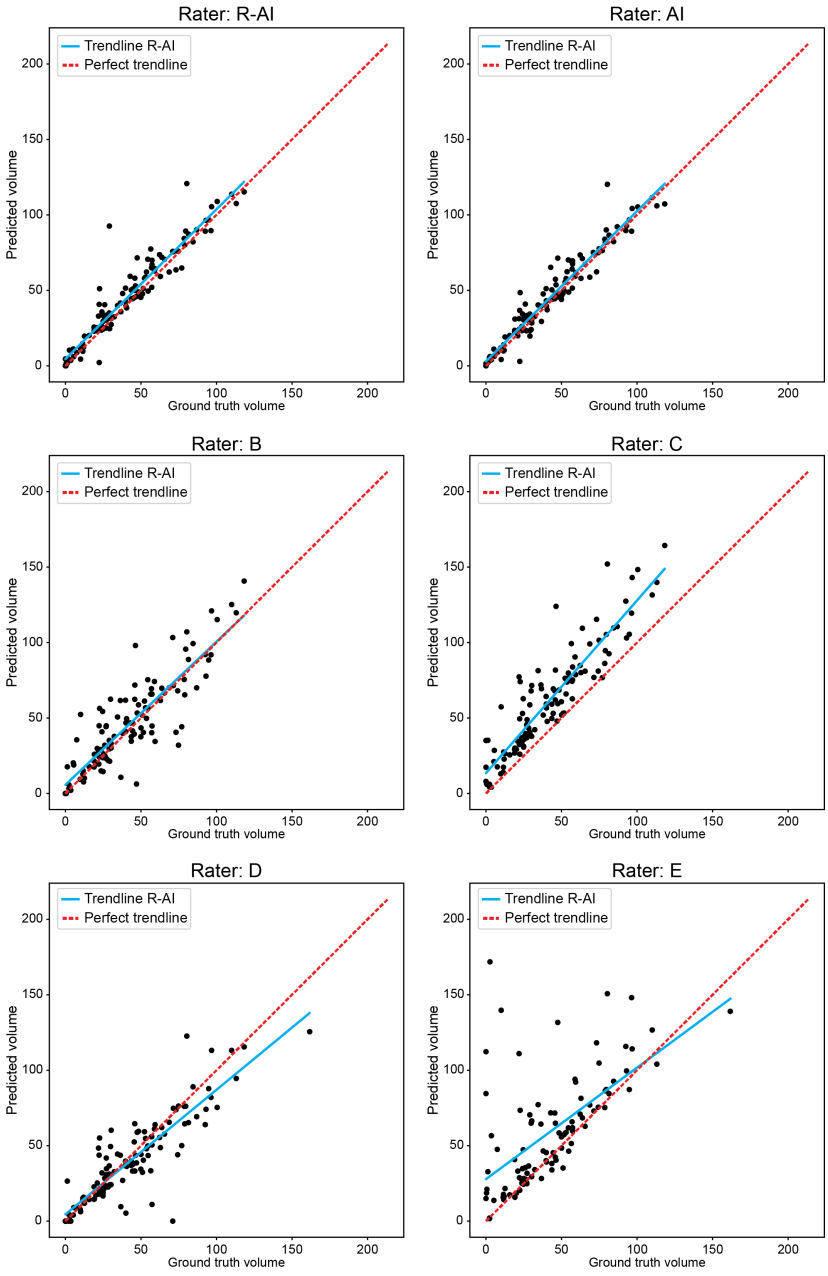


Figure 3: Predicted tumor core volumes versus ground truth for AI and human raters.Scatter plots show the relationship between predicted and ground truth volumes for eachrater and the AI model. The red line indicates perfect agreement (identity line), while the blue lines represent regression fits. The predicted tumor core volumes generated bythe AI and rater+AI correlated excellently with the ground truth. All human raters tended to oversegment compared with the ground truth.

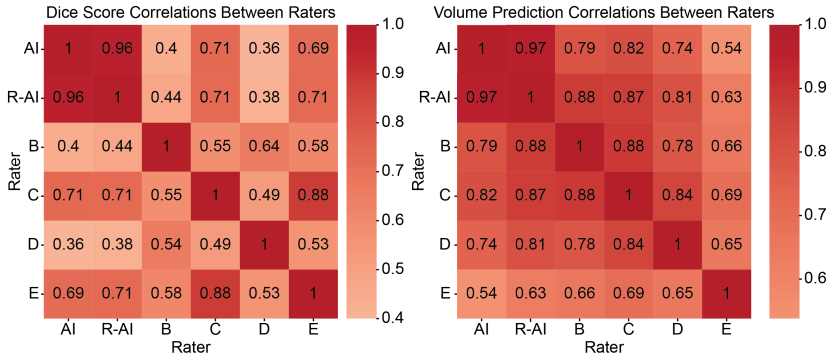


Figure 4: Pairwise correlations between human raters. Left: Heatmap of Dice score correlations between raters shows fair to substantial agreement. Right: Heatmap of predicted tumor core volume correlations demonstrates generally substantial agreement for volume compared with Dice.

Table 1: Tumor core segmentation performance across different raters: Four human raters (B-E) and one human rater refining AI GBCA-free segmentations (R-AI). The GBCA-free AI algorithm was trained on pre-contrast T1w, T2w, and T2-FLAIR images. Mean GT tumor volume was 43.35 mL. SD: standard deviation, HD: Hausdorff distance, ICC: Intraclass Correlation Coefficient.

Rater	Average volume (mL)	Average over(+) or under(-) estimation (mL)	Dice mean (± SD)	HD mean (± SD)	ICC
AI	46	2.65	0.83 ± 0.18	10.08 ± 9.84	0.98
R-AI	47.4	4.05	0.82 ± 0.19	9.49 ± 7.44	0.91
B	48.23	4.88	0.76 ± 0.20	14.05 ± 10.54	0.63
C	66.39	23.04	0.71 ± 0.21	16.54 ± 10.59	0.55
D	42.76	-0.59	0.73 ± 0.22	14.86 ± 14.07	0.52
E	61.02	18.35	0.71 ± 0.25	17.51 ± 13.64	0.46

Table 2: Inter-rater reliability for tumor core segmentation. Intraclass correlation coefficients (ICC) and Generalized Conformity Index were calculated across different groups of raters. *p-value was smaller than 0.0001.

Group	ICC	GCI (mean ± SD)
All raters + AI	0.71*	0.61±0.2
All raters	0.72*	0.6±0.19
Without E	0.77*	0.63±0.19
Without AI and without E	0.77*	0.6±0.19
AI and R-AI	0.97*	0.92±0.15

Table 3: Pairwise comparisons of sequence subsets deemed most helpful for segmentation. Only T2w vs T2w+DWI and T2w vs DWI remained significantly different after correction.

Comparison	p-value	p-value corrected	Effect size
T2w vs T2w + DWI	0.00	0.00	0.20
T2w vs DWI	0.00	0.02	0.34
T2w + DWI vs DWI	0.04	0.27	0.23
T1w + T2w vs DWI	0.30	1.79	0.38
T2w + DWI vs T1w + T2w	0.48	2.90	0.24
T2w vs T1w + T2w	0.80	4.83	0.09

Rater B consistently rated cases as more difficult than other raters ((Bowker $P \leq 10^{-6}$), while rater C tended to rate cases as easier. R-AI and rater D showed the most similar rating patterns.

Time–Quality Trade-off

Segmentation time analysis revealed significant differences between raters (ANOVA F=88.384, P< 0.001; Kruskal–Wallis H = 265.714, P<0.001). Rater D was the fastest (4.29±2.79 min per case), followed by R-AI (12.08±7.19 min), B (16.68±14.81 min), and C (33.37±22.83 min) (Figure 5 A, C). A significant decrease in time per case was observed (r=–0.304, P=0.001), consistent with a learning effect (Figure 5 B). Importantly, segmentation time did not correlate with DC scores (r=–0.001, P=0.977), indicating that spending more time did not result in higher segmentation accuracy.

Non-enhancing GBM identification

Identification of non-enhancing cases was difficult, as well as tumors with a small enhancing volume (<5mL; Figure 6). Supplementary table S1 contains an overview of missed or wrongly segmented cases. Figure 7 shows an enhancing lesion on the left that was consistently and accurately identified by all raters. In contrast, the right side shows a non-enhancing lesion that most raters misclassified.

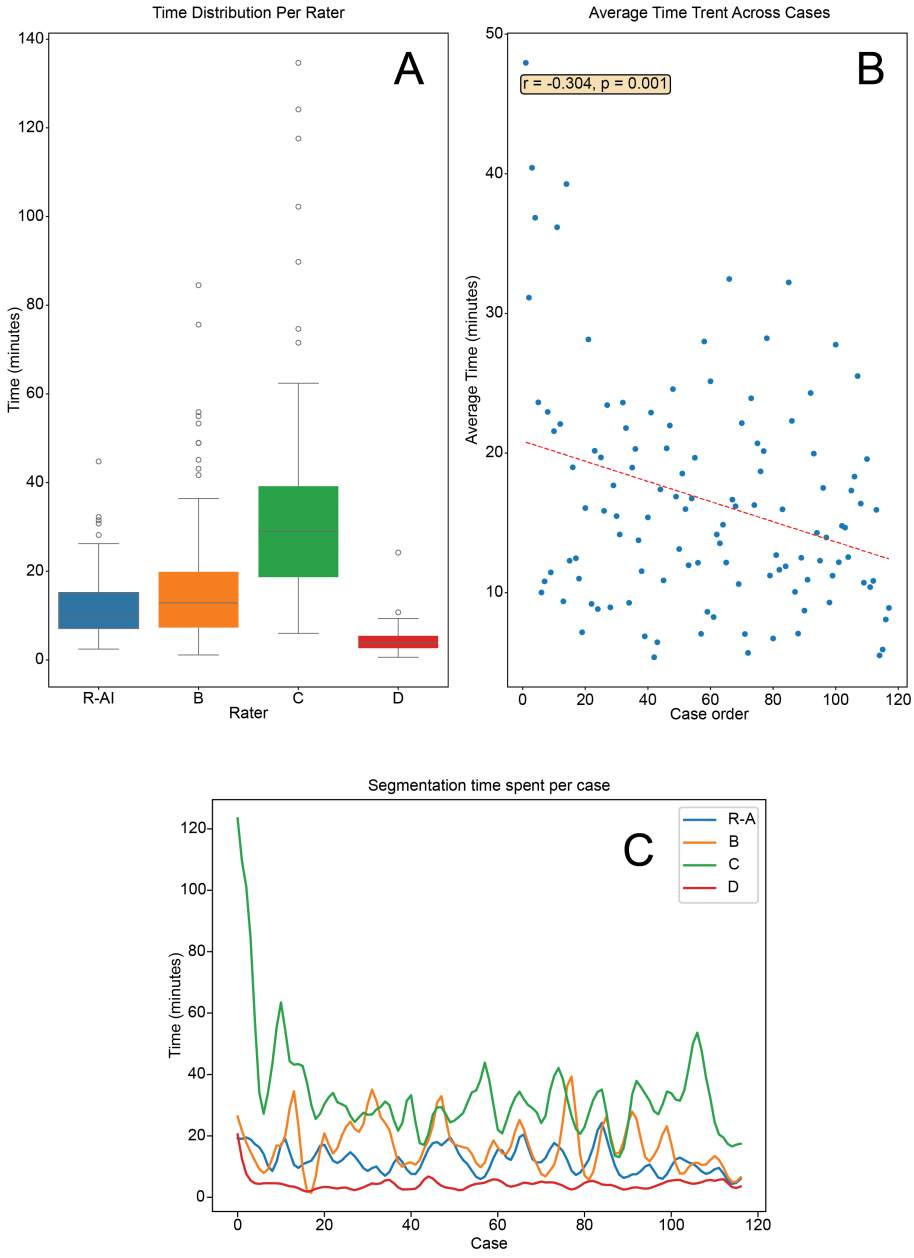


Figure 5: Segmentation time analysis across raters. (A) Distribution of segmentation times per case for individual raters, showing D as the fastest, followed by R-AI, B, and C, who was the slowest. (B) Average segmentation time per case across all raters demonstrates a significant decreasing trend consistent with a learning effect (r = -0.304, P = 0.001). (C) Segmentation times are shown for each rater across all cases. Joao was consistently the fastest, followed by R-AI, then B, while C required substantially longer times. The differences were very large, with near-complete separation between D and D (rank-biserial $\approx$ +1) and a strong effect between R-AI and C (rank-biserial $\approx$ -0.92).

Discussion

In this retrospective study, we evaluated whether contrast enhancement marking the tumor core of glioblastoma IDH-wt patients can be delineated on GBCA-free MR images by human raters alone and if an AI-algorithm trained exclusively on GBCA-free sequences (R-AI) can improve this delineation. R-AI achieved higher segmentation accuracy than human raters when only provided GBCA-free sequences, evidenced by the quantitative analysis, which shows statistically significantly higher DC scores, lower Hausdorff distances, and closer volumetric agreement with the ground truth. Human performance was varied and rater-dependent, with systematic volume bias (over- or undersegmentation) and moderate inter-rater agreement. Segmentation performance varied primarily with the decision-guiding images: T1w+T2w yielded the best performance, whereas DWI-guided segmentation decisions were most error-prone. Overall, GBCA-free delineation of the tumor core is feasible, and consistency and reliability improve with AI support and careful sequence selection.

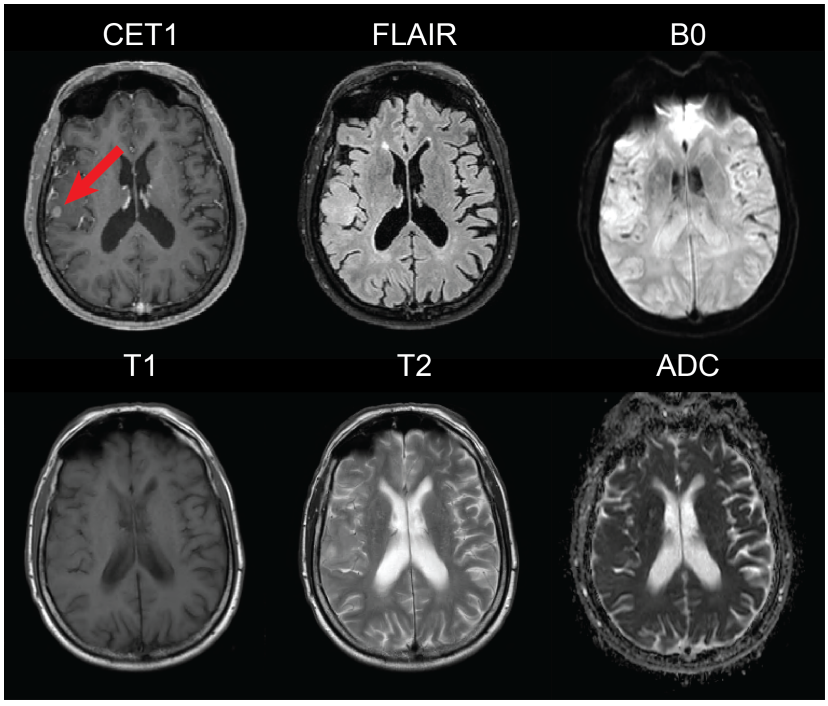


Figure 6: Small enhancing lesion in the right temporal lobe misclassified by raters. Theaxial post-contrast T1-weighted image shows a 0.6 mL enhancing lesion (red arrow) in theright temporal lobe. The AI correctly classified this case as enhancing, but the human raterrefining the AI output incorrectly reclassified it as non-enhancing. Two raters (B and D) also labeled this lesion as non-enhancing, while the other two raters segmented it withvolumes of 30.7 mL and 10.9 mL, illustrating substantial variability in rater performance.

Most prior brain tumor segmentation studies relied on post-contrast T1-weighted imaging to define enhancing tumor margins. In those settings, strong models achieve enhancing tumor margins DC values between 0.70–0.90 for strong models, with whole-tumor DC often higher due to non-enhancing/edema contributions.[13, 24–26] Our GBCA-free model, trained only on pre-contrast T1w, T2w, and T2-FLAIR, achieved a similar DC score to that of many contrast-dependent segmentation algorithms. This indicates that a substantial fraction of enhancing-margin information is encoded in structural sequences via secondary markers (e.g., necrosis, diffusion restriction, and T2 heterogeneity).[17] High average DCs for AI and R-AI (0.83 and 0.82, respectively) suggest that, for many cases, contrast may be unnecessary for defining the resectable tumor core margin in glioblastoma IDH-wt patients. This would be especially beneficial for patients with renal insufficiency, prior adverse reactions, vulnerable populations, including pregnant and breastfeeding women and children, and circumstances where shorter protocols and lower cost are prioritized.[27–29] However, our qualitative analysis of non-enhancing cases underscores that identifying these cases remains difficult for both humans and AI, even in our small sample.

The suggested segmentation protocol provides a robust and practical foundation for GBCA-free brain tumor segmentations and could serve as a standard for future studies. However, human raters tended to oversegment, and accuracy analysis revealed rater-specific differences in error patterns. Some raters consistently showed positive signed errors with larger absolute errors, while one rater tended to underestimate. This pattern aligns with prior reports of inter-rater variability and systematic bias in glioma delineation.[30] Dice, Hausdorff distance, and volumetric error improved significantly when a human was assisted by AI. Improvements to the protocol, such as advanced imaging biomarkers that are derived from (APT-CEST) and (ASL) might reduce these consistent errors.[31, 32]

The differences in volume definition compared to the standard clinical GT may have clinical consequences, given that the mean tumor volume was larger than tumor volume derived from CET1 sequences. Similarly, the mean resected volume is larger when approaching surgical removal of lesions based on T2-FLAIR (FLAIRectomy), which may extend far beyond contrast-enhancing regions.[9, 10] While resection or radiation of the enhancing tumor core remains the gold standard, the approach taken in this research is likely to have intermediate results between the gold standard (CE-based resection) and FLAIRectomy. Where with FLAIRectomy the additional morbidity is estimated from larger areas using FLAIR as a basis, the resection based on the approach suggested in this paper may not exceed the disease burden caused by FLAIR and will likely result in resection of all CE tumor, leading to clinically comparable results.

When raters used T1w+T2w sequences for segmentation, they obtained the highest DC; when T2w alone was used, the DC was intermediate; and DWI-led decisions were the least accurate. This is consistent with the biological and imaging expectations: structural heterogeneity, necrotic architecture, and mass effect are visible on T1w/T2w/FLAIR, whereas DWI signal can be nonspecific in GBM and may reflect cytotoxic edema, cellu-

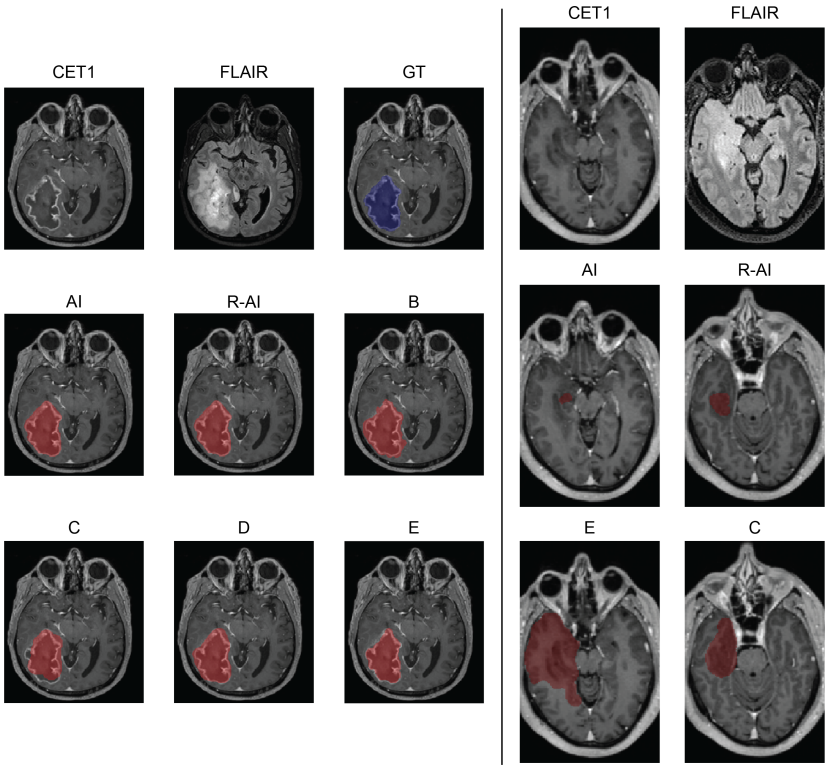


Figure 7: Left: Enhancing lesion consistently classified by all raters as enhancing. Theaxial post-contrast T1-weighted image shows an enhancing lesion (red arrow) that wascorrectly identified by all raters, with segmented volumes in close agreement andoverlapping with the gold-standard segmentation. Right: Non-enhancing lesionmisclassified by most raters as enhancing. The axial post-contrast T1-weighted imageshows a non-enhancing lesion (red arrow) that was incorrectly classified as enhancing by themajority of raters, in some cases (C and E) with substantial overestimation of the lesionvolume.

larity, or treatment effects heterogeneously.

All raters showed a learning effect, with time per case decreasing, but there was no correlation between time and accuracy. This suggests that segmentation efficiency gains were not at the expense of segmentation quality.

Identifying non-enhancing GBM cases was difficult. The AI alone correctly identified one of three; R-AI missed two and introduced one false non-enhancing label. Among human raters, identification of non-enhancing GBM cases ranged from failure to detect non-enhancing disease to marked overestimation of very small enhancing lesions. A retrospective review of misclassified cases by rater R-AI suggested a lack of canonical GBCA surrogates in non-enhancing patients: no T2 heterogeneity, no diffusion restriction, and no necrosis-like signal. While informative, this single-rater perspective does not explain why others confidently labeled the same lesions as enhancing and, in some

instances, inflated volumes. These error types are clinically relevant: small-volume overcalls may drive overtreatment or misleading response assessment, while misses increase the risk of undertreatment. Segmentation of small lesions and non-enhancing cases is notoriously difficult.

Several limitations should be acknowledged. The study was limited to glioblastoma, IDH-wt patients, a selection that may bias raters to find enhancement and could explain the poor identification of non-enhancing GBMs. While we observed good performance on this specific subset, additional research with other glioma subtypes is needed to assess performance in an average glioma population. Such broader analysis would strengthen the non-enhancing tumor analyses, particularly since lower-grade gliomas are generally non-enhancing.[33] In the questionnaire analyses, some subgroup analyses had small sizes after stratification, reducing power for pairwise contrasts even when global tests were significant.

Conclusion

A GBCA-free sequence-based delineation of the enhancing tumor core in GBM is feasible and, when augmented by a GBCA-free AI algorithm, achieves accuracy that approaches typical contrast-dependent benchmarks and improves correlation. Non-enhancing and small GBM lesions remain difficult to identify and segment. AI assistance offers improved tumor core identification and may assist in the safe reduction of contrast use in routine practice.

References

1 M. Weller et al. “EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood”. en. In: *Nat. Rev. Clin. Oncol.* 18 (3 Dec. 8, 2020), pp. 170–186. doi: 10.1038/s41571-020-00447-z.

2 I. J. H. G. Wamelink et al. “Brain tumor imaging without gadolinium-based contrast agents: Feasible or fantasy?” en. In: *Radiology* 310 (2 Feb. 2024), e230793. doi: 10.1148/radiol.230793.

3 G. Morana et al. “Use of contrast agents in oncological imaging: magnetic resonance imaging”. en. In: *Cancer Imaging* 13 (3 Sept. 23, 2013), pp. 350–359. doi: 10.1102/1470-7330.2013.9018.

4 *FDA Drug Safety Communication: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings*. May 16, 2018.

5 V. M. Runge. “Critical Questions Regarding Gadolinium Deposition in the Brain and Body After Injections of the Gadolinium-Based Contrast Agents, Safety, and Clinical Recommendations in Consideration of the EMA’s Pharmacovigilance and Risk Assessment Committee Recommendation for Suspension of the Marketing Authorizations for 4 Linear Agents”. en. In: *Invest. Radiol.* 52 (6 June 2017), pp. 317–323. doi: 10.1097/RLI.0000000000000374.

6 A. Scarciglia et al. “Gadolinium-based contrast agents (GBCAs) for MRI: A benefit-risk balance analysis from a chemical, biomedical, and environmental Point of View”. en. In: *Global Chall.* 9 (3 Mar. 2025), p. 2400269. doi: 10.1002/gch2.202400269.

7 L. M. Krohn et al. “Impacts of COVID-19 and climate change on wastewater-derived substances in urban drinking water: Evidence from gadolinium-based contrast agents in tap water from Berlin, Germany”. en. In: *Water Res.* 259 (121847 Aug. 1, 2024), p. 121847. doi: 10.1016/j.watres.2024.121847.

8 Z. Zhang et al. “Anthropogenic gadolinium contaminations in the marine environment and its ecological implications”. en. In: *Environ. Pollut.* 359 (124740 Oct. 15, 2024), p. 124740. doi: 10.1016/j.envpol.2024.124740.

9 F. Certo et al. “FLAIRectomy in supramarginal resection of glioblastoma correlates with clinical outcome and survival analysis: A prospective, single institution, case series”. en. In: *Oper. Neurosurg. (Hagerstown)* 20 (2 Jan. 13, 2021), pp. 151–163. doi: 10.1093/ons/opaa293.

10 A. F. Haddad et al. “FLAIRectomy: Resecting beyond the contrast margin for glioblastoma”. en. In: *Brain Sci.* 12 (5 Apr. 25, 2022), p. 544. doi: 10.3390/brainsci12050544.

11 G. Iliadis et al. “Volumetric and MGMT parameters in glioblastoma patients: survival analysis”. en. In: *BMC Cancer* 12 (1 Jan. 3, 2012), p. 3. doi: 10.1186/1471-2407-12-3.

12 B. M. Ellingson et al. “Volumetric measurements are preferred in the evaluation of mutant IDH inhibition in non-enhancing diffuse gliomas: Evidence from a phase I trial of ivosidenib”. en. In: *Neuro. Oncol.* 24 (5 May 4, 2022), pp. 770–778. doi: 10.1093/neuonc/noab256.

13 J. K. Ruffle et al. “Brain tumour segmentation with incomplete imaging data”. en. In: *Brain Commun.* 5 (2 Apr. 28, 2023), fcad118. doi: 10.1093/braincomms/fcad118.

14 J. Liu et al. “Automatic brain tissue and lesion segmentation and multi-parametric mapping of contrast-enhancing gliomas without the injection of contrast agents: A preliminary study”. en. In: *Cancers (Basel)* 16 (8 Apr. 17, 2024), p. 1524. doi: 10.3390/cancers16081524.

15 U. Baid et al. “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification”. In: *arXiv [cs.CV]* (July 5, 2021).

16 B. H. Menze et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. en. In: *IEEE Trans. Med. Imaging* 34 (10 Oct. 2015), pp. 1993–2024. doi: 10.1109/TMI.2014.2377694.

17 A. Azizova et al. “Human performance in predicting enhancement quality of gliomas using gadolinium-free MRI sequences”. en. In: *J. Neuroimaging* 34 (6 Nov. 2024), pp. 673–693. doi: 10.1111/jon.13233.

18 Ivar J.H.G Wamelink, Alle Meije Wink, Niels Verburg, Roeland S Eijgelaar, Emmanouil Koltsakis, Marcus Cakmak, Maarten Balder, Henk J.M.M Mutsaerts, Philip Witt Hamer Frederik Barkhof, Vera C. Keil. *IMAGO*. July 24, 2025. doi: 10.5281/zenodo.17582072.

19 H. G. Pemberton et al. “Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18911. doi: 10.1038/s41598-023-44794-0.

20 S. Bakas et al. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”. en. In: *Sci Data* 4 (Sept. 5, 2017), p. 170117. doi: 10.1038/sdata.2017.117.

21 E. Calabrese et al. “The university of California San Francisco preoperative diffuse glioma MRI dataset”. en. In: *Radiol. Artif. Intell.* 4 (6 Nov. 2022), e220058. doi: 10.1148/ryai.220058.

22 F. Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. en. In: *Nat. Methods* 18 (2 Feb. 2021), pp. 203–211. doi: 10.1038/s41592-020-01008-z.

23 T. K. Koo and M. Y. Li. “A guideline of selecting and reporting intraclass correlation coefficients for reliability research”. en. In: *J. Chiropr. Med.* 15 (2 June 2016), pp. 155–163. doi: 10.1016/j.jcm.2016.02.012.

24 A. Hatamizadeh et al. “Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images”. en. In: *Lecture Notes in Computer Science*. Lecture notes in computer science. Cham: Springer International Publishing, 2022, pp. 272–284. doi: 10.1007/978-3-031-08999-2_22.

25 X. Zeng et al. “DBTrans: A Dual-Branch Vision Transformer for Multi-Modal Brain Tumor Segmentation”. In: *Lecture Notes in Computer Science*. Lecture notes in computer science. Cham: Springer Nature Switzerland, 2023, pp. 502–512. doi: 10.1007/978-3-031-43901-8_48.

26 L. Wen et al. “A deep ensemble learning framework for glioma segmentation and grading prediction”. en. In: *Sci. Rep.* 15 (1 Feb. 6, 2025), p. 4448. doi: 10.1038/s41598-025-87127-z.

27 U. C. Anazodo et al. “A framework for advancing sustainable magnetic resonance imaging access in Africa”. en. In: *NMR Biomed.* 36 (3 Mar. 2023), e4846. doi: 10.1002/nbm.4846.

28 F. Proença et al. “Neuroimaging safety during pregnancy and lactation: a review”. en. In: *Neuroradiology* 63 (6 June 2021), pp. 837–845. doi: 10.1007/s00234-021-02675-1.

29 E. Blumfield et al. “Survey of gadolinium-based contrast agent utilization among the members of the Society for Pediatric Radiology: a Quality and Safety Committee report”. en. In: *Pediatr. Radiol.* 47 (6 May 2017), pp. 665–673. doi: 10.1007/s00247-017-3807-z.

30 M. Visser et al. “Inter-rater agreement in glioma segmentations on longitudinal MRI”. en. In: *NeuroImage Clin.* 22 (Feb. 22, 2019), p. 101727. doi: 10.1016/j.nicl.2019.101727.

31 O. M. Henriksen et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 1: Perfusion and Diffusion Techniques”. en. In: *Front. Oncol.* 12 (Mar. 3, 2022), p. 810263. doi: 10.3389/fonc.2022.810263.

32 T. C. Booth et al. “High-Grade Glioma Treatment Response Monitoring Biomarkers: A Position Statement on the Evidence Supporting the Use of Advanced MRI Techniques in the Clinic, and the Latest Bench-to-Bedside Developments. Part 2: Spectroscopy, Chemical Exchange Saturation, Multiparametric Imaging, and Radiomics”. en. In: *Front. Oncol.* 11 (2021), p. 811425. doi: 10.3389/fonc.2021.811425.

33 R. Guillevin et al. “Low-grade gliomas: the challenges of imaging”. en. In: *Diagn. Interv. Imaging* 95 (10 Oct. 2014), pp. 957–963. doi: 10.1016/j.diii.2014.07.005.

Supplementary Material

Supplementary Table S1.

Table S1: Missed non-enhancing cases and wrongly segmented enhancing lesions.

Rater	Non-enhancing lesion identification	Enhancing lesion identification and segmentation
R-AI	Identified 1/3 non-enhancing case	Classified one enhancing case (0.6mL) as non-enhancing (Figure 6)
B	Correctly identified all 3 non-enhancing cases	Mislabeled 2 enhancing lesions (<1mL each) as non-enhancing (Figure 6)
C	Identified all non-enhancing cases	Misclassified 5 enhancing lesions as non-enhancing: 2 <1mL, 2 between 2–3mL, 1 large lesion (297mL)
D	Did not identify any non-enhancing case	Markedly oversegmented small enhancing lesions: 5.2mL vs GT 0.5mL (10-fold), 30.8mL vs GT 1.1mL (27-fold)
E	Failed to identify all non-enhancing cases	Substantially oversegmented lesions <10mL; one case 30.7mL vs GT 0.6mL (50-fold overestimation)

Chapter 8

Gadolinium-free MRI using artificial intelligence in glioma: a clinically-oriented benchmark study

Ivar J.H.G. Wamelink, Rajeev A. Essed, Stefan.de Vries, Ellis Donders, Frederik Barkhof, Alle Meije Wink, Vera C. Keil, Stefano Trebeschi

As submitted to Investigative Radiology

Abstract

Objectives: To benchmark the technology readiness of deep learning algorithms designed to replace post-gadolinium contrast images in glioma imaging.

Materials and Methods: Using four independent open-source datasets in a retrospective design, four previously published contrast-enhanced T1-weighted image synthesis and six segmentation algorithms to generate contrast-enhanced volumes without relying on contrast-enhanced input sequences were trained. Clinically oriented performance evaluation included enhancement detection, prognostic assessment, and treatment response evaluation according to RANO 2.0 criteria.

Results: Segmentation algorithms demonstrated a higher sensitivity (97% vs 76%) and negative predictive value (NPV) (60% vs 35%) than synthesis algorithms, but lower specificity (18% vs 51%) for enhancement detection. Group-wise comparisons confirmed significantly higher sensitivity, NPV, and accuracy for segmentation algorithms (P=0.04). The intraclass correlation coefficient (ICC) between enhancing tumor volume and reference volume was poor for all synthesis models and two segmentation models (ICC 0.00–0.35), but moderate to good for the remaining four segmentation models (ICC 0.61–0.81). Enhancement volumes derived from segmentation models were prognostic in all but one model ($P < 0.001$). Adding age, sex, and surgery type as covariates further improved the c-index from 0.60 to 0.68. Agreement in response evaluation was significant in one of four synthesis algorithms and six of seven segmentation algorithms ($k = 0.38$-0.68).

Conclusion: Current GBCA-free glioma segmentation algorithms show a promising - but for clinical use still insufficient - diagnostic and prognostic performance. While approaching AI-derived ground truth annotations, they suffer from limited specificity and are biased towards training data.

Key words: MRI, Gadolinium based contrast agent reduction, AI, Glioma, Advanced MRI, Segmentation, CET1 synthesis

Introduction

Contrast-enhanced MRI using gadolinium-based contrast agents (GBCA) is widely used for adult-type diffuse glioma diagnosis and monitoring.[1] GBCAs are administered intravenously, reducing T1 relaxation time on T1-weighted post-contrast images (CET1), resulting in increased signal intensity in tumor structures that accumulate GBCA.[2] CET1 images are standard care to define resection and radiation margins without any clinically established alternative.[3, 4]

However, GBCA injection carries potential risks.[5] Studies have documented gadolinium accumulation in multiple tissues after follow-up imaging with GBCAs, prompting regulators to advise healthcare professionals to minimize repeated GBCA imaging studies when possible.[5–8] GBCAs from renal excretion cannot be removed by wastewater treatment plants, increasing their concentration in drinking water up to 30 times between 2009 and 2021.[9] Furthermore, GBCA reduction could save millions of dollars in healthcare annually.[10] Given the uncertainty surrounding the long-term effects of gadolinium retention, attention has also turned to the patient experience. Forty-one percent of patients reported some discomfort receiving a cannula, and 37% would prefer GBCA-free MRI protocols if diagnostic non-inferiority were assured.[11]

This situation has led researchers to explore artificial intelligence (AI)-based methods to avoid GBCA administration in diagnostic neuroradiology, especially neuro-oncological imaging, which is a major consumer of GBCA-based MRI.[3] AI-based models can be developed on standard neuro-oncological imaging protocols to increase their deployability. They should produce stable results despite variable training sequences, parameters, and scan orientations, allowing real-world clinical deployment. Two AI-based approaches are primarily pursued using non-GBCA-enhanced images: 1) generative algorithms trained to create synthetic CET1 images, or 2) segmentation algorithms trained to predict the enhancing area.[3] Both approaches rely on the assumption that 1) tumor appearance on non-GBCA sequences reflects blood-brain barrier disruption, therefore correlating with GBCA enhancement, and that 2) AI can extrapolate contrast enhancement from non-GBCA images.

Existing work reported technical metrics such as image reconstruction or voxel overlap that do not readily translate into clinical utility.[12, 13] It remains unexplored whether current GBCA-free AI approaches meaningfully capture endpoints directly relevant to patient management - such as the prognostic value of detected enhancement or agreement with RANO criteria in longitudinal scans.

Unlike prior methodological evaluations, we assess real-world technological readiness by comparing segmentation and synthesis approaches against the gold standard of automatic GBCA identification. We hypothesize that AI-based algorithms designed to reduce the use of GBCA in glioma imaging, remain inferior to segmentation algorithms trained with GBCA, and are not yet suitable for routine clinical implementation. To test this, we benchmark open-source, peer-reviewed algorithms for quantifying GBCA-equivalent tumor enhancement and assess their technological readiness level (TRL)

from non-contrast MRI scans, using multiple datasets and clinically meaningful endpoints. Our validation focuses on three clinical aspects: accuracy of enhancement detection, prognostic utility, and treatment response assessment - providing, to our knowledge, the first systematic evaluation of whether GBCA-free AI can match contrast-enhanced MRI in clinical neuro-oncology.

Materials and Methods

The local ethics committee approved this study (2021.0437). Workflow is depicted in Figure 1.

Dataset selection

We included four large open-source datasets: BraTS 2021[14–16], UCSF-PDGM[17], EGD[18], and IMAGO[19] with retrospective glioma MR imaging (Fig 1. Blue). The IMAGO subset used for response evaluation (first post-treatment scans of patients who underwent biopsy and radiochemotherapy) is not publicly accessible. We excluded 262 overlapping cases between BraTS and the UCSF-PDGM (Supplemental Digital Content Table 1). Table 1 shows data characteristics. The BraTS and UCSF-PDGM datasets were used as training data, and the EGD and IMAGO datasets were used as independent test sets to maximize generalizability (Fig 1. Green).

Table 1: Dataset demographics.

