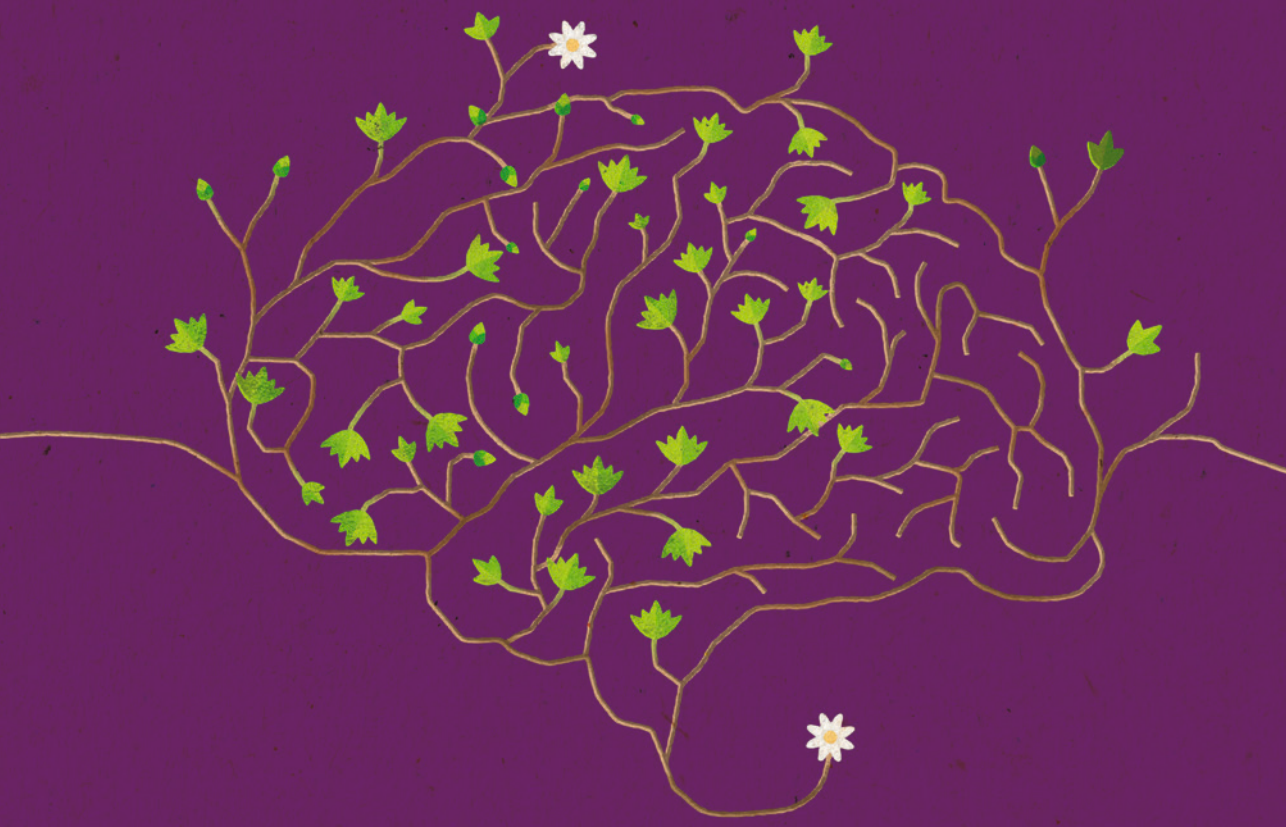
Reshaping glioma imaging via non-invasive non-enhanced MR imaging methods: the GLIOCARE project



Aynur Azizova

Reshaping glioma imaging via non-invasive non-enhanced MR imaging methods: the GLIOCARE project

Aynur Azizova

Financial support for printing of this thesis was kindly provided by Cancer Center Amsterdam.

Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Erwin Timmerman, persoonlijkproefschrift.nl

ISBN: 978-94-6522-208-0
https://doi.org/10.5463/thesis.1180

VRIJE UNIVERSITEIT

RESHAPING GLIOMA IMAGING VIA NON-INVASIVE NON-ENHANCED MR IMAGING METHODS: THE GLIOCARE PROJECT

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 woensdag 25 juni 2025 om 11.45 uur
in de universiteit

door

Aynur Azizova

geboren te Lerik, Azerbeidzjan

promotor:	prof.dr. F. Barkhof
copromotoren:	dr. J. Petr dr. V.C.W. Keil
promotiecommissie:	prof.dr. P. Wesseling prof.dr. M. Smits prof.dr.habil. T. Yousry dr. L. Douw prof.dr. E. Öztürk Işık dr. N. Sollmann

TABLE OF CONTENTS

Chapter 1

Introduction and Thesis Outline

Gliomas

Primary central nervous system (CNS) tumors rank as the eighth most common cancer in adults, with diffuse gliomas being the most frequent type of intra-axial malignant tumors[1]. Gliomas are thought to arise from neural stem or progenitor cells, though their exact etiology and contributing factors remain unknown[2]. Symptoms of gliomas are often nonspecific; headaches are more common in high-grade gliomas, while seizures are typical for low-grade gliomas[1]. Despite improvements in patient management aimed at early diagnosis, effective treatment, and continuous monitoring, outcomes remain poor. The median survival for glioblastoma, the most common and malignant subtype of adult-type diffuse gliomas, is approximately 12 months, while for low- and intermediate-grade gliomas, survival can range from 1 to 15 years[1,3].

Recent advancements in neuro-oncology have shed light on the complex molecular mechanisms driving glioma oncogenesis, significantly influencing patient outcomes, including disease progression and treatment responses. The World Health Organization (WHO) classification for CNS tumors incorporated these molecular alterations to more accurately identify glioma subtypes, starting from its fourth edition in 2016[4]. The final fifth edition of this classification, published in 2021, divides gliomas into three main categories: adult-type diffuse gliomas, pediatric-type diffuse low- and high-grade gliomas, and circumscribed astrocytic gliomas. Each group requires testing for specific molecular changes, with isocitrate dehydrogenase (IDH) mutations being critical for adult-type diffuse gliomas[5].

Adult-type diffuse gliomas are further classified into glioblastoma, IDH-wildtype, astrocytoma, IDH-mutant, and oligodendroglioma, IDH-mutant and 1p/19q-codeleted[5]. This subtyping corresponds to patient outcomes and guides treatment strategies. Histological features are still used for further grading within these subtypes. For example, necrosis and vascular proliferation define grade 3 in oligodendrogliomas and grade 4 in astrocytomas. However, molecular alterations now take precedence over histological features. For instance, a microscopically low-grade IDH-wildtype diffuse glioma, which lacks necrosis or vascular proliferation but shows molecular markers such as gain of chromosome 7 and loss of chromosome 10, telomerase reverse transcriptase promoter mutation, or epidermal growth factor receptor amplification, is reclassified as

grade 4 glioblastoma, IDH-wildtype. Additionally, the homozygous deletion of cyclin-dependent kinase inhibitor 2A/2B in astrocytomas automatically classifies the tumor as grade 4, regardless of histological features[1,5].

Magnetic resonance imaging (MRI) in glioma management

MRI is fundamental to managing glioma, serving as the primary non-invasive modality for detecting and characterizing these infiltrative brain tumors by providing structural and physiological insights from initial diagnosis to treatment planning and post-treatment monitoring[6,7]. MRI methods can be categorized into two complementary approaches: conventional and advanced[8,9]. Conventional MRI includes pre-contrast T1-weighted, T2-weighted, T2-FLAIR, and post-contrast-T1 weighted sequences, which offer detailed information on tumor morphology, such as location, necrosis, or enhancement pattern[6]. Advanced MRI techniques offer additional physiological insights into gliomas[10,11]. For instance, diffusion-weighted imaging (DWI) provides insight into cellular density[12], magnetic resonance spectroscopy (MRS) reveals the tumor's chemical composition[13], and dynamic susceptibility contrast perfusion imaging (DSC-MRI) assesses the vascular structure of the tumor[14].

Glioma MRI follows a structured, standardized brain tumor imaging protocol (BTIP) to ensure consistency and generalizability. Conventional MRI, including post-contrast T1-weighted images, and DWI are the core components of the BTIP. The interpretation and reporting of glioma MRI are further standardized through the Visually Accessible Rembrandt Images (VASARI) set, which provides a common language for describing glioma characteristics, such as enhancement features or diffusion properties[15]. Importantly, the response assessment in neuro-oncology criteria (RANO)[7] reinforce the necessity of MRI evaluation, alongside clinical findings and steroid treatment dose, in the post-treatment decision-making process.

Gadolinium-based contrast agents (GBCAs) as MRI contrast media

GBCAs have been the most widely used MRI contrast media since their introduction nearly four decades ago. These agents contain gadolinium, a rare earth element with seven unpaired electrons, which makes them highly effective in accelerating proton relaxation and increasing T1-weighted signal intensity in regions with a disrupted blood-brain barrier[16]. Of the seven FDA-approved

GBCAs, gadobutrol and gadoterate meglumine are the most commonly used in neuro-oncological imaging. Post-contrast T1-weighted sequence utilizing GBCA is integral to MRI protocols for glioma imaging, with the RANO 2.0 criteria identifying them as the most sensitive and reliable MRI sequence for assessing the treatment response, particularly in high-grade gliomas[6,7]. Furthermore, DSC-MRI, a widely used perfusion imaging technique, also uses GBCAs for glioma evaluation[10].

Implications of GBCAs

GBCAs are associated with various safety, environmental, and economic implications, necessitating careful consideration. Their safety concerns include allergic-like reactions, nephrogenic systemic fibrosis, and gadolinium deposition[16,17]. Acute allergic-like reactions are the most common, typically mild, but severe cases can occur infrequently. Nephrogenic systemic fibrosis, a rare yet potentially fatal condition, primarily affects patients with chronic kidney disease or acute kidney injury. Gadolinium deposition in the body, including the brain, has been observed even in patients with normal renal function, although its clinical impact remains unclear[16,18]. In response to these risks, the European Medicines Agency has restricted certain GBCAs, and clinical guidelines emphasize the need for scrutinized clinical justification for their use[16,18–20]. Additionally, special precautions are required when using GBCAs in vulnerable populations, such as pregnant or breastfeeding women and children, due to uncertain risk profiles[21–24].

Environmental implications are also significant, as gadolinium contamination of water bodies, called anthropogenic gadolinium, has been documented due to the disposal and excretion of GBCAs. This contamination negatively affects aquatic organisms, including plants and marine life[25–27]. Anthropogenic gadolinium has been detected in drinking water in countries such as Germany and Poland within Europe and in the surface waters of different canals and rivers in the Netherlands, raising concerns about potential adverse health effects on the human population[28–31]. Economically, using GBCAs contributes to the financial burden on healthcare systems. A cost-effectiveness analysis by Crowson et al.[32] found that incorporating GBCAs into MRI protocols for vestibular schwannoma resulted in a significant cost escalation from $2872 to $4089. The

economic considerations surrounding GBCAs also influence their usage in low- and middle-income countries, where MRI consumables are often limited[33].

The GLIOCARE project

GLIOCARE project (medical ethics review committee number: VUmc_2021-0437), which stands for "Glioma Imaging Omitting Contrast through Artificial Intelligence and Risk Evaluation," focuses on replacing GBCA-enhanced MRI with non-contrast MRI techniques, such as pre-contrast conventional MRI, DWI, arterial spin labeling (ASL) and amide-proton transfer chemical exchange saturation transfer (APT-CEST), for glioma imaging. This project has retrospective and prospective arms relying on human-rater performance and artificial intelligence (AI) analysis. The retrospective arm using human rater performance via conventional pre-contrast MRI sequences and DWI is a central focus of this thesis.

The research hypothesis for this thesis

Despite known risks to patient safety, economic concerns, and the need for careful risk-benefit assessment in vulnerable populations, GBCA-enhanced MRI remains the standard for glioma imaging supported by protocols like BTIP[6] and RANO criteria[7]. Yet, the evidence justifying this practice is surprisingly limited. Traditionally, contrast enhancement was associated with high-grade gliomas. However, research shows that enhancement is not exclusive to high-grade tumors as low-grade tumors may also enhance, and high-grade gliomas may lack enhancement[34,35]. This variability raises questions about the necessity and superiority of contrast-enhanced imaging, as hardly any literature demonstrates its clear advantage over GBCA-free imaging.

The RANO 2.0 criteria[7] underscore the importance of T2-weighted and FLAIR images in post-treatment evaluations. However, these are limited to non-enhancing gliomas, and post-contrast T1-weighted images remain the benchmark for enhancing tumors. Advances in AI have shown potential in generating synthetic post-contrast images from GBCA-free scans[36], but these innovations have yet to translate into clinical practice. Moreover, human raters' potential for predicting enhancement features or diagnostic information using pre-contrast images has not been investigated.

Given these issues, there is a growing interest in exploring the diagnostic potential of GBCA-free imaging in neuro-oncology. This thesis hypothesizes that human raters can make accurate diagnostic assessments based solely on GBCA-free images and that all imaging features typically seen in contrast-enhanced images can be inferred through a structured analysis of pre-contrast images.