Dataset	BraTS	UCSF	EGD	IMAGO
N	989*	501	774	500
Mean Age (SD)	N/A	56 (15)	unknown	55 (15)
Sex (Male in %)	N/A	299 (60%)	492 (64%)	303 (61%)
HGG (% of total)	N/A	445 (89%)	716 (93%)	420 (84%)
IDH status known (% of total)	N/A	501 (100%)	467 (60%)	500 (100%)
1p/19q co-deletion status known (% of total)	N/A	410 (82%)	259 (33%)	170 (34%)
Survival known (% of total)	N/A	500 (99%)	unknown	500 (100%)
ADC available (% of total)	N/A	501 (100%)	unknown	497 (99%)

*These are the number of included BraTS cases after excluding overlapping cases with UCSF. SD: standard deviation; IDH: isocitrate hydrogenase; ADC: apparent diffusion coefficient; HGG: high-grade glioma

Datasets were selected based on: (1) adult diffuse glioma cases, (2) availability of CET1, pre-contrast T1-weighted (T1w), T2-weighted (T2w), and T2-FLAIR images, and (3) Ground truth (GT) segmentations (1: non-enhancing tissue; 2: necrotic tissue; 3: enhancing tissue). Where expert-provided segmentations were missing, GT segmentations using the PICTURE network and CET1, T1w, T2w, and T2-FLAIR images were generated (validated software Fig 1. Lilac).[20, 21] The rationale is that models trained with

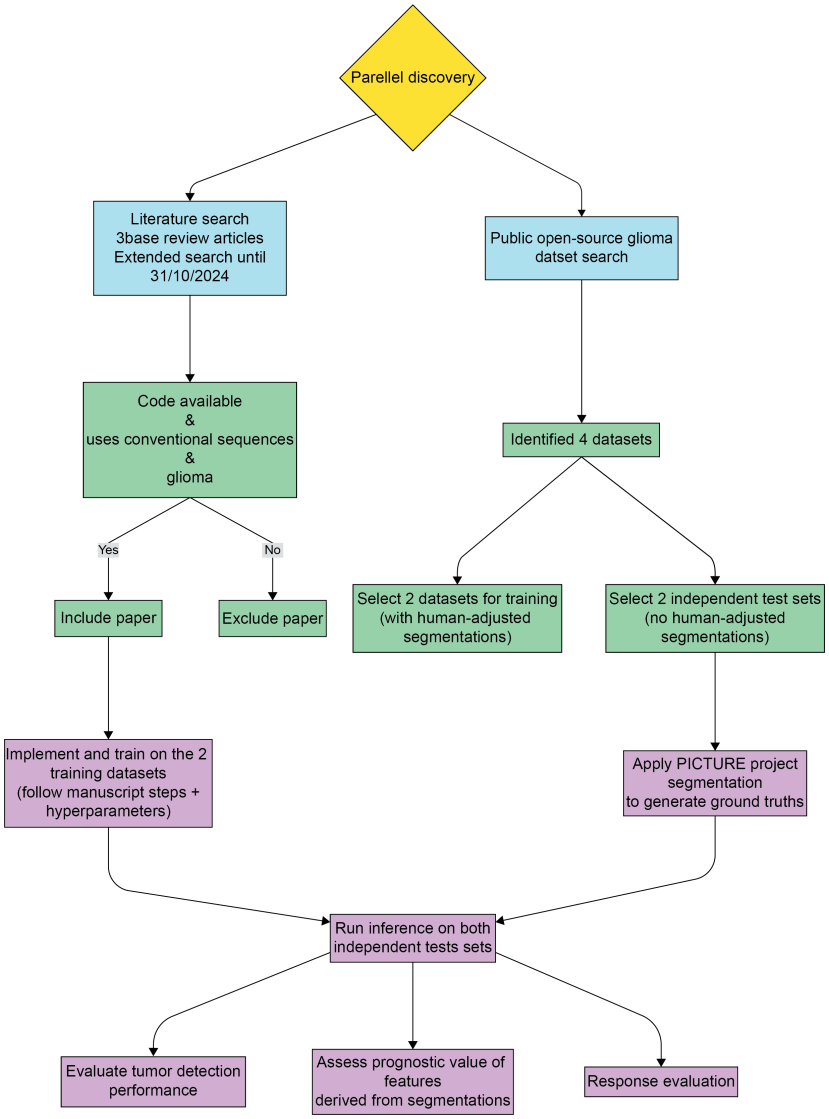


Figure 1: Workflow diagram of the benchmark study. Blue: Literature and database search;Green: paper inclusion and exclusion and dataset split; Lilac: model training, datasetpostprocessing, and inference: enhancement detection, prognostic performance, and responseevaluation.

CET1 images best reflect clinical practice, where CET1 is standard for delineating tumor extent, and provide an approximate upper bound of performance.

Selection and training of AI Algorithms

To ensure unbiased AI algorithm selection, we refer to three recent review studies for original publications: Pasquini et al.,[13] Moya-Sáez et al.,[12] Azad et al.[22] Using their terminology, we additionally abstract/title searched PubMed and Google Scholar using "synthetic postcontrast" OR "postcontrast generation" OR "sequence agnostic segmentation" OR "segmenting with missing modalities" on October 31, 2024 (Fig 1. Blue). Inclusion criteria were: (1) code available either publicly or on request, (2) use of at least CET1 and one of the other routine MRI sequences mentioned above. Exclusion criteria were: (1) usage of unconventional sequences such as susceptibility weighted imaging or low dose CET1, (2) non-glioma studies, (3) source code failed to execute (Fig 1. Green).

All algorithms were trained according to the authors' published setups, including preprocessing and hyperparameters. Entire datasets were assigned to training/validation or test cohorts to ensure completely independent test sets and prevent data leakage. Consistent training–validation splits were used across algorithms.
All segmentation models were combined into an ensemble model (Z) by applying majority voting to their outputs, under the assumption that consensus among models improves robustness and reduces the influence of individual model errors.

To support reproducibility, algorithm development, and standardized evaluation of future GBCA-equivalent tumor enhancement AI algorithms, we provide open access to our codebase and analytical framework.[23]

Experiments and statistics

Experiments were performed on complete datasets, which included oligodendroglioma, astrocytoma, and glioblastoma (GBM) subtypes. To further assess subtype-specific performance, a dedicated sub-analysis was carried out on glioblastoma cases. GBCA-free segmentation algorithms were directly evaluated against the GT segmentations of the test datasets (Fig 1. Lilac). GBCA-free CET1 synthesis algorithms were evaluated indirectly by generating synthetic CET1 images and segmenting those using the selected PICTURE algorithm, then comparing to GT.[20, 21]

For each benchmarked algorithm, we evaluated three key application tasks relevant to the clinical utility of CET1 imaging:

1. Enhancement detection performance

To determine whether an algorithm can detect contrast enhancement, we used specificity, sensitivity, positive predictive value, and negative predictive value. Enhancement was considered detected if at least one voxel was labeled as enhancing. Because reliable tumor volume estimation is essential for treatment planning, we assessed the agreement between the algorithm's predictions and the GT using the intraclass correlation coefficient (ICC). Pearson correlation was used to assess the inter-algorithm relationships, specifically by comparing the enhancing volumes and the Dice scores generated by each model. We used the same protocol for the tumor core, which consisted of the enhancing and necrotic tissue combined (Fig 2).

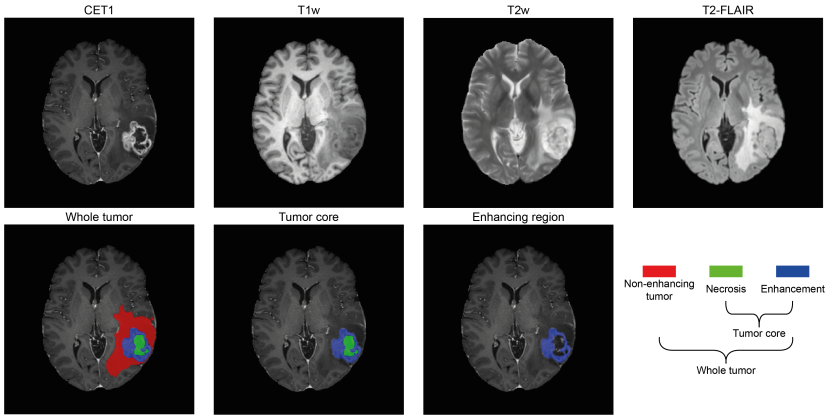


Figure 2: All four conventional sequences (top row) with segmentations of the enhancing tumor(blue), tumor core (blue + green), and total tumor (blue + green + red).

2. Prognostic performance

We quantified the prognostic value of segmentation-derived covariates (enhancing volume, tumor core volume, total tumor volume (enhancing, necrotic and peri-tumoral tissue combined, Fig. 2), and presence of enhancement) using hazard ratios (HR) from Cox proportional hazards models and the concordance index (c-index), to compare the prognostic performance of algorithms. Prognostic metrics were computed for the GT algorithm and for each algorithm independently. For all included algorithms, a mean c-index was calculated from 1,000 bootstrap iterations. The individual discriminatory performance of the ten predictive algorithms and the ground truth (GT) was investigated through pairwise comparisons.

3. Response evaluation

CET1 plays an important role in assessing response after chemo-radiotherapy. The current clinical standard is RANO 2.0, which we used to test whether GBCA-free outputs retain outcome-relevant performance.[24] Response evaluation requires follow-up scans, but these algorithms are not typically trained to handle post-surgical effects. To minimize this, we limited the analysis to biopsy-only patients.[25, 26] RANO response was determined on the follow-up scan using the PICTURE segmentation network, and agreement with the GT was assessed using Cohen's kappa.

The statistical significance level was set at 0.05. A Bonferroni correction was applied.

Results

AI Algorithms

Four algorithms for synthetic CET1 generation[27–30] and six algorithms for GBCA-free tumor enhancement segmentation could be successfully benchmarked (Fig. 3, Table 2).[31–36]

Supplemental Digital Content Table 2 contains all identified manuscripts and algorithms. Identified algorithms that generated synthetic CET1 often used generative adversarial networks (GAN) (Supplemental Digital Contental table 2), however, of the implemented algorithms, only 1 used a GAN. The U-Net was frequently used for both image synthesis and segmentation (N=8/10).[27, 28, 31–36] Only one segmentation algorithm, J, employed an additional synthesis loss (reconstruction loss; Weighted Mean Squared Error) to improve tumor segmentation.[32]

While most algorithms used all four routine MRI sequences (T1w, CET1, T2w, and T2-FLAIR), three relied on alternative combinations of input data.[29–31] One algorithm used T1w, T2w, and T2-FLAIR,[31] another algorithm used CET1, T1w, T2w, and apparent diffusion coefficient,[30] and one algorithm used only T1w and CET1.[29]

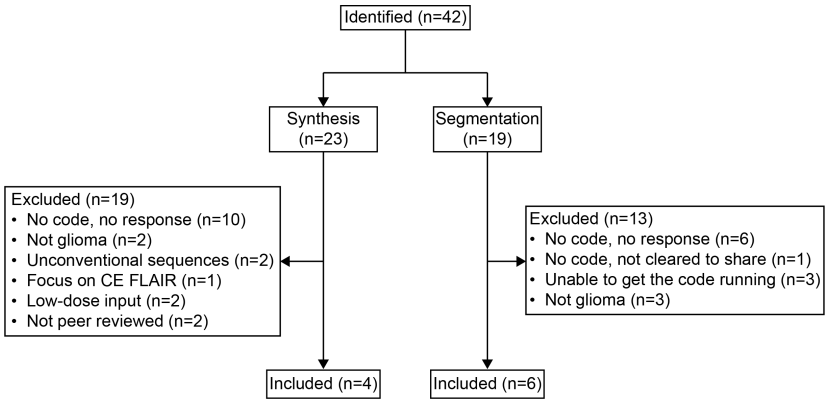


Figure 3: Consort diagram of included and excluded articles. All identified articles can be found in Supplemental Digital Content Table 2.

Table 2: Summary of the included articles. *These sequences were used during training. During inference a zero mask was inserted.

Pseudo name	Journal	1st author	Year	Algorithm group	Architecture	Input sequence	Output sequence
A	IEEE transactions on medical imaging	Sharma et al.27	2019	Synthesis	The algorithm can generate any combination of missing sequences in a single forward pass using implicit conditioning and curriculum learning, eliminating the need for multiple algorithms or retraining.	T1, T2, FLAIR	CET1
B	IEEE transactions on biomedical engineering	Chen et al.30	2023	Synthesis	They use an architecture that preserves high resolution multi scale representations and aggregates multi-scale information in parallel. They also use a local loss that improves the performance in tumor regions.	T1, T2, ADC	CET1
C	2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)	Piening et al.29	2024	Synthesis	They use an end-to-end unet with the mean absolute error as the loss.	T1	CET1

Continued on next page

Table 2 – *Continued from previous page*

Pseudo name	Journal	1st author	Year	Algorithm group	Architecture	Input sequence	Output sequence
D	Journal of Applied Clinical Medical Physics	Osman et al.28	2023	Synthesis	This article integrates residual blocks with densely connected dilated convolutions at the bottleneck to capture multi scale global context while preserving resolution, thereby improving segmentation accuracy and reducing overfitting.	T1, T2, FLAIR	CET1
E	Brain Communications	Ruffle et al.31	2023	Segmentation	They use the nnunet architecture.	T1, T2, FLAIR	3 labels
F	MIDL	Vadacchino et al.36	2021	Segmentation	They combine hierarchical discriminators and adversarial knowledge distillation, enabling a student network to learn multi scale latent representations from a full modality teacher and thus handle missing crucial modalities during inference.	CET1*, T1, T2, FLAIR	3 labels

Continued on next page

Table 2 – *Continued from previous page*

Pseudo name	Journal	1st author	Year	Algorithm group	Architecture	Input sequence	Output sequence
G	2021 IEEE/CVF International Conference on Computer Vision (ICCV)	Ding et al.33	2021	Segmentation	This article uses a region aware fusion module and segmentation based regularizer, which adaptively fuses multi modal MRI features based on tumor region probabilities to handle missing modalities and enhance brain tumor segmentation accuracy.	CET1*, T1, T2, FLAIR	3 labels
H	Proceedings of Machine Learning Research	Azad et al.34	2022	Segmentation	SMU-Net decomposes features into content and style spaces, using matching modules and mutual information modeling to transfer knowledge from full-modality to missing-modality algorithms for robust brain tumor segmentation.	CET1*, T1, T2, FLAIR	3 labels
I	Journal of Digital Imaging	Feng et al.35	2023	Segmentation	nnunet with sequence dropout	CET1*, T1, T2, FLAIR	3 labels

Continued on next page

Table 2 – *Continued from previous page*

Pseudo name	Journal	1st author	Year	Algorithm group	Architecture	Input sequence	Output sequence
J	Simulation and Synthesis in Medical Imaging	Mehta et al.32	2018	Segmentation	RS-Net leverages a shared 3D U-net latent representation to simultaneously synthesize missing MRI sequences and perform accurate multi-class tumor segmentation at full resolution.	CET1*, T1, T2, FLAIR	3 labels
Z	-	-	-	Segmentation	Ensemble segmentation by majority vote of all six included segmentation algorithms.	CET1*, T1, T2, FLAIR	3 labels

Application task 1. Enhancement detection performance

The GT segmentation algorithm identified 1,069 contrast-enhancing tumors (81% overall; independent pooled test set): 432 of 500 in IMAGO (86%) and 637 of 774 in EGD (82%).

All algorithms tended to over-detect the presence of tumor enhancement. Across algorithms, sensitivity was generally high but specificity was low (Table 3). Segmentation algorithms outperformed synthesis methods in sensitivity and negative predictive value (NPV), while synthesis methods had higher specificity (Table 3). Segmentation algorithms performed significantly better against synthesis algorithms (accuracy difference 0.05–0.38) (Supplemental Digital Content table 3). Group-wise statistical comparisons between synthesis and segmentation algorithms revealed significantly higher sensitivity, NPV, and accuracy for enhancement identification in segmentation algorithms ($p = 0.04$ for all three metrics) when applied to EGD data. For the IMAGO dataset, this trend was observed only for sensitivity and accuracy ($p = 0.06$ for both metrics). Detailed results are shown in Supplemental Digital Content Table 4.

Similar patterns were observed when detecting the entire tumor core. Five out of six segmentation algorithms in EGD were significantly better than the ensemble algorithm, and three algorithms (I, F, and J) significantly outperformed E. Detailed results in Supplemental Digital Content Tables 3, 4, and 5.

Estimating the exact enhancing tumor volume, most segmentation algorithms showed moderate-to-good agreement (intraclass correlation coefficient (ICC) 0.61–0.81), while synthesis methods performed poorly (ICC <0.35). Algorithms with low ICCs tended to oversegment substantially (Table 3). All synthesis algorithms underestimated volume (range: 9.35–20.21), whereas all segmentation algorithms overestimated it (3.92–9.34 mL). In particular, segmentation algorithms with poor ICCs showed the most extreme oversegmentation: H produced volumes of 257.13 mL (IMAGO) and 218.71 mL (EGD), and J produced 25.38 (IMAGO) and 38.66 (EGD) (Supplemental Digital Content table 4).

For tumor core, synthesis algorithms improved markedly (ICC up to 0.72), whereas segmentation performance decreased slightly except for H and I (Supplemental Digital Content table 6). Whereas all synthesis algorithms previously underestimated the enhancing tumor volume, performance on the tumor core was more heterogeneous: some algorithms continued to undersegment (5.46–14.42 mL), while others oversegmented (3.63–22.57 mL). Segmentation algorithms generally undersegmented the core (2.14–8.35 mL) and, in some cases, oversegmented (1.99–4.47 mL). Algorithms H and J continued to show substantial oversegmentation (24.32–262.38 mL). Results were consistent across datasets (Supplemental Digital Content Tables 4, 5, and 6).

Table 3: Enhancement detection performance per algorithm, presented as mean (95% CI). Bold indicates lowest and highest values. PPV: Positive Predictive Value; NPV: Negative Predictive Value; DSC: Dice Score Coefficient; ICC: Intraclass Correlation Coefficient

Model	Database	Sensitivity	Specificity	PPV	NPV	DSC	ICC	mAVE (mL)
A	IMAGO	0.78 (0.74 - 0.82)	0.58 (0.48 - 0.67)	0.89 (0.85 - 0.92)	0.39 (0.31 - 0.47)	0.18 (0 - 0.5)	0.15 (0.0 - 0.29)	-14.11
	EGD	0.80 (0.77 - 0.83)	0.51 (0.40 - 0.61)	0.92 (0.90 - 0.94)	0.25 (0.19 - 0.32)	0.17 (0 - 0.47)	0.15 (-0.01 - 0.3)	-17.73
C	IMAGO	0.95 (0.93 - 0.97)	0.10 (0.05 - 0.16)	0.81 (0.78 - 0.84)	0.33 (0.18 - 0.50)	**0.06** (0 - 0.23)	0.12 (0.0 - 0.23)	-13.53
	EGD	0.90 (0.88 - 0.93)	0.19 (0.12 - 0.28)	0.89 (0.87 - 0.91)	0.22 (0.13 - 0.31)	0.07 (0 - 0.27)	0.07 (-0.02 - 0.16)	-18.96
D	IMAGO	0.50 (0.46 - 0.55)	**0.92** (0.86 - 0.96)	0.96 (0.93 - 0.98)	0.31 (0.26 - 0.36)	0.21 (0 - 0.59)	0.13 (-0.0 - 0.26)	-15.21
	EGD	**0.45** (0.42 - 0.49)	0.88 (0.81 - 0.94)	**0.97** (0.94 - 0.98)	**0.18** (0.15 - 0.22)	0.18 (0 - 0.54)	0.04 (-0.02 - 0.1)	-20.21
B	IMAGO	0.92 (0.90 - 0.95)	0.28 (0.20 - 0.37)	0.84 (0.81 - 0.87)	0.48 (0.35 - 0.60)	0.13 (0 - 0.38)	0.35 (0.18 - 0.49)	-9.35
E	IMAGO	0.96 (0.94 - 0.97)	0.29 (0.21 - 0.38)	0.85 (0.81 - 0.88)	0.62 (0.49 - 0.75)	0.43 (0.1 - 0.76)	0.81 (0.73 - 0.86)	5.16
	EGD	0.95 (0.93 - 0.97)	0.22 (0.14 - 0.32)	0.90 (0.88 - 0.92)	0.38 (0.25 - 0.51)	**0.5** (0.17 - 0.83)	0.86 (0.8 - 0.9)	5.30
F	IMAGO	0.98 (0.97 - 0.99)	**0.01** (0.00 - 0.03)	**0.80** (0.77 - 0.84)	0.11 (0.00 - 0.43)	0.22 (0 - 0.46)	0.98 (0.97 - 0.99)	4.56
	EGD	**1.00** (1.00 - 1.00)	0.01 (0.00 - 0.04)	0.88 (0.86 - 0.90)	**1.00** (0.00 - 1.00)	0.33 (0.05 - 0.61)	1.00 (1.00 - 1.00)	5.38

Continued on next page

Table 3 – *Continued from previous page*

Model	Database	Sensitivity	Specificity	PPV	NPV	DSC	ICC	mAVE (mL)
G	IMAGO	0.96 (0.94 - 0.98)	0.14 (0.08 - 0.21)	0.82 (0.78 - 0.85)	0.45 (0.28 - 0.62)	0.32 (0.03 - 0.61)	0.96 (0.94 - 0.98)	5.32
	EGD	0.99 (0.98 - 0.99)	0.05 (0.01 - 0.10)	0.88 (0.86 - 0.91)	0.33 (0.10 - 0.60)	0.37 (0.08 - 0.66)	0.99 (0.98 - 0.99)	9.34
H	IMAGO	0.98 (0.96 - 0.99)	0.14 (0.08 - 0.21)	0.82 (0.79 - 0.85)	0.60 (0.40 - 0.79)	0.35 (0.03 - 0.67)	**0.00** (-0.08 - 0.08)	**257.13**
	EGD	0.98 (0.97 - 0.99)	0.06 (0.02 - 0.12)	0.89 (0.86 - 0.91)	0.31 (0.11 - 0.55)	0.42 (0.09 - 0.75)	**0.00** (-0.06 - 0.07)	**218.71**
I	IMAGO	0.94 (0.91 - 0.96)	0.43 (0.34 - 0.53)	0.87 (0.84 - 0.90)	0.62 (0.51 - 0.73)	0.44 (0.11 - 0.77)	0.75 (0.64 - 0.82)	6.73
	EGD	0.93 (0.91 - 0.95)	0.42 (0.32 - 0.52)	0.92 (0.90 - 0.94)	0.45 (0.35 - 0.56)	**0.5** (0.17 - 0.83)	0.81 (0.72 - 0.86)	6.40
J	IMAGO	0.99 (0.98 - 1.00)	0.11 (0.06 - 0.18)	0.82 (0.79 - 0.85)	0.80 (0.57 - 1.00)	0.24 (0 - 0.48)	0.33 (0.02 - 0.55)	25.38
	EGD	0.99 (0.98 - 1.00)	0.02 (0.00 - 0.06)	0.88 (0.86 - 0.90)	0.20 (0.00 - 0.56)	0.23 (0.01 - 0.45)	0.23 (-0.03 - 0.45)	38.66
Z	IMAGO	0.96 (0.94 - 0.98)	0.34 (0.25 - 0.43)	0.86 (0.82 - 0.89)	0.69 (0.56 - 0.82)	0.42 (0.09 - 0.75)	0.78 (0.73 - 0.83)	3.98
	EGD	0.96 (0.94 - 0.97)	0.20 (0.12 - 0.29)	0.90 (0.88 - 0.92)	0.39 (0.25 - 0.53)	0.47 (0.14 - 0.8)	**0.86** (0.82 - 0.89)	**3.92**

Application task 2. Prognostic performance
Median OS was 82 weeks (1.6 years; 1–789 weeks). Both enhancing and tumor core GT volumes were associated with OS (HR=1.26, CI: 1.15–1.39, p-value: <0.05; HR=1.23, CI: 1.12–1.36, p-value: <0.05). Both remained significant after adjusting for possible baseline confounding variables, age, sex, and surgery type. The c-index was 0.6 and 0.58 respectively, increasing to 0.68 after adjustment for confoundings.

All but one segmentation algorithm (H) performed comparably to GT in terms of HR and c-indices (Table 4), while all synthesis algorithms initially underperformed. Again, adjusting for confounders improved results, with two synthesis algorithms (A and B) now also yielding significant results (CI: 1.00–1.22, p-value: 0.04 and CI: 1.04–1.19, p-value: <0.05). Similar results were observed for tumor core.

Some algorithms, notably I and E, individually demonstrated high prognostic performance, even exceeding the GT in some instances (Fig. 4). However, bootstrapped hypothesis tests (with false discovery rate correction) confirmed no statistically significant differences between any pair of individual algorithms or between any algorithm and the ground truth.

Application task 3. Response Evaluation
Sixty-six patients had at least one follow-up MRI after biopsy to determine RANO response. Based on the GT, we observed that 39 patients showed PD (59%), 11 SD (17%), 6 PR (9%), and 10 CR (15%).

Three out of four synthesis algorithms and one out of five segmentation algorithms yielded non-significant kappas. In the remaining ones, we observed the lowest agreement for algorithm F (Kappa=0.38; Figure 5). All segmentation algorithms showed a kappa between 0.38 and 0.68.

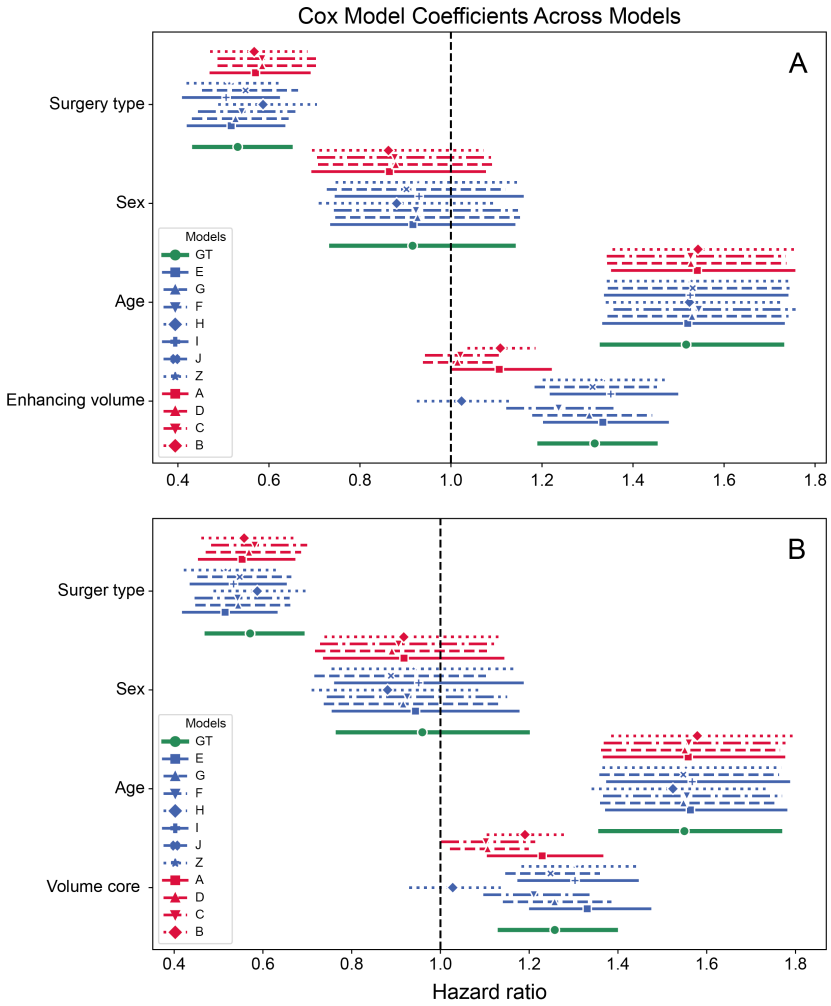


Figure 4: Cox proportional hazard models for enhancing volume (A) and tumor core volume (B) in mL with sex, age (in years) and surgery type (biopsy, partial resection, or gross totalresection) as additional covariates. Blue: segmentation models, red: synthesis models, green:ground truth

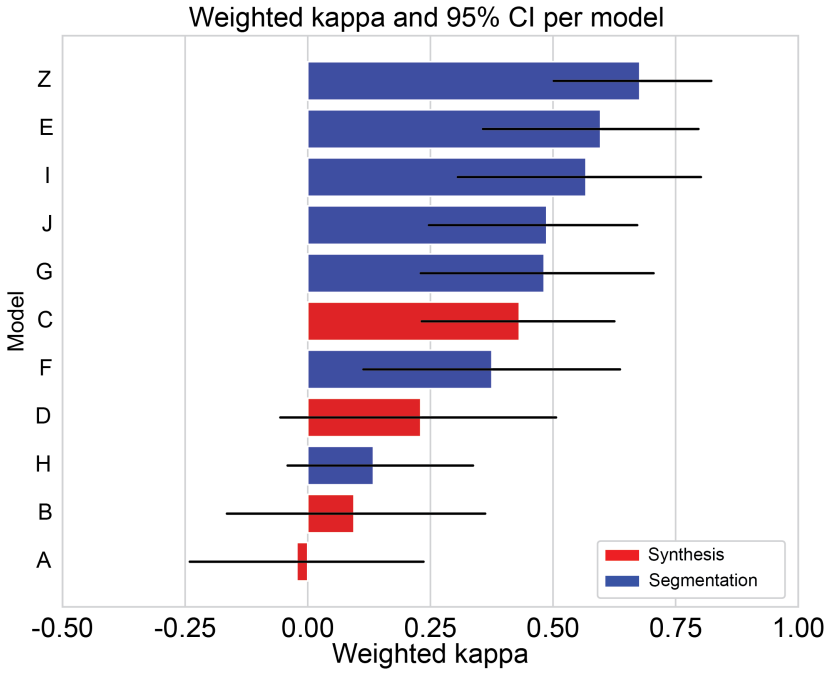


Figure 5: Weighted kappa for RANO response with 95% confidence interval for each model. CI:confidence interval

Table 4: Hazard ratios, c-index, and 95% confidence interval for overall survival prediction while only using segmentation volumes as prognostic markers. The tumor core contains the enhancing and necrotic tumor volume. CI: confidence interval HR: hazard ratio

	Enhancing volume					Tumor core					Total volume				
Model	c-index	HR	95% CI low	95% CI high	p-value	c-index	HR	95% CI low	95% CI high	p-value	c-index	HR	95% CI low	95% CI high	p-value
GT	0.60	1.26	1.15	1.39	<0.001	0.58	1.23	1.12	1.36	<0.001	0.57	1.24	1.13	1.37	<0.001
A	0.51	1.05	0.95	1.16	0.32	0.53	1.15	1.04	1.27	0.01	0.54	1.19	1.08	1.31	<0.001
B	0.55	1.07	0.99	1.15	0.07	0.55	1.12	1.03	1.21	<0.005	0.56	1.17	1.08	1.26	<0.001
C	0.51	1.01	0.93	1.09	0.79	0.50	1.03	0.94	1.13	0.56	0.54	1.10	1.00	1.21	0.04
D	0.49	1.00	0.92	1.09	0.96	0.50	1.04	0.95	1.13	0.43	0.54	1.09	1.01	1.18	0.02
E	0.59	1.27	1.15	1.40	<0.001	0.58	1.22	1.11	1.34	<0.001	0.57	1.26	1.14	1.39	<0.001
F	0.57	1.16	1.06	1.28	<0.001	0.57	1.13	1.02	1.24	0.01	0.57	1.21	1.10	1.34	<0.001
G	0.58	1.23	1.12	1.36	<0.001	0.57	1.20	1.09	1.32	<0.001	0.57	1.24	1.12	1.37	<0.001
H	0.58	1.04	0.94	1.15	0.48	0.57	1.04	0.94	1.15	0.46	0.57	1.04	0.94	1.15	0.42
I	0.60	1.25	1.13	1.38	<0.001	0.58	1.21	1.10	1.33	<0.001	0.57	1.26	1.14	1.40	<0.001
J	0.58	1.27	1.14	1.40	<0.001	0.57	1.21	1.11	1.33	<0.001	0.58	1.22	1.12	1.33	<0.001
Z	0.59	1.24	1.13	1.37	<0.001	0.58	1.19	1.08	1.31	<0.001	0.57	1.27	1.14	1.41	<0.001

Subanalysis. Glioblastoma
Segmentation dice was superior in higher-grade tumors. Out of all GBM patients, 627/678 had an enhancing tumor (328/366 IMAGO, 299/312 EGD). Changes in enhancement detection were mixed (Supplemental Digital Content Tables 7 and 8). In IMAGO, most models lost sensitivity (e.g. A decreased from 0.78 to 0.66, and E decreased from 0.96 to 0.83), with only B gaining specificity (0.28 to 0.37) and positive predictive value (PPV) (0.84 to 0.93). In contrast, in EGD, several models improved specificity (e.g. Z: 0.20 to 0.39, J: 0.02 to 0.18) while maintaining very high sensitivity ($\geq$ 0.95), indicating selective but not consistent benefits of subgroup restriction. No notable differences were observed in ICC, prognostic performance, or response evaluation.

Discussion

This multi-algorithm, multi-dataset benchmark systematically assessed the technology readiness of state-of-the-art AI algorithms for GBCA-free tumor enhancement detection (task 1), prognosis (task 2), and response assessment (task 3) in glioma. Segmentation algorithms consistently outperformed synthesis algorithms across all tasks, showing significantly higher sensitivity, NPV, and accuracy on the EGD (P=0.04) and a strong trend on the IMAGO (P=0.06) datasets. ICC for synthesis models and two segmentation models was poor (0.00–0.35), but moderate-to-good for four segmentation models (0.61–0.81). Segmentation-derived enhancement volumes were prognostic for survival (P<0.001), and including age, sex, and surgery type improved the c-index from 0.60 to 0.68.

Segmentation algorithms consistently outperformed synthesis algorithms in enhancement detection regarding sensitivity but showing lower specificity, resulting in a modest overall accuracy. This performance pattern highlights a key limitation in current training paradigms. Open-source datasets are heavily skewed toward enhancing tumors – most of which are high-grade lesions – >80% in the test datasets (IMAGO and EGD) and 93% in the training datasets (BraTS and UCSF-PDGM). This imbalance introduces systematic bias, possibly causing algorithms to overpredict enhancement, especially in lower-grade tumors where enhancement is less pronounced or entirely absent.

Prognostic performance using enhancement volume and covariates age, sex, and surgery type obtained a similar c-index as found in a study on progression.[37]

In the exploratory subset of patients with follow-up imaging, segmentation algorithms outperformed synthesis algorithms in RANO response evaluation. While these findings suggest that segmentation algorithms may approach the level of reliability required for response assessment in clinical trials and practice, the limited sample size warrants cautious interpretation. Prior reports highlight the potential of automated segmentation for RANO-based assessment in glioma, whereas the role of synthesis methods remains less well established.[38]

Lower performance in synthesis-driven pipelines can also be attributed to domain shift. Although synthesis algorithms may produce visually plausible reconstructions, the gener-

ated intensity values and contrast distributions often deviate from those of the GT images. Similar effects of intensity shifts and imperfect distribution matching on downstream tasks have been discussed in the literature on medical image synthesis and domain adaptation.[39, 40]

Several limitations should be acknowledged. The heterogeneous naming and goals of the articles made systematic searching challenging, and the limited response rate to code-sharing requests may have introduced selection bias. The algorithms evaluated were trained and tested on retrospective data and showed similar results on both independent test sets, but prospective validation in a clinical workflow remains necessary. A single voxel was used for enhancement detection, which might inflate sensitivity at the cost of specificity due to false-positive segmentations. However, there is no standard lesion volume found in the literature and therefore any selected minimum volume is arbitrary. According to RANO 2.0, the minimal measurable threshold is a 1 cm linear diameter (or 1 x 1 cm bidimensional). Although RANO does not define volumetric criteria, this corresponds approximately to 0.52 mL for a spherical lesion. Using a threshold of 0.52 mL, a lower sensitivity and PPV but higher specificity and NPV were obtained.[24] Another limitation is that, out of 42 published methods, only 10 algorithms could be successfully trained. Despite increasing efforts toward code sharing, most approaches had to be excluded due to lack of publicly available implementations; consequently, our analysis was restricted to methods with accessible codebases. Response assessment was performed on an exploratory dataset. Furthermore, the GT segmentations for the IMAGO and EGD datasets were obtained without additional human curation. Although this may affect absolute agreement rates, it does not affect the relative comparisons between algorithms, which is the primary focus of this study. Because the PICTURE network generates segmentations from CET1 – the clinical standard for delineating tumor extent – its output serves as a robust ground truth. Comparisons with models that exclude CET1 are therefore both fair and clinically meaningful.

In conclusion, GBCA-free segmentation methods show diagnostic and prognostic promise but remain limited for clinical use, currently at Technology Readiness Level 3/4 with movement toward Level 5. Future work should prioritize algorithm specificity, validation in diverse cohorts, and integration of imaging and clinical features to boost robustness. Ultimately, translation into clinical practice hinges more on clinical applicability than visual quality.

References

1 M. Weller et al. “EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood”. en. In: *Nat. Rev. Clin. Oncol.* 18 (3 Dec. 8, 2020), pp. 170–186. doi: 10.1038/s41571-020-00447-z.

2 M. A. Ibrahim, B. Hazhirkarzar, and A. B. Dublin. “Gadolinium magnetic resonance imaging”. In: (2018).

3 I. J. H. G. Wamelink et al. “Brain tumor imaging without gadolinium-based contrast agents: Feasible or fantasy?” en. In: *Radiology* 310 (2 Feb. 2024), e230793. doi: 10.1148/radiol.230793.

4 G. Morana et al. “Use of contrast agents in oncological imaging: magnetic resonance imaging”. en. In: *Cancer Imaging* 13 (3 Sept. 23, 2013), pp. 350–359. doi: 10.1102/1470-7330.2013.9018.

5 *FDA Drug Safety Communication: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings*. 2018. (Visited on 10/01/2021).

6 B. J. Guo, Z. L. Yang, and L. J. Zhang. “Gadolinium deposition in brain: Current scientific evidence and future perspectives”. en. In: *Front. Mol. Neurosci.* 11 (Sept. 20, 2018), p. 335. doi: 10.3389/fnmol.2018.00335.

7 V. Gulani et al. “Gadolinium deposition in the brain: summary of evidence and recommendations”. en. In: *Lancet Neurol.* 16 (7 July 2017), pp. 564–570. doi: 10.1016/S1474-4422(17)30158-8.

8 *Gadolinium-containing contrast agents - referral*. en. Mar. 18, 2016. (Visited on 05/16/2025).

9 L. M. Krohn et al. “Impacts of COVID-19 and climate change on wastewater-derived substances in urban drinking water: Evidence from gadolinium-based contrast agents in tap water from Berlin, Germany”. en. In: *Water Res.* 259 (121847 Aug. 1, 2024), p. 121847. doi: 10.1016/j.watres.2024.121847.

10 *MRI Contrast Media Agents Market Size, Share & Trends Analysis Report By Product (Paramagnetic Agents, Superparamagnetic Agents), By Application, By End-use, By Type, By Region, And Segment Forecasts, 2023 - 2030*.

11 I. J. H. G. Wamelink et al. “The patients’ experience of neuroimaging of primary brain tumors: a cross-sectional survey study”. en. In: *J. Neurooncol.* 162 (2 Apr. 2023), pp. 307–315. doi: 10.1007/s11060-023-04290-x.

12 E. Moya-Sáez, R. de Luis-García, and C. Alberola-López. “Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI”. In: *Frontiers in Neuroimaging* 2 (2023). doi: 10.3389/fnimg.2023.1055463.

13 L. Pasquini et al. “Synthetic Post-Contrast Imaging through Artificial Intelligence: Clinical Applications of Virtual and Augmented Contrast Media”. en. In: *Pharmaceutics* 14 (11 Nov. 4, 2022). doi: 10.3390/pharmaceutics14112378.

14 U. Baid et al. “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification”. In: *arXiv [cs.CV]* (July 5, 2021).

15 B. H. Menze et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. en. In: *IEEE Trans. Med. Imaging* 34 (10 Oct. 2015), pp. 1993–2024. doi: 10.1109/TMI.2014.2377694.

16 S. Bakas et al. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”. en. In: *Sci Data* 4 (Sept. 5, 2017), p. 170117. doi: 10.1038/sdata.2017.117.

17 E. Calabrese et al. “The university of California San Francisco preoperative diffuse glioma MRI dataset”. en. In: *Radiol. Artif. Intell.* 4 (6 Nov. 2022), e220058. doi: 10.1148/ryai.220058.

18 S. R. van der Voort et al. “The Erasmus Glioma Database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma”. en. In: *Data Brief* 37 (107191 Aug. 2, 2021), p. 107191. doi: 10.1016/j.dib.2021.107191.

19 Ivar J.H.G Wamelink, Alle Meije Wink, Niels Verburg, Roeland S Eijgelaar, Emmanouil Koltsakis, Marcus Cakmak, Maarten Balder, Henk J.M.M Mutsaerts, Philip Witt Hamer Frederik Barkhof, Vera C. Keil. *IMAGO*. July 24, 2025. doi: 10.5281/zenodo.17582072.

20 H. G. Pemberton et al. “Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18911. doi: 10.1038/s41598-023-44794-0.

21 F. Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. en. In: *Nat. Methods* 18 (2 Feb. 2021), pp. 203–211. doi: 10.1038/s41592-020-01008-z.

22 R. Azad et al. “Addressing missing modality challenges in MRI images: A comprehensive review”. In: *Comput. Vis. Media (Beijing)* 11 (2 2025), pp. 241–268. doi: 10.26599/CVM.2025.9450399.

23 *Disclosed in case accepted*.

24 P. Y. Wen et al. “RANO 2.0: Update to the Response Assessment in Neuro-Oncology criteria for high- and low-grade gliomas in adults”. en. In: *J. Clin. Oncol.* 41 (33 Nov. 20, 2023), pp. 5187–5199. doi: 10.1200/JCO.23.01059.

25 R. H. Helland et al. “Segmentation of glioblastomas in early post-operative multimodal MRI with deep neural networks”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18897. doi: 10.1038/s41598-023-45456-x.