Thesis aims and outline

The overall aim of this thesis is to assess the role of GBCAs in the management of adult-type diffuse glioma patients by evaluating the predictive accuracy of human raters using GBCA-free MRI techniques and exploring whether non-contrast MRI can safely replace contrast-enhanced MRI without compromising diagnostic precision. In Chapter 2, based on a comprehensive literature review, I provide state-of-the-art information on the current and future opportunities and challenges related to omitting or reducing GBCAs in primary brain tumors. In Chapter 3, I evaluate the performance of the VASARI set, a standardized glioma imaging vocabulary, through a systematic review and meta-analysis, as it has been utilized to develop human prediction algorithms in this thesis. In Chapter 4, I aim to determine whether the contrast enhancement patterns of gliomas, indicative of blood-brain barrier disruption, can be accurately predicted by human raters using GBCA-free MRI. In Chapter 5, I assess the preoperative diagnostic accuracy of human raters for adult-type diffuse glioma subtypes by directly comparing GBCA-free MRI with GBCA-enhanced MRI. In Chapter 6, I evaluate the application of DWI, a GBCA-free imaging technique, in preoperative glioma management by comparing visual and region-of-interest-based assessments to analyze their correlation, reproducibility, and impact on diagnostic decision-making. Finally, in Chapter 7, I summarize and discuss all the studies this thesis presents in the context of current literature, presenting alternative approaches and emphasizing future implications.

REFERENCES

1. van den Bent MJ, Geurts M, French PJ, et al. Primary brain tumours in adults. *Lancet* 2023;402:1564–79.
2. Weller M, Wen PY, Chang SM, et al. Glioma. *Nat Rev Dis Primers* 2024;10:33.
3. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RGW, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* 2015;372:2481–98.
4. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of tumors of the Central Nervous System: A summary. *Acta Neuropathol* 2016;131:803–20.
5. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the Central Nervous System: A summary. *Neuro Oncol* 2021;23:1231–51.
6. Sanvito F, Kaufmann TJ, Cloughesy TF, et al. Standardized brain tumor imaging protocols for clinical trials: current recommendations and tips for integration. *Front Radiol* 2023;3:1267615.
7. Wen PY, van den Bent M, Youssef G, et al. RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults. *J Clin Oncol* 2023;41:5187–99.
8. Hu LS, Smits M, Kaufmann TJ, et al. Advanced imaging in the diagnosis and Response Assessment of high-grade glioma: AJR Expert Panel Narrative review. *AJR Am J Roentgenol* https://doi.org/10.2214/AJR.23.30612.
9. Fouke SJ, Benzinger T, Gibson D, et al. The role of imaging in the management of adults with diffuse low grade glioma: A systematic review and evidence-based clinical practice guideline. *J Neurooncol* 2015;125:457–79.
10. Hirschler L, Sollmann N, Schmitz-Abecassis B, et al. Advanced MR Techniques for Preoperative Glioma Characterization: Part 1. *J Magn Reson Imaging* 2023;57:1655–75.
11. Hangel G, Schmitz-Abecassis B, Sollmann N, et al. Advanced MR Techniques for Preoperative Glioma Characterization: Part 2. *J Magn Reson Imaging* 2023;57:1676–95.
12. Huisman TAGM. Diffusion-weighted and diffusion tensor imaging of the brain, made easy. *Cancer Imaging* 2010;10 Spec no A:S163–71.
13. Suh CH, Kim HS, Jung SC, et al. 2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and meta-analysis using individual patient data. *Neuro Oncol* 2018;20:1573–83.
14. Pons-Escoda A, Garcia-Ruiz A, Naval-Baudin P, et al. Differentiating IDH-mutant astrocytomas and 1p19q-codeleted oligodendrogliomas using DSC-PWI: high performance through cerebral blood volume and percentage of signal recovery percentiles. *Eur Radiol* 2024;34:5320–30.
15. VASARI research project - the cancer imaging archive (TCIA) public access - cancer imaging archive wiki.
16. Starekova J, Pirasteh A, Reeder SB. Update on Gadolinium Based Contrast Agent Safety, From the AJR Special Series on Contrast Media. *AJR Am J Roentgenol* https://doi.org/10.2214/AJR.23.30036.
17. Cunningham A, Kirk M, Hong E, et al. The safety of magnetic resonance imaging contrast agents. *Front Toxicol* 2024;6:1376587.
18. Blomqvist L, Nordberg GF, Nurchi VM, et al. Gadolinium in Medical Imaging-Usefulness, Toxic Reactions and Possible Countermeasures-A Review. *Biomolecules* 2022;12.
19. McDonald RJ, Levine D, Weinreb J, et al. Gadolinium Retention: A Research Roadmap from the 2018 NIH/ACR/RSNA Workshop on Gadolinium Chelates. *Radiology* 2018;289:517–34.
20. Bendszus M, Laghi A, Munuera J, et al. MRI gadolinium-based contrast media: Meeting radiological, clinical, and environmental needs. *J Magn Reson Imaging* 2024;60:1774–85.
21. Proença F, Guerreiro C, Sá G, et al. Neuroimaging safety during pregnancy and lactation: a review. *Neuroradiology* 2021;63:837–45.

22. van Westrhenen A, Senders JT, Martin E, et al. Clinical challenges of glioma and pregnancy: a systematic review. *J Neurooncol* 2018;139:1–11.

23. Blumfield E, Moore MM, Drake MK, et al. Survey of gadolinium-based contrast agent utilization among the members of the Society for Pediatric Radiology: a Quality and Safety Committee report. *Pediatr Radiol* 2017;47:665–73.

24. Ouyang M, Bao L. Gadolinium contrast agent deposition in children. *J Magn Reson Imaging* https://doi.org/10.1002/jmri.29389.

25. Zhang Z, Jiang W, Gu T, et al. Anthropogenic gadolinium contaminations in the marine environment and its ecological implications. *Environ Pollut* 2024;359:124740.

26. Lenkinski RE, Rofsky NM. Contrast media-driven anthropogenic gadolinium: Knowns and unknowns. *Radiology* 2024;311:e240020.

27. Piarulli S, Riedel JA, Fossum FN, et al. Effects of gadolinium (Gd) and a Gd-based contrast agent (GBCA) on early life stages of zebrafish (Danio rerio). *Chemosphere* 2024;350:140950.

28. Dekker HM, Stroomberg GJ, Van der Molen AJ, et al. Review of strategies to reduce the contamination of the water environment by gadolinium-based contrast agents. *Insights Imaging* 2024;15:62.

29. Schmidt K, Bau M, Merschel G, et al. Anthropogenic gadolinium in tap water and in tap water-based beverages from fast-food franchises in six major cities in Germany. *Sci Total Environ* 2019;687:1401–8.

30. Wysocka IA, Rogowska AM, Kostrz-Sikora P. Investigation of anthropogenic gadolinium in tap water of polish cities: Gdańsk, Kraków, Warszawa, and Wrocław. *Environ Pollut* 2023;323:121289.

31. Klaver G, Verheul M, Bakker I, et al. Anthropogenic Rare Earth Element in rivers: Gadolinium and lanthanum. Partitioning between the dissolved and particulate phases in the Rhine River and spatial propagation through the Rhine-Meuse Delta (the Netherlands). *Appl Geochem* 2014;47:186–97.

32. Crowson MG, Rocke DJ, Hoang JK, et al. Cost-effectiveness analysis of a non-contrast screening MRI protocol for vestibular schwannoma in patients with asymmetric sensorineural hearing loss. *Neuroradiology* 2017;59:727–36.

33. Anazodo UC, Ng JJ, Ehiogu B, et al. A framework for advancing sustainable magnetic resonance imaging access in Africa. *NMR Biomed* 2023;36:e4846.

34. Castet F, Alanya E, Vidal N. Contrast-enhancement in supra- tentorial low-grade gliomas: a classic prognostic factor in the molecular age. *J Neurooncol* 2019;143:515–23.

35. Eichberg DG, Di L, Morell AA, et al. Incidence of high grade gliomas presenting as radiographically non-enhancing lesions: experience in 111 surgically treated non-enhancing gliomas with tissue diagnosis. *J Neurooncol* 2020;147:671–9.

36. Xie H, Lei Y, Wang T, et al. Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision. *Med Phys* 2022;49:3278–87.

Chapter 2

Brain Tumor Imaging without Gadolinium-Based Contrast Agents: Feasible or Fantasy?

Wamelink, I.J.H.G., Azizova, A., Booth, T.C., Mutsaerts, H.J.M.M., Ogunleye, A., Mankad, K., Petr, J., Barkhof, F., Keil, V.C.

Published: ***Radiology***
2024; vol. 310,2:e230793; doi: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.

Summary

Dose reduction and clinically driven imaging are measures to reduce gadolinium-based contrast agent exposure in primary brain tumor imaging without a negative impact on patient care; advanced MRI and artificial intelligence techniques will expand options for contrast agent-free imaging even more.

Essentials

- Some glioma imaging guidelines include suggestions on how to limit imaging frequency and gadolinium-based contrast agent (GBCA) use for select cases.
- Several studies have shown that GBCA dose reduction to 50-75% of the standard dose does not compromise diagnostic quality in gliomas and meningiomas.
- Contrast-enhanced (CE) MRI may provide hardly any additional information in pediatric lesions, according to retrospective studies.
- Advanced MR techniques, including diffusion-weighted imaging, amide-proton transfer chemical exchange saturation transfer, MR spectroscopy, and arterial spin labeling, seem to be promising alternatives to CE sequences.
- Synthetic postcontrast images generated using artificial intelligence from decreased-dose or non-GBCA sequences are promising alternatives to standard GBCA-enhanced scans.

INTRODUCTION

An MRI protocol for primary brain tumor imaging without contrast-enhanced (CE) sequences is generally considered insufficient for diagnostic purposes (1). By administering gadolinium-based contrast agents (GBCA), 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, relies on GBCA injection (Fig. 1).

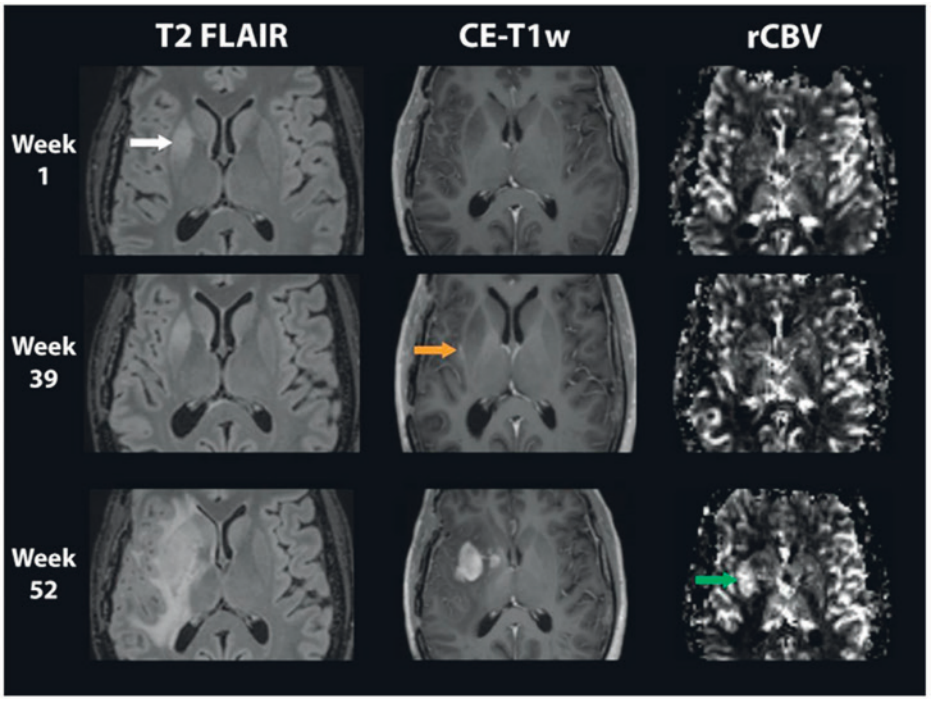


Figure 1: T2 fluid-attenuated inversion-recovery (FLAIR) images, contrast-enhanced T1-weighted (CE-T1w) images, and dynamic susceptibility contrast perfusion relative cerebral blood volume (rCBV) maps are pillars of primary brain tumor assessment. The example illustrates the independent value and behavior 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.

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-oncological 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).

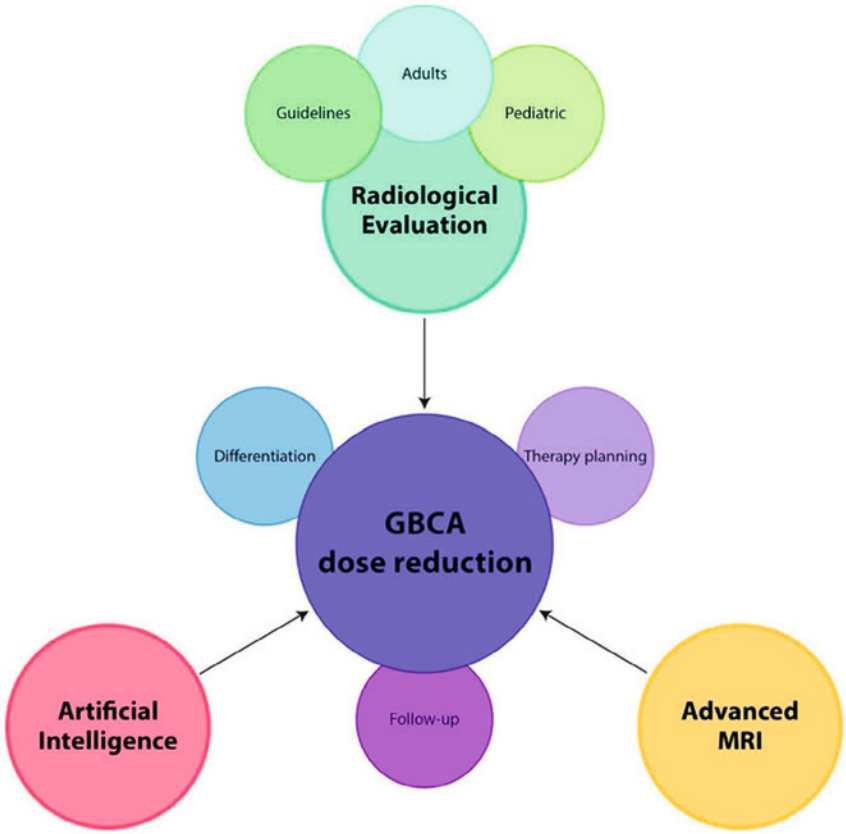


Figure 2: Diagram shows that radiologic evaluation practices (also expressed in guidelines), gadolinium-based contrast agent (GBCA)-free advanced MRI sequences, and artificial intelligence are the three pillars to help reduce GBCA use at all stages of neuro-oncologic MRI-based diagnostics.