26 A. Bianconi et al. “Deep learning-based algorithm for postoperative glioblastoma MRI segmentation: a promising new tool for tumor burden assessment”. en. In: *Brain Inform.* 10 (1 Oct. 6, 2023), p. 26. doi: 10.1186/s40708-023-00207-6.

27 A. Sharma and G. Hamarneh. “Missing MRI pulse sequence synthesis using multimodal generative adversarial network”. In: *arXiv [eess.IV]* (Apr. 27, 2019).

28 A. F. I. Osman and N. M. Tamam. “Contrast-enhanced MRI synthesis using dense-dilated residual convolutions based 3D network toward elimination of gadolinium in neuro-oncology”. en. In: *J. Appl. Clin. Med. Phys.* 24 (12 Dec. 2023), e14120. doi: 10.1002/acm2.14120.

29 M. Piening et al. “Conditional generative models for contrast-enhanced synthesis of T1w and T1 maps in brain MRI”. In: *arXiv [eess.IV]* (Oct. 11, 2024).

30 C. Chen et al. “Synthesizing MR image contrast enhancement using 3D high-resolution ConvNets”. en. In: *IEEE Trans. Biomed. Eng.* 70 (2 Feb. 2023), pp. 401–412. doi: 10.1109/TBME.2022.3192309.

31 J. K. Ruffle et al. “Brain tumour segmentation with incomplete imaging data”. en. In: *Brain Commun.* 5 (2 Apr. 28, 2023), fcad118. doi: 10.1093/braincomms/fcad118.

32 R. Mehta and T. Arbel. “RS-net: Regression-segmentation 3D CNN for synthesis of full resolution missing brain MRI in the presence of tumours”. In: *arXiv [cs.CV]* (July 28, 2018).

33 Y. Ding, X. Yu, and Y. Yang. “RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation”. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (Montreal, QC, Canada). IEEE, Oct. 2021. doi: 10.1109/iccv48922.2021.00394.

34 R. Azad, N. Khosravi, and D. Merhof. “SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities”. In: *arXiv [cs.CV]* (Apr. 6, 2022).

35 X. Feng et al. “Brain tumor segmentation for multi-modal MRI with missing information”. en. In: *J. Digit. Imaging* 36 (5 Oct. 2023), pp. 2075–2087. doi: 10.1007/s10278-023-00860-7.

36 S. Vadacchino et al. “HAD-Net: A hierarchical adversarial knowledge distillation network for improved enhanced tumour segmentation without post-contrast images”. In: *arXiv [eess.IV]* (Mar. 30, 2021).

37 C. Jayachandran Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *Lancet Digit Health* 3 (12 Dec. 2021), e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

38 P. Kickingereder et al. “Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study”. en. In: *Lancet Oncol.* 20 (5 May 2019), pp. 728–740. doi: 10.1016/S1470-2045(19)30098-1.

39 J. P. Cohen, M. Luck, and S. Honari. “Distribution matching losses can hallucinate features in medical image translation”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Lecture notes in computer science. Cham: Springer International Publishing, 2018, pp. 529–536. doi: 10.1007/978-3-030-00928-1_60.

40 R. Zhang, T. Pfister, and J. Li. “Harmonic unpaired image-to-image translation”. In: *arXiv [cs.CV]* (Feb. 25, 2019).

Supplementary Material

Table S1: List of the excluded BraTS21 patients that are also in the UCSF-PDGM dataset

BraTS21_ID	BraTS21_ID	BraTS21_ID	BraTS21_ID
BraTS2021_00000	BraTS2021_00084	BraTS2021_00525	BraTS2021_00638
BraTS2021_00002	BraTS2021_00085	BraTS2021_00526	BraTS2021_00639
BraTS2021_00003	BraTS2021_00087	BraTS2021_00528	BraTS2021_00640
BraTS2021_00005	BraTS2021_00088	BraTS2021_00529	BraTS2021_00641
BraTS2021_00006	BraTS2021_00089	BraTS2021_00530	BraTS2021_00642
BraTS2021_00008	BraTS2021_00090	BraTS2021_00532	BraTS2021_00645
BraTS2021_00009	BraTS2021_00094	BraTS2021_00533	BraTS2021_00646
BraTS2021_00011	BraTS2021_00095	BraTS2021_00537	BraTS2021_00649
BraTS2021_00012	BraTS2021_00096	BraTS2021_00538	BraTS2021_00650
BraTS2021_00014	BraTS2021_00097	BraTS2021_00539	BraTS2021_00651
BraTS2021_00016	BraTS2021_00098	BraTS2021_00540	BraTS2021_00652
BraTS2021_00017	BraTS2021_00099	BraTS2021_00542	BraTS2021_00654
BraTS2021_00018	BraTS2021_00103	BraTS2021_00543	BraTS2021_00655
BraTS2021_00019	BraTS2021_00464	BraTS2021_00544	BraTS2021_00656
BraTS2021_00020	BraTS2021_00466	BraTS2021_00545	BraTS2021_00657
BraTS2021_00021	BraTS2021_00468	BraTS2021_00547	BraTS2021_00658
BraTS2021_00022	BraTS2021_00469	BraTS2021_00548	BraTS2021_00659
BraTS2021_00024	BraTS2021_00470	BraTS2021_00549	BraTS2021_00661
BraTS2021_00025	BraTS2021_00472	BraTS2021_00550	BraTS2021_00663
BraTS2021_00026	BraTS2021_00477	BraTS2021_00551	BraTS2021_00667
BraTS2021_00028	BraTS2021_00478	BraTS2021_00552	BraTS2021_00668
BraTS2021_00030	BraTS2021_00479	BraTS2021_00554	BraTS2021_00674
BraTS2021_00031	BraTS2021_00480	BraTS2021_00555	BraTS2021_00675
BraTS2021_00032	BraTS2021_00481	BraTS2021_00556	BraTS2021_00676
BraTS2021_00033	BraTS2021_00483	BraTS2021_00557	BraTS2021_00677
BraTS2021_00035	BraTS2021_00485	BraTS2021_00558	BraTS2021_00679
BraTS2021_00036	BraTS2021_00488	BraTS2021_00559	BraTS2021_00680
BraTS2021_00043	BraTS2021_00491	BraTS2021_00601	BraTS2021_00682
BraTS2021_00044	BraTS2021_00493	BraTS2021_00602	BraTS2021_00683
BraTS2021_00045	BraTS2021_00494	BraTS2021_00604	BraTS2021_00684
BraTS2021_00046	BraTS2021_00495	BraTS2021_00605	BraTS2021_00685
BraTS2021_00048	BraTS2021_00496	BraTS2021_00606	BraTS2021_00686
BraTS2021_00049	BraTS2021_00498	BraTS2021_00607	BraTS2021_00687
BraTS2021_00052	BraTS2021_00499	BraTS2021_00608	BraTS2021_00688
BraTS2021_00053	BraTS2021_00500	BraTS2021_00610	BraTS2021_00689
BraTS2021_00054	BraTS2021_00501	BraTS2021_00611	BraTS2021_00690
BraTS2021_00056	BraTS2021_00502	BraTS2021_00612	BraTS2021_00691
BraTS2021_00058	BraTS2021_00504	BraTS2021_00613	BraTS2021_00692

Continued on next page

Table S1 – *Continued from previous page*

BraTS21_ID	BraTS21_ID	BraTS21_ID	BraTS21_ID
BraTS2021_00059	BraTS2021_00505	BraTS2021_00615	BraTS2021_00693
BraTS2021_00060	BraTS2021_00506	BraTS2021_00616	BraTS2021_00694
BraTS2021_00061	BraTS2021_00507	BraTS2021_00618	BraTS2021_00697
BraTS2021_00062	BraTS2021_00510	BraTS2021_00619	BraTS2021_00698
BraTS2021_00063	BraTS2021_00511	BraTS2021_00620	BraTS2021_00703
BraTS2021_00064	BraTS2021_00512	BraTS2021_00621	BraTS2021_00704
BraTS2021_00066	BraTS2021_00513	BraTS2021_00622	BraTS2021_00705
BraTS2021_00068	BraTS2021_00514	BraTS2021_00623	BraTS2021_00706
BraTS2021_00070	BraTS2021_00516	BraTS2021_00624	BraTS2021_00707
BraTS2021_00071	BraTS2021_00517	BraTS2021_00625	BraTS2021_00708
BraTS2021_00072	BraTS2021_00518	BraTS2021_00626	BraTS2021_00709
BraTS2021_00074	BraTS2021_00519	BraTS2021_00628	BraTS2021_00714
BraTS2021_00077	BraTS2021_00520	BraTS2021_00630	BraTS2021_00715
BraTS2021_00078	BraTS2021_00523	BraTS2021_00631	BraTS2021_00716
BraTS2021_00081	BraTS2021_00524	BraTS2021_00636	BraTS2021_00718
BraTS2021_00723	BraTS2021_00724	BraTS2021_00725	BraTS2021_00727
BraTS2021_00728	BraTS2021_00729	BraTS2021_00730	BraTS2021_00731
BraTS2021_00732	BraTS2021_00733	BraTS2021_00734	BraTS2021_00735
BraTS2021_00736	BraTS2021_00737	BraTS2021_00739	BraTS2021_00740
BraTS2021_00742	BraTS2021_00744	BraTS2021_00746	BraTS2021_00747
BraTS2021_00750	BraTS2021_00751	BraTS2021_00753	BraTS2021_00756
BraTS2021_00757	BraTS2021_00758	BraTS2021_00759	BraTS2021_00760
BraTS2021_00764	BraTS2021_00765	BraTS2021_00767	BraTS2021_00768
BraTS2021_00772	BraTS2021_00773	BraTS2021_00774	BraTS2021_00775
BraTS2021_00777	BraTS2021_00778	BraTS2021_01027	BraTS2021_01028
BraTS2021_01035	BraTS2021_01153	BraTS2021_01154	BraTS2021_01155
BraTS2021_01156	BraTS2021_01157	BraTS2021_01158	BraTS2021_01159
BraTS2021_01160	BraTS2021_01161		

Table S2: All identified manuscripts with title, journal, 1st author, model type (GAN, UNET, Transformer), reason for exclusion (if applicable) and link

Title	Journal	1st author	Year	Exclusion	Link
Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study	Lancet	Preetha	2021	No response from authors + no code	Link
Contrast-free MRI Contrast Enhancement with Deep Attention Generative Adversarial Network	ISMRM abstract 2019	Liu	2019	Not peer reviewed	Link
Deep learning of MRI contrast enhancement for mapping cerebral blood volume from single-modal non-contrast scans of aging and Alzheimer's disease brains	Frontiers in aging neuro-science	Liu	2022	Not glioma	Link
Multimodal MRI Synthesis Using Unified Generative Adversarial Networks	Medical physics	Dai	2021	No response from authors + no code	Link
Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium? A Feasibility Study	Investigative radiology	Kleesiek	2019	Too many sequences	Link
Missing MRI Pulse Sequence Synthesis Using Multi-Modal Generative Adversarial Network	IEEE transactions on medical imaging	Sharma	2019	Included	Link
Deep learning-based 3D MRI contrast-enhanced synthesis from a 2D noncontrast T2Flair sequence	Medical physics	Wang	2022	CE FLAIR	Link
Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision	Medical physics	Xie	2022	No response from authors + no code	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
Contrast-enhanced MRI Synthesis from Non-contrast MRI using Attention CycleGAN	Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging	Wang	2021	No response from authors + no code	Link
Synthesizing MR Image Contrast Enhancement Using 3D Highresolution ConvNets	IEEE transactions on biomedical engineering	Chen	2023	Included	Link
MIL normalization —— prerequisites for accurate MRI radiomics analysis	Computers in Biology and Medicine	Hu	2021	No response from authors + no code	Link
Feasibility of Simulated Postcontrast MRI of Glioblastomas and Lower-Grade Gliomas by Using Three-dimensional Fully Convolutional Neural Networks	Radiology AI	Calabrese	2021	Too many sequences	Link
Can Deep Learning Replace Gadolinium in Neuro-Oncology? A Reader Study	Investigative radiology	Ammari	2022	Low dose input	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain MRI	Magnetic Resonance in Medicine	Pasumart	2021	Low dose input	Link
Multi-modal Brain Tumor Segmentation via Missing Modality Synthesis and Modality-Level Attention Fusion	arXiv	Huang	2022	No response from authors + no code	Link
TISS-net: Brain tumor image synthesis and segmentation using cascaded dual-task networks and error-prediction consistency	Neurocomputir	Wu	2023	Focus on whole tumor segmentation	Link
Deep Learning for Post-Contrast T1-Weighted Brain MRI Synthesis	Master thesis	Oosterho	2023	Not peer reviewed	Link
Bi-directional Synthesis of Pre- and Post-contrast MRI via Guided Feature Disentanglement	Simulation and synthesis of medical imaging	Xue	2022	Not glioma	Link
CONDITIONAL GENERATIVE MODELS FOR CONTRAST-ENHANCED SYNTHESIS OF T1W AND T1 MAPS IN BRAIN MRI	2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)	Piening	2024	Included	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
Contrast-enhanced MRI synthesis using dense-dilated residual convolutions based 3D network toward elimination of gadolinium in neuro-oncology	Journal of applied clinical medical physics	Osman	2023	Included	Link
Synthesizing Contrast-Enhanced MR Images from Noncontrast MR Images Using Deep Learning	American journal of neuroradiology	Murugesa	2024	No response from authors + no code	Link
One Model to Synthesize Them All: Multi-contrast Multi-scale Transformer for Missing Data Imputation	IEEE transactions on medical imaging	Liu	2023	No response from authors + no code	Link
Medical Image Synthesis with Deep Convolutional Adversarial Networks	IEEE Transactions on Biomedical Engineering	Nie	2018	No response from authors + no code	Link
Brain tumour segmentation with incomplete imaging data	brain communications	Ruffle	2023	Included	Link
Handling missing MRI sequences in deep learning segmentation of brain metastases: a multicenter study	NPJ digital medicine	Grovik	2021	Not glioma	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
PIMMS: Permutation Invariant Multi-Modal Segmentation	Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support	Varsavsky	2018	Not glioma	Link
RS-Net: Regression-Segmentation 3D CNN for Synthesis of Full Resolution Missing Brain MRI in the Presence of Tumours	Simulation and Synthesis in Medical Imaging	Mehta	2018	Included	Link
Brain Tumor Segmentation on MRI with Missing Modalities	Arxiv	Shen	2019	No response from authors + no code	Link
HAD-Net: A Hierarchical Adversarial Knowledge Distillation Network for Improved Enhanced Tumour Segmentation Without Post-Contrast Images	MIDL	Vadacchi	2021	Included	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
RFNet: Region-aware Fusion Network for Incomplete Multi-modal Brain Tumor Segmentation	2021 IEEE/CVF International Conference on Computer Vision (ICCV)	Ding	2021	Included	Link
HeMIS: Hetero-Modal Image Segmentation	Medical Image Computing and Computer Assisted Intervention – MICCAI 2016	Havaei	2016	Not cleared to be published	Link
Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Gated Fusion	Medical Image Computing and Computer Assisted Intervention – MICCAI 2019	Chen	2020	Error could not be fixed	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
Brain Tumor Segmentation on MRI with Missing Modalities	Information processing in medical imaging	Shen	2019	No response from authors + no code	Link
SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities	Proceedings of Machine Learning Research	Azad	2022	Included	Link
Latent Correlation Representation Learning for Brain Tumor Segmentation with Missing MRI Modalities	IEEE transactions on image processing	Zhou	2021	No response from authors + no code	Link
Brain tumor segmentation with missing MRI modalities using edge aware discriminative feature fusion based transformer U-net	Applied Soft Computing	Jagadees	2024	No response from authors + no code	Link
Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images	Medical Image Analysis		2020	??	Link
Brain Tumor Segmentation for Multi-Modal MRI with Missing Information	Journal of Digital Imaging	Feng	2023	Included	Link

Continued on next page

Table S2 – *Continued from previous page*

Title	Journal	1st author	Year	Exclusion	Link
Hetero-Modal Variational Encoder-Decoder for Joint Modality Completion and Segmentation	Medical Image Computing and Computer Assisted Intervention – MICCAI 2019	Dorent	2019	Error could not be fixed	Link
A UNIFIED REPRESENTATION NETWORK FOR SEGMENTATION WITH MISSING MODALITIES		Lau	2019	No response from authors + no code	Link
Multimodal Learning with Incomplete Modalities by Knowledge Distillation	KDD 2020 conference paper	Wang	2020	No glioma	Link
CROSS-MODAL INFORMATION MAXIMIZATION FOR MEDICAL IMAGING: CMIM	2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	Tristan	2021	No response from authors + no code	Link
Missing Data Imputation via Conditional Generator and Correlation Learning for Multimodal Brain Tumor Segmentation	Pattern recognition letters	Zhou	2021	No response from authors + no code	Link

Table S3: Pairwise McNemar test results for enhancement and tumor core identification across all datasets and model combinations. Comparisons with B on the EGD dataset were not possible since this algorithm requires ADC input. Accuracy difference is defined by ACC_{Model1} - ACC_{Model2}

Model1	Model2	p-value enhancing	p-value core	Accuracy difference enhancing	Accuracy difference core
			EGD		
A	D	< 0.0009	< 0.0009	0.26	0.57
A	C	< 0.0009	< 0.0009	-0.06	0.09
A	B	1.00	1.00	N/A	N/A
A	E	< 0.0009	0.01	-0.10	0.02
A	G	< 0.0009	1.00	-0.11	0.00
A	F	< 0.0009	< 0.0009	-0.12	-0.02
A	H	< 0.0009	1.00	-0.11	0.00
A	I	< 0.0009	0.82	-0.11	0.00
A	J	< 0.0009	0.21	-0.11	-0.01
A	Z	< 0.0009	< 0.0009	-0.10	0.04
D	C	< 0.0009	< 0.0009	-0.32	-0.48
D	B	1.00	1.00	N/A	N/A
D	E	< 0.0009	< 0.0009	-0.36	-0.55
D	G	< 0.0009	< 0.0009	-0.37	-0.57
D	F	< 0.0009	< 0.0009	-0.38	-0.58
D	H	< 0.0009	< 0.0009	-0.37	-0.57
D	I	< 0.0009	< 0.0009	-0.37	-0.57
D	J	< 0.0009	< 0.0009	-0.37	-0.57
D	Z	< 0.0009	< 0.0009	-0.36	-0.52
C	B	1.00	1.00	N/A	N/A
C	E	< 0.0009	< 0.0009	-0.04	-0.07
C	G	< 0.0009	< 0.0009	-0.05	-0.09
C	F	< 0.0009	< 0.0009	-0.06	-0.11
C	H	< 0.0009	< 0.0009	-0.05	-0.09
C	I	< 0.0009	< 0.0009	-0.05	-0.09
C	J	< 0.0009	< 0.0009	-0.05	-0.10
C	Z	< 0.0009	< 0.0009	-0.05	-0.05
B	E	1.00	1.00	N/A	N/A
B	G	1.00	1.00	N/A	N/A
B	F	1.00	1.00	N/A	N/A
B	H	1.00	1.00	N/A	N/A
B	I	1.00	1.00	N/A	N/A
B	J	1.00	1.00	N/A	N/A
B	Z	1.00	1.00	N/A	N/A
E	G	0.34	0.01	-0.01	-0.02
E	F	0.08	< 0.0009	-0.02	-0.04

Continued on next page

Table S3 – *Continued from previous page*

Model1	Model2	p-value enhancing	p-value core	Accuracy difference enhancing	Accuracy difference core
E	H	0.46	< 0.0009	-0.01	-0.02
E	I	0.64	< 0.0009	-0.01	-0.02
E	J	0.35	< 0.0009	-0.01	-0.03
E	Z	0.88	0.01	0.00	0.02
G	F	0.21	< 0.0009	-0.01	-0.02
G	H	0.84	1.00	0.00	0.00
G	I	0.83	0.65	0.00	0.00
G	J	1.00	0.12	0.00	-0.01
G	Z	0.43	< 0.0009	0.01	0.04
F	H	0.12	< 0.0009	0.01	0.02
F	I	0.39	< 0.0009	0.01	0.01
F	J	0.11	0.03	0.01	0.01
F	Z	0.12	< 0.0009	0.02	0.06
H	I	1.00	0.65	0.00	0.00
H	J	0.84	0.14	0.00	-0.01
H	Z	0.63	< 0.0009	0.01	0.04
I	J	0.83	0.39	0.00	-0.01
I	Z	0.81	< 0.0009	0.00	0.05
J	Z	0.49	< 0.0009	0.01	0.05
			IMAGO		
A	D	< 0.0009	< 0.0009	0.16	0.16
A	C	0.08	0.08	-0.04	-0.04
A	B	0.01	0.01	-0.05	-0.07
A	E	< 0.0009	< 0.0009	-0.09	-0.07
A	G	0.02	< 0.0009	-0.06	-0.07
A	F	0.03	0.02	-0.05	-0.05
A	H	< 0.0009	< 0.0009	-0.07	-0.07
A	I	< 0.0009	< 0.0009	-0.10	-0.07
A	J	< 0.0009	< 0.0009	-0.08	-0.08
A	Z	< 0.0009	< 0.0009	-0.10	-0.10
D	C	< 0.0009	< 0.0009	-0.20	-0.20
D	B	< 0.0009	< 0.0009	-0.21	-0.22
D	E	< 0.0009	< 0.0009	-0.24	-0.23
D	G	< 0.0009	< 0.0009	-0.21	-0.22
D	F	< 0.0009	< 0.0009	-0.21	-0.21
D	H	< 0.0009	< 0.0009	-0.23	-0.23
D	I	< 0.0009	< 0.0009	-0.25	-0.23
D	J	< 0.0009	< 0.0009	-0.23	-0.23
D	Z	< 0.0009	< 0.0009	-0.26	-0.25

Continued on next page

Table S3 – *Continued from previous page*

Model1	Model2	p-value enhancing	p-value core	Accuracy difference enhancing	Accuracy difference core
C	B	0.46	0.03	-0.02	-0.03
C	E	0.01	0.03	-0.04	-0.03
C	G	0.39	0.07	-0.01	-0.02
C	F	0.53	0.43	-0.01	-0.01
C	H	0.04	0.01	-0.03	-0.03
C	I	< 0.0009	0.02	-0.05	-0.03
C	J	< 0.0009	< 0.0009	-0.04	-0.04
C	Z	< 0.0009	< 0.0009	-0.06	-0.06
B	E	0.11	0.84	-0.03	0.00
B	G	1.00	1.00	0.00	0.00
B	F	0.80	0.02	0.01	0.02
B	H	0.37	0.77	-0.02	0.00
B	I	0.03	0.45	-0.04	-0.01
B	J	0.16	0.23	-0.02	-0.01
B	Z	0.01	0.01	-0.04	-0.03
E	G	0.09	0.69	0.03	0.01
E	F	0.02	0.05	0.03	0.02
E	H	0.35	1.00	0.01	0.00
E	I	0.53	1.00	-0.01	0.00
E	J	0.78	0.72	0.01	-0.01
E	Z	0.30	0.02	-0.01	-0.03
G	F	0.73	0.07	0.01	0.01
G	H	0.28	0.63	-0.01	-0.01
G	I	0.04	0.63	-0.04	-0.01
G	J	0.10	0.31	-0.02	-0.01
G	Z	0.01	0.01	-0.04	-0.03
F	H	0.05	0.01	-0.02	-0.02
F	I	0.01	0.01	-0.04	-0.02
F	J	< 0.0009	< 0.0009	-0.03	-0.02
F	Z	< 0.0009	< 0.0009	-0.05	-0.04
H	I	0.16	1.00	-0.02	0.00
H	J	0.58	0.65	-0.01	-0.01
H	Z	0.02	0.03	-0.03	-0.03
I	J	0.34	0.61	0.02	-0.01
I	Z	0.90	0.03	0.00	-0.03
J	Z	0.14	0.09	-0.02	-0.02

Table S4: Groups-wise Mann-Whitney U tests between the synthesis and segmentation models for enhancement and tumor core identification.

Dataset	Label	Metric	Synthesis mean	Synthesis std	Segmentation mean	Segmentation std	Mann-Whitney p
IMAGO	enh	sensitivity	0.79	0.20	0.97	0.02	0.06
IMAGO	enh	specificity	0.47	0.35	0.21	0.15	0.41
IMAGO	enh	ppv	0.87	0.06	0.83	0.02	0.41
IMAGO	enh	npv	0.38	0.07	0.56	0.22	0.18
IMAGO	enh	accuracy	0.73	0.10	0.82	0.02	0.06
IMAGO	core	sensitivity	0.81	0.22	0.99	0.01	0.27
IMAGO	core	specificity	0.41	0.41	0.10	0.09	0.39
IMAGO	core	ppv	0.87	0.07	0.82	0.01	0.39
IMAGO	core	npv	0.46	0.23	0.66	0.25	0.39
IMAGO	core	accuracy	0.73	0.10	0.81	0.01	0.15
EGD	enh	sensitivity	0.72	0.24	0.97	0.03	0.06
EGD	enh	specificity	0.53	0.34	0.14	0.15	0.12
EGD	enh	ppv	0.93	0.04	0.89	0.01	0.12
EGD	enh	npv	0.22	0.04	0.44	0.26	0.06
EGD	enh	accuracy	0.69	0.17	0.87	0.01	0.06
EGD	core	sensitivity	0.76	0.31	0.98	0.02	0.24
EGD	core	specificity	0.33	0.58	0.00	0.00	0.24
EGD	core	ppv	1.00	0.00	1.00	0.00	1.00
EGD	core	npv	0.00	0.00	0.00	0.00	0.24
EGD	core	accuracy	0.76	0.30	0.98	0.02	0.24

Table S5: Tumor core detection performance per algorithm, presented as mean (95% CI). Bold indicates lowest and highest values.

Model	Dataset	Sensitivity	Specificity	PPV	NPV	DSC	ICC
A	IMAGO	0.78 (0.74–0.82)	0.58 (0.48–0.67)	0.89 (0.85–0.92)	0.39 (0.31–0.47)	0.49 (0.13–0.85)	0.72 (0.67–0.76)
A	EGD	0.99 (0.98–0.99)	0.43 (0.00–0.83)	0.99 (0.99–1.00)	0.25 (0.00–0.50)	0.53 (0.26–0.80)	0.61 (0.56–0.66)
D	IMAGO	0.50 (0.46–0.55)	0.92 (0.86–0.96)	0.96 (0.93–0.98)	0.31 (0.26–0.36)	0.42 (0.01–0.83)	0.53 (0.46–0.58)
D	EGD	**0.42** (0.38–0.45)	**1.00** (1.00–1.00)	**1.00** (1.00–1.00)	0.02 (0.00–0.03)	**0.23** (0.00–0.54)	0.46 (0.38–0.54)
C	IMAGO	0.95 (0.93–0.97)	0.10 (0.05–0.17)	0.81 (0.78–0.85)	0.33 (0.17–0.50)	0.38 (0.06–0.70)	0.60 (0.55–0.66)
C	EGD	0.89 (0.87–0.92)	0.29 (0.00–0.67)	0.99 (0.99–1.00)	0.02 (0.00–0.06)	0.40 (0.11–0.69)	0.56 (0.51–0.61)
B	IMAGO	**1.00** (0.99–1.00)	0.04 (0.01–0.08)	0.81 (0.77–0.84)	0.80 (0.25–1.00)	0.44 (0.12–0.76)	0.49 (0.32–0.62)
E	IMAGO	0.98 (0.97–1.00)	0.11 (0.06–0.18)	0.82 (0.79–0.85)	0.64 (0.40–0.85)	0.56 (0.20–0.92)	**0.78** (0.75–0.82)
E	EGD	0.97 (0.95–0.98)	0.43 (0.00–0.83)	0.99 (0.99–1.00)	0.10 (0.00–0.22)	0.63 (0.33–0.93)	0.74 (0.71–0.77)
G	IMAGO	0.99 (0.98–1.00)	0.06 (0.02–0.12)	0.81 (0.78–0.85)	0.58 (0.29–0.87)	0.49 (0.14–0.84)	0.72 (0.68–0.76)
G	EGD	0.98 (0.97–0.99)	**0.00** (0.00–0.00)	0.99 (0.98–1.00)	**0.00** (0.00–0.00)	0.57 (0.27–0.87)	0.73 (0.69–0.76)
F	IMAGO	0.99 (0.97–1.00)	0.01 (0.00–0.03)	**0.80** (0.77–0.84)	0.12 (0.00–0.50)	0.34 (0.01–0.67)	0.57 (0.51–0.63)

Continued on next page

Table S5 – Continued from previous page

Model	Dataset	Sensitivity	Specificity	PPV	NPV	DSC	ICC
F	EGD	**1.00** (1.00–1.00)	0.00 (0.00–0.00)	0.99 (0.98–1.00)	0.00 (0.00–0.00)	0.47 (0.16–0.78)	0.60 (0.54–0.65)
H	IMAGO	**1.00** (0.99–1.00)	0.06 (0.02–0.10)	0.81 (0.78–0.84)	**0.87** (0.50–1.00)	0.49 (0.13–0.85)	0.00 (-0.07–0.08)
H	EGD	0.99 (0.98–0.99)	0.27 (0.00–0.67)	0.99 (0.99–1.00)	0.14 (0.00–0.38)	0.58 (0.26–0.90)	0.01 (-0.06–0.08)
I	IMAGO	**1.00** (0.99–1.00)	0.06 (0.02–0.11)	0.81 (0.78–0.84)	0.87 (0.50–1.00)	0.50 (0.14–0.86)	0.72 (0.68–0.76)
I	EGD	0.99 (0.98–0.99)	**0.00** (0.00–0.00)	0.99 (0.98–1.00)	0.00 (0.00–0.00)	0.57 (0.24–0.90)	0.72 (0.68–0.75)
J	IMAGO	0.99 (0.98–1.00)	0.10 (0.05–0.16)	0.82 (0.79–0.85)	0.80 (0.56–1.00)	0.40 (0.09–0.71)	0.48 (0.27–0.62)
J	EGD	0.99 (0.99–1.00)	**0.00** (0.00–0.00)	0.99 (0.98–1.00)	**0.00** (0.00–0.00)	0.43 (0.16–0.70)	0.35 (0.11–0.52)
Z	IMAGO	0.97 (0.96–0.99)	0.28 (0.20–0.37)	0.85 (0.81–0.88)	0.73 (0.59–0.86)	0.54 (0.18–0.90)	0.71 (0.66–0.75)
Z	EGD	0.95 (0.93–0.96)	0.43 (0.00–0.83)	0.99 (0.99–1.00)	0.06 (0.00–0.15)	0.58 (0.24–0.92)	0.70 (0.64–0.75)

Table S6: Dice (DSC), hausdorff distance (HD95), mean average volume error (mAVE), and intraclass correlation coefficient (ICC) for the enhancing region and tumor core for each model and database.

Output	Model	Database	DSC	HD95 (mm)	mAVE (mL)	ICC
			Enhancing			
Synthesis	A	IMAGO	0.18	29.60	-14.11	0.153
Synthesis	A	EGD	0.17	31.77	-17.73	0.154
Synthesis	D	IMAGO	0.21	28.00	-15.21	0.132
Synthesis	D	EGD	0.18	30.39	-20.21	0.041
Synthesis	C	IMAGO	0.06	60.12	-13.53	0.118
Synthesis	C	EGD	0.07	59.22	-18.96	0.074
Synthesis	B	IMAGO	0.13	45.29	-9.35	0.354
Segmentation	E	IMAGO	0.43	19.70	5.16	0.810
Segmentation	E	EGD	0.50	17.97	5.30	0.862
Segmentation	G	IMAGO	0.32	37.02	4.56	0.731
Segmentation	G	EGD	0.37	41.76	5.38	0.782
Segmentation	F	IMAGO	0.22	56.02	5.32	0.630
Segmentation	F	EGD	0.33	60.33	9.34	0.714
Segmentation	H	IMAGO	0.35	32.21	257.13	0.001
Segmentation	H	EGD	0.42	29.09	218.71	0.002
Segmentation	I	IMAGO	0.44	19.37	6.73	0.753
Segmentation	I	EGD	0.50	18.83	6.40	0.808
Segmentation	J	IMAGO	0.24	44.75	25.38	0.334
Segmentation	J	EGD	0.23	57.37	38.66	0.234
Segmentation	Z	IMAGO	0.42	19.88	3.98	0.784
Segmentation	Z	EGD	0.47	20.10	3.92	0.864
			Tumor core			
Synthesis	A	IMAGO	0.49	25.42	5.38	0.719
Synthesis	A	EGD	0.53	36.35	7.72	0.614
Synthesis	D	IMAGO	0.42	22.18	-5.46	0.526
Synthesis	D	EGD	0.23	33.69	-14.42	0.463
Synthesis	C	IMAGO	0.38	51.94	3.63	0.604
Synthesis	C	EGD	0.40	56.98	-4.00	0.562
Synthesis	B	IMAGO	0.44	41.86	22.57	0.491
Segmentation	E	IMAGO	0.56	22.15	3.17	0.785
Segmentation	E	EGD	0.63	22.20	-3.86	0.743
Segmentation	G	IMAGO	0.49	38.94	1.99	0.724
Segmentation	G	EGD	0.57	42.29	-2.14	0.726
Segmentation	F	IMAGO	0.34	53.03	-6.31	0.572
Segmentation	F	EGD	0.47	63.36	-7.04	0.601
Segmentation	H	IMAGO	0.49	33.10	262.38	0.005
Segmentation	H	EGD	0.58	30.79	212.18	0.011
Segmentation	I	IMAGO	0.50	24.14	4.47	0.721
Segmentation	I	EGD	0.57	24.84	-4.54	0.716
Segmentation	J	IMAGO	0.40	44.14	24.32	0.479
Segmentation	J	EGD	0.43	60.32	38.59	0.350
Segmentation	Z	IMAGO	0.54	22.07	-2.82	0.709
Segmentation	Z	EGD	0.58	23.87	-8.35	0.704

Table S7: Sub analysis of glioblastoma patients. Enhancement and core detection performance per algorithm, presented as mean (95% CI). Bold indicates lowest and highest values. *Abbreviations:* PPV = Positive Predictive Value, NPV = Negative Predictive Value, DSC = Dice Score Coefficient, ICC = Intraclass Correlation Coefficient, mAVE = mean average volume error, HD95 = hausdorff distance

Model	Sensitivi	Specificity	PPV	NPV	DSC	ICC	mAVE	HD95
				Enhancement				
A	0.77 (0.73 - 0.82)	0.50 (0.33 - 0.67)	0.93 (0.90 - 0.96)	0.19 (0.12 - 0.28)	0.13 (0 - 0.37)	0.11 (-0.04 - 0.26)	-19.77	32.32
A	0.84 (0.80 - 0.88)	0.57 (0.35 - 0.79)	0.96 (0.94 - 0.98)	0.20 (0.11 - 0.31)	0.14 (0 - 0.36)	0.13 (-0.05 - 0.31)	-23.02	34.65
D	**0.51 (0.45 - 0.56)**	0.82 (0.68 - 0.93)	0.96 (0.93 - 0.99)	0.16 (0.11 - 0.21)	0.12 (0 - 0.41)	0.12 (-0.04 - 0.27)	-21.29	34.82
D	**0.47 (0.41 - 0.53)**	0.86 (0.69 - 1.00)	0.98 (0.95 - 1.00)	0.10 (0.06 - 0.15)	0.09 (0 - 0.33)	0.03 (-0.03 - 0.1)	-26.67	36.53
C	0.95 (0.92 - 0.97)	0.07 (0.00 - 0.18)	0.90 (0.87 - 0.93)	0.14 (0.00 - 0.33)	**0.05 (0 - 0.18)**	0.11 (-0.04 - 0.26)	-19.75	**59.44**
C	0.90 (0.87 - 0.94)	0.19 (0.04 - 0.38)	0.94 (0.91 - 0.97)	0.12 (0.03 - 0.25)	**0.05 (0 - 0.19)**	0.06 (-0.04 - 0.16)	-25.23	58.53
B	0.93 (0.90 - 0.96)	0.37 (0.22 - 0.53)	0.93 (0.90 - 0.95)	0.38 (0.22 - 0.54)	0.13 (0 - 0.35)	0.29 (0.05 - 0.47)	-14.0	45.02

Continued on next page

Table S7 – *Continued from previous page*

Model	Sensitivi	Specificity	PPV	NPV	DSC	ICC	mAVE	HD95
E	0.98 (0.97 - 0.99)	0.24 (0.11 - 0.38)	0.92 (0.89 - 0.95)	0.60 (0.33 - 0.85)	0.53 (0.26 - 0.8)	0.78 (0.66 - 0.84)	6.57	18.74
E	0.98 (0.97 - 1.00)	0.33 (0.13 - 0.55)	0.95 (0.93 - 0.98)	0.58 (0.29 - 0.87)	0.62 (0.37 - 0.87)	**0.88 (0.82 - 0.91)**	4.55	**15.31**
G	0.96 (0.94 - 0.98)	0.15 (0.05 - 0.28)	0.91 (0.88 - 0.94)	0.30 (0.11 - 0.52)	0.41 (0.16 - 0.66)	0.72 (0.66 - 0.77)	3.43	34.13
G	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.93 (0.90 - 0.96)	**0.00 (0.00 - 0.00)**	0.49 (0.24 - 0.74)	0.84 (0.8 - 0.87)	**1.83**	39.22
F	0.99 (0.97 - 1.00)	0.02 (0.00 - 0.09)	0.90 (0.87 - 0.93)	0.14 (0.00 - 0.50)	0.3 (0.06 - 0.54)	0.59 (0.51 - 0.66)	4.82	49.53
F	**1.00 (1.00 - 1.00)**	**0.00 (0.00 - 0.00)**	0.93 (0.90 - 0.96)	**0.00 (0.00 - 0.00)**	0.43 (0.17 - 0.69)	0.72 (0.59 - 0.8)	7.65	55.21
H	0.98 (0.97 - 0.99)	0.18 (0.07 - 0.31)	0.91 (0.88 - 0.94)	0.54 (0.25 - 0.82)	0.45 (0.16 - 0.74)	0.0 (-0.09 - 0.1)	**273.53**	31.09
H	0.99 (0.97 - 1.00)	0.19 (0.04 - 0.37)	0.94 (0.92 - 0.97)	0.50 (0.12 - 0.86)	0.54 (0.25 - 0.83)	-0.0 (-0.1 - 0.1)	227.23	26.8

Continued on next page

Table S7 – *Continued from previous page*

Model	Sensitivi	Specificity	PPV	NPV	DSC	ICC	mAVE	HD95
I	0.98 (0.97 - 0.99)	0.31 (0.17 - 0.47)	0.93 (0.90 - 0.95)	0.67 (0.43 - 0.88)	0.52 (0.26 - 0.78)	0.69 (0.52 - 0.79)	9.02	18.87
I	0.98 (0.96 - 0.99)	0.57 (0.35 - 0.78)	0.97 (0.95 - 0.99)	0.63 (0.40 - 0.85)	0.61 (0.36 - 0.86)	0.84 (0.77 - 0.88)	5.08	16.85
J	0.99 (0.98 - 1.00)	0.18 (0.07 - 0.32)	0.91 (0.88 - 0.94)	**0.79 (0.45 - 1.00)**	0.32 (0.09 - 0.55)	0.38 (0.03 - 0.6)	24.36	40.51
J	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.93 (0.90 - 0.96)	**0.00 (0.00 - 0.00)**	0.3 (0.09 - 0.51)	0.24 (-0.03 - 0.46)	36.45	56.32
Z	0.98 (0.96 - 0.99)	0.39 (0.24 - 0.56)	0.93 (0.91 - 0.96)	0.68 (0.47 - 0.87)	0.51 (0.25 - 0.77)	0.73 (0.66 - 0.79)	6.07	18.56
Z	0.97 (0.94 - 0.99)	0.33 (0.13 - 0.55)	0.95 (0.93 - 0.98)	0.41 (0.18 - 0.65)	0.59 (0.33 - 0.85)	**0.88 (0.85 - 0.9)**	2.33	17.25
				Tumor Core				
A	0.77 (0.73 - 0.82)	0.50 (0.33 - 0.67)	0.93 (0.90 - 0.96)	0.19 (0.12 - 0.28)	0.46 (0.13 - 0.79)	0.7 (0.64 - 0.75)	5.6	29.04
A	0.99 (0.97 - 1.00)	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.52 (0.27 - 0.77)	0.62 (0.5 - 0.7)	12.35	38.42