The accuracy of radiologists and surgeons relying only on GBCA-free (or reduced) images is examined, as well as guideline recommendations on GBCA usage and imaging intervals. We also consider advanced MR imaging techniques probing tumor physiology 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).

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 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.

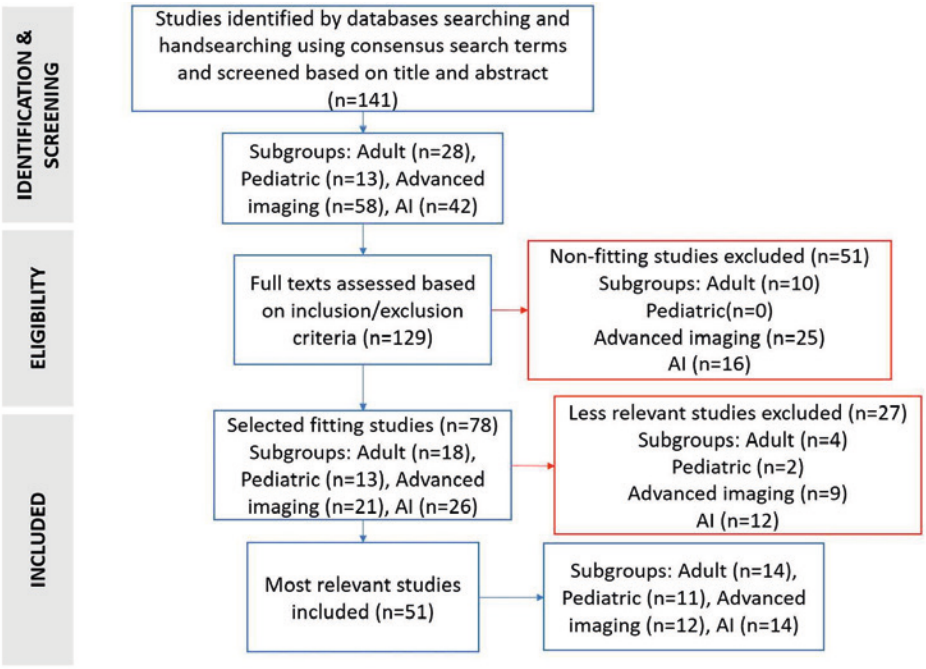


Figure 3: Flowchart visualizes the literature search conducted to shape the content of this review. Studies that did not contain a head-to-head comparison of a gadolinium-based contrast agent (GBCA)–dependent technique and a GBCA-free (or reduced) technique were considered less relevant. AI = artificial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

Table 1. Guidelines examined regarding GBCA

Guideline type and No.	Country or Society Name	Reference No.	GBCA-free or 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
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

Table 1. Guidelines examined regarding GBCA *(Continued)*

Guideline type and No.	Country or Society Name	Reference No.	GBCA-free or reduced Imaging or Longer Follow-up Intervals Mentioned
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
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)	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

Table 1. Guidelines examined regarding GBCA *(Continued)*

Guideline type and No.	Country or Society Name	Reference No.	GBCA-free or reduced Imaging or Longer Follow-up Intervals Mentioned
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
8	Joint meeting (January 30, 2014) among the FDA, NCI, clinical scientists, imaging experts, clinical trial 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 Neuro-oncology, 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.

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).

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 non-enhancing 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 last 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).

Arterial Spin Labeling

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.

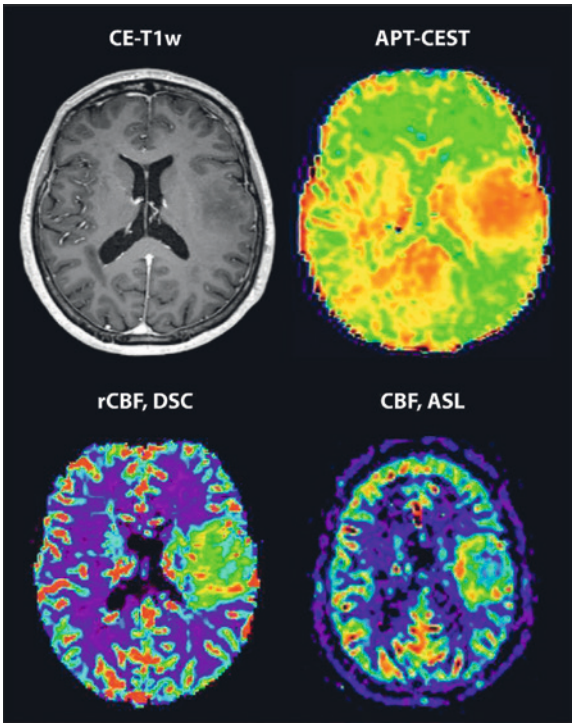


Figure 4: Contrast-enhanced T1-weighted (CE-T1w), amide-proton transfer chemical exchange saturation transfer (APT-CEST), dynamic susceptibility contrast (DSC), and arterial spin labeling (ASL) images in a nonenhancing left frontoinsular glioblastoma (isocitrate dehydrogenase wildtype). At the location 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.

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 K^{trans} (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).

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 (^{1}H) 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 ^{1}H 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 multi-center 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. Weller M, van den Bent M, Preusser M, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. Nat Rev Clin Oncol 2021;18(3):170–186. [Published correction appears in Nat Rev Clin Oncol 2022;19(5):357–358.]
2. Mehan WA Jr, González RG, Buchbinder BR, et al. Optimal brain MRI protocol for new neurological complaint. PLoS One 2014;9(10):e110803.
3. Falk Delgado A, Van Westen D, Nilsson M, et al. Diagnostic value of alternative techniques to gadolinium-based contrast agents in MR neuroimaging- a comprehensive overview. Insights Imaging 2019;10(1):84.
4. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro-oncol 2021; 23(8):1231–1251.
5. Consortium for Advancement of MRI Education and Research in Africa. CAMERA. https://www.cameramriafrica.org/aboutcamera. Accessed March 2023.
6. Deep L translator. https://www.deepl.com/translator. Accessed March 2023.
7. Gliomer hos voksne. DNOG. https://dnog.dk/assets/files/Retningslinier%20PDF/DNOG_gliomer_v3.0_AdmGodk250722.pdf. Published July 2022. Accessed March 2023.
8. Fangusaro J, Witt O, Hernáiz Driever P, et al. Response assessment in paediatric low-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. Lancet Oncol 2020;21(6):e305–e316.
9. Goldbrunner R, Stavrinou P, Jenkinson MD, et al. EANO guideline on the diagnosis and management of meningiomas. Neuro-oncol 2021;23(11):1821–1834.
10. Meningeomer. DNOG. https://dnog.dk/assets/files/Retningslinier%20 PDF/DNOG_Meningeom_v2.0_AdmGodk071221.pdf. Published December 2021. Accessed March 2023.
11. Liu BP, Rosenberg M, Saverio P, 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). AJR Am J Roentgenol 2021;217(5):1195–1205.
12. DeLano MC, Spampinato MV, Chang EY, et al. Dose-lowering in con- trast-enhanced MRI of the central nervous system: a retrospective, parallel-group comparison using gadobenate dimeglumine. J Magn Reson Imaging 2021;54(5):1660–1675.
13. Park SM, Kim HS, Jahng GH, Ryu CW, Kim SY. 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. Br J Radiol 2010;83(990):466–475.
14. Maynard J, Okuchi S, Wastling S, et al. World Health Organization grade II/III glioma molecular status: prediction by MRI morphologic features and apparent diffusion coefficient. Radiology 2020;296(1):111–121.
15. Leibetseder A, Leitner J, Mair MJ, et al. Prognostic factors in adult brain-stem glioma: a tertiary care center analysis and review of the literature. J Neurol 2022;269(3):1574–1590.
16. Wang QP, Lei DQ, Yuan Y, Xiong NX. Accuracy of ADC derived from DWI for differentiating high-grade from low-grade gliomas: systematic review and meta-analysis. Medicine (Baltimore) 2020;99(8):e19254.
17. Yao Y, Xu Y, Liu S, et al. Predicting the grade of meningiomas by clinical- radiological features: a comparison of precontrast and postcontrast MRI. Front Oncol 2022;12:1053089.
18. Tsai YT, Hung KC, Shih YJ, et al. Preoperative apparent diffusion coefficient values for differentiation between low and high grade meningiomas: an updated systematic review and meta-analysis. Diagnostics (Basel) 2022;12(3):630.

19. Li YM, Suki D, Hess K, Sawaya R. The influence of maximum safe resection of glioblastoma on survival in 1229 patients: can we do better than gross- total resection? J Neurosurg 2016;124(4):977–988.

20. Duma CM, Kim BS, Chen PV, 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. J Neurosurg 2016;125(Suppl 1):40–49.

21. Du X, He Q, Zhang B, Li N, Zeng X, Li W. Diagnostic accuracy of diffusion-weighted imaging in differentiating glioma recurrence from post-treatment-related changes: a meta-analysis. Expert Rev Anticancer Ther 2022;22(1):123–130.

22. Thompson G, Lawrie TA, Kernohan A, Jenkinson MD. Interval brain imaging for adults with cerebral glioma. Cochrane Database Syst Rev 2019;12(12):CD013137.

23. Mrowczynski OD, Zammar S, Bourcier AJ, et al. Utility of early postoperative magnetic resonance imaging after glioblastoma resection: implications on patient survival. World Neurosurg 2018;120:e1171–e1174.

24. Boto J, Guatta R, Fitsiori A, Hofmeister J, Meling TR, Vargas MI. Is contrast medium really needed for follow-up MRI of untreated intracranial meningiomas? AJNR Am J Neuroradiol 2021;42(8):1421–1428.

25. Tamrazi B, Mankad K, Nelson M, D'Arco F. Current concepts and challenges in the radiologic assessment of brain tumors in children: part 2. Pedi- atr Radiol 2018;48(13):1844–1860.

26. D'Arco F, Culleton S, De Cocker LJL, Mankad K, Davila J, Tamrazi B. Current concepts in radiologic assessment of pediatric brain tumors during treatment, part 1. Pediatr Radiol 2018;48(13):1833–1843.

27. Campion T, Quirk B, Cooper J, 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? Neuroradiology 2021; 63(6):953–958.

28. Dünger D, Krause M, Gräfe D, Merkenschlager A, Roth C, Sorge I. Do we need gadolinium-based contrast medium for brain magnetic resonance imaging in children? Pediatr Radiol 2018;48(6):858–864.

29. Marsault P, Ducassou S, Menut F, Bessou P, Havez-Enjolras M, Chateil JF. Diagnostic performance of an unenhanced MRI exam for tumor follow-up of the optic pathway gliomas in children. Neuroradiology 2019;61(6):711– 720. [Published correction appears in Neuroradiology 2019;61(6):721.]

30. Porto L, Kieslich M, Bartels M, Schwabe D, Zanella FE, Du Mesnil R. Leptomeningeal metastases in pediatrics: magnetic resonance image manifestations and correlation with cerebral spinal fluid cytology. Pediatr Int 2010;52(4):541–546.

31. Maloney E, Perez FA, Iyer RS, et al. Non-inferiority of a non-gadolinium-enhanced magnetic resonance imaging follow-up protocol for isolated optic pathway gliomas. Pediatr Radiol 2022;52(3):539–548.

32. Maloney E, Stanescu AL, Perez FA, et al. Surveillance magnetic resonance imaging for isolated optic pathway gliomas: is gadolinium necessary? Pediatr Radiol 2018;48(10):1472–1484.

33. Malbari F, Chintagumpala MM, Wood AC, et al. Gadolinium is not necessary for surveillance MR imaging in children with chiasmatic-hypothalamic low-grade glioma. Pediatr Blood Cancer 2021;68(10):e29178.

34. Dodgshun AJ, Maixner WJ, Hansford JR, Sullivan MJ. Low rates of recurrence and slow progression of pediatric pilocytic astrocytoma after gross- total resection: justification for reducing surveillance imaging. J Neurosurg Pediatr 2016;17(5):569–572.

35. Hirschler L, Sollmann N, Schmitz-Abecassis B, et al. Advanced MR techniques for preoperative glioma characterization: part 1. J Magn Reson Imaging 2023;57(6):1655–1675. [Published correction appears in J Magn Reson Imaging 2023 Aug 11.]

36. Qu Y, Kong D, Wen H, 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. Eur Radiol 2022;32(5):2976–2987.

37. Wang X, Cao M, Chen H, Ge J, Suo S, Zhou Y. 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. Radiol Oncol 2020;54(3):301–310.