Continued on next page

Table S7 – *Continued from previous page*

Model	Sensitivi	Specificity	PPV	NPV	DSC	ICC	mAVE	HD95
D	0.95 (0.92 - 0.97)	0.07 (0.00 - 0.18)	0.90 (0.87 - 0.93)	0.14 (0.00 - 0.33)	0.35 (0 - 0.72)	0.47 (0.39 - 0.54)	-3.87	28.95
D	0.90 (0.87 - 0.94)	**1.00 (0.00 - 1.00)**	**1.00 (1.00 - 1.00)**	0.06 (0.00 - 0.16)	**0.25 (0 - 0.57)**	0.53 (0.44 - 0.61)	-9.14	34.39
C	**1.00 (0.99 - 1.00)**	0.07 (0.00 - 0.18)	0.90 (0.87 - 0.93)	0.75 (0.00 - 1.00)	0.4 (0.11 - 0.69)	0.57 (0.5 - 0.64)	2.29	52.84
C	0.99 (0.97 - 1.00)	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.4 (0.12 - 0.68)	0.59 (0.51 - 0.66)	-1.64	55.11
B	**1.00 (0.99 - 1.00)**	0.10 (0.02 - 0.21)	0.91 (0.88 - 0.94)	0.80 (0.25 - 1.00)	0.48 (0.19 - 0.77)	0.41 (0.24 - 0.54)	25.0	42.92
E	0.99 (0.97 - 1.00)	0.15 (0.05 - 0.29)	0.91 (0.88 - 0.94)	0.55 (0.23 - 0.86)	0.67 (0.37 - 0.97)	0.84 (0.79 - 0.88)	6.33	22.05
E	0.98 (0.96 - 0.99)	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.73 (0.46 - 1.0)	0.87 (0.84 - 0.9)	0.4	**19.57**
G	0.99 (0.98 - 1.00)	0.07 (0.00 - 0.18)	0.90 (0.87 - 0.93)	0.43 (0.00 - 0.83)	0.59 (0.28 - 0.9)	0.8 (0.74 - 0.84)	6.15	38.59

Continued on next page

Table S7 – *Continued from previous page*

Model	Sensitivi	Specificity	PPV	NPV	DSC	ICC	mAVE	HD95
G	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.66 (0.38 - 0.94)	0.84 (0.81 - 0.87)	2.71	42.44
F	0.99 (0.97 - 1.00)	0.02 (0.00 - 0.09)	0.90 (0.87 - 0.93)	0.14 (0.00 - 0.50)	0.44 (0.12 - 0.76)	0.67 (0.61 - 0.73)	-1.67	49.11
F	**1.00 (1.00 - 1.00)**	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.57 (0.27 - 0.87)	0.79 (0.74 - 0.83)	0.86	57.29
H	0.51 (0.45 - 0.56)	0.82 (0.68 - 0.93)	0.96 (0.93 - 0.99)	0.16 (0.11 - 0.21)	0.58 (0.24 - 0.92)	0.0 (-0.09 - 0.1)	**278.56**	34.15
H	**0.45 (0.39 - 0.51)**	**1.00 (0.00 - 1.00)**	**1.00 (1.00 - 1.00)**	0.01 (0.00 - 0.03)	0.65 (0.33 - 0.97)	0.01 (-0.1 - 0.11)	226.63	29.25
I	**1.00 (1.00 - 1.00)**	0.05 (0.00 - 0.13)	0.90 (0.87 - 0.93)	**1.00 (0.00 - 1.00)**	0.63 (0.32 - 0.94)	0.78 (0.68 - 0.85)	9.69	22.91
I	0.99 (0.97 - 1.00)	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.7 (0.42 - 0.98)	**0.88 (0.85 - 0.9)**	**1.16**	21.88
J	0.99 (0.98 - 1.00)	0.15 (0.05 - 0.28)	0.91 (0.88 - 0.94)	0.75 (0.40 - 1.00)	0.47 (0.19 - 0.75)	0.5 (0.16 - 0.7)	29.46	43.19

Continued on next page

Table S7 – *Continued from previous page*

Model	Sensitivi	Specificity	PPV	NPV	DSC	ICC	mAVE	HD95
J	**1.00 (0.99 - 1.00)**	**0.00 (0.00 - 0.00)**	0.99 (0.98 - 1.00)	**0.00 (0.00 - 0.00)**	0.44 (0.18 - 0.7)	0.33 (0.01 - 0.55)	46.25	**58.63**
Z	0.98 (0.96 - 0.99)	0.37 (0.22 - 0.53)	0.93 (0.90 - 0.96)	0.67 (0.45 - 0.86)	0.66 (0.35 - 0.97)	0.82 (0.78 - 0.85)		
Z	0.95 (0.93 - 0.98)	0.50 (0.00 - 1.00)	**1.00 (0.99 - 1.00)**	0.06 (0.00 - 0.22)	0.69 (0.39 - 0.99)	0.87 (0.84 - 0.89)		

Table S8: Subanalysis of glioblastoma patients. Pairwise McNemar test results for enhancement and tumor core identification across all datasets and model combinations. Comparisons with B on the EGD dataset were not possible since this algorithm requires ADC input. Accuracy difference is defined by: $ACC_{Model1} - ACC_{Model2}$

Model1	Model2	p-value enhancing	p-value core	Accuracy difference enhancing	Accuracy difference core
			EGD		
A	D	< 0.0009	< 0.0009	0.32	0.53
A	C	0.53	< 0.0009	-0.04	0.08
A	B	N/A	N/A	N/A	N/A
A	E	< 0.0009	0.82	-0.12	0.01
A	G	< 0.0009	0.95	-0.11	-0.01
A	F	< 0.0009	0.24	-0.11	-0.01
A	H	< 0.0009	1.00	-0.11	0.00
A	I	< 0.0009	1.00	-0.13	0.00
A	J	< 0.0009	0.44	-0.11	-0.01
A	Z	< 0.0009	0.11	-0.10	0.03
D	C	< 0.0009	< 0.0009	-0.36	-0.45
D	B	N/A	N/A	N/A	N/A
D	E	< 0.0009	< 0.0009	-0.44	-0.52
D	G	< 0.0009	< 0.0009	-0.43	-0.53
D	F	< 0.0009	< 0.0009	-0.44	-0.54
D	H	< 0.0009	< 0.0009	-0.44	-0.53
D	I	< 0.0009	< 0.0009	-0.45	-0.53
D	J	< 0.0009	< 0.0009	-0.43	-0.54
D	Z	< 0.0009	< 0.0009	-0.43	-0.50
C	B	N/A	N/A	N/A	N/A
C	E	< 0.0009	< 0.0009	-0.08	-0.07
C	G	< 0.0009	< 0.0009	-0.07	-0.08
C	F	< 0.0009	< 0.0009	-0.08	-0.09
C	H	< 0.0009	< 0.0009	-0.08	-0.08
C	I	< 0.0009	< 0.0009	-0.09	-0.08
C	J	< 0.0009	< 0.0009	-0.07	-0.09
C	Z	0.01	0.08	-0.07	-0.05
B	E	N/A	N/A	N/A	N/A
B	G	N/A	N/A	N/A	N/A
B	F	N/A	N/A	N/A	N/A
B	H	N/A	N/A	N/A	N/A
B	I	N/A	N/A	N/A	N/A
B	J	N/A	N/A	N/A	N/A
B	Z	N/A	N/A	N/A	N/A
E	G	0.69	0.24	0.01	-0.02

Continued on next page

Table S8 – *Continued from previous page*

Model1	Model2	p-value enhancing	p-value core	Accuracy diff. enhancing	Accuracy diff. core
E	F	1.00	0.04	0.01	-0.02
E	H	1.00	0.73	0.01	-0.01
E	I	0.93	0.78	-0.01	-0.01
E	J	0.73	0.15	0.01	-0.02
E	Z	0.63	0.44	0.02	0.02
G	F	0.81	0.78	-0.01	-0.01
G	H	1.00	0.95	-0.01	0.01
G	I	0.40	0.95	-0.02	0.01
G	J	1.00	1.00	0.00	0.00
G	Z	1.00	0.01	0.00	0.04
F	H	1.00	0.24	0.00	0.01
F	I	0.66	0.24	-0.02	0.01
F	J	0.81	1.00	0.01	0.00
F	Z	0.94	< 0.0009	0.01	0.04
H	I	0.59	1.00	-0.02	0.00
H	J	1.00	0.63	0.01	-0.01
H	Z	0.93	0.06	0.01	0.03
I	J	0.40	0.63	0.02	-0.01
I	Z	0.09	0.08	0.03	0.03
J	Z	1.00	0.01	0.00	0.04
			IMAGO		
A	D	< 0.0009	< 0.0009	0.21	0.21
A	C	< 0.0009	< 0.0009	-0.12	-0.12
A	B	< 0.0009	< 0.0009	-0.13	-0.16
A	E	< 0.0009	< 0.0009	-0.16	-0.16
A	G	< 0.0009	< 0.0009	-0.13	-0.15
A	F	< 0.0009	< 0.0009	-0.14	-0.14
A	H	< 0.0009	< 0.0009	-0.16	-0.16
A	I	< 0.0009	< 0.0009	-0.17	-0.16
A	J	< 0.0009	< 0.0009	-0.17	-0.17
A	Z	< 0.0009	< 0.0009	-0.18	-0.17
D	C	< 0.0009	< 0.0009	-0.32	-0.32
D	B	< 0.0009	< 0.0009	-0.33	-0.37
D	E	< 0.0009	< 0.0009	-0.37	-0.36
D	G	< 0.0009	< 0.0009	-0.34	-0.36
D	F	< 0.0009	< 0.0009	-0.35	-0.35
D	H	< 0.0009	< 0.0009	-0.36	-0.36
D	I	< 0.0009	< 0.0009	-0.37	-0.36
D	J	< 0.0009	< 0.0009	-0.37	-0.37

Continued on next page

Table S8 – *Continued from previous page*

Model1	Model2	p-value enhancing	p-value core	Accuracy diff. enhancing	Accuracy diff. core
D	Z	< 0.0009	< 0.0009	-0.38	-0.38
C	B	0.67	< 0.0009	-0.01	-0.05
C	E	0.01	0.02	-0.05	-0.04
C	G	0.54	0.04	-0.02	-0.04
C	F	0.12	0.13	-0.03	-0.03
C	H	0.02	< 0.0009	-0.04	-0.04
C	I	< 0.0009	< 0.0009	-0.05	-0.04
C	J	< 0.0009	< 0.0009	-0.05	-0.05
C	Z	< 0.0009	< 0.0009	-0.06	-0.06
B	E	0.11	0.90	-0.03	0.01
B	G	1.00	0.50	0.00	0.01
B	F	0.59	0.07	-0.01	0.02
B	H	0.18	1.00	-0.03	0.00
B	I	0.04	1.00	-0.04	0.00
B	J	0.04	1.00	-0.04	0.00
B	Z	0.01	0.64	-0.05	-0.01
E	G	0.07	0.90	0.03	0.01
E	F	0.23	0.46	0.02	0.01
E	H	0.85	1.00	0.01	0.00
E	I	0.73	1.00	-0.01	0.00
E	J	0.85	0.79	-0.01	-0.01
E	Z	0.49	0.38	-0.01	-0.02
G	F	0.59	0.62	-0.01	0.01
G	H	0.14	0.65	-0.02	-0.01
G	I	0.04	0.65	-0.04	-0.01
G	J	0.04	0.46	-0.04	-0.01
G	Z	0.02	0.25	-0.04	-0.02
F	H	0.40	0.19	-0.01	-0.02
F	I	0.07	0.06	-0.03	-0.02
F	J	0.04	0.07	-0.02	-0.02
F	Z	0.03	0.05	-0.03	-0.03
H	I	0.47	1.00	-0.01	0.00
H	J	0.56	0.90	-0.01	-0.01
H	Z	0.20	0.49	-0.02	-0.01
I	J	1.00	0.82	0.00	-0.01
I	Z	0.85	0.51	-0.01	-0.01
J	Z	0.71	0.79	-0.01	-0.01

Chapter 9

To Gad or not-to-Gad: Leveraging AI for Selective Use of Gadolinium Contrast in Brain MR

Silvia Ingala*, Marko Bauer*, Ivar J.H.G. Wamelink, Mathias Perslev, Kaining Sheng, Mads Birch Sørensen, Sebastian Nørgaard Llambias, Arpit Kothari, Era Dwivedi, Martin Lillholm, Frederik Barkhof, Michael Bachmann Nielsen, Mads Nielsen, Bas M.M.S. Jasperse, Vera C. Keil, Akshay Pai

As submitted to Communications Medicine

Abstract

Background: Gadolinium-based contrast agents used for brain magnetic resonance imaging improve lesion detection but are associated with acute side effects, concerns about long-term tissue deposition, and economic and environmental costs. Regulatory bodies and radiological societies therefore recommend minimizing their use where possible. This study aims to develop and evaluate a real-time artificial intelligence algorithm, Apollo SmartGAD, that optimizes imaging workflows by recommending contrast-enhanced, reduced-dose contrast-enhanced, or non-enhanced imaging protocols during acquisition.

Methods: This retrospective, multicenter study evaluated 1,251 adult magnetic resonance imaging examinations (mean age 52.1 ± 20.4 years; 577 females). Apollo SmartGAD analyzes images acquired during scanning and recommends non-enhanced, reduced-dose contrast-enhanced, or standard-dose contrast-enhanced imaging protocols. Performance was assessed in two distinct cohorts: patients undergoing first-time brain assessment and patients undergoing glioma monitoring.

Results: Here we show that, in the first-time brain assessment cohort, Apollo SmartGAD identifies examinations not requiring contrast administration with a sensitivity of 73%, a specificity of 79%, and a negative predictive value of 86%. In the glioma monitoring cohort, the algorithm differentiate between non-enhanced, reduced-dose contrast-enhanced, and standard-dose contrast-enhanced protocols with an overall sensitivity of 77% and a specificity of 78%.

Conclusions: This study demonstrates that workflow-integrated, real-time artificial intelligence can support contrast administration decisions during brain magnetic resonance imaging. Such decision support may reduce unnecessary contrast use while preserving diagnostic performance, with potential benefits for patient safety, clinical efficiency, and sustainability.

Introduction

Gadolinium-based contrast agents (GBCA) are used in up to 35% of brain magnetic resonance (MR) imaging examinations[1] to visualize vessels or, most commonly, to exploit the consequences of blood-brain barrier (BBB) disruption thereby detecting certain brain pathologies. The disrupted BBB causes leakage of GBCAs into the extracellular space shown as enhancement on the T1 post-contrast MR images due to T1 shortening effects of the extravascular GBCA.[1] Based on the presence, location and pattern of contrast enhancement, it is possible to detect and characterize many brain pathologies for diagnostic and prognostic purposes, including brain tumors, inflammatory and infectious processes as well as some vascular pathologies.

The advantages of the administration of GBCA are offset by numerous drawbacks.[2] Linear-type GBCA administration is associated with nephrogenic systemic fibrosis (NSF) in patients with advanced renal disease.[3] Even with macrocyclic agents, varying degrees of deposition and retention of gadolinium in the brain and other organs, irrespective of renal function or integrity of the BBB, have culminated in the suspension of some GBCA in Europe in 2018 and restrictions of the use of GBCA to situations of strict clinical necessity.[4–6] On top of the minimization of the side effects,[7] the reduction of GBCAs administration could relieve patient burden,[8] positively impact healthcare cost and logistics within hospitals as well as reduce the environmental footprint.[4, 6, 9]

To address the overuse of GBCAs, two schools of thought have been explored in the literature, i.e., blanket reduction of GBCA dosage (without AI) or the use of AI methods to generate synthetic pseudo-contrast images based on zero-dosage or low-dosage brain MR sequences. The popularity of the approach with low-dosage brain MR and AI has led to the development of products such as SubtleGAD (for investigational use at the moment).[10] Despite demonstrating reasonable qualitative performance, the clinical usebase of such AI methods is limited. Most of the existing methodologies are developed for specific pathologies, and hence useful for follow-up imaging as opposed to first-line evaluation. In addition, these studies evaluate their methodology against “imaging” metrics of structural similarity or enhancing lesion volume correlation, overlooking clinical workflow-related metrics.[11–16] Assessing the actual necessity of GBCA administration at the point of the first-time brain MR examination may significantly reduce the number of GBCA-injections.

In this article, we hypothesize that reducing the use of GBCA in clinical practice should focus on supporting the protocol triage actions of radiologists via an adaptive intra-scan selection of appropriate patient-candidates for GBCA administration (either none, low-dose or standard-dose), rather than a blanket GBCA dose reduction. We propose a first-of-its-kind method, Apollo SmartGAD, which encapsulates two strategies: (i) in patients without a priori knowledge of any brain pathology, the method recommends whether the administration of GBCA is appropriate based on AI-generated synthetic pseudo-contrast images using non-contrast sequences; (ii) in patients with an a priori knowledge of a brain tumor diagnosis (both primary and secondary) obtained by the electronic health record, the method recommends the use of either no, low-dose or

standard-dose GBCA administration as defined by the GBCA vendor. We evaluated applications of our algorithm in two retrospective multicentric cohorts from real-life clinical scenarios in Denmark and India (application 1) and in a Dutch brain tumor dataset (application 2).

Methods

Algorithm for contrast recommendation for patients with unknown diagnosis

We developed and retrospectively tested an algorithm generating recommendations for GBCA administration in brain MR examinations (Apollo SmartGAD). Apollo SmartGAD is a proprietary tool developed by Cerebriu A/S (Copenhagen, Denmark). A graphical overview of the algorithm steps is outlined in Figure 1. Apollo SmartGAD is based on the interaction between four deep learning (DL) modules:

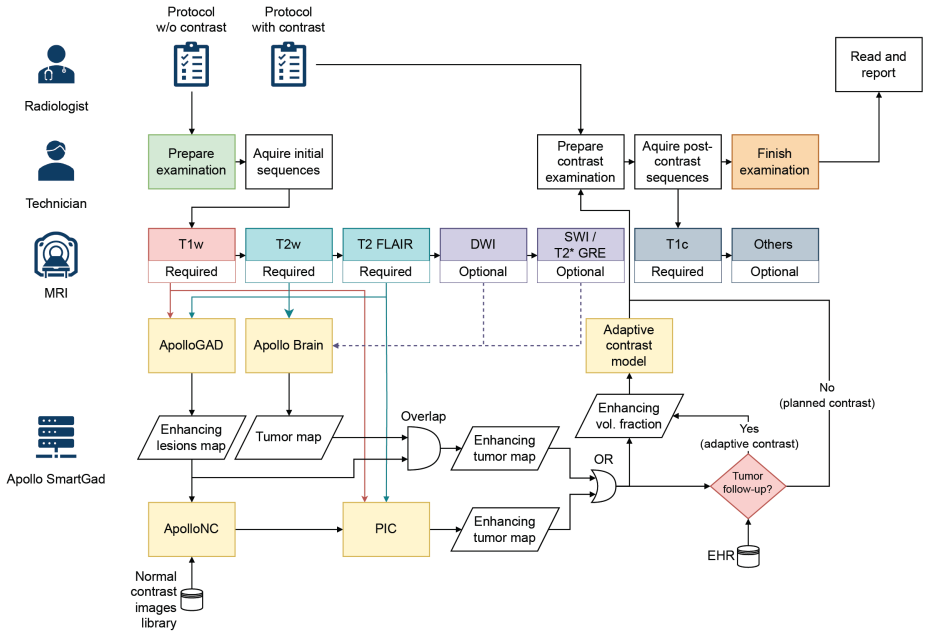


Figure 1: Overview of the modules of Apollo SmartGAD algorithm. The Apollo SmartGADalgorithm relies on four models: (i) space-occupying lesions detection model (Apollo Brain), (ii) a T1c synthesis model (ApolloGad), (iii) a normal contrast calibrator (ApolloNC), and (iv) an enhancing lesion segmentor (PIC, picture-segmentation based). The decision to administer GBCAs is based on the following algorithm combining outputs from Apollo Brain,ApolloGAD, ApolloNC and PIC. MR Sequence inputs to each module are indicated by solid lines for required sequences and dashed lines for optional sequences. The sequence input lines for sequences T2w and T2 FLAIR are colored blue to indicate both sequences are required inputs to all of the ApolloGAD, Apollo Brain and PIC components. The T1w sequence is differentially colored red to indicate that it is a required input only for the ApolloGAD and PIC modules (i.e., not Apollo Brain).

Apollo Brain: A proprietary DL based algorithm (Cerebriu A/S, Copenhagen, Denmark), which detects infarcts (except chronic), intracranial tumors, and intracranial hemorrhages (except isolated intraventricular hemorrhage, microhemorrhages, or those that are not hypointense on SWI/T2* GRE) based on 3 input MR sequences: a fluid attenuated inversion recovery (FLAIR) sequence, (optional) a b-1000 diffusion-weighted imaging (DWI) sequence, and (optional) a susceptibility weighted imaging (SWI) or T2* GRE sequence. In the present study, an investigational use version of Apollo Brain, which also analyzes a T2-weighted input sequence, was used.

ApolloGAD: a DL-based model developed specifically for this study to synthesize a pseudo-contrast T1-weighted (T1c) sequence synthesis. The model utilizes non-contrast T1-weighted, T2-weighted and FLAIR sequences as input. It is a custom DL-based pseudo contrast synthesis model developed for the present study described in detail below. ApolloGAD is trained to generate synthetic T1c images that may represent contrast enhancement in both normal and pathological tissue. The model architecture and training process are detailed below.

ApolloNC (normal contrast calibrator): a registration-based algorithm developed to attenuate contrast enhancement in anatomically normal regions within the synthetic T1c output produced by ApolloGAD. It does not override ApolloGAD's predictions in pathological tissue but rather aids in delineating normal from abnormal contrast enhancement for subsequent analysis by the PIC module. ApolloNC operates by registering the target synthetic T1c image to a fixed library of three normal reference scans (randomly selected ApolloGAD outputs from anatomically normal cases) using SynthMorph.[17] Voxels in the target image that overlap with enhancing regions in any of the registered templates are interpreted as normal and removed. This calibration step is illustrated in Figure 2.

PIC (Picture-nnU-Net for Lesion Segmentation): performs segmentation of contrast-enhancing lesions using the open-source picture-nnU-Net framework.[18] This model is a preoperative, multi-class glioma segmentation tool based on the nnU-Net architecture.[19] It segments whole tumor volume (including edema), tumor core, and enhancing tumor core components using T1-weighted, T1c (ground truth or synthetic), T2-weighted, and FLAIR sequences as input. The implementation used is the official version of the PICTURE model.[20]

In short, the algorithm initially relies on Apollo Brain software to determine the presence or absence of space-occupying lesions that may require GBCA administration. If Apollo Brain detects the presence of a lesion in the same region where the ApolloGAD algorithm predicts an enhancement, GBCA administration is recommended. In case Apollo Brain fails to identify any potentially enhancing lesions, ApolloGAD outputs are calibrated using ApolloNC to remove contrast uptake prediction in anatomically normal regions. Once calibrated, the PIC segmentation method is run on the pseudo-contrast T1c to confirm the presence or absence of one or more potentially enhancing lesion(s).

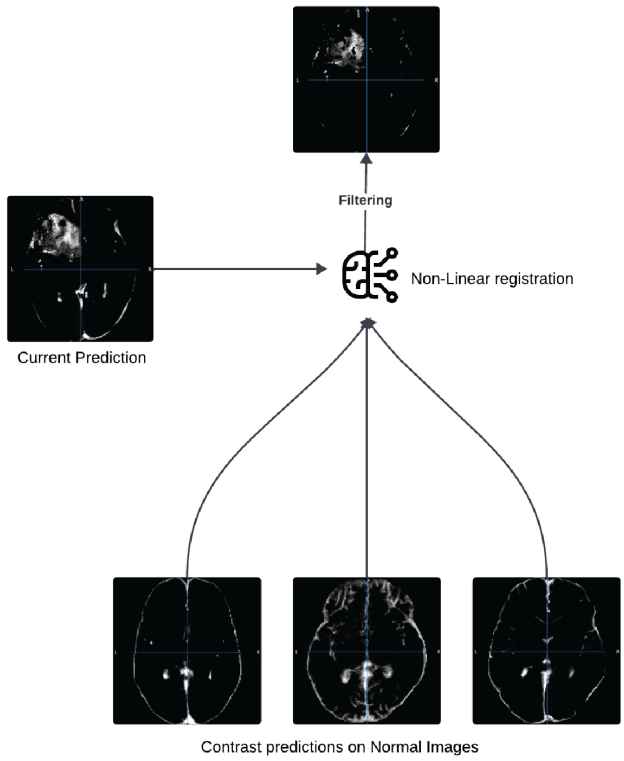


Figure 2: Illustration of the normal contrast calibrator. To reduce the prediction of contrast uptake signal in anatomically normal areas, the target synthesized contrast uptake image output by ApolloGAD is non-linearly registered to a fixed template library of 3 ApolloGAD outputs on anatomically normal scans. Contrast uptake from any of the registered library images is then removed from the target image.

Data

Ethics statement

This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki and current scientific guidelines. The Institutional Review Boards (IRBs) of hospital partners approved the study protocols or protocols for anonymous dataset generation. They waived the requirement for informed patient consent, given the minimal risk to subjects in the retrospective imaging study and the impracticality of obtaining informed consent from large numbers of patients retrospectively.

ApolloGAD training data

For the development of ApolloGAD, a total of 2,598 unique adult brain MRI studies with FLAIR, T2-weighted, pre- and post-contrast T1-weighted sequences and paired radiology reports were retrospectively collected from 8 different institutions across United States, Brazil, Israel, India, Ukraine and Denmark. An overview of the training data MRI

scanner vendors, sequence parameters, patient demographics, as well as diagnosed pathologies, can be found in Supplementary Tables S3 and S4. Additionally, training data was enriched with 7 open-source datasets: Brain Tumor Segmentation (BraTS) glioma[13], meningioma[14] and metastasis[15] challenge data, the University of Pennsylvania glioblastoma[21] cohort, the University of California San Francisco preoperative diffuse glioma MRI[22] dataset, Novosibirsk State Medical University diffuse glioma data[23], and Seoul National University Bundang Hospital supratentorial meningioma data[24].

Retrospective evaluation data
A retrospective evaluation of the clinical impact of the contrast decision algorithm introduced with Apollo SmartGAD was performed on a total of 1,251 unique adult brain MRI studies with T2-FLAIR, T2-weighted, pre- and post-contrast T1-weighted sequences and paired radiology reports, which were retrospectively collected from three institutions to generate three evaluation datasets:

Denmark. A subset of data was selected from a cohort of MR studies retrospectively collected from the Capital Region of Denmark between 2019 and 2021 originally described in Sagar et al.[25] All brain MR examinations of patients aged $\geq$ 18 were eligible for inclusion into the original cohort of 24,261 studies. Studies conducted between 1 January 1st, 2019 and 31st December, 2021 were retrieved. Inclusion criteria were availability of T2-FLAIR, T2-weighted, pre- and post-contrast T1-weighted sequences and availability of an associated radiological report. The final cohort included a total 469 studies with age 61.8 ± std 17.8 years at the time of scanning of which 255 (54.4%) were females, 208 (44.3%) were males and 6 (1.3%) were of unknown sex (missing or unavailable metadata) (Table 1).

India. A cohort of 350 MR studies from bodyScans, India (Kothari Diagnostics Pvt Ltd), were retrospectively randomly selected among those with at least the four required sequences (T2-FLAIR, T2-weighted, pre- and post-contrast T1-weighted) and associated radiological reports available. The cohort included patients with age 36.8 ± std 20.6 years at the time of scanning of which 145 (41.4%) were females, 190 (54.3%) were males and 15 (4.3%) were of unknown sex (missing or unavailable metadata) (Table 1).

Netherlands. The Amsterdam University Medical Centre provided a set of anonymized glioma and metastases cases. The glioma data was a subset of a larger dataset containing scans derived from a routine clinical setting obtained between June 2010 and July 2021. In total, 365 patients were included of which 222 were males [60.8%] and 143 were females [39.2%] with a mean age 53.0 ± std 15.4 years, all above the age of 18 and having all 4 sequences present (pre- and post-contrast T1-weighted, FLAIR, and T2-weighted). Tumor types and primary tumor origin for brain metastasis are reported in Table S2 of the Supplementary Materials and were determined by means of surgery or biopsy. The anonymized metastases dataset consisted of 67 patients of which 33 were males and 34 were females with a mean age 60.0 ± std 9.6 years, all treated in the Netherlands. Tissue diagnosis was performed by means of surgery or biopsy as part of routine clinical practice (Table 1).

Pathology prevalence. Radiological reports corresponding to MR scans from India and Denmark were reviewed by two MDs (S.I. and K.S.) and the presence or absence of contrast uptake was noted to define the ground truth (GT). In the case where the report mentioned an abnormality but contrast uptake was not mentioned (n=3), an expert rater (S.I., 8 years of experience in radiology) proceeded to visually inspect the images. Lesion type and subtype of the main radiological finding were recorded as well as extra findings and comorbidities.

Datasets from India (n=350) and Denmark (n=469) contained brain MR scans of patients with different pathologies across different centers, scanners, and scanning protocols. The Dutch dataset (n=432) contained anonymized preoperative brain MR scans of patients with either primary or secondary brain tumors. A summary of the demographics and lesions identified per cohort is available in Table 1. A visual overview is available in Figure 3A and 3B. Although the age distribution between the Indian and Danish cohorts was significantly different (India: age 36.8 ± std 20.6 years, 43.3% females; Denmark: age 61.8 ± std 17.8 years, 54.4% females), the proportion of patients that showed one or multiple abnormalities in the brain MRI was roughly two thirds of the total for both cohorts (India: n = 231 [66.0%]; Denmark: n = 287 [61.2%]) and the proportion of those showing some level of contrast enhancement was roughly one third of the total for both cohorts (India: n= 100 [28.6%]; Denmark: n=156 [33.3%]).

The distribution of pathologies was in line with expected prevalence[26–32] given the country and age groups of the Indian and Danish cohorts, with a similar proportion of normal cases (n=119 [34.0%], for India and n=182 [38.8%] for Denmark), but a higher proportion of brain tumors in the Danish cohort and a higher proportion of infectious disorders in the Indian cohort. There were very few cases of inflammatory or autoimmune disorders in the Indian cohort (n=4) and none in the Danish cohort. The proportion of vascular pathologies was slightly higher in the Danish cohort (India: n=45 [12.9%], Denmark: n=87 [18.6%]). The number of cases of neurodegeneration was 0 in the Danish cohort, as Danish guidelines do not indicate administration of GBCA for neurodegeneration protocols, and 18 in the Indian cohort. Remaining pathologies included a plethora of conditions including toxic and acquired metabolic disorders, genetic conditions and others.

The Dutch cohort (n=432, age 54.1 ± std 14.9 years, 40.7% females) was composed exclusively of patients with primary gliomas (n=365) or metastases (n=67). The majority (82.2%) of these tumors exhibited (some degree of) contrast uptake upon injection of GBCA.

Table 1: Demographics and pathology distributions per evaluation cohort. Continuous variables are expressed as mean ± standard deviation. Categorical variables are expressed as n, percentage (%). Age and sex were not available (NA) for n=14 and n=15 respectively for the India dataset and n=21 and n= 6 respectively for the Denmark dataset.

	India (n = 350)	**Denmark (n = 469)**	**Netherlands (n = 432)**
Age, yrs	36.8 ± 20.6	61.8 ± 17.8	53.0 ± 15.4
Sex, females	145 (43.3%)	255 (54.4%)	176 (39.2%)
GBCA uptake	100 (28.6%)	156 (33.3%)	355 (82.2%)
Normal	119	182	0
Brain Tumor	77	154	365
Cyst, mixed etiology	8	17	NA
Meningioma, WHO grade 1	10	26	NA
Craniopharyngioma	2	0	NA
Pituitary adenoma	13	0	NA
Schwannoma	4	1	NA
Astrocytoma, IDH-mutant, grade 2 and 3	1	0	39
Glioblastoma, IDH-wildtype, grade 4	12	6	230
Low-grade glioma, NOS, grade 2	9	8	75
Metastases, mixed primaries	9	84	67
Other/Unspecified	78	12	0
Infectious disorders	52	4	NA
Granuloma	36	0	NA
Cerebral abscess	5	0	NA
Meningitis	3	0	NA
Encephalitis	5	1	NA
Meningo-encephalitis	1	0	NA
Other	2	3	NA
Inflammatory and autoimmune disorders	4	0	NA
Multiple sclerosis	2	0	NA
Other	2	0	NA
Vascular pathology	45	87	NA
Acute infarct	15	19	NA
Hemorrhage	13	13	NA
Vascular malformations	1	3	NA
Other	16	52	NA
Neurodegeneration	18	NA	NA
Toxic and acquired metabolic conditions	4	NA	NA
Other pathologies	31	42	NA

Data Availability

While data sharing is subject to institutional and privacy regulations across different countries, we are exploring mechanisms to make a curated subset of anonymized data available to academic researchers upon reasonable request and under appropriate data use agreements.

ApolloGAD Model Development

Deep learning model architecture

Our contrast-difference synthesis algorithm, called ApolloGAD, is an ensemble of three U-Net[19] inspired models with an identical architecture, each adapted to a unique MRI sequence (detailed in Preprocessing of training data). The ensemble computes contrast-difference maps in the coordinate spaces of each of the three acquisitions, taking as input a concatenation of T2 FLAIR, T2-weighted, and pre-contrast T1-weighted sequences (pre-processed as described below), and produces a contrast-difference map between post- and pre-contrast T1-weighted sequences as an output. Each model instance operates on 1x1x1 mm^3 isotropic inputs of voxel space dimensionality 128x128x128 sampled at random within the input image volumes during training and from a grid to span all voxels of the input volume during inference. The outputs of the three model instances are averaged in native T1w sequence space to generate the final output map. The models have a fixed architecture based on the U-Net DL architecture, which largely follows that in Ronneberger et al.[33]. Key differences include the use of instance normalization layers due to small batch size (see Model training), valid padding to preserve image dimensions within a convolution block, transpose convolution to incorporate learnable up-sampling, a depth of five to increase the number of low-resolution feature channels, deep supervision to regularize feature maps at multiple resolutions, and a self-attention layer immediately following the input layer to assign importance weights to the three MRI sequences. Also of note is the employment of a sigmoid activation function at the end of the network rather than a softmax activation.

Training data preparation

ApolloGAD was trained on the “ApolloGAD training data” (see Methods, section Data) consisting of 2,598 unique adult brain MRI studies with T2-FLAIR, T2-weighted, pre- and post-contrast T1-weighted sequences and paired radiology reports. ApolloGAD is comprised of three individually trained instances of the U-Net base model described in Deep learning model architecture & ensemble strategy, each trained on the same training samples but preprocessed differently to be conformed to each of one of the three sequence spaces (T2-FLAIR, T2-weighted and pre-contrast T1w sequences). Each dataset had a “primary” sequence (e.g., T2-FLAIR) to which the other sequences were re-sampled and registered to using an affine transformation. A dedicated model instance was trained for each of the conformed dataset replicas, generating an ensemble of 3 models. The architecture and optimization hyperparameters were kept fixed for all three-model instances. However, each model used a random parameter initialization and optimization scheme, which, combined with different input dataset replicas, gave rise to small model variations in each model instance.

ApolloGAD was designed as an ensemble of three sequence specific sub-models to address the variability in MRI resolutions and acquisition planes across its three input sequences. Additionally, the slight variability of each trained model instance may further increase robustness of the ensemble when the models' errors are uncorrelated

The pre-processing steps were identical across the three sequence-conformed replica of the "ApolloGAD training data" except the use of a different primary sequence for each replica:

1. MRI sequences were skull-stripped using SynthStrip algorithm[34],
2. Primary sequence was resampled to isotropic space,
3. Non-primary sequences were resampled to isotropic primary sequence space,
4. Non-primary sequences were registered to primary sequence using SynthMorph[17] affine registration algorithm. In rare cases (<1% of samples) the SynthMorph tool was unable to successfully register one or more sequences due to severe motion artifacts or missing slices. These cases were manually identified and excluded from training. This methodological limitation is discussed in the Discussion section.
5. Sequences were concatenated and cropped to brain-only bounding box,
6. Sequence intensity values were scaled using min-max normalization to [0,1] range,
7. Contrast difference was computed as non-zero intensity difference between post- and pre-contrast T1-weighted sequences and was scaled using min-max normalization to [0,1] range.

Training data was split to training (85%) and validation (15%) according to data origin, MRI scanner and sequence parameters, as well as diagnosed pathologies. The same splits were used in all three sequence-conformed dataset replicas. External test sets comprised the clinical evaluation cohorts from India, Denmark, and the Netherlands (1,251 total cases), which were not used in training nor for validation, described in Methods, section Data.

Model training

Each model U-Net DL model instance in the ApolloGAD ensemble were initialized at random (with different seeds) using Xavier initialization and trained using a similar optimization hyperparameter set. A linear combination of a multi-scale structural similarity index measure (SSIM) loss, a global mean-absolute error (MAE) loss, as well as a local MAE loss, which was computed on regions of non-zero ground-truth contrast-difference connected components of at least 100 mm^3 in volume, excluding CSF and ventricles (segmentation of CSF and ventricles was obtained using SynthSeg algorithm[37]), was used to train all three model instances. The combination of global MAE, local MAE, and multiscale SSIM was selected to balance pixel-level accuracy and structural fidelity while avoiding the instability often observed with adversarial training[36], particularly when paired data quality varies across sources. Our use of SSIM helps capture perceptual structure, while local MAE targets enhancing regions specifically. Note that adversarial losses are also commonly used for synthetic image generation tasks, which could be investigated in future work.

An L2 regularization loss component with 10^{-5} regularization factor on the convolution kernel parameters was added. Gradient-based optimization was performed using the Adam optimizer with β_1=0.9, β_2=0.99 and a learning rate of 10^{-5}. A comprehensive suite of spatial and intensity data augmentation operations was applied probabilistically during training, which included, but were not limited to, axis reflection, resizing, rotations, elastic transformations, Gaussian blurring and Gaussian noising. Augmentations, in particular non-linear spatial augmentations such as elastic transformations, may generate biologically implausible images, but even strong applications of such augmentations have been shown able to yield more robust segmentation across unseen contrasts and resolutions.[35] In the present study, elastic transformations were included with conservative parameter settings primarily to encourage spatial robustness to scanner-induced distortions and minor patient movement, and augmentations were only applied to a subset of randomly selected batches with small application probability throughout training to not train primarily on augmented data. The outputs of the augmentation pipeline were manually audited prior to training to ensure that augmentations would not significantly hinder the interpretability or general information content of the image.

Each model was trained on a single 40GB NVIDIA A100 Tensor Core GPU. All training schemes used a batch size of 3, where the first element in the batch was either a normal study or a study which did not contain any contrast-enhancing pathology (e.g., a chronic infarction or an arachnoid cyst), the second element was a study which contained a glioma (e.g., astrocytoma or an oligodendroglioma), and the third element was a non-glioma contrast-enhancing pathology (e.g., meningioma or a schwannoma). It was not possible to use larger batch sizes than 3 given the available GPU memory for these models. When designing the model architecture and its complexity, it was ensured that a batch size of at least 3 could be used to at least enable the active sampling of batches to display a diverse pathology representation within each mini batch. Due to the small batch sizes, the model architecture was designed with instance normalization layers, which are commonly used in DL for 3D medical images and small batch sizes.

Each of the three U-Net models were first pre-trained for 750,000 steps (250 steps per epoch for 3000 epochs) on both the internal and the external training data, with all loss coefficients equal to 1, and were then fine-tuned for 250,000 steps on internal data only (250 steps per epoch for 1000 epochs), with local loss coefficient increased to 10.