38. Morana G, Tortora D, Staglianò S, et al. Pediatric astrocytic tumor grading: comparison between arterial spin labeling and dynamic susceptibility contrast MRI perfusion. Neuroradiology 2018;60(4):437–446.

39. Wang YL, Chen S, Xiao HF, et al. Differentiation between radiation-induced brain injury and glioma recurrence using 3D pCASL and dynamic susceptibility contrast-enhanced perfusion-weighted imaging. Radiother Oncol 2018;129(1):68–74.

40. Lavrova A, Teunissen WHT, Warnert EAH, van den Bent M, Smits M. Diagnostic accuracy of arterial spin labeling in comparison with dynamic susceptibility contrast-enhanced perfusion for brain tumor surveillance at 3T MRI. Front Oncol 2022;12:849657.

41. Seeger A, Braun C, Skardelly M, et al. Comparison of three different MR perfusion techniques and MR spectroscopy for multiparametric assessment in distinguishing recurrent high-grade gliomas from stable disease. Acad Ra- diol 2013;20(12):1557–1565.

42. Hou H, Diao Y, Yu J, 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. NMR Biomed 2023;36(1):e4821.

43. Choi YS, Ahn SS, Lee SK, 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. Eur Radiol 2017;27(8):3181–3189.

44. Togao O, Hiwatashi A, Yamashita K, et al. Grading diffuse gliomas without intense contrast enhancement by amide proton transfer MR imaging: comparisons with diffusion- and perfusion-weighted imaging. Eur Radiol 2017;27(2):578–588.

45. Wang Q, Zhang H, Zhang J, et al. The diagnostic performance of magnetic resonance spectroscopy in differentiating high-from low-grade gliomas: a systematic review and meta-analysis. Eur Radiol 2016;26(8):2670–2684.

46. Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. 2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and meta-analysis using individual patient data. Neuro-oncol 2018;20(12):1573–1583.

47. Henriksen OM, Del Mar Álvarez-Torres M, Figueiredo P, 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 diffu- sion techniques. Front Oncol 2022;12:810263.

48. Wang Y, Wu W, Yang Y, et al. Deep learning-based 3D MRI contrast-enhanced synthesis from a 2D noncontrast T2Flair sequence. Med Phys 2022;49(7):4478–4493.

49. Jayachandran Preetha C, Meredig H, Brugnara G, et al. Deep-learning- based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. Lancet Digit Health 2021;3(12):e784–e794.

50. Moya-Sáez E, de Luis-García R, Alberola-López C. Toward deep learning replacement of gadolinium in neuro-oncology: a review of contrast-enhanced synthetic MRI. Front Neuroimaging 2023;2:1055463.

51. Pasquini L, Napolitano A, Pignatelli M, et al. Synthetic post-contrast imaging through artificial intelligence: clinical applications of virtual and augmented contrast media. Pharmaceutics 2022;14(11):2378.

52. Dong J, Li S, Li L, 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. Br J Radiol 2022;95(1129):20201302.

53. Cheng J, Liu J, Yue H, Bai H, Pan Y, Wang J. Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric MRI images. IEEE/ACM Trans Comput Biol Bioinformatics 2022;19(2):1084–1095.
54. Hashido T, Saito S, Ishida T. A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas. Sci Rep 2020;10(1):6121.
55. Alis D, Bagcilar O, Senli YD, et al. The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas. Clin Radiol 2020;75(5):351–357.
56. Zhang J, Yao K, Liu P, et al. A radiomics model for preoperative prediction of brain invasion in meningioma non-invasively based on MRI: a multicentre study. EBioMedicine 2020;58:102933.
57. Cagney DN, Sul J, Huang RY, Ligon KL, Wen PY, Alexander BM. The FDA NIH Biomarkers, EndpointS, and other Tools (BEST) resource in neuro-oncology. Neuro-oncol 2018;20(9):1162–1172.
58. Chang K, Bai HX, Zhou H, et al. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. Clin Cancer Res 2018;24(5):1073–1081.
59. Mammadov O, Akkurt BH, Musigmann M, et al. Radiomics for pseudo-progression prediction in high grade gliomas: added value of MR contrast agent. Heliyon 2022;8(8):e10023.
60. Booth TC, Grzeda M, Chelliah A, et al. Imaging biomarkers of glioblastoma treatment response: a systematic review and meta-analysis of recent machine learning studies. Front Oncol 2022;12:799662. [Published correction appears in Front Oncol 2023;13:1217461.]
61. Bacchi S, Zerner T, Dongas J, et al. Deep learning in the detection of high-grade glioma recurrence using multiple MRI sequences: a pilot study. J Clin Neurosci 2019;70:11–13.
62. Brain tumours (primary) and brain metastases in over 16s. National Institute for Health and CareExcellence. https://www.nice.org.uk/guidance/ng99/chapter/Recommendations#investigation-of-suspected-brain-metastases. Published July 11, 2018. Updated January 29, 2021. Accessed March 2023.
63. Sepúlveda-Sánchez JM, Muñoz Langa J, Arráez MÁ, et al. SEOM clinical guideline of diagnosis and management of low-grade glioma (2017). Clin Transl Oncol 2018;20(1):3–15. [Published correction appears in Clin Transl Oncol 2018;20(1):108–109.]
64. Balañá C, Alonso M, Hernandez-Lain A, et al. SEOM clinical guidelines for anaplastic gliomas (2017). Clin Transl Oncol 2018;20(1):16–21. [Published correction appears in Clin Transl Oncol 2018 Jul;20(7):937.]
65. Martínez-Garcia M, Álvarez-Linera J, Carrato C, et al. SEOM clinical guidelines for diagnosis and treatment of glioblastoma (2017). Clin Transl Oncol 2018;20(1):22–28.
66. Gliome im Erwachsenenalter. Onkopedia. https://www.onkopedia.com/ de/onkopedia/guidelines/gliome-im-erwachsenenalter/@@guideline/ html/index.html. Published August 2021. Accessed March 2023.
67. Référentiel Glioblastome (grade IV OMS). ANOCEF. https://www. onco-grandest.fr/wp-content/uploads/2020/07/anocef_referentiel_ glioblastome2018.pdf. Published January 1, 2018. Accessed March 2023.
68. Roth P, Hottinger AF, Hundsberger T, et al. A contemporary perspective on the diagnosis and treatment of diffuse gliomas in adults. Swiss Med Wkly 2020;150:w20256.
69. Optimalcarepathwayforpeoplewithhigh-gradeglioma:secondedition.Cancer Council. https://www.cancer.org.au/assets/pdf/high-grade-glioma-cancer-optimal-cancer-care-pathway#_ga=2.49569006.995159714.1632182743- 372209673.1632182743. Published June 2021. Accessed March 2023.
70. Nationellt vårdprogram tumörer i hjärna, ryggmärg och dess hinnor. Regionala Cancercentrum I Samverkan. https://kunskapsbanken. cancercentrum.se/diagnoser/hjarna/vardprogram/. Accessed March 2023.
71. Jiang T, Nam DH, Ram Z, et al. Clinical practice guidelines for the management of adult diffuse gliomas. Cancer Lett 2021;499:60–72.
72. Santosh V, Sravya P, Gupta T, et al. ISNO consensus guidelines for practical adaptation of the WHO 2016 classification of adult diffuse gliomas. Neurol India 2019;67(1):173–182.

73. Gliomen. Federatie Medisch Specialisten. https://richtlijnendatabase.nl/rich- tlijn/gliomen/startpagina_-_gliomen_2023.html. Accessed March 2023.

74. Intracranieel Meningeoom. Federatie Medisch Specialisten. https:// richtlijnendatabase.nl/richtlijn/intracranieel_meningeoom/diagnostiek/ beeldvorming.html. Accessed March 2023.

75. 1 章 成人膠芽腫. Japan Society of Clinical Oncology. http://www.jsco- cpg.jp/brain-tumor/guideline/#I. Accessed March 2023.

76. Клинические Рекомендации По Диагностике И Лечению Больных С Первичными Опухолями Головного Мозга. Russian Oncology As- sociation. https://oncology-association.ru/wp-content/uploads/2020/09/opukholi-golovnogo-mozga.pdf. Published 2014. Accessed March 2023.

77. Kim YZ, Kim CY, Lim DH. The overview of practical guidelines for gliomas by KSNO, NCCN, and EANO. Brain Tumor Res Treat 2022;10(2):83–93.

78. Thust SC, Heiland S, Falini A, et al. Glioma imaging in Europe: a survey of 220 centres and recommendations for best clinical practice. Eur Radiol 2018;28(8):3306–3317.

79. Erker C, Tamrazi B, Poussaint TY, et al. Response assessment in paediatric high-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. Lancet Oncol 2020;21(6):e317–e329. [Published correction appears in Lancet Oncol 2020 Aug;21(8):e372.]

80. Mohile NA, Messersmith H, Gatson NT, et al. Therapy for diffuse astrocytic and oligodendroglial tumors in adults: ASCO-SNO guideline. J Clin Oncol 2022;40(4):403–426.

81. Redjal N, Venteicher AS, Dang D, et al. Guidelines in the management of CNS tumors. J Neurooncol 2021;151(3):345–359.

82. Stupp R, Brada M, van den Bent MJ, Tonn JC, Pentheroudakis G; ESMO Guidelines Working Group. High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2014;25 Suppl 3:iii93–iii101.

83. Davies E, Hopkins A. Good practice in the management of adults with malignant cerebral glioma: clinical guidelines. Working Group, Royal College of Physicians. Br J Neurosurg 1997;11(4):318–330.

84. Ellingson BM, Bendszus M, Boxerman J, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol 2015;17(9):1188–1198.

APPENDIX S1

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:

Imaging interval extension
Intermitting GBCA-free MRI
Reduced dose of GBCA per scan
AI approaches to avoid GBCA injection or reduce the dose
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

- original research articles (defined as 11 cases or more)
- 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 dose-reduction/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, 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

Table S1. Overview of the References

Guidelines					
Reference (PMID)	Year	Tumor type(s)	Purpose	National/Society	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

Table S1. Overview of the References *(Continued)*

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

Table S1. Overview of the References *(Continued)*

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

Table S1. Overview of the References *(Continued)*

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 3

Ten Years of VASARI Glioma Features: Systematic Review and Meta-Analysis of Their Impact and Performance

Azizova, A., *Prysiazhniuk, Y., Wamelink, I.J.H.G., Petr, J., Barkhof, F., Keil, V.C.*

Published: ***American Journal of Neuroradiology***
2024; vol. 45,8:1053-1062; doi:10.3174/ajnr.A8274

ABSTRACT

Background: Visually Accessible Rembrandt (Repository for Molecular Brain Neoplasia Data) Images (VASARI) features, a vocabulary to establish reproducible terminology for glioma reporting, have been applied for a decade, but a systematic performance evaluation is lacking.

Purpose: Our aim was to conduct a systematic review and meta-analysis of the performance of the VASARI features set for glioma assessment.

Data Sources: MEDLINE, Web of Science, EMBASE, and Cochrane Library were systematically searched until September 26, 2023.

Study Selection: Original articles predicting diagnosis, progression, and survival in patients with glioma were included.

Data Analysis: The modified Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was applied to evaluate the risk-of-bias. The meta-analysis used a random effects model and forest-plot visualizations, if ≥5 comparable studies with a low or medium risk of bias were provided.

Data Synthesis: Thirty-five studies (3,304 patients) were included. Risk-of-bias scores were medium (n=33) and low (n=2). Recurring objectives were overall survival (n=18) and isocitrate dehydrogenase mutation (IDH; n=12) prediction. Progression-free survival was examined in seven studies. In 4 studies (glioblastoma n=2, grade 2/3 glioma n=1, grade 3 glioma n=1), a significant association was found between progression-free survival and single VASARI features. The single features predicting overall survival with the highest pooled hazard ratios were multifocality (hazard ratio 1.80; 95%-CI 1.21-2.67; I^2 = 53%), ependymal invasion (hazard ratio 1.73; 95%-CI: 1.45-2.05; I^2=0%), and enhancing tumor crossing midline (hazard ratio 2.08; 95%-CI: 1.35-3.18; I^2=52%). IDH mutation-predicting models combining VASARI features rendered a pooled area under the receiver operating characteristic curve of 0.82 (95%-CI: 0.76-0.88) at considerable heterogeneity (I^2=100%). Combined input models using VASARI plus clinical, and/or radiomics features outperformed single data type models in all relevant studies (n=17).

Limitations: Studies were heterogeneously designed and often of a small sample size. Several studies used The Cancer Imaging Archive database, with likely overlapping cohorts. The meta-analysis for IDH was limited due to a high study heterogeneity.

Conclusions: Some VASARI features perform well in predicting overall survival and IDH mutation status, but combined models outperform single features. More studies with less heterogeneity are needed to increase the evidence level.

SUMMARY

Previous Literature: The VASARI feature set for standardized glioma reporting has been applied by numerous studies to address different research questions. Applications range from human radiological evaluations using single features to multivariable machine-learning approaches. Clinical research questions encompass glioma subtype discrimination and non-invasive image-based survival prediction. Prediction of overall survival and IDH mutation status was among the most explored research questions. Several studies have identified multifocality, ependymal invasion, and enhancing tumor crossing the midline as unfavorable overall survival predictors. Enhancement quality, proportion of enhancing tumor, proportion of necrosis, and proportion of edema were identified as the main indicators of IDH mutation status.