Algorithm for GBCA dosage recommendation for patients with known diagnosis of glioma

From a workflow perspective, we hypothesized that the decision to use low- vs. standard-dose GBCA is of higher benefit than a blanket call of reducing contrast. In lieu, we tested the performance of our algorithm in determining the need of standard-dose contrast and low-dose contrast in patients with known glioma diagnosis. We postulated that the ratio (enhancing tumor volume fraction) between the enhancing volume and the total lesion volume (defined by the non-enhancing lesion volume + edema volume + necrotic volume + enhancing lesion volume) is the most radiologically relevant discriminatory feature to recognize high-grade gliomas (HGGs) from LGGs, as LGGs by tendency enhance

less frequently and less extensively in the group of adult-type diffuse gliomas. Since algorithms tend to be biased by lesion volume size, as also observed in Jayachandran Preetha & Meredig et al.[11], we also inspected the discriminatory ability of tumor volume only.

Prediction of low- vs high-grade gliomas

We aimed to classify low- vs. high-grade gliomas based on patient demographics and image features computed from the ground truth- and pseudo-contrast uptake images, respectively. Logistic regression models were fitted with input features age (numeric), patient sex (female/male binary) and estimated enhancing tumor volume fraction (numeric in [0, 1]). Estimated enhancing tumor volume fractions were computed by applying the PIC model (see the Algorithm for contrast recommendation for patients with unknown diagnosis section above) to GT and pseudo-contrast images to predict segmentation masks for the whole and enhancing tumor areas, respectively. Their relative volumes were then computed. The logistic regression models were fitted to predict the likelihood of a glioma being high-grade based on the three input features {age, sex, enhancing volume fraction} using the scikit-learn (v.1.2.2) Python package with the "balanced" class weights mode, a maximum solver iteration count of 1000, a stopping criteria tolerance of 0.0001, a regularization strength of C = 1, and the "lbfgs" solver.

A total of 21 cases were excluded from all experiments where PIC failed to detect any tumor area (20 pseudo-contrast cases and 1 ground truth contrast case). The following experiments were performed for all remaining 344 cases:

1. A paired leave-one-out (LOO) experiment in which n=344 iterations of two logistic regression models were identically initialized and fit to identical subsets of n-1=343 training cases on the feature set computed from ground truth- and pseudo-contrast images, respectively, and applied to the features on the remaining testing case. In combination, this created a set of predictions of the probability of high-grade based on each feature set for all n=344 cases. The performance of the two classifiers were assessed by computing precision, recall and specificity metrics as well as the Area Under the Receiver Operating Characteristic Curve (ROC AUC) and Area Under the Precision-Recall curve (PR AUC). ROC AUCs were compared by computing 95% confidence intervals and a two-sided DeLong test for comparing AUC scores at a significance level of 5%.[38]

2. The entire LOO experiment (see point 1) was repeated 1,000 times with dataset bootstrapping. In each iteration, the dataset of 344 cases were uniformly resampled with replacements to construct a new dataset of 344 (possible repeated) cases and the LOO experiment executed on this dataset to produce a pair of PR AUC values; one for the classifier operating on features computed from ground truth- and pseudo contrast images, respectively. The mean and standard deviation of the obtained PR AUC scores were compared and their distributions visualized in histogram plots. The Coheen's d effect size was computed and interpreted as "small", "moderate", or "large" according to the thresholds of d=0.2, d=0.5, and d=0.8 as suggested by Coheen (1988).[39]

Results

To gad or not to gad

The out-of-the-box performance of the algorithm in detecting lesions with expected leakage of GBCA was tested across all lesions in both the Indian and Danish cohorts against the GT annotations. A summary of the evaluation metrics per cohort as well as the pooled results are reported in Table 2. Results were comparable in the two cohorts. When pooling the data, the proportion of patients who had a true contrast uptake and were correctly recommended by the algorithm for GBCA administration was 73% (sensitivity), while the proportion of patients who did not have contrast uptake and were correctly not recommended for GBCA administration ranged 79% (specificity). The proportion of patients recommended for GBCA administration who actually had contrast uptake was 62% (positive predictive value, PPV), while the proportion of patients not recommended for GBCA administration that indeed did not show any GBCA uptake was 86% (negative predictive value, NPV). The proportion of patients who did not have GBCA uptake but were incorrectly recommended for GBCA administration by the algorithm ranged 21% (false positive rate, FPR). The proportion of patients who showed uptake but were incorrectly not recommended for GBCA administration by the algorithm was 27% (false negative rate, FNR). Finally, the proportion of patients recommended for GBCA administration who did not actually have any GBCA uptake was 38% (false discovery rate, FDR). Examples of correct predictions are shown in Figure 4.

An overview of the incorrectly classified cases, i.e., false negatives (FN) and false positives (FP), is depicted in Table 3. Example images are shown in Figure 5. Most of the false negatives were in the brain tumor category. Out of the 190 tumors that showed real contrast uptake, 41 were FN. Among these 41 cases, the most missed were metastasis (20) and meningiomas (11). No associated edema was mentioned in the radiological reports for these lesions. The maximum lesion diameter among missed lesions was 2.8 cm (a meningioma), as manually measured and mentioned in the radiological report. Apollo Brain identified 4 out of 11 meningiomas as tumors, but this collided with insufficient contrast enhancement in the pseudo-contrast images produced, hence, due to the decision system of the algorithm, they were deemed to

Table 2: Summary of evaluation metrics assessing the algorithm performance in predicting contrast uptake (yes/no). For each of the evaluation metrics, 95% confidence intervals are indicated in brackets. Acronyms: positive predictive value (PPV), negative predictive value (NPV), false positive rate (FPR), false discovery rate (FDR), false negative rate (FNR).

Metric	India	Denmark	Pooled
Sensitivity	0.74 (0.65, 0.82)	0.73 (0.66, 0.80)	0.73 (0.68, 0.78)
Specificity	0.82 (0.77, 0.87)	0.76 (0.72, 0.81)	0.79 (0.75, 0.82)
PPV	0.63 (0.54, 0.71)	0.61 (0.54, 0.68)	0.62 (0.56, 0.67)
NPV	0.89 (0.84, 0.92)	0.85 (0.80, 0.89)	0.86 (0.83, 0.89)
FPR	0.18 (0.13, 0.23)	0.24 (0.19, 0.28)	0.21 (0.18, 0.25)
FDR	0.37 (0.29, 0.46)	0.39 (0.32, 0.45)	0.38 (0.33, 0.44)
FNR	0.26 (0.18, 0.35)	0.27 (0.20, 0.34)	0.27 (0.22, 0.32)

not be recommended for GBCA administration. Among the 20 FN metastases, 18 were from the Danish cohort and all of them were reported as small lesions. Among the 20 FN metastases, the largest lesion had a maximum diameter of 2.1 cm. Once more, Apollo Brain predicted 17 out of 20 cases as tumors but since ApolloGAD showed no enhancement due small lesion size, no GBCA recommendation was given for these cases.

Table 3: Definition of false negatives (FN) and false positives (FP) for predicting enhancement. For FN we report (i) the total number of FN per lesion type and subtype FN_{lesion} (n), (ii) the total number of lesions of that type and subtype in the dataset FN_{lesion}+TPtextsubscriptlesion (total) and (iii) the ratio of FN per lesion type and subtype divided by the total lesions of that type and subtype in the dataset (ratio = n/total = $FN_{lesion}/(FN_{lesion}+TP_{lesion})$). For FP, we report (i) the total number of FP per lesion type and subtype FP_{lesion} and (ii) the ratio between the number of FP of that type and subtype in the dataset FP_{lesion} divided by the total number of FP in the dataset FP_{total}=116 (ratio = FP_{lesion}/FP_{total}).

	False Negatives		**False Positives**	
Lesion type	n	Total (Ratio %)	n	Ratio %
Brain tumor	41	190 (21.6)	17	14.7
Cyst	4	5 (80.0)	7	6.0
Pituitary adenoma	4	10 (40.0)	0	0.0
Meningioma	11	36 (30.6)	0	0.0
Low-grade glioma	0	0 (0.0)	3	2.7
Ependymoma	0	0 (0.0)	1	0.9
Glioblastoma	0	0 (0.0)	1	0.9
Primary CNS lymphoma	1	6 (16.7)	2	1.7
Metastasis	20	90 (22.2)	3	2.6
Other	1	2 (50.0)	0	0.0
Vascular pathology	11	23 (47.8)	45	38.8
Acute/subacute infarct	6	13 (43.2)	6	5.2
Hemorrhage	3	6 (50.0)	12	10.3
Other	2	2 (100.0)	20	17.2
Vascular Malformations	0	2 (0.0)	0	0.0
Infectious disorders	12	40 (30.0)	4	3.5
Granuloma	7	26 (26.9)	0	0.0
Cerebral abscess	1	4 (25.0)	0	0.0
Meningitis	1	3 (33.3)	0	0.0
Meningo-encephalitis	1	1 (100.0)	4	3.5
Other	1	3 (33.3)	0	0.0
Inflammatory and autoimmune disorders	1	1 (100.0)	1	0.9
Other	3	8 (37.5)	11	9.5
Normal	0	0 (0.0)	37	31.9

Additionally, 11 out of 132 vascular pathologies were marked to show some degree of GBCA uptake. However, they were due to the ischemia-induced blood-brain barrier breakdown and would clinically likely not warrant GBCA administration. Of these 11 cases, 3 cases had DWI b-1000 images available and Apollo Brain predicted recent

infarctions in all three. A total of 12 out of 56 infections were missed, of which 7 were granulomas either due to tuberculosis or neurocysticercosis, both conditions more prevalent in the Indian population.

Of the 819 cases, 116 were false positives, of which 37 cases were normal (31.9% of FPs). All cases from the Danish cohort had SWI and DWI images missing, which could be the reason for such a large number of FP. In addition, 17 FP cases across Danish and Indian datasets showed a brain tumor; among them, 3 were low-grade gliomas (LGGs), 7 were benign cysts, 3 non-enhancing metastases, and 2 central nervous system (CNS) tumors not otherwise specified (NOS). Vascular lesions accounted for 45 FP (38.8%). A substantial amount of them could have been addressed by using Apollo Brain predictions of infarcts and hemorrhages. Among the vascular lesions, 6 were acute infarcts, 12 hemorrhages, and 20 were a mix of different vascular malformations, for which MR angiography was indicated. Finally, infectious disorders accounted for a small number of false positives (n=4).

Classification performance metrics for the Apollo Brain sub-component in detecting tumors (including metastases) and infections (as measured by at least 1 voxel predicted in an image of tumor ground truth) on the Denmark, India and pooled (Denmark + India) cohorts are shown in Supplementary Table S5. Apollo had high sensitivity (0.84) on the Denmark cohort and lower on India cohort (0.65) while a high specificity on the India cohort (0.90) and lower on the Denmark cohort (0.56). The low specificity on the Denmark cohort is likely driven by false positives due to missing input DWI and SWI sequences as mentioned above. The low sensitivity on India cohort is likely driven by higher relative fraction of infections of which many are small and more likely to be missed. Performance metrics for the PIC sub-component in predicting contrast enhancement (as measured by at least 1 voxel predicted in an image with within any enhancing lesion), on the Denmark, India and pooled (Denmark + India) cohorts are shown in Supplementary Table S6. PIC had a high specificity on the Denmark cohort compared to Apollo (0.81 vs. 0.56) but lower sensitivity on Denmark (0.42) and lower sensitivity and specificity on India (0.67 and 0.66, respectively).

GBCA dosage for known brain tumors (low- vs. standard dose)

Absolute enhancing volume showed only moderate discriminatory power for distinguishing HGGs from LGGs, yielding a high false-positive rate (specificity 28% with true contrast images). This suggests that the ratio of enhancing-to-total lesion volume provides a more informative and robust metric for classification.

The sensitivity and specificity of the detection of contrast-enhancing lesions in the Dutch cohort are reported in Table 4, together with the performance of ApolloGAD with PIC segmentation model. Since the diagnosis of primary or secondary tumors is known a priori for the Dutch dataset as per the study design, running Apollo Brain on these cases was deemed redundant (all tumors were detected, data not shown). The mean structural similarity index measure (SSIM) between true post-contrast T1-weighted images and pseudo-contrast images generated by ApolloGAD ranged between 0.77 and 0.85 across

all cohorts (Table 5). The Indian cohort showed lower SSIM due to an overall lower resolution and quality of images (Table S1 Supplementary Materials). For the Dutch dataset, limited to adult-type diffuse gliomas and metastases, we obtained a score of 0.83.

Table 4: Sensitivity and specificity in predicting presence or absence of contrast enhancement using Apollo SmartGAD and ApolloGAD with PIC without employment of Apollo Brain in the Dutch cohort (tumor diagnosis known a priori). GT is defined as presence or absence of contrast after radiological evaluation of the images.

	ApolloGAD + ApolloNC	ApolloGAD + ApolloNC + PIC
Sensitivity (95% CI)	68.2% (63.2%, 72.8%)	73.5% (68.9%, 78.1%)
Specificity (95% CI)	59.7% (48.6%, 70.0%)	67.5% (57.0%, 78.0%)

Table 5: Mean SSIM across cohorts with 95% confidence intervals.

	India (n=350)	Denmark (n=469)	Netherlands (n=432)
SSIM score	0.77 (0.71, 0.83)	0.85 (0.80, 0.89)	0.83 (0.78, 0.88)

Figure 3C illustrates the enhancing volume to total lesion volume ratio for LGGs and HGGs. The true enhancing volume was defined as the volume computed by the PIC model on true post-contrast T1 weighted images. All cases where ApolloGAD did not predict any enhancement were excluded. The mean Dice for cases with glioma tumors was 0.27 ± 0.19 (median Dice = 0.25). For brain metastases, the mean Dice was 0.41 ± 0.22 (median Dice = 0.43). Figure 3C confirms that GT contrast uptake is a strong predictor of LGGs vs HGGs. However, we proposed to use the enhancing volume to total lesion volume ratio to classify LGGs and HGGs in order to mitigate potential bias from contrast overprediction in larger LGGs of synthetic pseudo-contrast images. Two logistic regression classifiers were used to distinguish HGGs from LGGs using the ratio between the estimated enhancing tumor volume and whole tumor volume as predictor after correction for age and sex. Then, performance of these classifiers was assessed in comparison to the biopsy results dichotomised for HGGs and LGGs.

Afterwards, we performed a leave-one-out (LOO) experiment. Figure 3D and 3E show respectively receiver operating characteristic curve (ROC) area under the curve (AUC) and precision-recall (PR) AUC for the classifiers operating on post-contrast T1-weighted, defined as GT, and pseudo-contrast images. The ROC AUC of the classifier operating on features computed from post-contrast T1-weighted images was 0.90 (95% CI: 0.86, 0.93) while the precision recall (PR) AUC was 0.98 (95 % CI: 0.98, 0.98), with a sensitivity of 0.80, specificity of 0.85, PPV of 0.96, and NPV of 0.51 at operating point 0.50 (10 FP & 55 FN with HGGs as positive class). The ROC AUC of the classifier operating on features computed from pseudo-contrast images was 0.80 (95% CI: 0.74, 0.86) while the PR AUC was 0.93 (95 % CI 0.93, 0.93), with a sensitivity of 0.77, specificity of 0.78, PPV of 0.93, and NPV of 0.46 at operating point 0.50 (15 FP & 63 FN w. high-grade

glioma as positive class). There was a significant difference between the AUC scores of the two classifiers (DeLong test, P-value < 0.001, α= 0.05).

The experiment was repeated 1,000 times with bootstrapping to assess the reliability of the results reported in relation to variations in the training dataset composition as well as to estimate an effect size on the obtained PR AUC scores when using GT post-contrast T1-weighted vs. pseudo-contrast images for feature extraction (Figure 3G). The PR AUC scores were 0.98 ± std 0.00 and 0.93 ± std 0.01 using post-contrast T1-weighted and pseudo contrast image features, respectively.

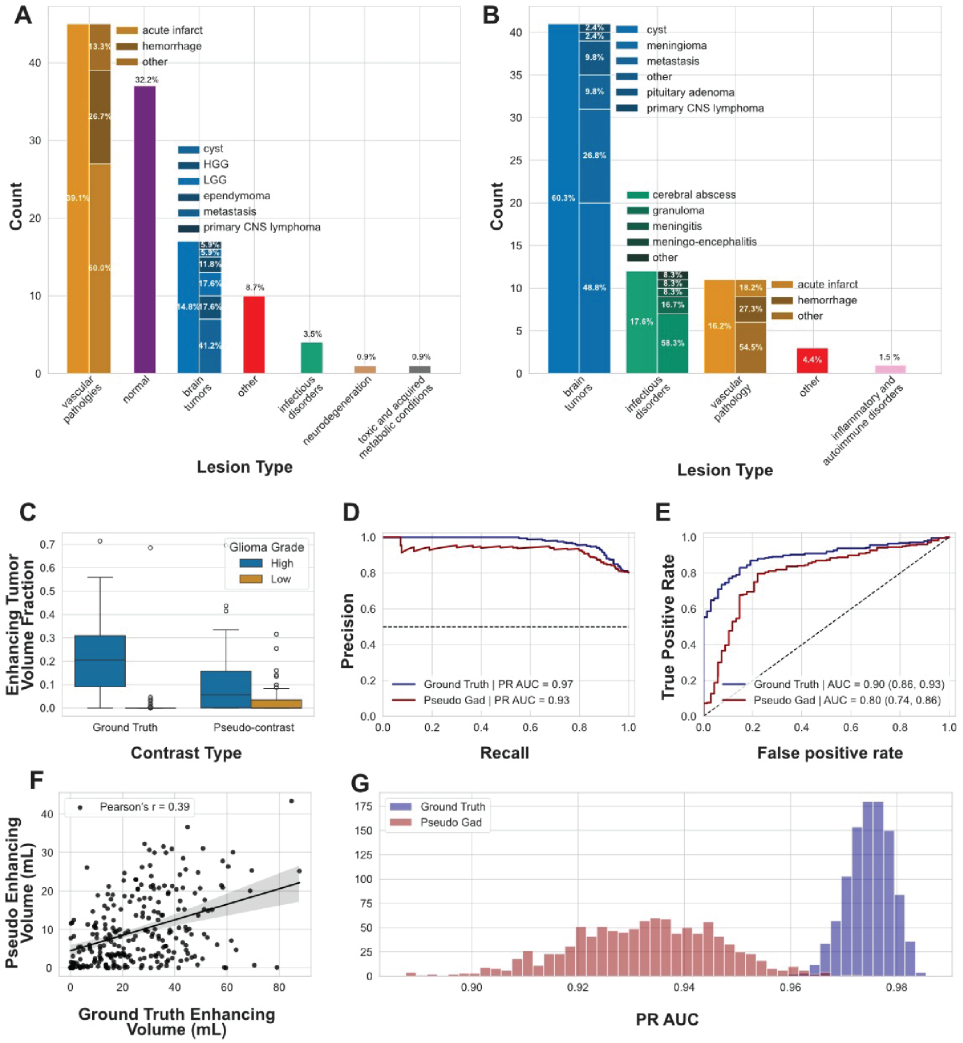


Figure 3: Results overview. A: False positive lesion types and subtypes. Unlike Table 3, where the proportions are based on contrast positive lesion type, the figure illustrates proportions with respect to all of the lesion type or subtype irrespective of the contrast uptake. B: False negative lesion types and subtypes. Unlike the Table 3, where the proportions are based on contrast positive lesion type, the figure illustrates proportions with respect to all of the lesion type or subtype irrespective of the contrast uptake. C. Enhancing volume to total lesion volume ratio box plots for low grade and high grade gliomas using true post-contrast T1 weighted images and ApolloGAD. D. PR AUC and E. ROC AUC obtained in a LOO high- vs. low-grade glioma classification experiment using post-contrast T1-weighted images (blue line) vs. pseudo contrast images (red line) for feature extraction. A 95% confidence interval is shown for the ROC AUC scores. F. Scatter plot showing the correlation between the True enhancing volume and Pseudo enhancing volume. The Pearson's correlation coefficient is r=0.39 with a p-value < 0.00001. LGGs were excluded from this analysis due to the lack of enhancing tumor components in the majority of the cases. G. Observed distributions of PR AUC scores obtained in 1,000 bootstrap repetitions of the paired leave-one-out HGGs vs. LGGs classification experiments using post-contrast T1-weighted (blue histogram) vs. pseudo-contrast (red histogram) images for feature extraction.

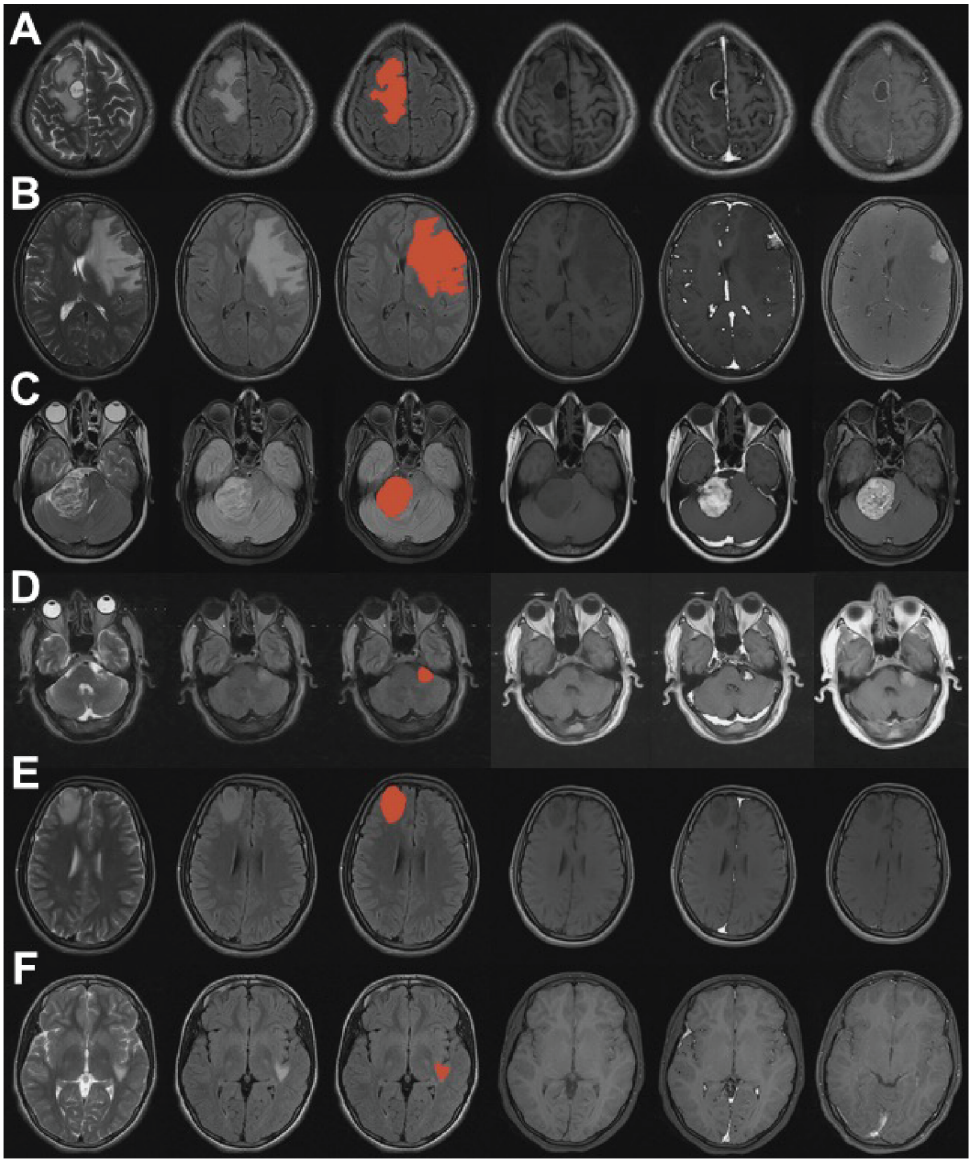


Figure 4: Examples of true positive (TP) and true negative (TN) evaluation studies. Each row represents selected 2D slices from a unique brain MRI study, with columns corresponding to the following MRI sequences: T2-weighted, T2-FLAIR, T2-FLAIR with overlaid Apollo Brain tumor segmentation (red), pre-contrast T1-weighted, pre-contrast T1-weighted with overlaid ApolloGAD contrast-difference map (grayscale), and post-contrast T1-weighted. A. Case of neurocysticercosis, correctly predicted to be showing GBCA uptake on post-contrast T1-weighted MRI sequence (TP). B. Case of meningioma, correctly predicted to be showing GBCA uptake (TP). C. Case of schwannoma, correctly predicted to be showing GBCA uptake (TP). D. Case of schwannoma, correctly predicted to be showing GBCA uptake (TP). E. Case of low-grade glioma, correctly predicted to not be showing GBCA uptake (TN). F. Case of dysembryoplastic neuroepithelial tumor, correctly predicted to not be showing GBCA uptake (TN).

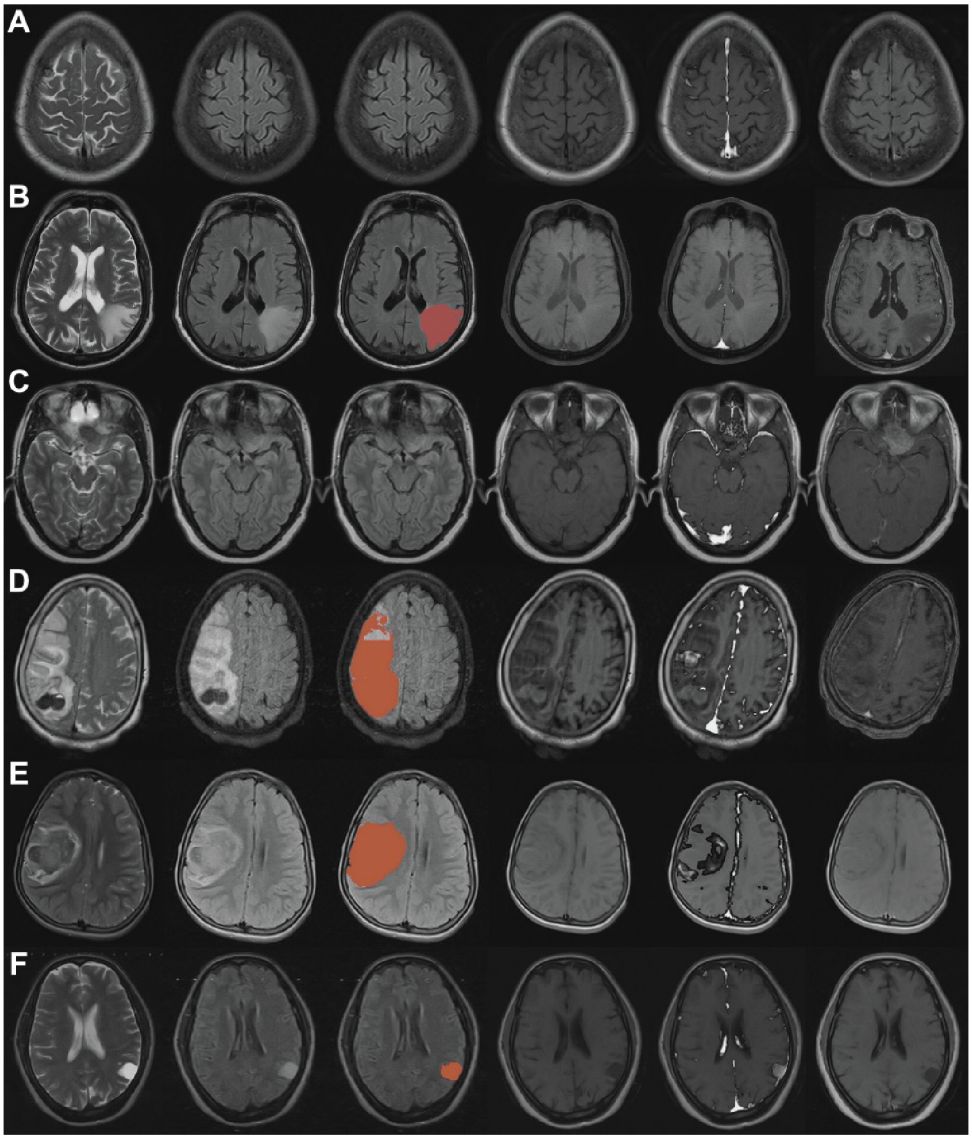


Figure 5: Examples of Apollo SmartGAD false negative (FN) and false positive (FP) evaluation studies. Each row represents selected slices from a unique brain MRI study. Columns correspond to the following MRI sequences: T2-weighted, T2-FLAIR, T2-FLAIR with overlaid Apollo Brain tumor segmentation (red), pre-contrast T1-weighted, pre-contrast T1-weighted with overlaid ApolloGAD contrast-difference map (grayscale), and post-contrast T1-weighted. A. Study with meningioma, incorrectly predicted to not be showing GBCA uptake on post-contrast T1-weighted MRI sequence (FN). B. Study with brain metastasis, incorrectly predicted to not be showing GBCA uptake (FN). C. Study with meningioma, incorrectly predicted to not be showing GBCA uptake (FN). D. Study with subacute infarct and intra-axial hemorrhage, incorrectly predicted to be showing GBCA uptake (FP). E. Study with intra-axial hemorrhage, incorrectly predicted to be showing GBCA uptake (FP). F. Study with epidermoid cyst, incorrectly predicted to be showing GBCA uptake (FP).

Discussion

In this study, we examined how predicting the necessity of gadolinium-based contrast agents (GBCAs) during a brain MR examination could meaningfully reduce the number of contrast scans performed globally, which would in turn minimize patients discomfort and potential adverse effects related to GBCAs, while also decreasing the environmental footprint of GBCAs.[40] Radiologists assess the need for GBCA when assigning a scanning protocol to a patient. Hence, we approached the problem from a workflow perspective and developed an algorithm designed to reassess the need for GBCA administration on the actual MR data in real time.[41] The recommendation to administer GBCA was automatically determined, without prior knowledge of patient diagnosis, through an algorithm that uses pre-contrast brain MR sequences as input and generates synthetic pseudo-contrast images. Moreover, in the case of a priori knowledge of a glioma diagnosis, the algorithm recommends whether no, low- or standard GBCA dose administration is required. Our findings indicate that real-time decision-making regarding GBCAs administration based on pre- and pseudo-contrast images is feasible but requires fine-tuning on smaller lesions. This approach could enhance patient safety, streamline radiology workflows, and contribute to more efficient and sustainable medical imaging practices, hence offering significant clinical and economic benefits.

A crucial aspect of our proposed algorithm is its high negative predictive value (NPV), which was 86%. The NPV indicates the proportion of negative results (i.e., no GBCA administration recommended) that are correctly identified as true negatives (i.e., cases where GBCA administration is genuinely unnecessary). By reliably determining when GBCAs are unnecessary, the use of this software can significantly improve the efficiency and safety of brain MR imaging, as it could ensure that patients who do not require GBCAs are correctly identified thereby minimizing the potential risks associated with GBCAs and contributes to the efficient use of medical resources, reducing costs and improving the logistical efficiency of radiology clinics. Nevertheless, a workflow-efficiency and cost analysis were beyond the scope of this study and should be addressed in future work. Further, while a high NPV was observed, 14% of cases GBCA uptake were erroneously classified as showing no uptake. This finding underscores the importance of implementing SmartGAD as a clinical decision-support tool rather than as a substitute for radiologist interpretation. In settings where the cost of missing subtle enhancement is high (e.g., suspicion of metastasis), radiologists would retain discretion to override the algorithm's recommendation. Additionally, flagged uncertainties could trigger a follow-up scan or radiologist review.

A significant number of false negatives (FN) were observed in predicting the presence of enhancement in metastases (22% missed) and in meningiomas (30% missed). Our findings suggest that omitting GBCA administration in patients with suspicion of metastases may not be justified, rendering the use of Apollo SmartGAD unsuitable for cases where scans are performed to confirm or rule out the presence of metastasis, as these can be very small. Although full-dose contrast to accurately detect the lesions seems appropriate at this stage, further research is needed to establish whether low-dose is feasible in this context.[6]

On the other hand, the relatively high percentage of missed meningiomas (30%) usually does not carry severe consequences. As a matter of fact, it can be argued that in clinical practice patients with a meningioma detected as a part of a routine brain MR get recalled for further imaging using GBCAs. The possibility of recalls could have economic implications. In our study, the largest missed meningioma had a maximum diameter of 2.8 cm which was called “small” in the corresponding radiological report. Furthermore, both Boto et al. and Sathraju et al. showed excellent correlation between meningiomas’ size measured on non-contrast T2w- and post-contrast T1 weighted sequences, questioning the need for contrast especially in small untreated meningiomas.[42, 43] Some authors suggest more conservative approaches with a lowered contrast dosage[44], but this might not always be necessary especially if the non-contrast images are typical. All in all, the assessment of FNs indicates that future developments of the algorithm should focus on further fine-tuning to account for smaller lesions.

Among all the false positives (FPs), 32% of the patients where the radiological reports indicating nothing clinically abnormal were recommended for post-contrast imaging by the algorithm which was a significant contributor to a moderate positive predictive value (PPV) of 62% However, among all patients marked as clinically normal, only 12% of cases (37 out of 301) had a recommendation of GBCA administration by the algorithm. In fact, a high number of FPs was due to the presence of vascular lesions such as ischemic infarcts or hemorrhages, where GBCA administration is not routinely performed but may be warranted for angiography in a subset of cases. However, it must be noted that the prediction of post-contrast angiographic images was beyond the scope of the intended use of the software. Furthermore, the detection of ischemic or hemorrhagic lesions, may be addressed using peripheral outputs of the Apollo Brain software, which was not directly addressed in this study.

The real-life benefit of the proposed algorithm will vary depending on the specific workflow of different radiology units. For instance, in workflows where CT imaging is the first choice, patients landing with an MR order will either have a CT negative scan with persistent symptoms or a lesion requiring further investigation with advanced MR imaging. In the first case, the likelihood of detecting contrast-enhancing lesions will be low because of the negative CT. Hence, due to its high NPV, our algorithm could be beneficial in sparing GBCA administration in centers where blanket-use of GBCA is routinely preferred. Instead, in centers where technologists flag lesions potentially requiring GBCA administration during the scan, the algorithm could support technologists in lesions detection.

Nearly 15% of the FPs were space-occupying lesions including cysts, metastases, low-grade gliomas (LGGs), and post-operative high grade gliomas (HGGs). In most of these cases, the administration of contrast may actually be warranted, with the notable exception of some types of simple fluid-filled benign cysts, e.g., arachnoid or pineal cysts. All in all, predictions of the algorithm might be improved for space occupying lesions by increasing the size and variety of the training dataset and by adding some

options for feature extraction from the imaging data such as texture analysis and radiomic features to better differentiate lesions requiring GBCA administration. It should also be noted that the evaluation of the algorithm was based on the presence or absence of contrast enhancement in post-contrast T1w images. However, in first-line evaluations and/or in the absence of a clear clinical suspicion, the decision on the administration of GBCA usually follows the detection of unclear lesions on pre-contrast sequences, hence determining many recalls.

The second aim of this study was to investigate the feasibility of lowering the dose of GBCA in patients with known gliomas and metastases. We tested our algorithm on a glioma dataset by hypothesizing that the lesion-to-contrast ratio may be used to classify high-grade (HGG) from low-grade (LGG) gliomas thereby supporting informed recommendations for either reduced or standard GBCA dosage in brain MRI protocols.[8] While the presence of contrast enhancement is more commonly associated with HGGs, we specifically used the ratio of enhancing to total lesion volume as the classification metric to minimize potential bias arising from absolute lesion size. We tested the hypothesis with both synthetically generated T1 post-contrast images from our algorithm and the real post-contrast T1w images. The underlying rationale was that, since LGGs usually require long-term follow-up and typically demonstrate no contrast enhancement[Larsen2017-ed, Jo2022-bd], a reliable algorithm capable of distinguishing LGG vs HGG could facilitate safe and targeted GBCA dose reduction. Such dose optimization decisions can also be applied in the context of HGGs, although they must be approached with caution as changes in GBCA dosage might affect the quantification of enhancing tumor volume, which has been shown to be a reliable prognostic indicator[45]. These hypotheses warrant further validation in future studies. In fact, GBCA dosage lowering may not be advisable in cases where optimal diagnostic or prognostic assessment or better treatment planning may be achieved with standard dosage. Current literature does not yet provide definitive guidance on the minimum effective GBCA dose required for optimal diagnostic confidence.

We confirmed that the presence of any contrast enhancement is an excellent predictor of possible HGGs. The PR-AUC of predicting high grade gliomas based on enhancement-to-lesion volume ratio was 0.97, with a sensitivity of 80% and specificity of 85%. Pseudo-contrast images generated a PR-AUC of 0.93 with sensitivity of 77% and specificity of 78%. The closeness in discriminatory ability of our algorithm in recommending lower or standard-dosage compared to the ground truth indicates the feasibility of using this solution in clinical workflows. Nevertheless, our algorithm tended to overpredict LGGs, suggesting that further evidence, especially in the form of prospective multicenter studies and randomized control trials, is needed to translate this model into a clinic-ready tool.

We obtained high structural similarity index measure scores between contrast-difference maps computed using true post-contrast T1-weighted images and pseudo-contrast images across all cohorts (range 0.77–0.85), with the lowest scores in the Indian cohort due to an overall lower resolution and quality of images. In comparison, the SSIM score reported by Jayachandran Preetha & Meredig et al. was 0.82, but their dataset was limited to large glioblastomas, similar to the score of 0.83 that we obtained in the Dutch cohort across adult-type diffuse gliomas and metastases.[11] While the SSIM scores between predicted and ground-truth contrast-difference maps are high, reflecting strong structural and intensity-level agreement, the Dice scores are significantly lower due to the effects of thresholding. Binarization discards continuous intensity information and is highly sensitive to small shifts near the threshold boundary, leading to lower overlap even when overall structure and location match well. This discrepancy highlights the complementary nature of SSIM and Dice: the former captures nuanced structural similarity, while the latter penalizes even minor differences in predicted lesion extent after hard thresholding.

This study has several strengths. We validated our algorithm for using retrospective multicentric cohorts from Denmark and India and we examined its utility in follow-up scenarios for known brain tumors using a brain tumor dataset retrospectively obtained over a period of >10 years from one single academic site in the Netherlands (Amsterdam), specifically assessing the decision for low or standard GBCA doses. The algorithm performance was stable across a wide range of technical parameters, i.e., scanner vendor and model types as well as sequences, acquisition protocols, and across a varied array of pathologies that presented in line with the expected prevalence per country. This confirms that the algorithm is robust and reliable. No bias was detected in relation to technical or clinical parameters.

These findings support the feasibility of employing AI-driven strategies to optimize and reduce the use of gadolinium-based contrast agents (GBCAs) in brain MRI, aligning with current trends in radiological innovation. Mallio et al. highlighted the significant potential of artificial intelligence to minimize or even obviate the need for GBCAs; however, they underscored the importance of further validation through prospective, multicenter studies prior to widespread clinical adoption.[46] While earlier investigations, such as that of Kleesiek et al.[47], have demonstrated the capability of AI models to infer contrast-enhanced imaging features in glioma patients, the present study provides additional evidence that a clinically integrated, multicenter AI-based contrast recommendation system is not only feasible, but also scalable across a broader range of neuropathologies and clinical scenarios. Unlike lesion-specific approaches, our model enables real-time triage and protocol adaptation at the point of care, advancing the potential for personalized and sustainable imaging workflows.