Key Findings: This meta-analysis identified three robust VASARI features (multifocality, ependymal invasion, enhancing tumor crossing midline) to predict overall survival. Meta-analysis for IDH-predicting models showed a pooled AUC of 0.82 with considerable heterogeneity between studies. Combined models, including VASARI features next to clinical, genomics, and radiomics features, usually outperformed VASARI-only models.

Knowledge Advancement: Future studies should adhere to the original VASARI scoring definitions to minimize between-study heterogeneity. Given the time-consuming nature of manual extraction, it is crucial to develop automatic extraction technology. Reducing the feature set to the most promising ones can decrease the workload when radiologist input is required.

INTRODUCTION

MRI is the essential pillar of preoperative glioma diagnosis and later therapy assessment. While there is consensus regarding a standardized neuro-oncological imaging protocol[1], reporting of image aspects of glioma (radiophenotype) is less standardized. Criteria, as defined by response assessment in neuro-oncology (RANO 2.0)[2], comprise only limited criteria relying on quantitative measurements dedicated to the follow-up setting. In a study setting, however, a controlled reporting vocabulary is needed to identify reproducible imaging glioma biomarkers or to generate data input suitable for artificial intelligence approaches.

Approximately a decade ago, The Cancer Genome Atlas project of the National Cancer Institute addressed this problem by suggesting a controlled vocabulary for glioma imaging called Visually Accessible Rembrandt (Repository for Molecular Brain Neoplasia Data) Images (VASARI)[3]. The VASARI set combines different MRI features (Online Supplemental Data), such as enhancement pattern and tumor location. A consensus group defined the features based on expert opinion and literature. The VASARI project incorporated standard MRI sequences, including DWI, but excluded advanced imaging techniques, like PWI. The current set comprises 30 semantic features, three of which apply to postoperative situations. All features are rated on the basis of scoring systems (Online Supplemental Data).

Since its proposal, numerous studies have applied the VASARI set[4–38]. Applications range from human radiological evaluations using single features[5–8] to multivariable machine-learning approaches[21,27–29]. Clinical research questions cover glioma subtype discrimination[13,14,16,17] and non-invasive image-based survival prediction[7,10,13,22]. Various VASARI features were identified as prognostic factors, including tumor location, involvement of eloquent brain areas, ependymal or pial invasion, as well as diagnostic indicators, such as the definition of enhancing margin and proportion of necrotic or enhancing tumor, in these studies. However, these applications have never been systematically evaluated, perhaps explaining why using the VASARI features set is not recommended by any clinical or scientific guideline. As complete feature rating is time-consuming, a critical analysis identifying the most powerful features and models is

pivotal for future study designs and potential clinical use. Recent, however not VASARI-centered, publications indicated, e.g., pial and subependymal invasion (features 18 and 19) as negative for survival[39]. A meta-analysis may confirm the hypothesis of features 18 and 19 as particularly promising predictive VASARI features.

This systematic review and meta-analysis aims to gauge the performance of the VASARI set for glioma evaluation to identify a subset of the most diagnostic and prognostic predictive features to warrant usage in trials or even clinically.

MATERIALS AND METHODS

This study was registered in PROSPERO (ID for the published protocol: CRD42023392548) and was conducted according to Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)[40]. The research question was "What is the VASARI features' performance in predicting diagnosis, progression, and survival in patients with glial tumors?" using the participants, index tests, and target conditions (PIT) criteria[41].

Data sources and screening step

The search string had two components: VASARI and tumor types - including high-grade and low-grade oligodendroglioma, astrocytoma, and glioblastoma. Inclusion criteria were the following: 1) participants: adult or pediatric patients with glial tumors; 2) intervention/index test: human or automated methods using multiparametric conventional MRI as a source for VASARI; 3) comparison: standard interpretation of images with or without VASARI; 4) outcomes: prediction of the diagnosis, including histologic tumor grades and certain well-known mutations (isocitrate dehydrogenase (IDH), 1p/19q codeletion, telomerase reverse transcriptase promoter, and O(6)-methylguanine-DNA methyltransferase methylation status), progression, or survival; and 5) study design: original articles with a retrospective or prospective design.

Exclusion criteria were the following: 1) studies not performed on living humans; 2) studies with <10 participants (considered case series); 3) studies exclusively predicting other genetic/molecular alterations than indicated in the inclusion

criteria; 4) non-peer-review journals, conference abstracts, review papers, preprints; and 5) publications not in English.

A database search was performed on September 26, 2023, using MEDLINE (PubMed), Web of Science, EMBASE, and Cochrane Library. The search protocol is presented as Online Supplemental Data. One author screened the titles and abstracts (A.A., 7 years of radiology experience) for inclusion criteria after excluding duplicates using Rayyan software (https://rayyan.ai/reviews). Additionally, the reference lists of included articles were screened by title and abstract (hand search). Uncertain cases were evaluated in consultation with another author (V.C.K., 11 years of neuroradiology experience).

Data extraction

To guarantee identical rating standards, we used 5 studies to pilot the extraction process by 2 authors (A.A., V.C.K). Two authors then performed data extraction for the remaining studies (A.A., for radiological content, and Y.P., a second-year Ph.D. student in neuroscience with expertise in statistics). The Online Supplemental Data list the data extraction components.

Systematic review quality assessment

The quality of each study was evaluated by two authors (A.A., V.C.K) using the modified "Quality Assessment of Diagnostic Accuracy Studies-2" (QUADAS-2) tool[42] encompassing 5 domains for assessing the risk-of-bias and 3 domains for evaluating applicability concerns (Online Supplemental Data). Discrepancies were resolved by consensus. Five studies, also used to pilot data extraction, were piloted for quality control assessment to identify systematic discrepancies in QUADAS-2 tool use.

Meta-analysis and statistics

Studies examining patient cohorts with a similar tumor type, statistical models with comparable inputs, and reported identical endpoints of interest were grouped. The availability of ≥5 studies with a majority of QUADAS-2 categories scoring low or medium risk of bias was the liberal minimum for a meta-analysis. Otherwise, a narrative synthesis summarized the findings. If studies with overlapping cohorts were available, the total sample size was determined by including the largest cohort from overlapping studies (Online Supplemental Data). In

case of missing data, other reported metrics (standard errors, sample sizes)[43] served to derive parameters where feasible. Alternatively, corresponding authors were contacted and requested to provide supplementary information. Meta-analyses were conducted in R (Version 4.3.0.; http://www.r-project.org/) using generic inverse-variance and random-effect models to account for methodological between-study heterogeneity. To mitigate interpretability concerns arising from potential collinearity within VASARI datasets, the meta-analysis of hazard ratios (HRs) utilized metrics derived solely from univariable Cox proportional models. Heterogeneity was assessed with Higgins I^2 statistics, considering values above 50% as significant heterogeneity. All tests were two-sided. Because this is a meta-analysis, no correction was done for multiple testing. Forest plots were used to visualize results, complemented by pooled statistics with corresponding confidence and prediction intervals. The best-performing VASARI model was chosen if more than one model had been tested. The criteria for performance of meta-regression, sensitivity analysis, and non-reporting bias analysis was the availability of at least 10 studies to ensure the reliability of conclusions drawn from these analyses according to the Cochrane Handbook for Systematic Reviews of Interventions[44].

RESULTS

Overview

Figure 1 visualizes the literature selection. Thirty-five studies (29-335 patients) fulfilled the inclusion criteria (Online Supplemental Data). Eleven studies used The Cancer Imaging Archive cohort, possibly covering overlapping cohorts. The overlap-corrected number of recruited patients was 3,304 (Online Supplemental Data). The overall risk-of-bias scores were medium (n=33) and low (n=2; Online Supplemental Data). Only two studies applied the 2021 World Health Organization Classification of Central Nervous System Tumors (WHO CNS 5)[38] or reported IDH-wildtype glioblastoma only[16]. The remaining 33 studies explicitly used the 2016 or 2007 WHO CNS classifications (n=5[19,22,26,27,30]) or did not mention it (n=28[4–15,17,18,20,21,23–25,28,29,31–37]), resulting in applicability uncertainties regarding patient selection. One study[38] was vague in the use of the reference standard.

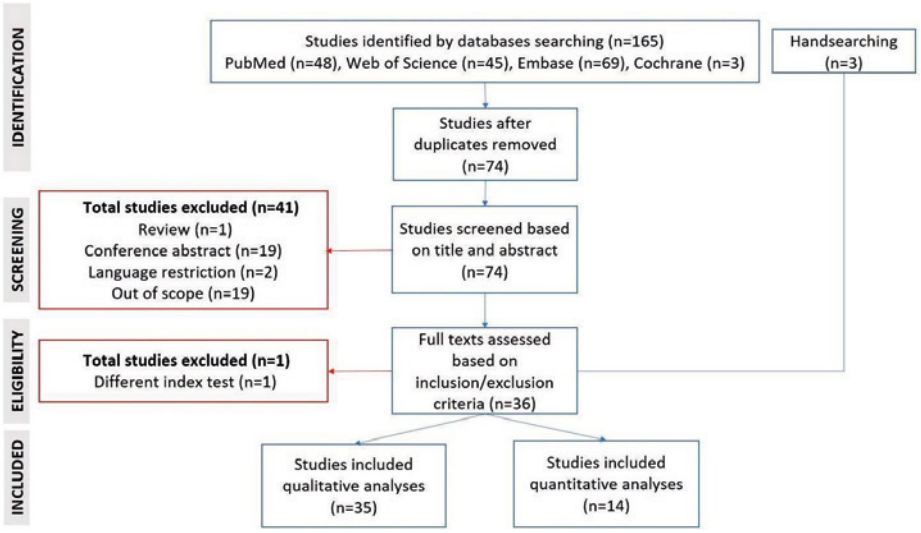


Figure 1: PRISMA flowchart describing the literature selection process.

Survival prediction

Fifteen[4,5,7–11,15,18,21,24,27,29,34,35] out of 18 overall survival (OS)-predicting studies exclusively involved glioblastoma. Results for univariable analyses indicating the per-feature performance of the VASARI set are shown in Figure 2a. Eleven glioblastoma studies found a significant association between OS and different VASARI features[5,7–10,15,18,24,29,34,35]. Comparing single data type models with only one input data type (e.g., clinical, pathological, FET-PET/CT, or genomics data), VASARI-based models outperformed others in two studies[21,27]. However, a clinical model (age and Karnofsky performance status) slightly outperformed a VASARI-only model in the study by Peeken et al. (C-indices 0.64; 95%-CI: 0.55-0.72 vs. 0.66; 95%-CI: 0.58-0.73)[18]. Compared to VASARI-only models, several studies[9,15,18,21,27] stressed the superiority of predictive models trained on combined features, including VASARI plus clinical, genomics, treatment, and/or radiomics features. Peeken et al.[18] concluded that a combined model (VASARI, clinical (age and Karnofsky performance status), and pathological features (O(6)-methylguanine-DNA methyltransferase methylation)) performed better than the VASARI-only model (C-indices 0.72; 95%-CI: 0.61-0.80 vs. 0.64; 95%-CI: 0.55-0.72). Mazurowski et al.[4] showed the added value of VASARI features (VASARI+clinical: area under the receiver operating characteristic curve [AUC] 0.81; 95%-CI: 0.71-0.90; C-index 0.69; 95%-CI: 0.63-0.75) compared to clinical only

including age, gender and Karnofsky performance status (AUC 0.62; 95%-CI: 0.49-0.74; C-index 0.58; 95%-CI: 0.50-0.66; both p< 0.01).

Three OS-predicting studies recruited grade 2/3 gliomas (Fig. 2a). Zhou et al.[13] found a significant association between OS and ten VASARI features in a univariable analysis of which only one feature, definition of the non-enhancing margin, was significant in multivariable analyses. They demonstrated that a well-defined non-enhancing margin correlated with longer overall survival than an ill-defined non-enhancing margin. Lee et al.[22] showed that the performance of clinical features (age, Karnofsky performance status and extent of resection) plus molecular subtype (IDH mutation and 1p/19q co-deletion status) model increased when extending it with VASARI features (C-indices 0.84; 95%-CI: 0.75-0.90 vs. 0.91; 95%-CI: 0.86-0.96). Similar results were found by Park et al.[26], improving the integral-AUC (initially 0.74; 95%-CI: 0.69-0.81) of the clinical model (age, Karnofsky performance status, extent of resection and histological grade) by adding VASARI features (LASSO model 0.77; 95%-CI: 0.74-0.85, elastic net model 0.78; 95%-CI: 0.74-0.85).

Glioblastoma was included in five of seven progression-free survival predicting studies, with two studies[10,18] showing a significant association with single VASARI features (Fig. 2b). Peeken et al.[21] evaluated the performance of a combined model (VASARI plus clinical (age, gender and Karnofsky performance status) plus pathological (O(6)-methylguanine-DNA methyltransferase methylation, IDH mutation status and Ki 67% proliferation index) showing superior performance compared with single data type models including their VASARI-only model (C-indices validation set 0.68; 95%-CI: 0.57-0.78 vs. 0.61; 95%-CI: 0.50-0.72, p=0.014 respectively). In that study, VASARI features showed the highest performance in the validation set among single data type models (C-index 0.61; 95%-CI: 0.50-0.72). They were also the dominant variable in the combined model (VASARI plus clinical plus pathological features).

Two of seven progression-free survival-predicting studies involved grade 2/3[13] or grade 3 gliomas[22] (Fig. 2b). Lee et al.[22] showed that adding VASARI features increased the performance of the combined model (clinical plus molecular) with C-indices of 0.79; 95%-CI: 0.71-0.85 vs. 0.84; 95%-CI: 0.79-0.91.

Figure 2: The association between different VASARI features and (a), progression-free survival (b), and IDH mutation status (c) listed per study involving glioblastoma and grade 1-4 glioma cohorts. Green indicates tested and significant feature; red, tested and non-significant feature or feature with statistical results not provided; gray, features not tested. The list of VASARI features and their detailed descriptions, along with respective scoring systems for each feature are in the Online Supplemental Data.

IDH mutation status prediction

Reported tumor entities were glioblastoma[5,14,29,30] and grade 2 to 4 gliomas[13,17,20,25,28,32,36,37] (Online Supplemental Data). Individually evaluated VASARI features are shown in Figure 2c. Five studies[5,13,28,29,32] found no association between IDH mutation and individual VASARI features. Among other studies, the main features that more consistently identified as significant predictors of IDH mutation status were enhancement quality (feature 4)[17,20,30,36,37], proportion of enhancing tumor (feature 5)[17,20,30,37], proportion of necrosis (feature 7)[17,30,36,37], and proportion of edema (feature 14)[14,17,20,36]. There was a negative correlation between these features and the presence of IDH mutation.

Higher performance of combined models was shown in several studies[13,20,28,32]. Su et al.[20] showed that a VASARI model (using feature 6: proportion of non-enhancing tumor) outperformed the diffusion texture analysis model with an AUC of 0.92; 95%-CI: 0.80-0.98 vs. 0.72; 95%-CI: 0.57-0.85. The proportion of non-enhancing tumor was significantly higher in IDH-mutant high-grade gliomas compared to IDH wild-type high-grade gliomas. However, the combined model (proportion of non-enhancing tumor plus ADC entropy) was the best-performing model (AUC 0.95; 95%-CI: 0.85-0.99). Sun et al.[32] determined an imaging model (VASARI features plus T2-FLAIR mismatch sign) as the best single data type model (AUC 0.75; 95%-CI: 0.60-0.89) in the test set.

Other studies

Six studies predicted histological tumor grade[12,13,19,33,37,38] or further mutations [1p/19q co-deletion, n=3[13,17,32]; O(6)-methylguanine-DNA methyltransferase methylation, n=2[16,29]; telomerase reverse transcriptase promoter, n=3[16,23,30]], or presence of true progression (n=3)[6,13,31], respectively. The characteristics of these studies are summarized in the Online Supplemental Data.

Meta-analysis

Survival prediction

The meta-analysis of individual VASARI features to predict OS in glioblastoma studies identified eight features with an HR above 1.25 (Fig. 3; HR range 1.32-2.08) and seven equal to or below it (Fig. 4; HR range 0.89-1.25). The HR cut-off of 1.25 was arbitrarily defined for visual representation. Enhancing tumor crossing

midline (feature 23, pooled HR 2.08; 95%-CI: 1.35-3.18), multifocality (feature 9, pooled HR 1.80; 95%-CI: 1.21-2.67), and ependymal invasion (feature 19, pooled HR 1.73; 95%-CI: 1.45-2.05) were the strongest predictors.

Figure 5 gives a detailed overview of HRs for individual VASARI features, including those features that were not selected for the meta-analysis, lacking the mandatory minimum number of 5 articles. Meta-regressions were not possible due to a lack of homogeneous articles.

Meta-analysis of progression-free survival-predicting studies could not be conducted, because only 2 studies[10,18] met the necessary prerequisites for this analysis. An overview of HRs for individual VASARI features used in these 2 studies is provided in the Online Supplemental Data.

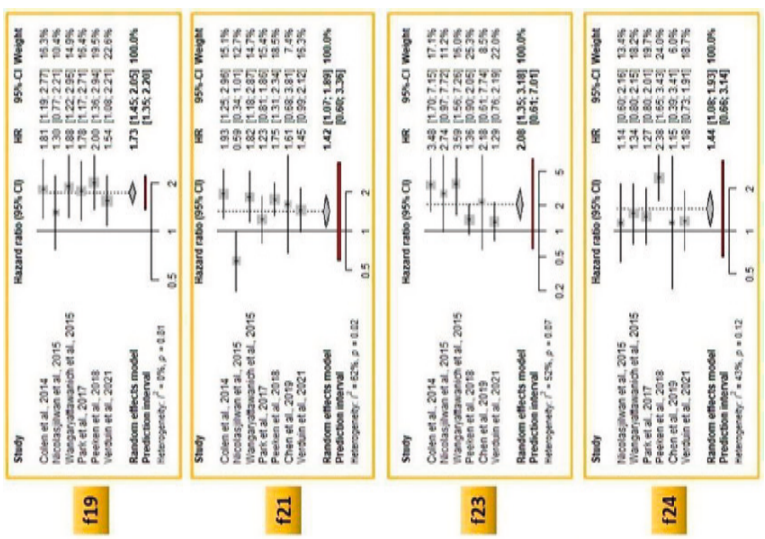


Figure 3: Meta-analysis of OS-predicting studies with a pooled HR >1.25. Note that the list of VASARI features and their detailed descriptions, along with respective scoring systems for each feature are provided in Online Supplemental Data. f indicates VASARI feature; f3, eloquent brain; f9, multifocality; f11, thickness of enhancing margin; f13, definition of non-enhancing margin; f19, ependymal invasion; f21, deep white matter invasion; f23, enhancing tumor crosses midline; f24, satellites.

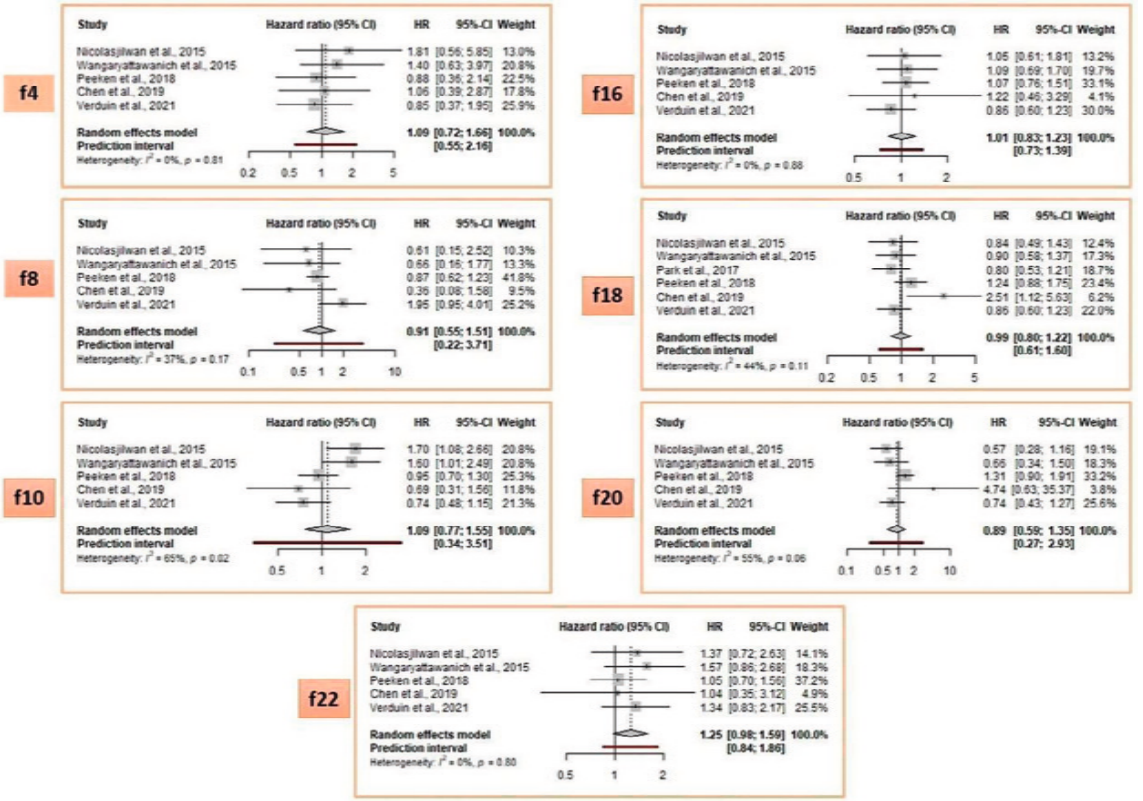


Figure 4: Meta-analysis of OS-predicting studies with a pooled HR ≤1.25. Note that the list of VASARI features and their detailed descriptions, along with respective scoring systems for each feature are provided in Online Supplemental Data. f indicates VASARI feature; f4, enhancement quality; f8, cysts; f10, T1/FLAIR ratio; f16, hemorrhage; f18, pial invasion; f20, cortical involvement; f22, non-enhancing tumor crosses midline.

Study	Colen et al. 2014	Nicolasjilwan et al. 2015	Wangaryattawa-nich et al. 2015	Park et al. 2017	Peeken et al. 2018	Chen et al. 2019	Verduin et al. 2021
Tumor type	glioblastoma	glioblastoma	glioblastoma	glioblastoma	glioblastoma	glioblastoma	glioblastoma
Sample size	104	102	94	108	189	127	188
feature 1		**0.54**	**2.28**	0.99	frontal: 1.00, temporal: 0.92, insular: 2.09, parietal: 0.67, occipital: 1.27, brainstem: 0.40	0.63	
feature 2		**0.59**	0.75		1.52	0.56	
feature 3		0.62	1.45	**1.55**	speech motor: 0.85, speech: 1.16, receptive: 0.73, motol/vision: 0.70	1.23	1.36
feature 4		1.81	1.40		0.88	1.06	0.85
feature 5		**1.89**			0.80	1.32	
feature 6		0.86		(34-67%) 1.15 (≥68%) 1.41	0.87	0.56	
feature 7		1.19		(34-67%) 0.83 (≥68%) 0.72	0.94	0.94	
feature 8		0.61	0.66		0.87	0.36	1.95
feature 9		0.58	**2.54**	**2.15**	**2.50**	2.14	1.49
feature 10		**1.70**	**1.60**		0.95	0.69	0.75
feature 11		1.82	1.41		1.48	1.13	1.39
feature 12			1.15		1.14	1.13	1.00
feature 13		0.84	1.52		**1.80**	0.95	1.15
feature 14		0.79		(34-67%) 0.90 (≥68%) 0.68	**1.48**	2.02	
feature 15					1.21	1.04	
feature 16		1.05	1.09		1.07	1.22	0.86
feature 17		0.76	1.07		facilitated: 0.68, restricted: 0.96, neither/equal: 0.77		
feature 18		0.84	0.90	0.80	1.24	**2.51**	0.86
feature 19	**1.81**	1.30	**1.88**	**1.78**	**2.00**		**1.54**
feature 20		0.57	0.66		1.31	4.74	0.74
feature 21	**1.93**	0.59	**1.82**	1.23	**1.75**	1.61	1.45
feature 22		1.37	1.57	1.59	1.05	1.04	1.34
feature 23	**3.48**	2.74	**3.59**		1.36	2.18	1.29
feature 24		1.14	1.34	1.27	**2.38**	1.15	1.18
feature 25		0.87	0.89		0.92	2.78	
feature 26					**0.66**		
feature 27					**0.73**		
feature 28					**0.64**		
feature 29		1.01	**1.02**		**1.03**	1.82	1.07
feature 30		1.00	1.01		1.03	0.65	1.04
FWER correction	Bonferroni	No	No	No	No	No	No

Figure 5: Color table describing HRs of single VASARI features in OS-predicting studies using univariable Cox proportional models. Thresholds of single HR were defined arbitrarily, and each color represents a different range: blue, HR <0.75; green, HR 0.75 to <0.85; light red, HR 0.85 - 1.25; orange, HR >1.25 - 2.5; dark red, HR >2.5. Values highlighted in bold indicate HRs with statistical significance (*P* value < .05). The list of VASARI features and their detailed descriptions, along with respective scoring systems for each feature are provided in Online Supplemental Data.

IDH status prediction

IDH status studies displayed a very heterogeneous design regarding prediction models and VASARI features (Fig. 6). None of these studies used the same feature combination, allowing only a general approach ("Are multivariable VASARI feature models powerful in predicting IDH status?"). These studies[13,17,20,25,28,32,37] included grade 1-4 gliomas. The AUC of these models was good, ranging from 0.73 to 0.92 (Fig. 6). Meta-regressions were not possible due to a lack of homogeneous articles.

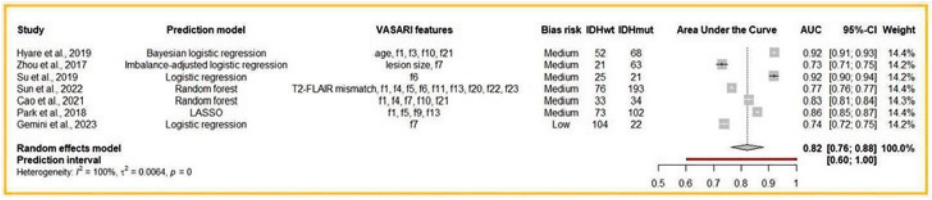


Figure 6: Meta-analysis of glioma IDH mutation status-predicting studies. The list of VASARI features and their detailed descriptions, along with respective scoring systems for each feature are provided in Online Supplemental Data. IDHwt indicates IDH wild-type; IDHmut, IDH-mutant.