However, some limitations of the study warrant consideration. First, Apollo SmartGAD operates exclusively on pre-contrast structural sequences, which do not directly reflect pathophysiological processes such as BBB breakdown or GBCA kinetics. Nevertheless, the aim of this study was not to synthetically reproduce true post-contrast imaging but

rather to predict the likelihood of enhancement based on patterns statistically correlated with contrast uptake. This is achieved through a workflow-oriented design and a large, annotated training dataset. While it is important to note that certain pathologies, such as meningitis, may present with minimal or no discernible abnormalities on pre-contrast structural imaging, both meningitis cases in our dataset were accurately identified as requiring contrast, likely due to subtle sulcal hyperintensities on FLAIR sequences. However, we acknowledge that this performance may not generalize to broader clinical populations without incorporating additional information such as clinical metadata or advanced non-contrast imaging features and further research is needed to evaluate and expand the system's robustness in these contexts. Second, the annotation of ground truth was primarily based on human-made radiology reports. Visual inspection was performed only in case a space occupying lesion was mentioned in the report without any comment regarding contrast uptake. Even though such an annotation method is relatively fast and cost-effective, in a small proportion of cases (n=5) the reports did not match the radiological findings. Therefore, a larger study with multi-reader consensus ground truth annotation is necessary. Third, for the cohort assessed for GBCA administration recommendation (Indian and Danish cohorts), the ground truth was based on true contrast images as opposed to the necessity of GBCA administration based on clinical history or physical examination, which were unavailable in the case of our datasets, and will set the basis for future prospective studies. Fourth, the current version of Apollo SmartGAD does not explicitly account for relaxivity differences among various GBCA during model training. This may affect the precision of recommendations for low- or no-contrast protocols. Although the model was developed using a diverse dataset drawn from multiple institutions, vendors, and GBCA types—potentially enabling it to learn generalizable enhancement patterns rather than agent-specific signal characteristics—future iterations could benefit from incorporating explicit metadata on GBCA type to enhance calibration accuracy. Fifth, the methodology has certain technical constraints, particularly the dependence on accurate sequence registration within the ApolloGAD and ApolloNC components. In rare instances involving severe motion or other imaging artifacts, registration may fail. While such cases were manually excluded from the current analysis, further research is warranted to evaluate the potential impact of undetected registration errors on system performance in routine clinical settings.

The reliance on the ApolloNC and PIC components could be viewed as a potential limitation of the system. However, it is important to emphasize that their integration significantly enhances the overall specificity of Apollo SmartGAD, particularly in scans without lesions, making them essential to the system's performance. ApolloNC was introduced to mitigate background false positives generated by the PIC open-source model when applied to the synthetically generated T1c sequence. As the ApolloGAD model is trained to highlight regions of high likelihood of general contrast uptake – rather than to generate visually realistic post-contrast uptake images or highlight exclusively pathological enhancement – it may produce elevated signal intensity in anatomically normal structures, such as vessels or the choroid plexus or outputs that do not visually resemble true post contrast images. Consequently, when the PIC model (originally trained on real T1c scans) is applied to these synthetic sequences, it may erroneously interpret

normal enhancement patterns as pathological findings. ApolloNC addresses this issue by selectively suppressing signal in regions associated with physiologic enhancement while preserving contrast in tumor regions, thereby improving lesion-specificity without compromising sensitivity (see Figure 2). This targeted calibration was designed to improve the system's overall performance without compromising sensitivity.

The results presented, together with the strengths and limitations of the Apollo SmartGAD contrast recommendation system, should be interpreted in the context of ongoing developments in GBCA technology. Notably, the recent approval of high-relaxivity agents such as Gadopiclenol represents an important advance, offering the potential to achieve diagnostic-quality imaging at lower doses.[48] These innovations are highly compatible with the goals of Apollo SmartGAD, which aims to optimize GBCA usage without compromising diagnostic yield. Importantly, because the model is trained on difference images—capturing the relative enhancement between pre- and post-contrast sequences—it is likely to be agnostic to the absolute relaxivity of the agent used, provided that measurable enhancement is present. In this sense, Apollo SmartGAD may remain robust across a range of current and emerging contrast agents. In particular, for patients with known gliomas, reduced-dose imaging protocols incorporating high-relaxivity agents could serve as a clinically viable alternative to standard-dose regimens, pending further clinical validation.

In practice, SmartGAD runs during acquisition on the pre-contrast series and returns a recommendation (no GBCA / low dose / standard dose) with an uncertainty score. The output can be presented to users in two ways: (i) on-console as a protocol suggestion that the technologist can accept or override before the contrast decision point, and (ii) PACS/RIS as a non-diagnostic alert attached to the study. The system will not alter images or reports; it only provides time836 sensitive protocol guidance that remains under radiologist/technologist control.

In conclusion, detecting findings during the ongoing MRI scan can support a more targeted and personalized approach, speeding up the diagnostic process, increasing the quality of healthcare and reducing expenses due to inappropriate imaging protocols.[48] Moreover, the selective use of GBCA can result in saving MR scan time which in turn could result in financial benefits for the radiology units using this sort of conditional methodology.[4] Apollo SmartGAD shows potential in reducing the use of GBCA by providing a real-time recommendation of no, low-dose, or standard GBCA injection based on a limited set of non-contrast brain MR images thus positively impacting clinical workflow.

References

1 B. Tsui et al. “Reducing Gadolinium Contrast With Artificial Intelligence”. en. In: *Journal of Magnetic Resonance Imaging* n/a (n/a). doi: 10.1002/jmri.29095.

2 J. Starekova, A. Pirasteh, and S. B. Reeder. “Update on Gadolinium Based Contrast Agent Safety, From the AJR Special Series on Contrast Media”. en. In: *AJR Am J Roentgenol* (Oct. 18, 2023). PMID: 37850581. doi: 10.2214/AJR.23.30036.

3 M. Mathur, J. R. Jones, and J. C. Weinreb. “Gadolinium Deposition and Nephrogenic Systemic Fibrosis: A Radiologist’s Primer”. In: *RadioGraphics* 40 (1 Jan. 2020). publisher: Radiological Society of North America, pp. 153–162. doi: 10.1148/rg.2020190110.

4 M. Bendszus et al. “MRI Gadolinium-Based Contrast Media: Meeting Radiological, Clinical, and Environmental Needs”. en. In: *Journal of Magnetic Resonance Imaging* n/a (n/a). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.29181. doi: 10.1002/jmri.29181.

5 J. C. Weinreb et al. “Use of Intravenous Gadolinium-based Contrast Media in Patients with Kidney Disease: Consensus Statements from the American College of Radiology and the National Kidney Foundation”. In: *Radiology* 298 (1 Jan. 1, 2021), pp. 28–35. doi: 10.1148/radiol.2020202903.

6 I. J. H. G. Wamelink et al. “Brain tumor imaging without gadolinium-based contrast agents: Feasible or fantasy?” en. In: *Radiology* 310 (2 Feb. 2024), e230793. doi: 10.1148/radiol.230793.

7 R. J. McDonald et al. “Gadolinium Retention: A Research Roadmap from the 2018 NIH/ACR/RSNA Workshop on Gadolinium Chelates”. en. In: *Radiology* 289 (2 Nov. 2018), pp. 517–534. doi: 10.1148/radiol.2018181151.

8 I. J. H. G. Wamelink et al. “The patients’ experience of neuroimaging of primary brain tumors: a cross-sectional survey study”. en. In: *J Neurooncol* 162 (2 Apr. 1, 2023), pp. 307–315. doi: 10.1007/s11060-023-04290-x.

9 K. Inoue et al. “Impact on gadolinium anomaly in river waters in Tokyo related to the increased number of MRI devices in use”. In: *Marine Pollution Bulletin* 154 (May 1, 2020), p. 111148. doi: 10.1016/j.marpolbul.2020.111148.

10 *SubtleGAD™*.

11 C. J. Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *The Lancet Digital Health* 3 (12 Dec. 1, 2021). publisher: Elsevier PMID: 34688602, e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

12 J. Cheng et al. “Prediction of Glioma Grade Using Intratumoral and Peritumoral Radiomic Features From Multiparametric MRI Images”. en. In: *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19 (2 Apr. 1, 2022), pp. 1084–1095. doi: 10.1109/TCBB.2020.3033538.

13 U. Baid et al. “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification”. In: *arXiv [cs.CV]* (July 5, 2021).

14 D. LaBella et al. “The ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge 2023: Intracranial Meningioma”. arXiv:2305.07642 [cs, stat]. May 12, 2023.

15 A. W. Moawad et al. “The Brain Tumor Segmentation (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI”. arXiv:2306.00838 [eess, q-bio]. June 1, 2023.

16 F. G. Gonçalves, S. D. Serai, and G. Zuccoli. “Synthetic Brain MRI: Review of Current Concepts and Future Directions”. en. In: *Top Magn Reson Imaging* 27 (6 Dec. 2018). PMID: 30516691, pp. 387–393. doi: 10.1097/RMR.0000000000000189.

17 M. Hoffmann et al. “SynthMorph: Learning Contrast-Invariant Registration Without Acquired Images”. In: *IEEE Transactions on Medical Imaging* 41 (3 Mar. 2022). event-title: IEEE Transactions on Medical Imaging, pp. 543–558. doi: 10.1109/TMI.2021.3116879.

18 *picture-production / picture-nnunet-package · GitLab*. en. May 27, 2024.

19 F. Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. en. In: *Nat. Methods* 18 (2 Feb. 2021), pp. 203–211. doi: 10.1038/s41592-020-01008-z.

20 H. G. Pemberton et al. “Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18911. doi: 10.1038/s41598-023-44794-0.

21 S. Bakas et al. “The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics”. en. In: *Sci Data* 9 (1 July 29, 2022). publisher: Nature Publishing Group, p. 453. doi: 10.1038/s41597-022-01560-7.

22 E. Calabrese et al. “The university of California San Francisco preoperative diffuse glioma MRI dataset”. en. In: *Radiol. Artif. Intell.* 4 (6 Nov. 2022), e220058. doi: 10.1148/ryai.220058.

23 E. Filimonova et al. “Utilizing the Amide Proton Transfer Technique to Characterize Diffuse Gliomas Based on the WHO 2021 Classification of CNS Tumors”. en. page: 2023.08.24.23294427. Aug. 25, 2023.

24 K. Hwang et al. “Fully Automated Segmentation Models of Supratentorial Meningiomas Assisted by Inclusion of Normal Brain Images”. In: *J Imaging* 8 (12 Dec. 15, 2022). PMID: 36547492 PMCID: PMC9782766, p. 327. doi: \hfill\break10.3390/jimaging8120327.

25 M. V. Sagar et al. “COVID-19-associated cerebral microbleeds in the general population”. In: *Brain Communications* 6 (3 June 1, 2024), fcae127. doi: 10.1093/braincomms/fcae127.

26 S. Challa. “Granulomatous diseases of the central nervous system: Approach to diagnosis”. en. In: *Indian J. Pathol. Microbiol.* 65 (Supplement May 2022), S125–S134. doi: 10.4103/ijpm.ijpm_1067_21.

27 A. Dasgupta, T. Gupta, and R. Jalali. “Indian data on central nervous tumors: A summary of published work”. en. In: *South Asian J. Cancer* 5 (3 July 2016), pp. 147–153. doi: 10.4103/2278-330X.187589.

28 S. P. Jones et al. “Stroke in India: A systematic review of the incidence, prevalence, and case fatality”. en. In: *Int. J. Stroke* 17 (2 Feb. 2022), pp. 132–140. doi: 10.1177/17474930211027834.

29 S. Kamalakannan et al. “Incidence & prevalence of stroke in India: A systematic review: A systematic review”. en. In: *Indian J. Med. Res.* 146 (2 Aug. 2017), pp. 175–185. doi: 10.4103/ijmr.IJMR_516_15.

30 S. M. Hald et al. “Trends in incidence of intracerebral hemorrhage and association with antithrombotic drug use in Denmark, 2005-2018”. en. In: *JAMA Netw. Open* 4 (5 May 3, 2021), e218380. doi: 10.1001/jamanetworkopen.2021.8380.

31 S. V. Vestergaard et al. “Occurrence, mortality and cost of brain disorders in Denmark: a population-based cohort study”. en. In: *BMJ Open* 10 (11 Nov. 17, 2020), e037564. doi: 10.1136/bmjopen-2020-037564.

32 B. K. Rasmussen et al. “Epidemiology of glioma: clinical characteristics, symptoms, and predictors of glioma patients grade I-IV in the the Danish Neuro-Oncology Registry”. en. In: *J. Neurooncol.* 135 (3 Dec. 2017), pp. 571–579. doi: 10.1007/s11060-017-2607-5.

33 O. Ronneberger, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv [cs.CV]* (May 2015).

34 A. Hoopes et al. “SynthStrip: skull-stripping for any brain image”. In: *NeuroImage* 260 (Oct. 15, 2022), p. 119474. doi: 10.1016/j.neuroimage.2022.119474.

35 B. Billot et al. “Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets”. In: *Proceedings of the National Academy of Sciences* 120 (9 Feb. 28, 2023). publisher: Proceedings of the National Academy of Sciences, e2216399120. doi: 10.1073/pnas.2216399120.

36 F. Farnia and A. Ozdaglar. “Do GANs always have Nash equilibria?” en. In: *Proceedings of the 37th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, Nov. 21, 2020, pp. 3029–3039.

37 BBillot. *SynthSeg*. original-date: 2020-04-21T00:28:42Z. Sept. 12, 2023.

38 E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson. “Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach”. In: *Biometrics* 44 (3 1988). publisher: [Wiley, International Biometric Society], pp. 837–845. doi: 10.2307/2531595.

39 J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. DOI: 10.4324/9780203771587. New York: Routledge, July 1, 1988. 567 pp.

40 V. Gulani et al. “Gadolinium deposition in the brain: summary of evidence and recommendations”. en. In: *Lancet Neurol.* 16 (7 July 2017), pp. 564–570. doi: 10.1016/S1474-4422(17)30158-8.

41 *Revolutionizing Brain MRI Acquisition • APPLIED RADIOLOGY*.

42 J. Boto et al. “Is Contrast Medium Really Needed for Follow-up MRI of Untreated Intracranial Meningiomas?” In: *AJNR Am J Neuroradiol* 42 (8 Aug. 2021). PMID: 34117017 PMCID: PMC8367609, pp. 1421–1428. doi: 10.3174/ajnr.A7170.

43 S. Sathraju et al. “Reducing Gadolinium Exposure in Patients Undergoing Monitoring for Meningiomas”. In: *Cureus* 15 (4). PMID: 37187666 PMCID: PMC10180544, e37492. doi: 10.7759/cureus.37492.

44 T. Dowers et al. “Comparative Evaluation of Lower Gadolinium Doses for MR Imaging of Meningiomas: How Low Can We Go?” en. In: *American Journal of Neuroradiology* (June 10, 2024). publisher: American Journal of Neuroradiology section: ORIGINAL RESEARCH. doi: 10.3174/ajnr.A8375.

45 B. M. Ellingson et al. “Baseline pretreatment contrast enhancing tumor volume including central necrosis is a prognostic factor in recurrent glioblastoma: evidence from single and multicenter trials”. In: *Neuro Oncol* 19 (1 Jan. 2017). PMID: 27580889 PMCID: PMC5193027, pp. 89–98. doi: 10.1093/neuonc/now187.

46 C. A. Mallio et al. “Artificial Intelligence to Reduce or Eliminate the Need for Gadolinium-Based Contrast Agents in Brain and Cardiac MRI: A Literature Review”. en. In: *Investigative Radiology* 58 (10 Oct. 2023), p. 746. doi: 10.1097/RLI.0000000000000983.

47 J. Kleesiek et al. “Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?: A Feasibility Study”. en. In: *Invest. Radiol.* 54 (10 Oct. 2019), pp. 653–660.

48 C. Kuhl et al. “Efficacy and Safety of Half-Dose Gadopiclenol versus Full-Dose Gadobutrol for Contrast-enhanced Body MRI”. In: *Radiology* 308 (1 July 2023). publisher: Radiological Society of North America, e222612. doi: 10.1148/radiol.222612.

Supplementary Material

Table S1: Scanner vendors & sequence parameters overview for evaluation data.

	Netherlands (n = 432)	**India (n = 350)**	**Denmark (n = 469)**	**Total (n = 1251)**
Scanner Vendor	Number of, %			
Siemens	150, 34.7%	144, 41.1%	7, 1.5%	301, 24.1%
Phillips	107, 24.8%	32, 9.1%	451, 96.2%	590, 47.2%
GE	167, 38.7%	158, 45.1%	11, 2.3%	336, 26.9%
Toshiba	8, 1.9%	0, 0.0%	0, 0.0%	8, 0.6%
Hitachi	0, 0.0%	16, 4.6%	0, 0.0%	16, 1.3%
Magnetic Field Strength	Number of, %			
0.3 T	0, 0.0%	22, 6.3%	0, 0.0%	22, 1.8%
1.0 T	0, 0.0%	9, 2.6%	7, 1.5%	16, 1.3%
1.5 T	178, 41.2%	314, 89.7%	343, 73.8%	835, 67.0%
3.0 T	254, 58.8%	5, 1.4%	115, 24.7%	374, 30.0%
Sequences				
T2 FLAIR				
Acquisition Type (3D)	354, 81.9%	0, 0.0%	11, 2.4%	365, 29.4%
Slice thickness	1.83 $\pm$ 1.39	5.11 $\pm$ 0.37	4.36 $\pm$ 0.71	3.70 $\pm$ 0.95
Slice spacing	2.07 $\pm$ 1.99	6.62 $\pm$ 0.52	5.30 $\pm$ 0.90	4.55 $\pm$ 1.33
T2w				
Acquisition Type (3D)	1, 0.2%	0, 0.0%	0, 0.0%	1, 0.08%
Slice thickness	4.88 $\pm$ 0.54	5.11 $\pm$ 0.38	4.19 $\pm$ 0.42	4.69 $\pm$ 0.45
Slice spacing	5.11 $\pm$ 0.64	6.63 $\pm$ 0.56	5.16 $\pm$ 0.50	5.55 $\pm$ 0.57
T1w				
Acquisition Type (3D)	56, 13.0%	58, 17.0%	79, 17.0%	193, 15.6%
Slice thickness	4.45 $\pm$ 1.31	4.59 $\pm$ 1.26	3.69 $\pm$ 1.16	4.20 $\pm$ 1.24
Slice spacing	4.94 $\pm$ 1.01	5.81 $\pm$ 2.07	4.44 $\pm$ 1.54	5.00 $\pm$ 1.56
T1w +Gad				
Acquisition Type (3D)	418, 96.8%	116, 34.0%	458, 97.9%	992, 79.9%
Slice thickness	1.15 $\pm$ 0.72	3.79 $\pm$ 1.89	1.76 $\pm$ 0.56	2.12 $\pm$ 1.14
Slice spacing	1.26 $\pm$ 1.06	5.21 $\pm$ 2.48	1.03 $\pm$ 0.74	2.28 $\pm$ 1.52

Table S2: Tumor types and primary tumor origin for brain metastasis as determined by means of surgery or biopsy.

Glioma type	Grade	Number of patients
Dutch Dataset		
Glioblastoma (IDH-wt)	IV	259
Astrocytoma (IDH-mutated, not 1p/19q codeleted)	IV	5
Astrocytoma (IDH-mutated, not 1p/19q codeleted)	III	18
Astrocytoma (IDH-mutated, not 1p/19q codeleted)	II	40
Oligodendroglioma (IDH-mutated, 1p/19q codeleted)	III	16
Oligodendroglioma (IDH-mutated, 1p/19q codeleted)	II	28
Opticus glioma	I	1
Pilocytic astrocytoma	I	1
Isomorphic diffuse glioma	I	1
Metastasis type		**Number of patients**
Melanoma		20
Mamma Carcinoma		1
Adenocarcinoma mamma		3
Adenocarcinoma pancreas		2
Small-cell lung carcinoma		1
Adenocarcinoma lung		8
Adenocarcinoma NSCLC		19
Lung or mamma carcinoma		1
Adenocarcinoma colorectal		1
Squamous cell NSCLC		1
LCNEC lung		3
Adenocarcinoma non-small cell sinonasal carcinoma		1
Adenocarcinoma prostate		1
Large cell esophageal carcinoma		1
Hypopharynx carcinoma		1
Adenocarcinoma stomach		1
Neuroendocrine tumor		1
Clear cell renal carcinoma		1

Table S3: Scanner vendors & sequence parameters overview for training data (Part 1 of 2). Categorical variables are expressed as n, percentage (%). Continuous variables are expressed as mean ± standard deviation.

	United States 1 (n=178)	United States 2 (n=6)	Brazil (n=800)	Israel (n=781)
Scanner Vendor				
Siemens	84, 47.2%	5, 83.3%	0, 0.0%	0, 0.0%
Phillips	84, 47.2%	0, 0.0%	558, 69.8%	781, 100.0%
GE	9, 5.1%	1, 16.7%	242, 30.2%	0, 0.0%
Fujifilm	1, 0.6%	0, 0.0%	0, 0.0%	0, 0.0%
Magnetic Field Strength				
1.0 T	0, 0.0%	0, 0.0%	0, 0.0%	0, 0.0%
1.2 T	1, 0.6%	0, 0.0%	2, 0.2%	0, 0.0%
1.5 T	77, 43.3%	4, 66.7%	0, 0.0%	420, 53.8%
3.0 T	100, 56.2%	2, 33.3%	798, 99.8%	361, 46.2%
Sequences				
T2 FLAIR				
Acquisition Type (3D)	0, 0.0%	0, 0.0%	25, 3.1%	15, 1.9%
Slice thickness	4.4 ± 0.7	4.7 ± 0.5	4.9 ± 0.7	4.5 ± 0.7
Slice spacing	5.7 ± 0.9	5.4 ± 0.6	5.7 ± 1.0	5.4 ± 0.8
T2w				
Acquisition Type (3D)	17, 9.6%	0, 0.0%	0, 0.0%	0, 0.0%
Slice thickness	3.7 ± 1.2	4.5 ± 0.5	5.1 ± 0.2	4.5 ± 0.5
Slice spacing	4.8 ± 1.4	0.2 ± 0.5	5.8 ± 0.5	5.5 ± 0.5
T1w				
Acquisition Type (3D)	49, 27.8%	1, 16.7%	19, 2.4%	0, 0.0%
Slice thickness	3.6 ± 1.7	4.2 ± 1.5	5.0 ± 0.5	4.5 ± 0.5
Slice spacing	5.9 ± 1.0	5.6 ± 0.4	5.7 ± 0.8	5.5 ± 0.5
T1w +Gad				
Acquisition Type (3D)	85, 53.8%	1, 16.7%	629, 79.3%	358, 45.8%
Slice thickness	2.7 ± 1.8	4.2 ± 1.5	2.1 ± 1.6	3.2 ± 2.0
Slice spacing	4.4 ± 2.5	5.6 ± 0.5	1.9 ± 2.0	3.7 ± 2.5

Table S3: Scanner vendors & sequence parameters overview for training data (Part 2 of 2). Categorical variables are expressed as n, percentage (%). Continuous variables are expressed as mean ± standard deviation.

	India 1 (n=202)	India 2 (n=158)	Ukraine (n=22)	Denmark (n=451)	Total (n=2598)
Scanner Vendor					
Siemens	202, 100.0%	8, 5.1%	2, 9.1%	23, 5.1%	324, 12.5%
Phillips	0, 0.0%	1, 0.6%	20, 90.9%	412, 91.3%	1856, 71.4%
GE	0, 0.0%	149, 94.3%	0, 0.0%	16, 3.6%	417, 16.1%
Fujifilm	0, 0.0%	0, 0.0%	0, 0.0%	0, 0.0%	1, 0.0%
Magnetic Field Strength					
1.0 T	0, 0.0%	0, 0.0%	0, 0.0%	0, 0.0%	2, 0.1%
1.2 T	0, 0.0%	0, 0.0%	0, 0.0%	0, 0.0%	1, 0.0%
1.5 T	202, 100.0%	157, 99.4%	22, 100.0%	98, 21.8%	1778, 68.5%
3.0 T	0, 0.0%	1, 0.6%	0, 0.0%	352, 78.2%	816, 31.4%
Sequences					
T2 FLAIR					
Acquisition Type (3D)	0, 0.0%	9, 15.0%	12, 54.5%	176, 39.0%	237, 9.5%
Slice thickness	5.0 ± 0.0	4.7 ± 1.3	1.0 ± 0.0	2.7 ± 1.4	4.3 ± 1.2
Slice spacing	6.1 ± 0.3	5.9 ± 2.2	2.5 ± 1.9	2.9 ± 2.0	5.1 ± 1.6
T2w					
Acquisition Type (3D)	0, 0.0%	0, 0.0%	0, 0.0%	1, 0.2%	18, 0.7%
Slice thickness	5.0 ± 0.0	4.9 ± 0.2	5.0 ± 0.0	3.7 ± 0.6	4.5 ± 0.7
Slice spacing	6.1 ± 0.2	6.9 ± 0.7	5.1 ± 0.4	4.5 ± 1.0	5.5 ± 0.9
T1w					
Acquisition Type (3D)	0, 0.0%	1, 1.7%	21, 95.4%	376, 83.4%	467, 18.7%
Slice thickness	5.0 ± 0.2	4.9 ± 0.5	3.0 ± 2.0	1.8 ± 1.1	4.1 ± 1.4
Slice spacing	6.2 ± 0.6	6.8 ± 0.7	1.1 ± 1.2	1.6 ± 4.4	4.9 ± 1.9
T1w +Gad					
Acquisition Type (3D)	45, 22.3%	8, 13.3%	22, 100.0%	443, 98.2%	1591, 64.3%
Slice thickness	4.1 ± 1.6	4.5 ± 1.3	1.0 ± 0.0	1.3 ± 0.6	2.6 ± 1.8
Slice spacing	6.3 ± 0.6	6.6 ± 1.2	0.8 ± 0.0	1.0 ± 0.6	2.9 ± 2.5

Table S4: Demographics and pathology distributions per training cohort (Part 1 of 2). Continuous variables are expressed as mean ± standard deviation. Categorical variables are expressed as n, percentage (%).

	US 1 (n=178)	US 2 (n=6)	Brazil (n=800)	Israel (n=781)	India 1 (n=202)
Age, yrs	44.7 ± 20.5	65.2 ± 9.9	56.9 ± 18.4	53.7 ± 16.8	43.6 ± 19.0
Sex, females	81, 50.3%	3, 50.0%	427, 53.4%	405, 51.9%	102, 55.4%
Normal	34	0	7	200	26
Brain Tumor	60	6	310	231	97
Cyst	7	0	18	7	10
Meningioma	16	5	67	20	20
Pituitary adenoma	2	0	4	0	12
Schwannoma	1	1	20	1	13
Glioma	10	0	32	0	21
Metastases	13	0	51	0	9
Other/Unspeci	11	0	118	203	12
Infectious disorders	3	0	51	2	14
Vascular pathology	65	0	508	308	16
Acute infarct	31	0	221	199	3
Hemorrhage	34	0	287	109	13
Other pathologies	108	4	652	39	142

Table S4: Demographics and pathology distributions per training cohort (Part 2 of 2). Continuous variables are expressed as mean ± standard deviation. Categorical variables are expressed as n, percentage (%).

	India 2 (n=158)	Ukraine (n=22)	Denmark (n=451)	Total (n=2598)
Age, yrs	NA	NA	60.5 ± 17.0	53.8 ± 18.3
Sex, females	NA	NA	NA	1018, 52.7%
Normal	7	0	128	402
Brain Tumor	108	23	150	985
Cyst	8	0	18	68
Meningioma	6	19	40	193
Pituitary adenoma	5	1	7	31
Schwannoma	5	3	4	48
Glioma	38	0	17	118
Metastases	14	0	32	119
Other/Unspecified	32	0	32	408
Infectious disorders	19	0	3	92
Vascular pathology	22	0	87	1006
Acute infarct	7	0	26	487
Hemorrhage	15	0	61	519
Other pathologies	90	8	244	1287

Table S5: Tumor classification performance of Apollo Brain. Classification performance metrics for the Apollo Brain sub-module on Denmark (n=469), India (n=350) and Pooled (Denmark + India, n=819) data cohorts. Apollo Brain was measured with regards to its ability to predict tumors (including metastases) and infections (n=158 in the Denmark cohort, n=123 in the India cohort, and n=281 in the Pooled cohort). Acronyms: positive predictive value (PPV), negative predictive value (NPV), false positive rate (FPR), false discovery rate (FDR), false negative rate (FNR).

Metric	Denmark (n=469)	India (n=350)	Pooled (n=819)
Sensitivity	0.84 (0.78, 0.89)	0.65 (0.56, 0.73)	0.76 (0.71, 0.80)
Specificity	0.56 (0.51, 0.62)	0.90 (0.85, 0.93)	0.70 (0.67, 0.74)
PPV	0.49 (0.44, 0.55)	0.78 (0.69, 0.85)	0.57 (0.52, 0.62)
NPV	0.88 (0.82, 0.91)	0.83 (0.77, 0.87)	0.85 (0.81, 0.88)
FPR	0.44 (0.38, 0.49)	0.10 (0.07, 0.15)	0.30 (0.26, 0.34)
FDR	0.51 (0.45, 0.57)	0.22 (0.15, 0.31)	0.43 (0.38, 0.48)
FNR	0.16 (0.11, 0.22)	0.35 (0.27, 0.44)	0.24 (0.12, 0.23)

Table S6: Enhancing lesion classification performance of PIC. Enhancing lesion classification performance metrics for the PIC sub-module on Denmark (n=469), India (n=350) and Pooled (Denmark + India, n=819) data cohorts. PIC was measured with regards to its ability to predict at least 1 voxel of contrast enhancement within an image with any contrast enhancing lesion (n=156 in the Denmark cohort, n=100 in the India cohort, and n=256 in the Pooled cohort). Acronyms: positive predictive value (PPV), negative predictive value (NPV), false positive rate (FPR), false discovery rate (FDR), false negative rate (FNR).

Metric	Denmark (n=469)	India (n=350)	Pooled (n=819)
Sensitivity	0.42 (0.35, 0.50)	0.67 (0.57, 0.75)	0.52 (0.46, 0.58)
Specificity	0.81 (0.76, 0.85)	0.66 (0.60, 0.72)	0.74 (0.70, 0.78)
PPV	0.52 (0.43, 0.61)	0.44 (0.36, 0.52)	0.48 (0.42, 0.54)
NPV	0.74 (0.69, 0.78)	0.83 (0.76, 0.88)	0.77 (0.74, 0.81)
FPR	0.12 (0.16, 0.24)	0.34 (0.28, 0.40)	0.26 (0.23, 0.30)
FDR	0.48 (0.40, 0.57)	0.56 (0.48, 0.64)	0.52 (0.47, 0.58)
FNR	0.58 (0.50, 0.65)	0.33 (0.25, 0.43)	0.48 (0.42, 0.54)

Chapter 10

Summary and Discussion

Summary

Despite the continued reliance on CET1 imaging for glioma diagnosis, treatment monitoring, and disease follow-up, growing concerns about the health risks, environmental impact, and financial costs of repeated gadolinium-based contrast agent (GBCA) use have prompted the search for alternative imaging biomarkers. Moreover, contrast enhancement alone appears insufficient to fully capture viable tumor tissue, underscoring the need for novel biomarkers that can better influence patient outcomes.

This thesis, conducted within the GLIOCARE project, explores two complementary pathways to reduce GBCA dependence: 1) advanced MRI techniques, specifically amide proton transfer chemical exchange saturation transfer (APT-CEST) and arterial spin labeling (ASL), which provide non-invasive insights into tumor biochemistry and physiology, and 2) artificial intelligence (AI), which leverages large imaging datasets to extract subtle features and enable GBCA-free or reduced-GBCA imaging workflows.

The PENGUIN survey study was introduced in **Chapter 2** to assess patient attitudes towards GBCA use and MRI follow-up. The study revealed that while patients trust current protocols, there is a preference for GBCA-free imaging if diagnostic accuracy is maintained. The study also highlights that demographic and clinical factors such as sex, age, and tumor grade influence patient MRI experiences, suggesting a need for more personalized and patient-centered imaging approaches. We also found that patients were positive regarding the current clinical follow-up frequency. This finding should be considered if follow-up frequency is decreased.

A narrative literature review was performed in **Chapter 3** to identify the available GBCA reduction research and was categorized into 1) guidelines, 2) advanced imaging, and 3) AI approaches. Despite the growing interest towards GBCA reduction, most studies were retrospective. Key insights from our findings include the feasibility of universal or subgroup-targeted GBCA reduction, the potential for clinically driven personalized imaging intervals, that GBCA-free radiomic features can match or exceed CET1 derived biomarkers in diagnostic performance, and that AI and advanced imaging based solutions are appropriate tools toward GBCA reduction. In this study we also found that ASL is a promising GBCA-free perfusion technique comparable to DSC. AI methods show potential in synthesizing contrast-like images from non-contrast inputs and achieve comparable diagnostic accuracy to AI methods that use CET1 images.

Regarding advanced MRI, APT-CEST shows promise for non-invasive glioma grading and differentiation between tumor progression and treatment effects. **Chapter 4** demonstrates that a clinically feasible whole-brain APT-CEST sequence at 3 Tesla exhibits good short- and long-term reproducibility in both healthy volunteers and glioma patients, with tumor-to-normal tissue contrast exceeding scan-rescan variability. We found low reproducibility in the orbitofrontal gyri, likely due to B_0 inhomogeneity and areas with large veins which is consistent with literature. Six glioma patients were included in this reproducibility study and we found higher MTR_{asym} values in the tumor ROIs (mean 1.59 ± 0.67) than in the contralateral ROI (mean 0.50 ± 0.46). These results

motivated the further inclusion of prospective patients (N=100) in the GLIOCARE project.

Descriptive and Radiomics features were extracted from these prospective GLIOCARE patients and used to predict IDH status, 1p/19q status, and tumor grade in **Chapter 5**. In this analysis, 48 GLIOCARE patients, 42 Russian patients, and 129 Chinese patients were included. 52/100 GLIOCARE patients were excluded due to missing molecular information, non-gliomas, or post-operative status. We found that MTR_{asym} univariate and Radiomics features were consistently associated with molecular subtypes. Similar to Chapter 4, we found higher MTR_{asym} values in the tumor ROIs than in the contralateral ROI. We found that tumor CBF was only increased in glioblastoma patients. Combining CBF features with MTR_{asym} features improved the predictive performance, though minimally. MTR_{asym} and CBF values differed significantly between datasets. APT cross-vendor generalizability was poor while same vendor APT had good generalizability. Tissue normalization by contralateral ROI removed cross-dataset differences.

The creation of the IMAGO (IMAging in GliOma) database in **Chapter 6** provides additional resources for training and validation of AI algorithms. We developed a large in-house database of $\sim$1700 patients and an open-access version with 500 patients (available through the Health-RI XNAT platform and supported by the EUCAIM 2024 open call). Both datasets are continuously being expanded with new patients and additional clinical and histological data. The in-house database is currently being expanded with additional histological data and radiotherapy scans and maps. All 500 patients in the online database are diagnosed with adult-type diffuse gliomas according to the WHO CNS5 (2021) criteria. We found that patient survival data and characteristics are similar as those reported elsewhere in the literature.

In **Chapter 7**, 100 glioblastoma IDH-wildtype patients from the IMAGO database are segmented by four raters using only non-contrast sequences. An AI algorithm that also only used non-contrast sequences was developed and an additional rater finetuned the segmentation produced by the AI algorithm. We found that human raters could segment the tumor core with moderate accuracy. The segmentations of the AI-assisted rater were significantly better in Dice, Hausdorff distance, and segmentation volume error. We also found that segmentation error decreased per case, which suggests that raters learn over time. There was no correlation between Dice scores and segmentation time, indicating that spending more time did not result in improved segmentations.

Chapter 3 also highlighted limitations in current AI studies. First, most AI GBCA reduction studies focus on technical improvements rather than clinical applicability, and second, most AI manuscripts are limited to small sample sizes. **Chapter 8** addresses these two shortcomings in an AI benchmark study. In this study we validated peer-reviewed GBCA-free segmentation and synthesis algorithms on a large open-source dataset. This benchmark study focused on three clinical tasks: 1) enhancement detection; 2) prognostic performance; and 3) therapy response evaluation. We were able to include six segmentation and four synthesis algorithms and found that segmentation algorithms consistently outperform synthesis-based algorithms in all three clinical tasks. Segmenta-

tion algorithms showed significantly higher sensitivity, NPV, and accuracy on one of the test datasets and a trend on the other test dataset. Intraclass correlation coeficient with the ground truth was poor for the synthesis algorithms and moderate-to-good for the segmentation algorithms. Segmentation-derived enhancement volumes were prognostic for survival, and including age, sex, and surgery type improved the c-index from 0.60 to 0.68. While segmentation algorithms were better than synthesis algorithms, neither are currently ready for clinical integration.

The AI pipeline developed in **Chapter 9**, Apollo SmartGAD, is a clinically oriented triage tool that categorizes patients based on the necessity of GBCA administration (none, low-dose, or standard-dose). While demonstrating promising sensitivity and specificity, the tool still misclassifies some cases, particularly small lesions and typically enhancing tumors like metastases and meningiomas, indicating that it should serve as an assisting tool rather than a replacement for current protocols. We also found that the presence of any contrast enhancement on the true contrast as well as the synthetic contrast images, is a good predictor of possible HGGs (80% and 85% vs 77 and 78% sensitivity and specificity respectively).

Conclusion

This thesis has contributed to the field of glioma research, by investigating APT-CEST, ASL, and AI techniques to reduce the usage of GBCA. The findings in this thesis highlight that APT-CEST and to a lesser extent ASL may be used to find novel clinical biomarkers to characterize glioma tumors. The findings also highlight that AI tools are not yet ready for direct clinical implementation and that their development should focus more on clinical applicability than on technological improvements. Future studies should explore whether tissue normalization could sufficiently mitigate APT-CEST vendor variability and whether APT-CEST can better delineate the true extent of the tumor than CET1 imaging. Designing AI studies around the clinical workflow will not only have clinical impact, but will also have a positive impact on patient care.

Discussion

Contrast enhanced T1 images remain a cornerstone of glioma diagnosis, treatment evaluation, and disease monitoring. However, concerns about health, the environment, and the costs associated with repeated GBCA administration have become increasingly relevant.[1–3] It also appears that contrast enhancement is an inadequate marker for identifying viable tumor in its entirety, and that novel biomarkers are needed to improve survival. To address these limitations and reduce reliance on GBCA-based imaging, new imaging strategies and analytical tools are required.
This thesis is part of the GLIOCARE project, which aims to reduce GBCA use by evaluating two approaches: 1) advanced MRI techniques, specifically amide proton transfer chemical exchange saturation transfer (APT-CEST) and arterial spin labeling (ASL), which non-invasively measure tumor biochemistry and physiology, and 2) artificial intelligence, whose capacity to learn from large imaging datasets and extract subtle image features can facilitate the transition from GBCA-dependent workflows to reduced- or GBCA-free imaging protocols.

This discussion presents the two approaches separately and concludes with suggestions for future directions.