DISCUSSION

This study shows that VASARI features have primarily been used to predict OS and IDH mutation status. A meta-analysis of OS-predicting studies revealed that the three most robust single features in determining OS were multifocality (pooled HR = 1.80), ependymal invasion (pooled HR = 1.73), and enhancing tumor crossing midline (pooled HR = 2.08) confirming our hypothesis that some features are stronger predictors than others. In a meta-analysis of IDH mutation-predicting VASARI models that incorporated different combinations of single VASARI features, the pooled AUC was 0.82, with considerable variability between single studies. Combined models incorporating different non-imaging data types outperformed single data type models, including VASARI-only models, in determining survival, mutation status, or grades of glial tumors.

The survival rates of glioma patients remain low despite an aggressive treatment strategy. This is likely the main reason why prognosis prediction was

found to be the primary objective of the included studies. Multifocal tumor distribution (feature 9) and ependymal invasion (feature 19) were among the most unfavorable OS predictors[10,15,18,29], which could be prioritized in studies when evaluating all 30 features is impractical. Thomas et al.[45] also demonstrated a significant correlation between multiple lesions and other negative prognostic indicators, such as low Karnofsky performance score and resection volume. Lim et al.[46] found that newly diagnosed glioblastoma with ependymal invasion and cortical involvement were more likely to have multifocal distribution and noncontiguous tumor recurrence with the initial lesion. The midline crossing enhancing tumor (feature 23) had the highest pooled HR of 2.08 and was the strongest OS limiting predictor. Wangaryattawanich et al.[10] observed a 9.2-month OS difference (4.8 months vs. 14 months, p=0.001) and a 4.2-month progression-free survival difference (2.4 months vs. 6.6 months, p<0.0001) of cases separated by whether the enhancing tumor was crossing midline (feature 23); similar results were shown by Colen et al.[8], with mean OS of 5.9 months vs. 14.3 months (p<0.0003).

According to the 2021 WHO CNS classification, the IDH mutation status is crucial to classifying adult-type diffuse gliomas[47]. The term "IDH-mutant glioblastoma" has been changed to "IDH-mutant astrocytoma" requiring IDH-wild-type status for glioblastoma. Our study identified four VASARI features (enhancement quality[17,20,30,36,37], proportion of enhancing tumor[17,20,30,37], proportion of necrosis[17,30,36,37], and proportion of edema[14,17,20,36]) as the main indicators of IDH mutation status. However, it was not possible to quantitatively evaluate the impact of these features on IDH status prediction due to the limited number of studies and their methodological differences.

Combined feature models, including various variables such as clinical, imaging/VASARI, radiomics, genomics, or pathological features, predicting OS or other objectives, outperformed single data type models showing the importance of a multidisciplinary approach in decision-making. The inputs used for these models varied among studies. However, certain factors were frequently chosen, such as age and Karnofsky performance score for clinical models[4,18,22,26], IDH mutation, 1p/19q codeletion, O(6) -methylguanine-DNA methyltransferase or histological grade[18,21,22] for pathological/genomics models, and shape or texture features for radiomics models[24,31,32,34]. Nevertheless, the methodological het-

erogeneity between those studies, such as different sample sizes and VASARI features or differently structured multivariable models, lowers evidence and precludes translation from the research field to clinical application.

On the basis of the results of our study, several recommendations can be made to improve the generalizability of glioma research using VASARI features. The number and choice of applied VASARI features differed between studies, and some VASARI features were independently modified from their original definition in different studies. The postoperative features 26-28 and feature 15 (edema crossing midline) were rarely used, making their significance challenging to estimate. Some studies modified VASARI features (using features 4-7, 14, and 17) from the original scoring system, hindering generalizability. Future studies should adhere to the original VASARI scoring system and - if feasible - evaluate all VASARI features to find relevant features for a particular objective. Since the manual extraction of VASARI features is time-consuming. Therefore, automatic extraction technology should be developed, which may also help to minimize the heterogeneity described. Where a radiologist's input is needed, reducing the entire feature set to only the most promising features- such as multifocality, ependymal invasion, and enhancing tumor crossing the midline for overall survival prediction, as identified in this meta-analysis- may alleviate the workload. It lies in the nature of the exhaustive VASARI feature set to include collinearity between several features and features which leave a higher chance for inter-rater disagreement. A reduction of features for research purposes according to reproducibility and predictive value appears to be the proximate consequence.

Our study has several limitations. Studies were heterogeneous in design, and some had a small sample size, with a mean of 124 cases. Some studies used the TCIA database, leading to partially overlapping cohorts, for which we attempted to correct. The meta-analysis was limited due to a high study heterogeneity. Meta-regression, sensitivity analysis, and non-reporting bias analysis could not be performed due to the limited number of studies and the diversity in reported predictive models and used feature sets. Studies reporting the performance of VASARI-based OS-predicting models had substantial methodological heterogeneity, as each study used a different set of VASARI features. Model endpoints were also heterogeneous. Most of the studies reported a regression model predicting a continuous OS parameter (in days or months); some

used classification with different thresholds to predict a more or less favorable OS outcome. Other reasons for heterogeneity were using different statistical models (random survival forest, Cox proportional regression, or clustering analysis) and inconsistent utilization of VASARI feature selection methods for single data type models. Statistical data - needed for a meta-analysis, such as confidence intervals or standard errors - were inconsistently reported. Although we attempted a meta-analysis for IDH-predictive VASARI models, these models used different combinations of VASARI features, making it difficult to establish an added value of the VASARI model. Additionally, most studies did not provide molecular diagnostics for glioma grading. Although expected for studies published before 2021, this was one of the main reasons most studies were assigned the "medium" QUADAS-2 category for risk of bias assessment with applicability concerns in patient selection.

CONCLUSIONS

This meta-analysis reveals that certain features in the VASARI set have promise in predicting overall survival and IDH mutation status. However, the added value of VASARI for predicting tumor grade, true progression, and the status of other mutations remains uncertain, mainly due to insufficiently comparable studies. The discriminatory power of individual VASARI features differs considerably. A core set of promising, as emerged from this meta-analysis, may be worth prioritizing for scientific evaluation and considered in clinical use, to avoid using the exhaustive list of VASARI features, which is too time-consuming for daily practice.

REFERENCES

1. Ellingson BM, Bendszus M, Boxerman J, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol 2015;17:1188–98. DOI: https://doi.org/10.1093/neuonc/nov095
2. Wen PY, van den Bent M, Youssef G, et al. RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults. J Clin Oncol 2023;41:5187–99. DOI: https://doi.org/10.1200/JCO.23.01059
3. VASARI Research Project - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. The Cancer Imaging Archive. https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project Accessed 26 Sep, 2023
4. Mazurowski MA, Desjardins A, Malof JM. Imaging descriptors improve the predictive power of survival models for glioblastoma patients. Neuro Oncol 2013;15:1389–94. DOI: https://doi.org/10.1093/neuonc/nos335
5. Gutman DA, Cooper LAD, Hwang SN, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology 2013;267:560–9. DOI: https://doi.org/10.1148/radiol.13120118
6. Agarwal A, Kumar S, Narang J, et al. Morphologic MRI features, diffusion tensor imaging and radiation dosimetric analysis to differentiate pseudo-progression from early tumor progression. J Neurooncol 2013;112:413–20. DOI: https://doi.org/10.1007/s11060-013-1070-1
7. Jain R, Poisson LM, Gutman D, et al. Outcome prediction in patients with glioblastoma by using imaging, clinical, and genomic biomarkers: focus on the nonenhancing component of the tumor. Radiology 2014;272:484–93. DOI: https://doi.org/10.1148/radiol.14131691
8. Colen RR, Vangel M, Wang J, et al. Imaging genomic mapping of an invasive MRI phenotype predicts patient outcome and metabolic dysfunction: a TCGA glioma phenotype research group project. BMC Med Genomics 2014;7:30. DOI: https://doi.org/10.1186/1755-8794-7-30
9. Nicolasjilwan M, Hu Y, Yan C, et al. Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. J Neuroradiol 2015;42:212–21. DOI: https://doi.org/10.1016/j.neurad.2014.02.006
10. Wangaryattawanich P, Hatami M, Wang J, et al. Multicenter imaging outcomes study of The Cancer Genome Atlas glioblastoma patient cohort: imaging predictors of overall and progression-free survival. Neuro Oncol 2015;17:1525–37. DOI: https://doi.org/10.1093/neuonc/nov117
11. Rao A, Rao G, Gutman DA, et al. A combinatorial radiographic phenotype may stratify patient survival and be associated with invasion and proliferation characteristics in glioblastoma. J Neurosurg 2016;124:1008–17. DOI: https://doi.org/10.3171/2015.4.JNS142732
12. Yu J, Wang M, Song J, et al. Potential Utility of Visually AcceSAble Rembrandt Images Assessment in Brain Astrocytoma Grading. J Comput Assist Tomogr 2016;40:301–6. DOI: https://doi.org/10.1097/RCT.0000000000000352
13. Zhou H, Vallières M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 2017;19:862–70. DOI: https://doi.org/10.1093/neuonc/now256
14. Lasocki A, Tsui A, Gaillard F, et al. Reliability of noncontrast-enhancing tumor as a biomarker of IDH1 mutation status in glioblastoma. J Clin Neurosci 2017;39:170–5. DOI: https://doi.org/10.1016/j.jocn.2017.01.007
15. Park M, Lee S-K, Chang JH, et al. Elderly patients with newly diagnosed glioblastoma: can preoperative imaging descriptors improve the predictive power of a survival model? J Neurooncol 2017;134:423–31. DOI: https://doi.org/10.1007/s11060-017-2544-3

16. Ersoy TF, Keil VC, Hadizadeh DR, et al. New prognostic factor telomerase reverse transcriptase promotor mutation presents without MR imaging biomarkers in primary glioblastoma. Neuroradiology 2017;59:1223–31. DOI: https://doi.org/10.1007/s00234-017-1920-1

17. Park YW, Han K, Ahn SS, et al. Prediction of IDH1-Mutation and 1p/19q-Codeletion Status Using Preoperative MR Imaging Phenotypes in Lower Grade Gliomas. AJNR Am J Neuroradiol 2018;39:37–42. DOI: https://doi.org/10.3174/ajnr.A5421

18. Peeken JC, Hesse J, Haller B, et al. Semantic imaging features predict disease progression and survival in glioblastoma multiforme patients. Strahlenther Onkol 2018;194:580–90. DOI: https://doi.org/10.1007/s00066-018-1276-4

19. Su C-Q, Lu S-S, Han Q-Y, et al. Intergrating conventional MRI, texture analysis of dynamic contrast-enhanced MRI, and susceptibility weighted imaging for glioma grading. Acta radiol 2019;60:777–87. DOI: https://doi.org/10.1177/0284185118801127

20. Su C-Q, Lu S-S, Zhou M-D, et al. Combined texture analysis of diffusion-weighted imaging with conventional MRI for non-invasive assessment of IDH1 mutation in anaplastic gliomas. Clin Radiol 2019;74:154–60. DOI: https://doi.org/10.1016/j.crad.2018.10.002

21. Peeken JC, Goldberg T, Pyka T, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. Cancer Med 2019;8:128–36. DOI: https://doi.org/10.1002/cam4.1908

22. Lee M, Han K, Ahn SS, et al. The added prognostic value of radiological phenotype combined with clinical features and molecular subtype in anaplastic gliomas. J Neurooncol 2019;142:129–38. DOI: https://doi.org/10.1007/s11060-018-03072-0

23. Ivanidze J, Lum M, Pisapia D, et al. MRI Features Associated with TERT Promoter Mutation Status in Glioblastoma. J Neuroimaging 2019;29:357–63. DOI: https://doi.org/10.1111/jon.12596

24. Chen X, Fang M, Dong D, et al. Development and Validation of a MRI-Based Radiomics Prognostic Classifier in Patients with Primary Glioblastoma Multiforme. Acad Radiol 2019;26:1292–300. DOI: https://doi.org/10.1016/j.acra.2018.12.016

25. Hyare H, Rice L, Thust S, et al. Modelling MR and clinical features in grade II/III astrocytomas to predict IDH mutation status. Eur J Radiol 2019;114:120–7. DOI: https://doi.org/10.1016/j.ejrad.2019.03.003

26. Park CJ, Han K, Shin H, et al. MR image phenotypes may add prognostic value to clinical features in IDH wild-type lower-grade gliomas. Eur Radiol 2020;30:3035–45. DOI: https://doi.org/10.1007/s00330-020-06683-2

27. Lu Y, Patel M, Natarajan K, et al. Machine learning-based radiomic, clinical and semantic feature analysis for predicting overall survival and MGMT promoter methylation status in patients with glioblastoma. Magn Reson Imaging 2020;74:161–70. DOI: https://doi.org/10.1016/j.mri.2020.09.017

28. Cao M, Suo S, Zhang X, et al. Qualitative and Quantitative MRI Analysis in IDH1 Genotype Prediction of Lower-Grade Gliomas: A Machine Learning Approach. Biomed Res Int 2021;2021:1235314. DOI: https://doi.org/10.1155/2021/1235314

29. Verduin M, Primakov S, Compter I, et al. Prognostic and Predictive Value of Integrated Qualitative and Quantitative Magnetic Resonance Imaging Analysis in Glioblastoma. Cancers 2021;13. DOI: https://doi.org/10.3390/cancers13040722

30. Ahn SS, An C, Park YW, et al. Identification of magnetic resonance imaging features for the prediction of molecular profiles of newly diagnosed glioblastoma. J Neurooncol 2021;154:83–92. DOI: https://doi.org/10.1007/s11060-021-03801-y

31. Wang J, Yi X, Fu Y, et al. Preoperative Magnetic Resonance Imaging Radiomics for Predicting Early Recurrence of Glioblastoma. Front Oncol 2021;11:769188. DOI: https://doi.org/10.3389/fonc.2021.769188