Patient Perspectives and Emerging Approaches to Reducing GBCA Use in Glioma MRI

As a first step, this thesis investigates patient perspectives in the Netherlands regarding GBCA use and neuro-oncological MRI follow-up as part of the PENGUIN survey (**Chapter 2**). Patients expressed overall confidence in the current MRI follow-up regimen but preferred a GBCA-free MRI protocol, provided diagnostic accuracy remains equivalent. The motivation to reduce GBCA use is backed by patient interest, aligning with the broader movement towards safer, patient-centered imaging. Beyond the patient's GBCA experience, **Chapter 2** also investigates the general MRI experience of patients with glioma. They reported positive attitudes towards the current clinical follow-up frequency, which is relevant to prolonged and more personalized scan intervals. While reducing imaging frequency could improve resource allocation, it may initially increase anxiety for some patients.[4, 5] Given the overall trust patients place in clinical expertise, such anxiety will likely diminish once new evidence-based follow-up protocols prove reliable. The PENGUIN survey also revealed that demographic and clinical factors such as sex, age, and tumor grade influence the MRI experience.[6] These findings highlight the need for a more personalized and patient-sensitive scanning process, as improved comfort could reduce motion artifacts and yield higher-quality images.[7]

To understand existing approaches towards GBCA reduction I performed a narrative literature review in **Chapter 3**. The research field was divided into three domains: 1) guidelines and standard radiologic evaluation; 2) advanced imaging techniques; and 3) artificial intelligence Approaches. Most GBCA reduction studies are retrospective, yet they still provide valuable insights. These insights are:
a. that GBCA reduction can be pursued universally or targeted to specific subgroups, such as patients with non-enhancing postoperative gliomas;[8]

b. that follow-up imaging can be clinically driven rather than using fixed time intervals, moving more towards personalized treatment;
c. that combinations of GBCA-free radiomic features generally outperform or are noninferior to CET1 in diagnostic biomarker machine learning studies;[9, 10] and
d. that evidence suggests that advanced imaging techniques and AI-based solutions could lead clinical practice toward reduced GBCA dependence.
ASL provides GBCA-free perfusion metrics that match GBCA-dependent DSC for grading and treatment surveillance and a more recent study reported similar findings.[11] AI methods can synthesize visually plausible post-contrast images from non-contrast/low-dose inputs or achieve comparable diagnostic performance using GBCA-free sequences.[12–14] ASL still lacks widespread adoption due to differences in optimal cutoff values across vendors and sites. Targeted multicenter, multivendor validation studies could help build the necessary standardization.[15, 16]

Findings of Advanced MRI

Although APT-CEST is not yet implemented clinically, its potential to advance glioma imaging is promising: reported applications include non-invasive glioma grading, differentiation between true progression and radiation necrosis, and distinction between tumor and edema.[17–19] These use cases often assume that viable tumor tissue corresponds to the contrast-enhancing region, even though histopathological and stereotactic biopsy studies have repeatedly demonstrated extensive tumor cell infiltration in the surrounding non-enhancing FLAIR-hyperintense regions.[20, 21] Surgical studies using a FLAIR-based resection approach (also referred to as “FLAIRectomy”) reported improved prognosis, suggesting that viable tumor extends beyond what is captured by CE imaging.[22, 23] During a FLAIRectomy, tumor delineation is based on the T2 FLAIR hyperintense areas either resected in their entirety or in part. This inherently also may include edematous areas that are virtually tumor-free, as both regions, the non-enhancing tumor part and edema are visually not discernible. APT-CEST may help quantify viable tumor components near the margins as studies suggest that APT-CEST can better differentiate viable tumor from edema.[24, 25] Before APT-CEST can be clinically used to derive quantitative biomarkers, reproducibility in healthy and tumorous tissue must be established.
Chapter 4, investigates the short- and long-term reproducibility of a clinically feasible whole-brain APT-CEST sequence at 3 Tesla. Sufficiently high short- and long-term APT-CEST reproducibility in healthy volunteers and in glioma patients was observed, and the tumor-to-normal tissue contrast exceeded APT-CEST scan-rescan variability. Other studies found similar results of within- and between-session differences.[26, 27] The lowest reproducibility was found in the orbitofrontal gyri, likely due to B_0 inhomogeneity, a known limitation near air cavities.[28] Although the number of glioma patients in Chapter 4 was limited (n=6), APT-CEST signal intensity was higher in the tumor ROI (mean 1.59 ± 0.67) than in the contralateral ROI (mean 0.50 ± 0.46). This motivated the continued prospective acquisition of APT-CEST data, ultimately yielding a larger database of 100 glioma cases. In addition to APT-CEST, these patients also underwent ASL imaging.

In **Chapter 5**, descriptive and radiomic features were extracted from APT-CEST MTR_{asym} and ASL CBF maps of these 48 glioma patients and used to classify molecular glioma

subtypes (IDH wild-type vs. mutated, 1p/19q non-codeleted vs. codeleted, and WHO grades 2, 3, and 4). Alongside the prospectively scanned GLIOCARE patients, patients from Russian and Chinese datasets were also included. APT-CEST univariate analyses and radiomics features seemed best suited for predicting glioma subtypes. MTR_{asym} features were significantly associated with the molecular subtypes. Combining CBF and MTR_{asym} may improve predictive performance. However, we found significantly different MTR_{asym} and CBF values between the datasets, and cross-vendor generalizability was poor. Poor generalizability may originate from scanner-dependent factors such as B_0/B_1 inhomogeneities, labeling efficiency, and sequence implementation.
Mean absolute CBF and MTRasym values in the tumour differed between datasets likely due to a general intensity bias. Tissue normalization to contralateral ROI improved the between dataset generalizability and the values after normalization showed no significant differences between datasets anymore. After normalization, CBF had comparable univariate diagnostic performance to MTR_{asym} for IDH mutation and 1p/19q codeletion prediction. Thus appropriate normalization could account for inter-subject variability. In particular, normalization may reduce variability related to factors that influence absolute CBF values, such as labeling efficiency and physiological differences in baseline CBF, thereby improving both inter-subject comparability and reproducibility across scanners. However, normalization cannot fully correct for all between-center differences. For example, discrepancies in spatial resolution remain: ASL voxels are relatively large, whereas the enhancing tumor rim can be relatively thin. This mismatch means that partial volume effects within the enhancing ROI may still bias tumor-derived CBF values and limit comparability across scanners and sequence implementations. These findings align with some studies showing equal performance of ASL and DSC.[11, 29, 30] Though, as of yet, there are no studies that compare performance across different ASL sequences or centers.
While these findings demonstrate the potential of APT-CEST MTR_{asym} and ASL CBF to predict IDH status, 1p/19q status, and WHO grade, the relatively small sample size limits the robustness and conclusions. The radiomics features obtained higher predictive performance than the univariate features.
To further test reproducibility and generalizability, and to determine whether radiomics features truly offer greater predictive value than univariate features, these analyses should be expanded using a larger dataset. Under typical assumptions for hypothesis testing (α = 0.05, power = 0.8) and effect sizes in the small-to-moderate range expected for quantitative MRI biomarkers (e.g. Cohen's d $\approx$ 0.3–0.4), sample sizes on the order of a few hundred subjects are commonly required to detect clinically relevant differences.[31, 32] For example, a two-sample comparison with d = 0.4 requires approximately 100 participants per group. Taking into account the need to analyse three major glioma subtypes separately, the expected data loss, and the use of multivariable models, a total target of $\sim$400 participants ($\sim$ 133 participants in each of the three common glioma types) should be feasible and provide sufficient statistical power.

Besides the mentioned imaging biomarkers that can be extracted from APT-CEST, APT-CEST also has the potential to identify viable tumor components beyond the enhancing margins. However, at present, APT-CEST has limitations and may not yet

be able to define the true margins accurately. Firstly, the spatial resolution of clinical APT-CEST images is generally low. The APT-CEST sequence used in this thesis has a native in-plane pixel size of 2.8x2.8 mm with a slice thickness of 2.8 mm interpolated to 1.4x1.4x2.8 mm. Interpolation improves image appearance but does not increase true spatial resolution. In comparison, clinical T1w and T2-FLAIR images at our hospital have a much higher resolution with voxel sizes 0.9 mm isovoxel. Because APT-CEST's in-plane and through-plane resolutions are lower, partial volume effects are more pronounced. With partial-volume effect, the voxel size is large relative to the structure being imaged and can therefore contain signals from multiple tissue types.[33] Voxels containing signals from viable tumor tissue as well as healthy tissue, blurring tumor-to-healthy tissue boundaries and limiting the accuracy of true margin definition. The low APT-CEST resolution is generally due to scan time, whole-brain coverage, and SNR. Secondly, the implemented sequence lacks fluid suppression. In contrast, research has shown that fluid suppression can improve visual readability and tumor delineation by reducing spurious hyperintensities in cystic or necrotic regions and CSF spaces.[34] Fluid suppression selectively reduces the contribution of such fluid compartments while preserving the APT-CEST signal from solid or semi-solid tissue.[34] In the study by Schüre et al., fluid suppression is achieved in post-processing using a spillover-corrected APT-CEST metric. It therefore does not intrinsically increase scan time or reduce spatial resolution. In practice, robust fluid-suppressed APT-CEST often benefits from richer Z-spectral sampling and high-quality B_0/B_1 mapping. This, however, can make protocols more complex and may limit the feasibility of whole-brain, high-resolution acquisitions within standard clinical time slots.

Findings using Artificial Intelligence

In Chapter 3, literature on AI-based GBCA reduction revealed several studies synthesizing full-dose CET1 images from low-dose or even GBCA-free inputs. Studies were limited to small sample sizes (60-640) and were often private.[35–38] Many studies relied entirely on the BraTS glioma imaging database, which is limited in its clinical diversity. Most studies focused on optimization of technical metrics rather than clinical applicability. To expand the available open-source glioma data and to obtain an independent test set for training AI algorithms, the IMAGO (IMAging in GliOma) database was created, as described in **Chapter 6**. IMAGO exists in two forms: 1) in-house, with MRI and clinical data of $\sim$1,700 glioma patients; and 2) an open-access database, accessible through the Health-RI XNAT and supported through the EUCAIM 2024 open call.[39] The 500 patients in the online database are all diagnosed with adult-type diffuse gliomas according to the WHO CNS5 (2021) criteria. Patient survival data and characteristics are similar as those reported elsewhere in the literature.[40, 41] Both IMAGO versions (in-house and online) are continuously expanding with new patients and additional clinical and histological data. Several of the patients in the in-house database were diagnosed before assessment of IDH status was routinely performed (2014), leaving molecular information incomplete. Retrospective immunohistochemical staining for the IDH1 R132H mutant protein is currently being performed on stored samples of adult-type diffuse gliomas to address this gap. Because this approach detects approximately 90% of IDH-mutant gliomas, additional molecular testing should ideally be performed, especially of tumors from patients

younger than 55 years at diagnosis but with negative staining for the IDH1 R132H mutant protein. The decision was made to perform immunohistochemical staining first because of its price-to-identification ratio. Currently, at our institute, such staining costs €6.40 per case, whereas the costs for more advanced molecular analysis are €250-600 per case. Of note, in tumors that appear to be IDH-mutant based on immunohistochemistry, the 1p/19q status should ideally be assessed as well.
Furthermore, continued refinement of automated segmentations through manual correction should be prioritized. These curated labels will benefit external users of IMAGO and in-house projects such as the PICTURE post-operative segmentation framework.

Studies on human-AI collaborations were not specifically targeted in the Chapter 3 literature search. A recent decision tree study reported sufficient diagnostic accuracy for identifying tumor enhancement but lacked spatial information about enhancing areas, limiting its utility for surgery and response assessment, where precise margin delineation is crucial.[42]
To address this gap, **Chapter 7** investigates whether humans can identify the tumor core without using GBCA imaging and whether their segmentation performance improves with AI support. Four raters segmented the tumor core of 100 glioblastoma IDH-wildtype patients, using only non-contrast sequences. One additional rater was aided by a GBCA-free AI segmentation algorithm. Human raters could segment the enhancing tumor areas using GBCA-free sequences with moderate accuracy. AI-based GBCA-free segmentation assistance significantly improved Dice, Hausdorff distance, and segmentation volume error. Segmentation time decreased per case, indicative of a learning effect for the raters. Spending more time per case did not result in higher segmentation time as there was no correlation between Dice scores and segmentation time.

Chapter 3 also showed that most AI GBCA reduction studies focus on technical improvements rather than clinical applicability. To address these two shortcomings, an AI benchmark study was conducted in Chapter 8, in which validated peer-reviewed GBCA-free segmentation and synthesis algorithms were trained on a large open-source dataset. Validation focused on three clinical tasks: 1) enhancement detection; 2) prognostic performance; and 3) therapy response evaluation. Unfortunately, of the 42 identified published algorithms (21 synthesis, 21 segmentation), only 11 were publicly available, and only 2 others were shared upon request, despite most articles stating that data and algorithms were available upon request. Four synthesis and six segmentation algorithms using BraTS[43–45] and UCSF-PDGM[46] were trained and validated on two independent test sets (EGD[47] and IMAGO (**Chapter 6**)). This study used a relatively large patient cohort (2,764 glioma patients) compared to the studies of the included algorithms(90-1,251). Segmentation algorithms outperformed synthesis-based algorithms in all three clinical tasks. This is most likely due to domain shift: although synthesis algorithms generate visually convincing CE-like images, their voxel intensity distributions deviate from genuine CET1 contrasts, negatively affecting subsequent segmentations. Similar effects have been discussed in the literature on medical image synthesis and domain adaptation.[48, 49] Despite its decent-to-moderate performance, GBCA-free segmentation

algorithms remain insufficient for imminent clinical integration.
We investigated the added value of ASL for GBCA-free segmentation. Although CET1 is sensitive to blood-brain barrier (BBB) disruption, and cerebral perfusion is increased in those areas, the results showed no improvement in Dice coefficient.[50] A possible explanation is the combinatorial effect of a large voxel size and imperfect ASL-to-T1w registration, leading to spatial misalignment despite reflecting the same pathophysiological mechanism.

Chapter 9 also tackles these limitations (overreliance on a single dataset, a limited number of included cases, and a tendency to focus on technical improvements) by developing and evaluating Apollo SmartGAD. Unlike previous studies, this tool was designed to serve as a supporting role in radiologists' clinical workflow.[12, 51] Rather than focusing solely on reducing GBCA use, Apollo SmartGAD triages the patients into three categories: 1) No GBCA is needed, 2) low-dose GBCA is needed, and 3) standard-dose GBCA is needed. The included algorithms were trained on a large multicenter database and validated on three datasets. Training data was representative of a real-world scenario, where not only glioma patients are present but also other brain tumors, infectious disorders, vascular and other pathologies. The algorithm was able to predict the need for GBCA use, but, like other proposed synthesis and segmentation methods, it had difficulty with smaller lesions.[52] While the tool had a high negative predictive value, several cases were still incorrectly classified as showing no uptake. Especially in tumors that generally show enhancement (metastases, meningiomas), the number of false negatives was high, and use of this tool should therefore have an auxiliary role, as missing small enhancing lesions may have severe consequences. Rather than guideline- or indication-based GBCA reduction strategies, which typically operate at the protocol level, Apollo SmartGAD leverages real-time image information. In **Chapter 9** we also learned that training exclusively on carefully curated single-pathology cohorts (e.g., glioblastoma-only datasets) leads to over-optimistic performance estimates that do not generalize to routine practice.[53] The multicenter design and broad inclusion criteria for the dataset revealed that space-occupying but non-enhancing lesions and vascular pathology are important sources of false positive recommendations, which would likely have been underestimated in more restricted datasets. Future expansions as well as new datasets should also include other primary brain tumors as well as brain metastases to ensure AI-algorithm generalizability across clinically relevant entities. Additionally, including healthy participants could help deep learning algorithms learn normal tissue variability and anatomical boundaries, improving robustness in clinical deployment.[54, 55]

Future perspectives

This thesis and a number of studies have demonstrated advantages of APT-CEST, but the limitations discussed above also indicate important directions for improvement. Despite the relatively small dataset used in this work, the results suggest that APT-CEST can be of clinical value as a supportive tool for the radiologist's differential diagnosis. However, the presence of an elevated APT-CEST signal cannot be straightforwardly interpreted as proof of viable tumor tissue. To resolve this, studies are needed that systematically correlate APT-CEST findings with histopathology, for instance by obtaining targeted biopsies from regions with increased APT-CEST signal. Only with such validation can we reliably determine whether elevated APT-CEST truly corresponds to viable tumor.
In the meantime, APT-CEST may still be used as a complementary imaging modality in the differential diagnosis. When interpreted alongside conventional MRI sequences such as T1-weighted and T2-FLAIR, APT-CEST could help clinicians form a more refined hypothesis about the location and extent of viable tumor tissue.

We also observed that most GBCA-free AI research focuses on technical advancements rather than clinical translation. While this thesis already incorporated a clinically oriented validation study - assessing enhancement detection, prognostic power, and RANO-based response evaluation - and a study that focuses on supporting the clinical workflow of radiologists, additional studies should explicitly examine clinical decision-making impact: 1) Diagnostic equivalence studies, comparing tumor grading and treatment planning decisions made with synthetic versus true CE images; 2) Workflow studies, evaluating neurosurgical or radiotherapy planning based solely on synthetic images to assess whether they offer adequate guidance for pre-operative or target-volume delineation. The proposed triage tool could be further expanded by improving the created algorithms, for example, by using more data, implementing more extensive loss functions, and incorporating clinical metadata in the decision-making process.
Finally, while synthetic CET1 images may serve as an appropriate transitional step toward GBCA-free workflows in some cases, they often fail to capture subtle enhancement. Using low-dose CET1 inputs instead to guide synthesis could potentially preserve critical spatial information. Rather than inferring enhancement from non-contrast sequences, AI algorithms could refine enhancement to address one of the shortcomings of the shown synthesis approaches in this thesis. This strategy may also reduce domain shift between the synthetic and true CET1 intensities observed in Chapter 8, and so improve downstream segmentation and clinical-task performance (e.g. enhancement detection sensitivity, and decision making tasks such as surgical and radiotherapy planning). This approach is being explored within the REDUCE study, a new component of the GLIO-CARE project, in which CET1 and DSC images will be acquired with 50% of the standard GBCA dose. After the low-dose images, the remaining 50% of the standard dose is administered so that REDUCE provides a setting to compare within-session diagnosis to determine when half-dose imaging is sufficient and when full-dose is warranted. The study design offers an opportunity to prospectively evaluate the triage philosophy proposed in **Chapter 9**: standard dose for uncertain/high-risk cases, half-dose protocols when some enhancement information is deemed necessary, and GBCA-free imaging for low-risk. Because REDUCE also includes saturation bolus-free half-dose DSC perfu-

sion, it can clarify how perfusion-based information changes with dose reduction and omission of T1 saturation, and whether ASL can substitute or complement reduced-dose DSC, directly extending the ASL-versus-DSC considerations discussed earlier. Thus, REDUCE functions as a translational bridge, directly testing whether the limitations observed in GBCA-free synthesis and segmentation can be mitigated through partial-dose acquisition while still achieving a meaningful reduction in cumulative GBCA exposure.

Figure 1 provides a visual summary of the workflow proposed in this thesis, illustrating how advanced MRI and AI approaches could be integrated to reduce reliance on gadolinium-based contrast agents (GBCA) in glioma imaging. The workflow can be divided into three phases. In phase 1 (acquisition), MRI data are obtained using GBCA-free sequences, including conventional non-contrast imaging and advanced techniques such as ASL and APT-CEST. In phase 2 (information extraction), these MRI images are analyzed to derive quantitative imaging biomarkers (optionally including radiomics) and to generate AI outputs (e.g., segmentation or synthesis), with evaluation focused on clinically relevant tasks such as tumor characterization, delineation, synthetic images, treatment response assessment, and prognostication. As a future direction, phase 2 may evolve from producing independent biomarkers toward providing clinically oriented decision support that directly addresses specific clinical questions, while remaining interpretable and validated. In phase 3 (clinical decision-making), clinicians integrate the extracted biomarkers and AI outputs to determine whether no GBCA, low-dose GBCA, or standard-dose GBCA is warranted for the individual patient. In selected cases, the workflow may also allow escalation to additional contrast administration when clinically necessary (e.g., when subtle enhancement is suspected), reinforcing a clinician-in-the-loop approach.

Conclusions

In this thesis, I demonstrated that APT-CEST and ASL can characterize tumor types and have matured to the point where integration into standard clinical workflows could reasonably be considered to enhance diagnostic decision-making. While direct clinical implementation of AI algorithms still requires refinement - particularly regarding fine spatial detail and generalizability- the proposed clinician-in-the-loop triaging algorithm represents a more immediately translatable solution. Although this triaging tool itself will benefit from further optimization, it is substantially closer to readiness for routine clinical use. Ultimately, GBCA reduction should be seen as a means to an end. The primary objective remains optimizing patient care. Given the culminating evidence highlighting the limitations and potential risks of GBCA use, there is a growing rationale to adopt novel, non-invasive biomarkers - such as those derived from APT-CEST, ASL, and AI - to more accurately delineate the true extent of gliomas and, in doing so, improve clinical outcomes.

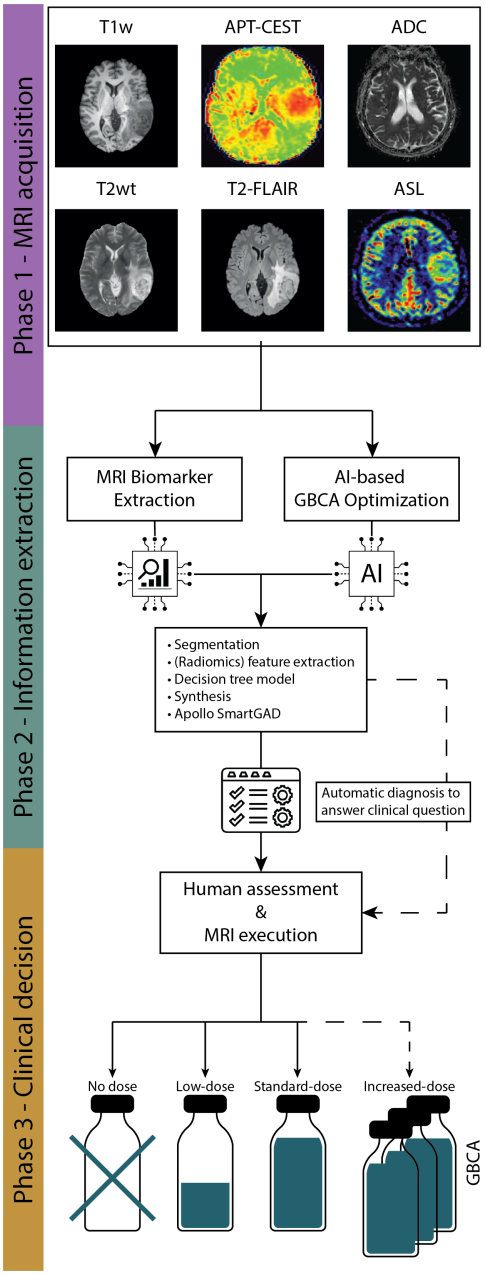


Figure 1: Overall workflow for integrating advanced MRI (ASL, APT-CEST) and AI methods to reduce GBCA use in clinical glioma imaging. Phase 1: GBCA-free data acquisition. Phase 2: information extraction (quantitative biomarkers and AI outputs) supporting clinical tasks (e.g., characterization, delineation, response, prognosis), with potential future expansion toward clinically oriented decision support. Phase 3: clinician-in-the-loop triage of GBCA administration (no, low-dose, or standard-dose), with optional escalation when clinically indicated. *abbreviations:* APT-CEST = amide proton trasnfer chemical exchange saturation transfer, ADC = apparent diffusion coefficient, ASL = arterial spin labeling, GBCA = gadolinium based contrast agent.

References

1 P. Y. Wen et al. “RANO 2.0: Update to the Response Assessment in Neuro-Oncology criteria for high- and low-grade gliomas in adults”. en. In: *J. Clin. Oncol.* 41 (33 Nov. 20, 2023), pp. 5187–5199. doi: 10.1200/JCO.23.01059.

2 N. Iyad et al. “Gadolinium contrast agents- challenges and opportunities of a multidisciplinary approach: Literature review”. en. In: *Eur. J. Radiol. Open* 11 (100503 Dec. 2023), p. 100503. doi: 10.1016/j.ejro.2023.100503.

3 Y. Fujita et al. “Impacts of anthropogenic gadolinium on the activity of the ammonia oxidizing bacterium Nitrosomonas europaea”. en. In: *Chemosphere* 257 (127250 Oct. 2020), p. 127250. doi: 10.1016/j.chemosphere.2020.127250.

4 C. L. Monroe et al. “Does Surveillance-Detected Disease Progression Yield Superior Patient Outcomes in High-Grade Glioma?” en. In: *World Neurosurg.* 135 (Mar. 2020), e410–e417. doi: 10.1016/j.wneu.2019.12.001.

5 S. Y. Ji et al. “Radiological assessment schedule for high-grade glioma patients during the surveillance period using parametric modeling”. en. In: *Neuro. Oncol.* 23 (5 May 5, 2021), pp. 837–847. doi: 10.1093/neuonc/noaa250.

6 S. A. S. Koh et al. “Interethnic variation in the prevalence of claustrophobia during MRI at Singapore General Hospital: does a wider bore MR scanner help?” en. In: *Proc. Singap. Health.* 26 (4 Dec. 2017), pp. 241–245. doi: 10.1177/2010105817695819.

7 M. Dostál et al. “What factors affect a patient’s subjective perception of MRI examination”. en. In: *Sci. Rep.* 14 (1 Sept. 30, 2024), p. 22731. doi: 10.1038/s41598-024-74231-9.

8 *Gliomer hos voksne*. Version 3. July 25, 2022.

9 T. Hashido, S. Saito, and T. Ishida. “A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas”. In: *Sci. Rep.* 10 (1 Apr. 9, 2020), p. 6121. doi: 10.1038/s41598-020-62658-9.

10 D. Alis et al. “The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas”. en. In: *Clin. Radiol.* 75 (5 May 2020), pp. 351–357. doi: 10.1016/j.crad.2019.12.008.

11 Y. Prysiazhniuk et al. “Diffuse glioma molecular profiling with arterial spin labeling and dynamic susceptibility contrast perfusion MRI: A comparative study”. en. In: *Neurooncol. Adv.* 6 (1 Jan. 2024), vdae113. doi: 10.1093/noajnl/vdae113.

12 C. Jayachandran Preetha et al. “Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study”. en. In: *Lancet Digit Health* 3 (12 Dec. 2021), e784–e794. doi: 10.1016/S2589-7500(21)00205-3.

13 E. Moya-Sáez, R. de Luis-García, and C. Alberola-López. “Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI”. In: *Frontiers in Neuroimaging* 2 (2023). doi: 10.3389/fnimg.2023.1055463.

14 K. Chang et al. “Residual Convolutional Neural Network for the Determination of IDH Status in Low- and High-Grade Gliomas from MR Imaging”. en. In: *Clin. Cancer Res.* 24 (5 Mar. 1, 2018), pp. 1073–1081. doi: 10.1158/1078-0432.CCR-17-2236.

15 Y. Qu et al. “Perfusion measurement in brain gliomas using velocity-selective arterial spin labeling: comparison with pseudo-continuous arterial spin labeling and dynamic susceptibility contrast MRI”. en. In: *Eur. Radiol.* 32 (5 May 2022), pp. 2976–2987. doi: 10.1007/s00330-021-08406-7.

16 A. Lavrova et al. “Diagnostic accuracy of arterial spin labeling in comparison with dynamic susceptibility contrast-enhanced perfusion for brain tumor surveillance at 3T MRI”. en. In: *Front. Oncol.* 12 (May 20, 2022), p. 849657. doi: 10.3389/fonc.2022.849657.

17 O. Togao et al. “Grading diffuse gliomas without intense contrast enhancement by amide proton transfer MR imaging: comparisons with diffusion- and perfusion-weighted imaging”. en. In: *Eur. Radiol.* 27 (2 Feb. 2017), pp. 578–588. doi: 10.1007/s00330-016-4328-0.

18 J. Zhou et al. “Differentiation between glioma and radiation necrosis using molecular magnetic resonance imaging of endogenous proteins and peptides”. en. In: *Nat. Med.* 17 (1 Jan. 2011), pp. 130–134. doi: 10.1038/nm.2268.

19 S. Jiang et al. “Identifying recurrent malignant glioma after treatment using amide proton transfer-weighted MR imaging: A validation study with image-guided stereotactic biopsy”. en. In: *Clin. Cancer Res.* 25 (2 Jan. 15, 2019), pp. 552–561. doi: 10.1158/1078-0432.CCR-18-1233.

20 O. Eidel et al. “Tumor infiltration in enhancing and non-enhancing parts of glioblastoma: A correlation with histopathology”. en. In: *PLoS One* 12 (1 Jan. 19, 2017), e0169292. doi: 10.1371/journal.pone.0169292.

21 A. Lasocki and F. Gaillard. “Non-Contrast-Enhancing Tumor: A New Frontier in Glioblastoma Research”. en. In: *AJNR Am. J. Neuroradiol.* 40 (5 May 2019), pp. 758–765. doi: 10.3174/ajnr.A6025.

22 F. Certo et al. “FLAIRectomy in supramarginal resection of glioblastoma correlates with clinical outcome and survival analysis: A prospective, single institution, case series”. en. In: *Oper. Neurosurg. (Hagerstown)* 20 (2 Jan. 13, 2021), pp. 151–163. doi: 10.1093/ons/opaa293.

23 A. F. Haddad et al. “FLAIRectomy: Resecting beyond the contrast margin for glioblastoma”. en. In: *Brain Sci.* 12 (5 Apr. 25, 2022), p. 544. doi: 10.3390/brainsci12050544.

24 K. Kamimura et al. “Amide proton transfer imaging of tumors: theory, clinical applications, pitfalls, and future directions”. en. In: *Jpn. J. Radiol.* 37 (2 Feb. 2019), pp. 109–116. doi: 10.1007/s11604-018-0787-3.

25 J. E. Park et al. “Pre- and posttreatment glioma: Comparison of amide proton transfer imaging with MR spectroscopy for biomarkers of tumor proliferation”. en. In: *Radiology* 278 (2 Feb. 2016), pp. 514–523. doi: 10.1148/radiol.2015142979.

26 Y. Wu et al. “Reproducibility of APT-weighted CEST-MRI at 3T in healthy brain and tumor across sessions and scanners”. en. In: *Sci. Rep.* 13 (1 Oct. 23, 2023), p. 18115. doi: 10.1038/s41598-023-44891-0.

27 J. B. Lee et al. “Repeatability of amide proton transfer-weighted signals in the brain according to clinical condition and anatomical location”. en. In: *Eur. Radiol.* 30 (1 Jan. 2020), pp. 346–356. doi: 10.1007/s00330-019-06285-7.

28 A. Matakos, J. Balter, and Y. Cao. “Estimation of geometrically undistorted B(0) inhomogeneity maps”. en. In: *Phys. Med. Biol.* 59 (17 Sept. 7, 2014), pp. 4945–4959. doi: 10.1088/0031-9155/59/17/4945.

29 E. S. Bayraktar et al. “Comparison of ASL and DSC perfusion methods in the evaluation of response to treatment in patients with a history of treatment for malignant brain tumor”. en. In: *BMC Med. Imaging* 24 (1 Mar. 22, 2024), p. 70. doi: 10.1186/s12880-024-01249-w.

30 G. Yamin et al. “Arterial spin-labeling and DSC perfusion metrics improve agreement in neuroradiologists’ clinical interpretations of posttreatment high-grade glioma surveillance MR imaging-an institutional experience”. en. In: *AJNR Am. J. Neuroradiol.* 45 (4 Apr. 8, 2024), pp. 453–460. doi: 10.3174/ajnr.A8190.

31 J. Cohen. *Statistical power analysis for the behavioral sciences*. 2nd ed. London, England: Routledge, May 13, 2013. 567 pp. doi: 10.4324/9780203771587.

32 K. S. Button et al. “Power failure: why small sample size undermines the reliability of neuroscience”. en. In: *Nat. Rev. Neurosci.* 14 (5 May 2013), pp. 365–376. doi: 10.1038/nrn3475.

33 M. A. González Ballester, A. P. Zisserman, and M. Brady. “Estimation of the partial volume effect in MRI”. en. In: *Med. Image Anal.* 6 (4 Dec. 2002), pp. 389–405. doi: 10.1016/s1361-8415(02)00061-0.

34 J.-R. Schüre et al. “Fluid suppression in amide proton transfer-weighted (APTw) CEST imaging: New theoretical insights and clinical benefits”. en. In: *Magn. Reson. Med.* 91 (4 Apr. 2024), pp. 1354–1367. doi: 10.1002/mrm.29915.

35 E. Gong et al. “Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI”. In: *J. Magn. Reson. Imaging* 48 (2 2018), pp. 330–340. doi: 10.1002/jmri.25970.

36 S. Pasumarthi et al. “A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain MRI”. In: *Magn. Reson. Med.* 86 (3 2021), pp. 1687–1700. doi: 10.1002/mrm.28808.

37 H. Luo et al. “Deep learning-based methods may minimize GBCA dosage in brain MRI”. en. In: *Eur. Radiol.* 31 (9 Sept. 2021), pp. 6419–6428. doi: 10.1007/s00330-021-07848-3.

38 S. Ammari et al. “Can Deep Learning Replace Gadolinium in Neuro-Oncology?: A Reader Study”. In: *Invest. Radiol.* (July 28, 2021). doi: 10.1097/RLI.0000000000000811.

39 Ivar J.H.G Wamelink, Alle Meije Wink, Niels Verburg, Roeland S Eijgelaar, Emmanouil Koltsakis, Marcus Cakmak, Maarten Balder, Henk J.M.M Mutsaerts, Philip Witt Hamer Frederik Barkhof, Vera C. Keil. *IMAGO*. July 24, 2025. doi: 10.5281/zenodo.17582072.

40 M. Salvati et al. “Glioblastoma: Molecular profile and immunophenotypic analysis as prognostic tools for tailored therapy and decision making in a recent surgical series”. en. In: *Interdiscip. Neurosurg.* 20 (100697 June 2020), p. 100697. doi: 10.1016/j.inat.2020.100697.

41 A. Olar et al. “IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas”. en. In: *Acta Neuropathol.* 129 (4 Apr. 2015), pp. 585–596. doi: 10.1007/s00401-015-1398-z.

42 A. Azizova et al. “Human performance in predicting enhancement quality of gliomas using gadolinium-free MRI sequences”. en. In: *J. Neuroimaging* 34 (6 Nov. 2024), pp. 673–693. doi: 10.1111/jon.13233.

43 U. Baid et al. “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification”. In: *arXiv [cs.CV]* (July 5, 2021).

44 B. H. Menze et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. en. In: *IEEE Trans. Med. Imaging* 34 (10 Oct. 2015), pp. 1993–2024. doi: 10.1109/TMI.2014.2377694.

45 S. Bakas et al. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”. en. In: *Sci Data* 4 (Sept. 5, 2017), p. 170117. doi: 10.1038/sdata.2017.117.

46 E. Calabrese et al. “The university of California San Francisco preoperative diffuse glioma MRI dataset”. en. In: *Radiol. Artif. Intell.* 4 (6 Nov. 2022), e220058. doi: 10.1148/ryai.220058.

47 S. R. van der Voort et al. “The Erasmus Glioma Database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma”. en. In: *Data Brief* 37 (107191 Aug. 2, 2021), p. 107191. doi: 10.1016/j.dib.2021.107191.

48 J. P. Cohen, M. Luck, and S. Honari. “Distribution matching losses can hallucinate features in medical image translation”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Lecture notes in computer science. Cham: Springer International Publishing, 2018, pp. 529–536. doi: 10.1007/978-3-030-00928-1_60.

49 R. Zhang, T. Pfister, and J. Li. “Harmonic unpaired image-to-image translation”. In: *arXiv [cs.CV]* (Feb. 25, 2019).

50 A. R. Deibler et al. “Arterial spin-labeling in routine clinical practice, part 3: hyperperfusion patterns”. en. In: *AJNR Am. J. Neuroradiol.* 29 (8 Sept. 2008), pp. 1428–1435. doi: 10.3174/ajnr.A1034.

51 J. Kleesiek et al. “Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?: A Feasibility Study”. In: *Invest. Radiol.* 54 (10 Oct. 2019), pp. 653–660. doi: 10.1097/RLI.0000000000000583.

52 R. H. Helland et al. “Segmentation of glioblastomas in early post-operative multimodal MRI with deep neural networks”. en. In: *Sci. Rep.* 13 (1 Nov. 2, 2023), p. 18897. doi: 10.1038/s41598-023-45456-x.

53 H. F. Tbahriti et al. “Machine learning and deep learning in glioblastoma: a systematic review and meta-analysis of diagnosis, prognosis, and treatment”. en. In: *Discov. Oncol.* 16 (1 Aug. 7, 2025), p. 1492. doi: 10.1007/s12672-025-03303-7.

54 H. Tang, N. Sun, and S. Shen. “Improving generalization of deep learning models for diagnostic pathology by increasing variability in training data: Experiments on osteosarcoma subtypes”. en. In: *J. Pathol. Inform.* 12 (1 Aug. 4, 2021), p. 30. doi: 10.4103/jpi.jpi_78_20.

55 A. T. Tran, T. Zeevi, and S. Payabvash. “Strategies to improve the robustness and generalizability of deep learning segmentation and classification in neuroimaging”. en. In: *BioMedInformatics* 5 (2 June 2025), p. 20. doi: 10.3390/biomedinformatics5020020.


A

Appendix

Nederlandse Samenvatting

Ondanks dat het diagnostiseren, monitoren van de behandeling en het in de gaten houden van het verloop van de ziekte bij glioom patiënten afhankelijk is van CET1-beelden, zijn er groeiende zorgen rondom gezondheidsrisico's, de impact op het milieu en financiële kosten door het herhaaldelijk gebruik van gadoliniumhoudende contrastmiddelen (GBCA's). Deze zorgen hebben de zoektocht naar alternatieve beeldvormingsbiomarkers aangewakkerd. Bovendien lijkt contrastopname op zichzelf onvoldoende om tumorweefsel volledig in kaart te brengen, wat de behoefte van het vinden van nieuwe biomarkers onderstreept om zo de patiëntenzorg te verbeteren.

Dit proefschrift, uitgevoerd binnen het GLIOCARE-project, verkent twee complementaire routes om de afhankelijkheid van GBCA's te verminderen: 1) geavanceerde MRI-technieken, met name amide proton transfer chemical exchange saturation transfer (APT-CEST) en arterial spin labeling (ASL), die niet-invasieve inzichten bieden in de biochemie en fysiologie van tumoren, en 2) kunstmatige intelligentie (AI), die grote beeldvormingsdatasets kan gebruiken om subtiele kenmerken te extraheren en GBCA-vrije of GBCA-gereduceerde beeldvormingsworkflows mogelijk te maken.

De PENGUIN-enquêtestudie werd geïntroduceerd in **Hoofdstuk 2** om de houding van patiënten ten aanzien van het gebruik van GBCA en MRI-follow-up te beoordelen. De studie liet zien dat patiënten weliswaar vertrouwen hebben in de huidige protocollen, maar de voorkeur geven aan GBCA-vrije beeldvorming als de diagnostische nauwkeurigheid behouden blijft. De studie benadrukt ook dat demografische en klinische factoren, zoals geslacht, leeftijd en tumorgraad, de MRI-ervaringen van patiënten beïnvloeden, wat wijst op een behoefte aan meer gepersonaliseerde en patiëntgerichte beeldvormingsbenaderingen. We vonden ook dat patiënten positief waren over de huidige frequentie van klinische follow-up. Met deze bevinding dient rekening te worden gehouden als de follow-up frequentie wordt verlaagd.

In **Hoofdstuk 3** werd een literatuurreview uitgevoerd om het beschikbare onderzoek naar GBCA-reductie te identificeren; dit werd ingedeeld in 1) richtlijnen, 2) geavanceerde beeldvorming en 3) AI-benaderingen. Ondanks de groeiende interesse in GBCA-reductie waren de meeste studies retrospectief. Belangrijke inzichten uit onze bevindingen zijn onder meer: de haalbaarheid van universele of op subgroepen gerichte GBCA-reductie; het potentieel voor klinisch gedreven, gepersonaliseerde beeldvormingsintervallen; dat GBCA-vrije radiomische kenmerken even goed of beter kunnen presteren dan CET1-afgeleide biomarkers wat betreft diagnostische prestaties; en dat AI- en geavanceerde beeldvormingsoplossingen geschikte instrumenten zijn voor GBCA-reductie. In deze studie vonden we ook dat ASL een veelbelovende GBCA-vrije perfusietechniek is die vergelijkbaar is met DSC. AI-methoden tonen potentieel om contrast-achtige beelden te synthetiseren uit niet-contrast beelden en bereiken een vergelijkbare diagnostische nauwkeurigheid als AI-methoden die CET1-beelden gebruiken.

Wat betreft geavanceerde MRI toont APT-CEST potentie voor niet-invasieve gliomagradering en differentiatie tussen tumorprogressie en behandelingseffecten. **Hoofdstuk 4** laat zien dat een klinisch haalbare whole-brain APT-CEST-sequentie op 3 Tesla een goede korte- en langetermijnreproduceerbaarheid heeft bij zowel gezonde vrijwilligers als glioompatiënten. Hierbij overstijgt het contrast tussen tumor en normaal weefsel de scan-rescan-variabiliteit. We vonden een lage reproduceerbaarheid in de orbitofrontale gyri, waarschijnlijk door B_0-inhomogeniteit en gebieden met grote venen, wat overeenkomt met de literatuur. In deze reproduceerbaarheidsstudie werden zes glioompatiënten geïncludeerd en vonden we hogere MTR_{asym}-waarden in de tumor-ROI's (gemiddelde 1,59 ± 0,67) dan in de contralaterale ROI (gemiddelde 0,50 ± 0,46). Deze resultaten motiveerden de verdere inclusie van prospectieve patiënten (N=100) in het GLIOCARE-project.

Uit deze prospectieve GLIOCARE-patiënten werden descriptieve en radiomische kenmerken geëxtraheerd en gebruikt om de IDH-status, 1p/19q-status en tumorgraad te voorspellen in **Hoofdstuk 5**. In deze analyse werden 48 GLIOCARE-patiënten, 42 Russische patiënten en 129 Chinese patiënten geïncludeerd. 52/100 GLIOCARE-patiënten werden geëxcludeerd wegens ontbrekende moleculaire informatie, non-gliomen of post-operatieve status. We vonden dat univariate MTR_{asym}- en radiomische kenmerken geassocieerd waren met moleculaire subtypen. Net als in Hoofdstuk 4 vonden we hogere MTR_{asym}-waarden in de tumor-ROI's dan in de corresponderende contralaterale ROI. We vonden dat de tumor-CBF uitsluitend verhoogd was bij glioblastoompatiënten. Het combineren van CBF-kenmerken met MTR_{asym}-kenmerken verbeterde de voorspellende prestaties, zij het minimaal. MTR_{asym}- en CBF-waarden verschilden significant tussen datasets. De cross-vendor generaliseerbaarheid van APT was gering, terwijl APT binnen dezelfde vendor een goede generaliseerbaarheid had. Weefselnormalisatie ten opzichte van de contralaterale ROI verwijderde verschillen tussen datasets.

Het opzetten van de IMAGO (IMAging in GliOma) database in **Hoofdstuk 6** zorgt voor aanvullende middelen om AI algoritmen te trainen en te valideren. We hebben een grote in-huis database van ongeveer 1700 patiënten opgezet waarvan de informatie van 500 patiënten publiekelijk toegankelijk is (beschikbaar via het Health-RI XNAT platform en ondersteund door de EUCAIM 2024 open call). Beide datasets worden continu uitgebreid met nieuwe patiënten en aanvullende klinische en histologische data. De in-huis database wordt op het moment uitgebreid met extra histologische data en radiotherapie beelden. Alle 500 patiënten uit de online database hebben adult-type diffuse gliomen volgens de WHO CNS5 (2021)-criteria. We vonden dat overlevingsgegevens en patiëntkenmerken vergelijkbaar zijn met die elders in de literatuur worden gerapporteerd.

In **Hoofdstuk 7** worden 100 glioblastoom IDH-wildtype patiënten uit de IMAGO database gesementeerd door vier beoordelaars, waarbij uitsluitend niet-contrastsequenties worden gebruikt. Een AI algoritme dat eveneens alleen niet-contrastsequenties gebruikte om te segmenteren was ontwikkeld. Een extra beoordelaar verbeterde de segmentaties van dit algoritme. We vonden dat menselijk beoordelaars de tumor kern met redelijke

nauwkeurigheid konden segmenteren. De segmentaties van de beoordelaar die ondersteund werd door het AI algoritme waren significant beter wat betreft Dice, Hausdorff distance, en segmentatievollume-error. We vonden ook dat de segmetnatiefout per casus afnam, wat suggereert dat beoordelaars leren in de loop van de tijd. Er was geen correlatie tussen Dice-scores en segmentatie tijd, wat aangeeft dat meer tijd spenderen aan een casus niet resulteerde in betere segmentaties.

Hoofdstuk 3 liet ook beperkingen van huidige AI-studies zien. Ten eerste richten de meeste AI-studies naar GBCA-reductie zich op technische verbeteringen in plaats van op klinische toepasbaarheid, en ten tweede zijn de meeste AI-manuscripten beperkt tot kleine steekproefgroottes. **Hoofdstuk 8** pakt deze twee tekortkomingen aan in een AI benchmark studie. In deze studie hebben we peer-reviewed gadolinium vrije segmentatie en synthesis algoritmen gevalideerd op een grote open-source dataset. Deze benchmark studie focust zich op drie klinische taken: 1) het detecteren van enhancement; 2) de prognostische performance; en 3) het evalueren van therapie response. We konden zes segmentatie en 4 synthesis algoritmen includeren en vonden dat segmentatie algoritmen consistent beter presteerden dan synthesis algoritmen in alle drie de klinische taken. Segmentatie algoritmen hadden een significant hogere sensitiviteit, NPV, en nauwkeurigheid op een van de test datasets en een trend op de andere testdataset. De intraclass correlatie coefficiënt tussen het segmentatie volume en het werkelijke volume was slecht voor de synthesis algoritmen en matig tot goed voor de segmentatie algoritmen. Aankleurende volumes die verkregen waren via segmentatie algoritmen waren prognostisch voor overleving. Het toevoegen van leeftijd, geslacht en type chirurgie verbeterde de c-index van 0.6 naar 0.68. Hoewel segmentatie algoritmen beter presteerden dan synthesis algoritmen, zijn beide momenteel nog niet klaar voor klinische implementatie.

De AI-pijplijn die is ontwikkeld in **Hoofdstuk 9**, Apollo SmartGAD, is een klinisch georiënteerde triagetool die patiënten categoriseert op basis van de noodzaak tot toediening van GBCA (geen, lage dosis of standaarddosis). Hoewel de tool veelbelovende sensitiviteit en specificiteit vertoonde, werden sommige casussen nog steeds verkeerd geclassificeerd. Met name kleine laesies en typisch aankleurende tumoren zoals metastasen en meningeomen werden verkeerd geclassificeerd, wat aangeeft dat de tool als hulpmiddel moet dienen en niet als vervanging van de huidige protocollen. We vonden ook dat de aanwezigheid van enige contrastopname op de echte contrastbeelden als op de synthetische contrastbeelden, een goede voorspeller is van mogelijke hooggradige gliomen (respectievelijk 80% en 85% versus 77 en 78% sensitiviteit en specificiteit).

Conclusie

Deze thesis heeft bijgedragen aan het veld van glioma onderzoek door APT-CEST, ASL en AI technieken te onderzoeken om het gebruik van gadolinium houdende contrastmiddelen te verminderen. De bevindingen in dit proefschrift laten zien dat APT-CEST en in mindere maten ASL mogelijk gebruikt kunnen worden om nieuwe klinische biomarkers te vinden om gliomen te karakteriseren. De bevindingen tonen ook aan dat AI-tools nog niet klaar zijn voor directe klinische implementatie en dat hun ontwikkeling meer gericht moet

zijn op klinische toepasbaarheid in plaats van technische verbeteringen. Toekomstige studies moeten onderzoeken of weefsel normalisatie de ontwikkelaars-/merk-variabiliteit van APT-CEST voldoende kan verminderen en of APT-CEST de werkelijke omvang van de tumor beter kan identificeren dan wanneer CET1 beeldvorming wordt gebruikt. Het ontwerpen van AI-studies rondom de klinische workflow zal niet alleen klinische impact hebben, maar zal ook een positieve impact hebben op de patiënten zorg.

List of publications

This thesis

Ivar J.H.G. Wamelink, Joost P.A. Kuijer, Beatriz E. Padrela, Yi Zhang, Frederik Barkhof, Henk J.M.M. Mutsaerts, Jan Petr, Eslmarieke van de Giessen, Vera C. Keil. Reproducibility of 3T APT-CEST in Healthy Volunteers and Patients With Brain Glioma; Journal of Magnetic Resonance Imaging; 28-03-2022; doi: 10.1002/jmri.28239.

Ivar J.H.G. Wamelink, Hugo L. Hempel, Elsmarieke van de Giessen, Mark H.M. Vries, Philip De Witt Hamer, Frederik Barkhof, Vera C. Keil. The patients experience of neuroimaging of primary brain tumors a cross-sectional survey study; Journal of Neuro-Oncology; 28-03-2023; doi: 10.1007/s11060-023-04290.

Ivar J.H.G. Wamelink, Aynur Azizova, Thomas C. Booth, Henk J.M.M. Mutsaerts, Afolabi Ogunleye, Kshitij Mankad, Jan Petr, Frederik Barkhof, Vera C. Keil. Brain tumor imaging without gadolinium based contrast agents: feasible or fantasy?; Radiology; 06-02-2024; doi: 10.1148/radiol.230793.

Silvia Ingala*, Marko Bauer*, **Ivar J.H.G. Wamelink**, Mathias Perslev, Kaining Sheng, Mads Birch Sørensen, Sebastian Nørgaard Llambias, Arpit Kothari, Era Dwivedi, Martin Lillholm, Frederik Barkhof, Michael Bachmann Nielsen, Mads Nielsen, Bas M.M.S. Jasperse, Vera C. Keil, Akshay Pai. To Gad or not-to-Gad: Leveraging AI for Selective Use of Gadolinium Contrast in Brain MR, in submission. * These authors have contributed equally to this work.

Ivar J.H.G. Wamelink, Alle Meije Wink, Niels Verburg, Roelant S. Eijgelaar, Emmanouil Koltsakis, Marcus Cakmak, Henk J.M.M. Mutsaerts, Philip de Witt Hamer, Mathilde Kouwenhoven, Pieter Wesseling, Frederik Barkhof, Vera C. Keil. Amsterdam Imaging and Clinical Glioma Database; IMAGO, in submission.

Ivar J.H.G. Wamelink, Aynur. Azizova, Elif Kaya, E. Koltsakis, Joao N. Ramos, Aziz A. Tan, Normann Kornemann, Frederik Barkhof, Alle Meije Wink , Vera C. Keil. Contrast-Enhancing Tumor Margin Detection in Gliomas using Non-Contrast MRI: From Human-Only to Human-AI Assisted Assessment, in submission.

Ivar J.H.G. Wamelink, Rajeev A. Essed, Stefan.de Vries, Ellis Donders, Frederik Barkhof, Alle Meije Wink, Vera C. Keil, Stefano Trebeschi. Gadolinium-free MRI using artificial intelligence in glioma: a clinically-oriented benchmark study, in submission.

Rajeev A. Essed, **Ivar J.H.G. Wamelink**, Jan Petr, Joost P.A. Kuijer, Alle Meije Wink, Shuncong Wang; Frederik Barkhof, Vera C. Keil. Amide Proton Transfer and Arterial Spin Labeling for Non-Invasive Molecular Stratification of Glioma: A Multi-Dataset Imaging Biomarker Study, in submission.

Other publications

Esther Warnert, Soetkin Beun, Nils Broeckx, Vera Keil, Camille Maumet, Kathy Oliver, Jan Petr, Joana Pinto, Katya Driessche, Ivar. J.H.G. Wamelink, Patricia Clement. Medical data in science: patient information matters!; GliMR conference; 12-01-2022; doi: 10.13140/RG.2.2.26796.80008.

Aynur Azizova, **Ivar J. H. G. Wamelink**, Yeva Prysiazhniuk, Marcus Cakmak, Elif Kaya, Jan Petr, Frederik Barkhof, Vera C. Keil. Human performance in predicting enhancement quality of gliomas using gadolinium-free MRI sequences; Journal of Neuroimaging; 11-2024; 34(6):673-693. doi: 10.1111/jon.13233.

Marcus Cakmak, Sepehr Mohammadian, Vera C. W. Keil, Joost W. Schouten, Philip C. de Witt Hamer, Thijs van der Vaart, Rutger K. Balvers, **Ivar J. H. G. Wamelink**, Frederik Barkhof, Martin van den Bent, Mark Vries, Marion Smits. How useful is contrast-enhanced MRI in the long-term surveillance of glioma? A multicentre retrospective longitudinal cohort study; European Radiology; 26-11-2024; doi: 10.1007/s00330-024-11333-y.

Aynur Azizova, Yeva Prysiazhniuk, **Ivar J. H. G. Wamelink**, Marcus Cakmak, Elif Kaya, Pieter Wesseling Philip C. de Witt Hamer, Niels Verburg, Jan Petr, Frederik Barkhof, Vera C. Keil. Preoperative prediction of diffuse glioma type and grade in adults: a gadolinium-free MRI-based decision tree; European Radiology; 03-2025; 35(3):1242-1254. doi: 10.1007/s00330-024-11140-5.

Aynur Azizova, Yeva Prysiazhniuk, Marcus Cakmak, Elif Kaya, Jan Petr, Frederik Barkhof, **Ivar J. H. G. Wamelink**, Vera C. Keil. Visual versus region-of-interest based diffusion evaluation and their diagnostic impact in adult-type diffuse gliomas; Neuroradiology; 08-11-2025; vol 35, 4257–4266; doi: 10.1007/s00234-025-03832-6.

Rajeev Essed, Yeva Prysiazhniuk, **Ivar J.H.G. Wamelink**, Aynur Azizova, Vera C. Keil. Performance of amide proton transfer imaging to differentiate true progression from therapy-related changes in gliomas and metastases. European Radiology; 02-2025; 35(2):580-591. doi: 10.1007/s00330-024-11004-y.

Portfolio

Department	Radiology and Nuclear Medicine
PhD period	September 2021 - December 2025
Promotor	prof.dr. F. Barkhof
Copromotors	dr. V.C. Keil
	dr. A.M. Wink

Courses	Year	ECTS
MRI voor klinisch fysici	2021	1
How to write research papers	2022	1
Parallel and GPU Programming in Python	2022	0.57
GNK–AMS integrity course	2022	2
Project Management	2023	0.61
Computing in R	2023	0.71
Didactical Skills	2023	0.4
BioBusiness	2023	3
GNK–AII–BROK (WMO)	2025	1.5
		10.79

(Inter)national conferences, meetings, extended stays	Year	ECTS
GliMR networking event Brno	2021	0.4
Medical Image Computing Summer School	2022	1.79
GliMR AI Training School	2022	1.43
ESMRMB–GREC	2022	0.71
CCA conference social activity organisation	2022	0.75
Organization GliMR conference	2022	1
Organization GliMR social event	2023	1
RNG research day Amsterdam	2022	0.64
Wetenschappelijke dag LWNO	2022	0.32
ISMRM Benelux	2022	0.36
ISMRM	2022	1.79
GliMR 3rd annual meeting	2022	1.1
Annual meeting Amsterdam Neuroscience	2022	0.6
Organization GliMR social event	2023	1
ISMRM Benelux	2023	0.4
GliMR Porto conference	2023	1
ISMRM 2023	2023	1.79

(Inter)national conferences, meetings, extended stays	Year	ECTS
CCA retreat	2023	1.5
OOA retreat	2023	2
OHBM Seoul	2024	1.4
Research stay China	2025	3.43
		24.41

Teaching and Student Supervision	Year	ECTS
Student supervisie	2021–2025	3
Practicum Anatomie en Neuroimaging	2023	0.11
Practicum Anatomie en Neuroimaging	2024	0.11
Practicum Neuroimaging	2025	0.11
		3.33

Other Academic Activities	Year	ECTS
Journal club RNG	2021–2025	4.5
Grant writing	2021–2025	1.5
Creation of PhD starters package	2023	0.5
		6.5
		Total 45.03

About the author

Ivar Johannes Hendrikus Guillaume Wamelink was born on April 2, 1996, in 's-Hertogenbosch, the Netherlands. He went to De Opstap primary school in 's-Hertogenbosch, and after obtaining his gymnasium diploma at Het Stedelijk Gymnasium van 's-Hertogenbosch he went to live in Enschede to study Technical Medicine at the University of Twente (2014). Technical Medicine is a 60:40 combination of technical courses such as mathematics, physics, and other engineering courses and anatomy, pathology, and soft skills. The studies prides itself on the heavy study load while focusing also on patient contact and research. After 3 years Ivar obtained his Bachelor's degree and, after his study abroad in Seoul, South-Korea, he graduated in 2020 from the Master's track Medical Imaging and Intervention. During his final Master's internship at the Antoni van Leeuwenhoek (10 monhts) his interests in the brain, artificial intelligence, programming, and medical imaging were deepened. During this internship he researched the prediction of immunotherapy outcaome in melanoma patients with brain metastases. During this internship he helped publish 2 manuscripts and a book chapter. After his Master's he shortly worked as a vaccination administer during the worldwide covid pandemic. But, having enjoyed his final internship on AI and brain imaging, he sought a continuoation in this field. The GLIOCARE PhD project which focuses on AI, advanced MRI imaging, and brain cancer was therefore a perfect fit. Starting his PhD at Amsterdam University Medical Centers, location VUmc, in September 2021, Ivar first investigated the reproducibility of APT-CEST MRI imaging and the patient experience towards the MRI and gadolinium use. During his PhD he actively recruited patients for the prospective scanning of APT-CEST and ASL. He also built the IMAGO database, a large (>1600 patients) multimodel glioma database that contains prospective MRI scans, clinical data, and histological data. He worked on AI projects and even collaborated with Tiantan University Hospital in Beijing China. Aside from his academic pursuits, Ivar Crossfits intensively and plays fieldhockey for fun. He enjoys cooking difficult or large recipes but probably likes eating them more.

Dankwoord

Het dankwoord, het enige hoofdstuk dat door vrijwel iedereen "bekeken" wordt. Op de weg naar het behalen van een PhD word je gesteund door veel mensen. Soms gebeurt dit actief en soms gebeurt dit terwijl je het zelf, of zij, niet door hebben, maar aan iedereen zeg ik alvast: dankjewel! Het aantal mensen dat een bijzondere rol heeft gespeeld is groot en ik zal niet iedereen kunnen benoemen. Ook aan hen: bedankt! Dit proefschrift was natuurlijk niet mogelijk geweest zonder de deelname van de vele mensen met een hersentumor. Ik ben hen ontzettend dankbaar dat zij, ondanks hun eigen situatie, bereid waren om bij te dragen aan dit onderzoek.

Een aantal personen wil ik hier persoonlijk bedanken voor hun steun en vriendschap:

Frederik, die als 'grote baas', constant een prettige bird's-eye view had over mij en mijn projecten. Jouw begeleiding, expertise, connecties en inzichten hebben veel voor mijn projecten betekend. Hierdoor heb ik geleerd verder te kijken dan wat je op het eerste oog ziet. Zeker het hoog houden van meerdere ballen is iets waar jij ontzettend goed in bent en waar je mij in hebt begeleid. Je hebt me geleerd om op bepaalde momenten te focussen op één of twee projecten als dat nodig was. Op deze manier was het makkelijker om stappen te maken en het overzicht te bewaren. Jouw kritische blik op projecten of manuscripten heb ik zeer gewaardeerd en hebben mijn projecten verbeterd. En ondanks dat je een zeer drukke agenda hebt, maak je altijd tijd vrij als iemand daar om vraagt. Dit siert jou, omdat dit iets is wat niet iedere promotor wil of kan doen. Toen Vera in de laatste periode van mijn PhD met zwangerschapsverlof ging, heb je ook een actievere rol in de afronding gespeeld. Je wist me duidelijk te vertellen hoe ik mijn situatie moest relativeren en het maakte niet uit wanneer ik belde (of het nu in de rij op Schiphol was of op een vrijdag om 18:00). Kortom, dankjewel voor alles!

Lieve **Vera** de woorden die ik hier schrijf zullen te kort schieten, maar ik zal toch een poging wagen. Ik was jouw 'eerste' (PhD-student :)). Ondanks dat ik 15 minuten te laat was op mijn eerste sollicitatiegesprek, heb je me toch aangenomen op het GLIOCARE-project. Het was een groot avontuur met veel reisjes, veel lachen en veel leren. Op de afdeling zijn er meerdere co-promotoren, maar jij was en bent de beste. Er werd niet voor niks tegen mij gezegd dat men jaloers was op mij dat zij jou niet in hun promotieteam hadden. Jij hebt een ongekende inzet voor jouw werk en dit is iets wat ik bewonder en waar ik hopelijk zelf ook iets van opgestoken heb. We zaten altijd op één lijn over het ziekenhuis, maar je bleef altijd innoveren en ongestoord doorwerken om door de soms ietwat onnodige bureaucratie heen te komen. Door jou heb ik meer oog voor detail, weet ik dat het soms nodig is om te accepteren en te blijven focussen en dat je altijd tijd voor iemand kan vrijmaken. Je wist tijdens onze meetings vaak in korte tijd te achterhalen wat ik moest doen. Sommige dingen nam jij op je, ondanks dat ik soms dacht dat ik deze hoorde te doen :). Tijdens onze meetings hadden we het ook vaak over andere dingen dan de projecten waar we mee bezig waren. En jouw Duitse directheid kon deze

gesprekken vaak een interessante en soms grappige wending geven. Die directheid is overigens iets wat ik ontzettend heb gewaardeerd tijdens mijn promotie en is ook iets waar ik op kon bouwen. Jouw relaxte en vrijgevige houding heeft er ook voor gezorgd dat ik ontzettend veel plekken in de wereld heb weten te bezoeken. Het is natuurlijk ongebruikelijk om naar een resort in Turkije te gaan voor werk... Zonder jou was mijn PhD traject ongekend veel zwaarder geweest en ik wil je daarvoor ontzettend bedanken!

Proooofessooorreee **Alle Meije**, AM, vanaf het begin van mijn PhD was jij aanwezig bij mijn voortgangsmeetings. Officieel was je toen nog geen deel van mijn promotie team. Eigenlijk was dit vreemd, omdat je niet alleen altijd aanwezig was, maar ook altijd tijd vrij maakte als ik ergens hulp mee nodig had. Soms hadden wij een verschillende kijk op de academie en dat heeft tot mooie en interessante discussies geleid. Dit was niet 'you are selling your soul' zoals sommige zeggen, maar over projecten zelf of over normen en waarden. Over dit laatste waren we het eigenlijk vrijwel altijd eens en konden we soms lang door praten. Van jouw coding wizardry en expertise op het gebied van registratie heb ik veel geleerd. Als je niet klaar stond om direct te kunnen helpen, dan maakte je er wel tijd voor. Overigens ben jij niet de 'klassieke' co-promotor. Je sluit je altijd aan bij ons (de broekie PhD studenten) om plezier te hebben. Of het nu karaoke zingen in Korea is of een Surinaamse dans, je wil altijd leuk mee doen en dat is iets waar sommigen van kunnen leren. Voor al jouw hulp en al het lachen wil ik je graag bedanken!

My many thanks to the reading committee **prof.dr. D.P. Noske, prof. N. Sollmann, prof.dr.ir A.J. Nederveen, prof.dr. J.J.C. Verhoeff, prof.dr. A. Van Der Hoorn, prof.dr. S. Bisdas**. Thank you very much for taking the time to read my thesis and for being present during my defense. I hope the questions won't be too difficult and that I can answer them to your satisfaction.

The dynamic duo **Henk-Jan** en **Jan** I want to thank for answering my ASL questions, and, if they were not ASL questions, for answering my ExploreASL questions. Thank you also for assisting Vera and myself during grant writing. Something the two of you are very good at, but unfortunately have to spend too much time on. I also enjoyed our many trips abroad immensely. The list of places I visited with the two of you is long and, once there, you brought me in contact with many different people and disciplines. Your vast networks have been very useful throughout my PhD. While I was not officially part of your 'peacock' family I always felt welcome and appreciated when you invited me. De fietslegende van het VUmc, **Joost**. Zeker aan het begin van mijn PhD hebben we veel contact gehad over de APT-CEST sequentie en het daarbij horende manuscript. Naast dat je ontzettend veel kennis hebt op het gebied van MRI (zeer bewondernswaardig), ben je ook heel geduldig. Je wist altijd op een heldere en rustige manier uit te leggen hoe de vork in zijn steel zit als ik met MRI vragen aankwam. Ik wil je ook nog bedanken voor het regelen van de Siemens diners. Deze social events waren de kers op de taart van de toch al onvergetelijke ISMRM trips.

Paranimfen, **Mathijs**, dr Dijsseldorf, of Mr ASL, en **Erik-Jan**, brovarius paranimfus maximus, EJ, Mr Boston, super bedankt dat jullie allebei jezelf mooi wilden maken en

naast mij wilden zitten op deze dag. Een ongekende prestatie van jullie, dat mag gezegd worden. **Mathijs**, vanaf het moment dat ik in het VU kwam werken wisten we dat we het met elkaar konden vinden. Je bent serieus als het moet, maar eigenlijk ben je dat nooit en dat was een perfecte match. Onze ideeën en filosofieën zijn dan ook vaak hetzelfde, waardoor we mooie en interessante gesprekken hebben, maar vooral elkaars persoonlijkheid kunnen versterken. Of het nu op werk was, op een congres, of gewoon chillen, het is altijd een succes. Althans, totdat ik samen met jou de trein pak, want dat was echt altijd drama (ik geloof nog steeds dat je ongeluk brengt). Maar ik denk vooral dat we het allebei tot het einde hebben gered omdat we elkaar wisten te steunen als het even tegen zat. Dan konden we heerlijk venten over wat er dan al niet gebeurd was, en dat is goed. Overigens weet ik zeker dat ik het zonder jou niet tot het einde zou hebben gered... hoe had ik anders aan mijn ASL uitwerkingen moeten komen? Je hebt een ongekend doorzettingsvermogen (je weet wat ik bedoel) en bent een zeer goede mentor; ik weet dat dat je erg ver gaat brengen. Hopelijk zie je dat zelf ook in, want je vindt het allemaal 'normaal', maar dat is het niet. Erik-Jan, nee... dat klinkt raar, **EJ**, ik ken je al vanuit Enschede, maar we zijn pas echt goede vrienden geworden in Utrecht. Dat verbaast me trouwens niks, want in Enschede konden we ook al nerden over van alles en nogwat. Dat vrienden worden heb je wel een beetje opgedrongen. Dichtbij wonen, nog dichterbij wonen, samen hockeyen, samen naar Boston... we leken op een gegeven moment wel een getrouwd stel. Het hielp ook niet dat jij bij ons kind aan huis was, of zoals Wouter dat zo mooi kan zeggen: 'de parasiet'. Ik heb lang nagedacht over wat ik wil schrijven en uiteindelijk kwam ik tot de volgende conclusie: het maakt niet uit wat ik schrijf bij jou. Gaan we koken, dan is het gezellig en lekker. Gaan we gamen, dan kunnen we dat tot 3 uur 's nachts doen en denken dat het een goed idee is om door te gaan tot 7. Bij de hockey zijn we sportief en kunnen we lachen (en soms huilen). Is er iets gebeurd, dan steunen we elkaar. Heb ik jouw hulp nodig, dan ben je er voor mij en andersom. Doen we een drankje, dan loopt het vaak uit de hand. Praten we over iets interessants, dan komen de echte nerds naar boven en blijven we maar door gaan. Je bent geïnteresseerd en doordacht. Kortom, het maakt niet uit wat we doen, het eindigt altijd in een succes (wellicht zijn we echt een getrouwd stel...). En zonder dit alles had ik die PhD nooit gehaald.
Dankjulliewel voor jullie steun, interesse en vriendschap!

My international and culturally diverse office friends (which was absolutely in my favor) **dr. Esteeeveeees** and 'malaka' **Sasha**. It is safe to say, we had the best and most interesting office. To the left of me, I had the woman who made the most interesting noises and who tried to teach me that 'sharing is caring', while across from me, I had the guy who taught me AI but also didn't know that a rollercoaster park had rollercoasters. We will leave out the grumpy robot, the other person in our office. Bea, thanks for taking time out of your sometimes too busy schedule to listen to my rants or my problems. I always appreciated our talks, as they allowed us to briefly forget the shit we were working on at that time. Your enthusiasm is absolutely contagious. If there is a party? Let's go. Science? Let's do it. "Ivar, I talked to this guy, listen, it's sooo cool'. Please hold on to this mindset because I know it will bring you far. Alex, you were my partner in crime during the GliMR conferences. You will be forever known as the king of all social

events. Your openness and care are things I treasure in you as a friend, and I think a lot of people can learn from you. Besides, you were always yourself and didn't care what others thought of you when you did something silly. I wish, at times, I could borrow that from you :). When we shared hotel rooms during our trips, we could talk for a long time before closing our eyes, after which you didn't mind me waking you up when you were cutting down a forest (thanks!).

Aynur and Rajeev, Vera's other 'kids'. **Aynur**, I have seen you sink your teeth into your work and in getting your certificate to practice medicine in the UK. I appreciate how much time and effort this must have taken you and I admire that. Your clinical knowledge is something I could always count on, and you did not mind answering all my questions. I hope your calm, polite, and always happy attitude will soon visit us again! Our trips with GliMR together with Vera and Alex I will always remember. **Rajeev**, mr polyglot, who remembers everything he reads. Our trip to Beijing was so interesting and fun. While we didn't really know each other well before going there, I quickly realized we were cut from the same cloth but also not. I'm glad I spent this time with you where we learned about Chinese culture and you taught me so much about LLMs and the Raspberry Pi.

Het doen van een PhD kan soms zwaar zijn. Het hebben van mensen om je heen die ook in het soms bijna zinkende bootje zitten, helpt de lasten dragen. Ik heb ontzettend geluk gehad met mijn 'collegas'. Want collega's kan ik ze eigenlijk niet noemen en velen zullen nooit zo een relatie met hun collega's opbouwen. Gedurende mijn PhD zijn dit vrienden geworden waar ik erg dankbaar voor ben. **Maisita**, mijn mede-foodie van de afdeling! Die passie hebben we gelukkig kunnen delen op de cocktail rooftopbar in Dublin of het kleine vis restaurantje daar aan het water. Uiteraard mogen we het werktripje naar Lipari met Alex niet vergeten. Ik bewonder jouw kijk op de wereld en hoe je daar ook naar handelt. **Luigi**, you and I hit it off from the first time we met. During our plane ride to London, we bonded and became Bonnie and Clyde. We had (which not many people know) dinner at Gordon Ramsay's and held Nerf gun fights in the office. **Janine**, het was altijd een mysterie voor mij wanneer je er was (omdat ik onregelmatig naar kantoor ging), maar als we beiden op kantoor waren was het altijd gezellig en kon ik je over je huisdieren uitvragen. **David**, wiens MRI en onderzoekskennis zeer groot is en waar iedereen met open oren naar luistert. **Lyduine**, jij bent een ongekende drijfveer geweest die onze afdeling zo gezellig en hecht heeft gemaakt. Jij als ervaren researcher weet iedereen altijd op de juiste manier te helpen en advies te geven. **Marito**, the guy who taught me Italian and who had really dry humor. **Lonneke**, jij bent degene die er voor zorgt dat we niet allemaal vast roesten achter ons bureau en af en toe een wandeling maken tijdens de lunch.

To all my co-authors and colleagues, thank you for all your efforts and feedback. You did not only help improve my manuscripts, but also helped me become a better researcher and person.

During my PhD, there is one thing I consistently did almost every day: CrossFit. Where most people would call me crazy, I always enjoyed the workouts where my heart rate

would jump to 200 and I felt like I couldn't continue anymore (well, I enjoyed it afterwards). When I started Crossfi,t I always went to the 7 AM class. **Peter** you always managed to push me just a little bit further, and thank you for knowing what to tell me to grab that extra kg or do that extra rep. I was never alone in those classes. **Santi**, I could always tell when you were present. First of all, because you are just as competitive as me (I always kept an eye on you if I were ahead of you) and second of all, because you are pretty loud ;). We were usually accompanied by **Namki** who, only after 2.5 years, told us how to correctly pronounce his name. It was fun to have a buddy to learn all the movements with. Where all the 'pros' could do the movements correctly, you and I usually did them strict and fell behind after 5 reps. I was honored when you asked me to watch Bangle and Liz and I really enjoyed staying at your hyper-modern house and playing with him during summer. Every class I generally had one goal: trying to beat **Fred**. Because we are both competitive I learned a lot. Besides, your knowledge and skills about all the movements really helped me. Thank you for teaching me so much and for designing all our open gym workouts. It was really relaxing to be able to arrive and simply be told what to do! **Pablito** we share our passion for snowboarding and I really hope we will keep doing this for many years to come. Japow is coming!!!! Your relaxed attitude is very welcome when are together with many loud people. Finalmente il mio amico italiano **Enrico**. È davvero bello parlare della nostra passione per il cibo. Grazie per aver condiviso le tue ricette e per avermi permesso di farti tante domande. Inoltre, grazie per avermi aiutato con l'italiano, so che non deve essere facile sentirmi parlare di pistasjoos. Becoming more fit also meant I had to become strongerm, and who better to do that with than with **Ånne** and **Simon** while being coached by our own personal mountain goatje **Chris**. The Tuesday and Thursday evenings became three things: 1. A moment to let off steam and throw around some heavy weights; 2. A time to turn off the thoughts of that day and laugh so much (while of course being told to not be a **** and to lift heavier by Chris); 3. To move so many times to a new house in PR city. To all of you, thank you so much for the talks during the workouts, after the workouts, and for keeping me sane. Without this routine and without the ability to clear my head before or after a day, I would definitely not have managed to reach the end.

Naast CrossFit was hockey mijn andere uitlaatklep. Aan al mijn teamgenoten, dank jullie wel voor de vele sportieve jaren en voor het accepteren van mijn soms iets te fanatieke persoonlijkheid. **Tom**, Duff, je was zo een prettige aanvoerder die altijd eerlijk en duidelijk was. Ik heb veel bewondering voor je hoe je steeds met twee petten in het veld stond en altijd fair handelde (tot soms mijn frustratie aan toe). Maar je weet vaak op een tactische manier mijn frustratie om te zetten naar iets positiefs. Dankje voor alle sportiviteit, je ongekend droge humor, je perfecte nieuwe namen voor oefeningen en het delen in de passie voor lekker koken en eten.

De **#MAMA** boys zijn er al lange tijd bij. Het begon uiteraard met el parasito die bijna dagelijks spontaan bij Erik, Wouter en mij op de stoep stond. De groep werd al snel uitgebreid met Lars, Nico, Jelle en Harri. **Lars**, **Wouter** en **Erik**, door jullie heb ik in Enschede toch nog een beetje van Audentis kunnen proeven, ondanks dat ik er niet bij zat. Ik ben jullie hier ontzettend dankbaar voor dat jullie mij zo vaak op sleeptouw mee

hebben genomen. Al was ik officieel geen deel van jullie jaarclubs, ik voelde me wel altijd welkom en gewaardeerd. Dus oprechte dank daarvoor. En tot mijn geluk wilden twee van jullie (en soms **Martine** ook) met mij in Utrecht gaan wonen. Het was fijn dat ze er soms was, want dan kon ze met haar soms kritische en zorgzame vragen ons zo nu en dan in het gareel houden. Ik heb een ontzettend fijne tijd met jullie gehad. Het maakte nooit uit wat we deden, onderweg met el tankino om te zeilen in Heeg, rum cake eten met de politie of over Ibiza te zingen terwijl we Risk speelden, het was altijd leuk en gezellig dan wel interessant. En dan had ik ook nog het geluk om met Wouter in het hockeyteam te spelen. De wedstrijden waren leuk, maar de TDs en feestjes naderhand waren vaak legendarisch. Nog nooit heb ik iemand zo vaak zien scoren via LinkedIn. En als het dan @maandag avond mannen avond was, zorgde **Jelle** er wel voor dat we een locatie hadden waar we konden genieten van foto's van 'mr vakantiedagen' **Nico**. Boys, aan jullie allemaal, dank je voor het aanhoren van mijn geklaag. Thanks dat ik nooit in het midden van de auto hoefde te zitten. Thanks voor de feestjes, borrels, cabaret avonden en toekan uitjes. Ik hoop echt dat we de maandag traditie er nog jaren in zullen houden!

Lieve **pap**, **mam**, **Yannick** en **Arvid** jullie zullen zeggen dat ik dit zelf heb gedaan, maar we weten allemaal dat ik dit zonder jullie dit nooit had gehaald. Jullie continue onvoorwaardelijke steun en liefde hebben mij gevormd tot de persoon die ik nu ben. Het gaat me niet lukken om uit te leggen wat jullie voor me betekenen en daarom is dit het belangrijkste: Jullie zijn geweldig, ik houd van jullie, dankjulliewel <3. Het is fijn om een 'team' achter je te hebben waar je altijd op kan rekenen. Moet ik iets economisch weten, dan klop ik bij de Excel wizard aan. Wil ik iets meer wiskundigs weten, dan ga ik naar Vido. Wil ik even lachen of over iets persoonlijks praten, dan klop ik weer bij **Yannick** aan. En wil ik meer over de toekomst nadenken, dan ga ik weer naar **Arvid**. Papa en mama hebben het geweldig gedaan om ons te vormen en 'volwassen' te laten worden en we weten elkaar perfect aan te vullen. Broertjes, ga zo door, jullie komen er wel! **Mam**, best gek om iets naar je te schrijven terwijl ik weet dat je dit nooit gaat lezen. Zonder jouw levensleus 'nee heb je, ja kan je krijgen', was ik zo ontzettend veel dingen misgelopen. Het heeft vele deuren voor me geopend die anders door mijn soms ietwat verlegen karakter nooit open waren geweest. Waar ik soms meer introvert ben, was jij juist levendig en lekker gek. Jouw kunstige, energieke en zorgzame karakter zal ik blijven koesteren en voor altijd bij me houden, waarbij ik ook hopelijk wat van jouw doorzettingsvermogen en vastberadenheid heb geërfd. Ik mis je nog steeds elke dag. I love you. **Pap**, zodra ik het allemaal echt niet meer weet, klop ik bij jou aan: "het orakel". Je weet altijd raad, want advies zal je nooit geven ;). En ik weet dat jij vindt dat die dankwoorden van tegenwoordig veel te lang zijn en ik zal je dan ook sparen. Je hebt er voor gezorgd dat ik goed nadenk en mijzelf voorbereid voordat ik iets doe. Door jou kan ik dingen 'zelf' doen en kijk ik op een scherpe en realistische manier om me heen. Jouw eigenschappen en wat ik van jou geleerd heb, zijn bijzonder complementair aan die van mama. Ik hoop dat, als ik later groot ben, ik net zo ben als jullie. Kortom, ik heb geluk gehad en ik houd van jullie vier.

Lieve **Liz**, wellicht ben jij het beste wat ik heb gevonden tijdens mijn PhD. Zonder jouw liefde en zorgzaamheid was het onmogelijk geweest mijn PhD af te ronden. Waar ik vaak analytisch nadenk en minder mijn gevoel laat spreken, wist jij op de juiste momenten door te vragen zodat daar juist ruimte voor was. Ik weet dat ik daar niet altijd makkelijk in ben en dat je je momenten dan ook zorgvuldig uitkoos. Op deze manier weet je precies te sturen tussen 'laat hem maar ff' en 'dit denk/wil je echt'. Je zorgt er ook voor dat ik rustig word op momenten dat het te veel voor me is, waardoor ik kan relativeren. Je bent slim, gedreven en een ontzettende nerd (net als ik), je bent sportief, competitief en rustig. Je bent zorgzaam en meedenkend waarbij ik je soms probeer te laten inzien dat je ook wel eens voor jezelf mag kiezen. Je laat me inzien wanneer ik fout zit, al gebeurt dat natuurlijk vrijwel nooit. En wat ik wellicht het fijnst vind, is het warme, gelukkige en gewaardeerde gevoel dat ik krijg als je naar me lacht (iets wat gelukkig vaak gebeurt). Waar ik altijd rustig moet opstaan en 'grumpy' het bed uit kom, ben jij meteen vrolijk en energiek, wat voor een gezellige dynamiek zorgt. Maar als we gaan slapen, zijn de rollen omgedraaid en hoop je maar al te graag dat ik rustig naast je kom liggen. Waar ik het leuk vind om grote reizen te maken, vind jij dat een beetje spannend. Waar ik graag binnen wil blijven, zorg jij ervoor dat we naar buiten gaan. Waar we allebei graag een hondje willen, weten we gelukkig ook dat dat er nog wel komt. Het helpt dat we soms verschillend zijn, want daardoor kunnen we elkaar perfect aanvullen. Het maakt niet uit wat we doen, waar we zijn of hoe we ons voelen, als we samen zijn is het altijd een prachtig avontuur. Dankjewel voor alles, ik houd van jou!