32. Sun C, Fan L, Wang W, et al. Radiomics and Qualitative Features From Multiparametric MRI Predict Molecular Subtypes in Patients With Lower-Grade Glioma. Front Oncol 2021;11:756828. DOI: https://doi.org/10.3389/fonc.2021.756828

33. Ge X, Wang M, Ma H, et al. Investigated diagnostic value of synthetic relaxometry, three-dimensional pseudo-continuous arterial spin labelling and diffusion-weighted imaging in the grading of glioma. Magn Reson Imaging 2022;86:20–7. DOI: https://doi.org/10.1016/j.mri.2021.11.006

34. Ruan Z, Mei N, Lu Y, et al. A Comparative and Summative Study of Radiomics-based Overall Survival Prediction in Glioblastoma Patients. J Comput Assist Tomogr 2022;46:470–9. DOI: https://doi.org/10.1097/RCT.0000000000001300

35. Sacli-Bilmez B, Firat Z, Topcuoglu OM, et al. Identifying overall survival in 98 glioblastomas using VASARI features at 3T. Clin Imaging 2023;93:86–92. DOI: https://doi.org/10.1016/j.clinimag.2022.10.011

36. Sahu A, Patnam NG, Goda JS, et al. Multiparametric Magnetic Resonance Imaging Correlates of Isocitrate Dehydrogenase Mutation in WHO high-Grade Astrocytomas. J Pers Med 2022;13. DOI: https://doi.org/10.3390/jpm13010072

37. Gemini L, Tortora M, Giordano P, et al. Vasari Scoring System in Discerning between Different Degrees of Glioma and IDH Status Prediction: A Possible Machine Learning Application? J Imaging Sci Technol 2023;9. DOI: https://doi.org/10.3390/jimaging9040075

38. You W, Mao Y, Jiao X, et al. The combination of radiomics features and VASARI standard to predict glioma grade. Front Oncol 2023;13:1083216. DOI: https://doi.org/10.3389/fonc.2023.1083216

39. Andersen BM, Miranda C, Hatzoglou V, et al. Leptomeningeal metastases in glioma: The Memorial Sloan Kettering Cancer Center experience. Neurology 2019;92:e2483–91. DOI: https://doi.org/10.1212/WNL.0000000000007529

40. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Int J Surg 2021;88:105906. DOI: https://doi.org/10.1136/bmj.n160

41. Salameh J-P, Bossuyt PM, McGrath TA, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ 2020;370:m2632. DOI: https://doi.org/10.1136/bmj.m2632

42. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36. DOI: https://doi.org/10.7326/0003-4819-155-8-201110180-00009

43. Cortes C, Mohri M. Confidence Intervals for the Area Under the ROC Curve. In: Saul L, Weiss Y, Bottou L, eds. Advances in Neural Information Processing Systems.Vol 17. MIT Press; 2004. DOI: https://dl.acm.org/doi/10.5555/2976040.2976079

44. Cochrane handbook for systematic reviews of interventions. https://training.cochrane.org/handbook#how-to-access Accessed 28 Dec, 2023

45. Thomas RP, Xu LW, Lober RM, et al. The incidence and significance of multiple lesions in glioblastoma. J Neurooncol 2013;112:91–7. DOI: https://doi.org/10.1007/s11060-012-1030-1

46. Lim DA, Cha S, Mayo MC, et al. Relationship of glioblastoma multiforme to neural stem cell regions predicts invasive and multifocal tumor phenotype. Neuro Oncol 2007;9:424–9. DOI: https://doi.org/10.1215/15228517-2007-023

47. Kurokawa R, Kurokawa M, Baba A, et al. Major Changes in 2021 World Health Organization Classification of Central Nervous System Tumors. Radiographics 2022;42:1474–93. DOI: https://doi.org/10.1148/rg.210236

ONLINE SUPPLEMENTAL DATA

Supplement A: Systematic search protocol

Search sources:

- **Database search:** PubMed, Web of Science, EMBASE, and Cochrane Library
- **Hand-searching:** Performed using a review of references from the publications identified in databases

The two components of the search string: VASARI and tumor type

- First component: Visually Accessible Rembrandt Images and VASARI
- Second component: Brain neoplasm, glial tumor, glioma, oligodendroglioma, astrocytoma, glioblastoma, GBM, high-grade glioma, HGG, low-grade glioma, LGG

Search term combinations for database search:

1. PubMed (title and abstract with MeSH terms)

((Visually Accessible Rembrandt Images[Title/Abstract]) OR (VASARI[Title/Abstract])) AND ((glioma[MeSH Terms]) OR (brain neoplasm[MeSH Terms]) OR (glial tumor[Title/Abstract]) OR (glioma[Title/Abstract]) OR (LGG[Title/Abstract]) OR (low-grade glioma[Title/Abstract]) OR (HGG[Title/Abstract]) OR (high-grade glioma[Title/Abstract]) OR (GBM[Title/Abstract]) OR (glioblastoma[Title/Abstract]) OR (oligodendroglioma[Title/Abstract]) OR (astrocytoma[Title/Abstract]))

2. Web of Science (only title and abstract)

(TI="Visually Accessible Rembrandt Images" OR TI=VASARI OR AB="Visually Accessible Rembrandt Images" OR AB=VASARI)

AND

((TI="brain neoplasm*" OR AB="brain neoplasm*") OR (TI=glioma* OR AB=glioma*) OR

(TI="glial tumor*" OR AB="glial tumor*") OR (TI=glioblastoma* OR AB=glioblastoma*) OR

(TI=oligodendroglioma* OR AB=oligodendroglioma*) OR (TI=astrocytoma* OR AB=astrocytoma*) OR (TI=GBM* OR AB=GBM*) OR (TI="high-grade glioma*" OR AB="high-grade glioma*") OR (TI=HGG* OR AB=HGG*) OR (TI="low-grade glioma*" OR AB="low-grade glioma*") OR (TI=LGG* OR AB=LGG*))

3. EMBASE (major focus)

('Visually Accessible Rembrandt Images'/mj OR 'VASARI'/mj)

AND

('brain neoplasm'/mj OR 'glioma'/mj OR 'glial tumor'/mj OR 'glioblastoma'/mj OR 'oligodendroglioma'/mj OR 'astrocytoma'/mj OR 'GBM'/mj OR 'high-grade glioma'/mj OR 'HGG'/mj OR 'low-grade glioma'/mj OR 'LGG'/mj)

4. Cochrane Library (title, abstract, and keywords)

VASARI OR Visually Accessible Rembrandt Images in Title Abstract Keyword AND brain neoplasm OR glioma OR glial tumor OR glioblastoma OR oligodendroglioma OR astrocytoma OR GBM OR high-grade glioma OR HGG OR low-grade glioma OR LGG in Title Abstract Keyword - (Word variations have been searched)

Supplement B: Calculation of publicly available The Cancer Imaging Archive (TCIA) overlapping cohorts

The list of the studies with sample sizes using the TCIA cohort

1. Mazurowski et al. *n=82*
2. Gutman et al. *n=75*
3. Jain et al. *n=45*
4. Colen et al. *n=104*
5. Nicolasjilwan et al. *n=102*
6. Wangaryattawanich et al. *n=94*
7. Rao et al. *n=92*
8. Zhou et al. *n=165*

9. Chen et al. *n=127*
10. Park et al. 2020 *n=158*
11. **Ruan et al. *n=200* (the largest cohort)**

To consider possibly overlapping cohorts, the corrected number of recruited patients was calculated using only the study with the largest cohort (Ruan et al.), and the other ten studies were excluded from the calculation.
Calculation: 4,348 - 82 - 75 - 45 - 104 - 102 - 94 - 92 - 165 - 127 - 15 **= 3,304**
4,348*= not corrected recruited patient numbers from all 35 included studies*
3,304*=corrected recruited patient numbers from 25 included studies*

SUPPLEMENTAL TABLES

Table S1. Descriptions and scoring system of VASARI features set

		Descriptions
VASARI	**Names**	**Definitions**
f1	Tumor location	Location of lesion geographic epicenter; the largest component of the tumor (either CET or nCET)
f2	Side of tumor epicenter	Side of lesion epicenter irrespective of whether the lesion crosses into the contralateral hemisphere
f3	Eloquent brain	Does any component of the tumor (CET or nCET) involve the eloquent cortex or the immediate subcortical white matter of the eloquent cortex (motor, language, vision)?
f4	Enhancement Quality	A qualitative degree of contrast enhancement is defined as having all or portions of the tumor that demonstrate a significantly higher signal on the postcontrast T1WI compared to precontrast T1WI. Mild = when barely discernible but unequivocal degree of enhancement is present relative to pre-contrast images. Marked = obvious tissue enhancement
f5	Proportion Enhancing	When scanning through the entire tumor volume, what proportion of the entire tumor is estimated to be enhancing? (Assuming that the entire abnormality may be comprised of: (1) an enhancing component, (2) a non-enhancing component, (3) a necrotic component, and (4) an edema component)
f6	Proportion nCET	When scanning through the entire tumor volume, what proportion of the entire tumor is estimated to represent a nonenhancing tumor or nCET (not edema)? Non-enhancing tumor (nCET) is defined as regions of T2W hyperintensity (less than the intensity of cerebrospinal fluid, with corresponding T1W hypointensity) that are associated with mass effect and architectural distortion, including blurring of the grey-white interface. (Assuming that the entire abnormality may be comprised of: (1) an enhancing component, (2) a non-enhancing component, (3) a necrotic component, and (4) an edema component)

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

		Descriptions
VASARI	**Names**	**Definitions**
f7	Proportion Necrosis	Visually, when scanning through the entire tumor volume, what proportion of the tumor is estimated to represent necrosis? Necrosis is defined as a region within the tumor that does not enhance, is high on T2WI and proton density images, is low on T1WI, and has an irregular border). (Assuming that the entire abnormality may be comprised of: (1) an enhancing component, (2) a non-enhancing component, (3) a necrotic component, and (4) an edema component)
f8	Cyst(s)	Cysts are well-defined, rounded, often eccentric regions of very bright T2W signal and low T1W signal essentially matching CSF signal intensity, with very thin, regular, smooth, nonenhancing, or regularly enhancing walls, possibly with thin, regular, internal septations
f9	Multifocal or Multicentric	Multifocal is defined as having at least one region of the tumor, either enhancing or nonenhancing, which is not contiguous with the dominant lesion and is outside the region of signal abnormality (edema) surrounding the dominant mass. This can be defined as those resulting from dissemination or growth by an established route, spread via commissural or other pathways, or via CSF channels or local metastases, whereas Multicentric are widely separated lesions in different lobes or different hemispheres that cannot be attributed to one of the previously mentioned pathways. Gliomatosis refers to the generalized neoplastic transformation of the white matter of most of a hemisphere
f10	T1/FLAIR ratio	The T1/FLAIR ratio is a gross comparison of the overall lesion size between pre-contrast T1 and FLAIR (in the same plane). Select T1~FLAIR when pre-contrast T1 abnormality (exclusive of signal intensity) approximates the size of FLAIR abnormality; Select T1<FLAIR when the size of T1 abnormality is moderately smaller than the surrounding FLAIR envelope; or select T1<<FLAIR when the size of the pre-contrast T1 abnormality is much smaller than the size of FLAIR abnormality. (If no FLAIR images are provided, select No FLAIR images)
f11	Thickness of enhancing margin	First version (round 1): If most of the enhancing rim is thin, regular, and has a homogenous enhancement, the grade is thin. If most of the rim demonstrates nodular and/or thick enhancement, the grade is thick. If there is only solid enhancement and no rim, the grade is none Second version (round 2): If most of the enhancing rim is thin, regular, measures < 3mm in thickness, and has a homogenous enhancement, the grade is minimal. If most of the rim demonstrates nodular and/or thick enhancement measuring 3mm or more, the grade is thick/nodular. If there is only solid enhancement and no rim, the grade is solid

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

		Descriptions
VASARI	**Names**	**Definitions**
f12	Definition of the enhancing margin	Assess if most of the outside margin of the enhancement is well-defined (i.e., sharply marginated) or poorly defined (fluffy or indistinct). Are you able to easily trace the margin of enhancement?
f13	Definition of the non-enhancing margin (e.g., Grade III)	If most of the outside non-enhancing (nCET) margin of the tumor is well-defined (i.e., sharply marginated) and smooth (geographic) versus if the margin is poorly defined (fluffy or indistinct)
f14	Proportion of Edema	Visually, when scanning through the entire tumor volume, what proportion of the entire abnormality is estimated to represent vasogenic edema? Edema should be greater in signal than nCET and somewhat lower in signal than CSF. Pseudopods are characteristic of edema. (Assuming that the entire abnormality may be comprised of: (1) an enhancing component, (2) a non-enhancing component, (3) a necrotic component, and (4) an edema component)
f15	Edema Crosses Midline	Edema spans white matter commissures extending into the contralateral hemisphere (exclusive of herniated ipsilateral tissue)
f16	Hemorrhage	Intrinsic hemorrhage anywhere in the tumor matrix. Any intrinsic foci of low signal on T2WI or high signal on T1WI. Select cannot determine if findings are indistinct or may actually represent mineral instead of hemorrhage
f17	Diffusion	Predominantly facilitated or restricted diffusion in the enhancing or nCET portion of the tumor. (Based on ADC map). Rate CET alone when present, otherwise, use nCET. Select facilitated if the enhancing (when present) or nCET portion of the tumor is predominantly hyperintense to normal tissue on ADC maps. Select restricted if the enhancing (when present) or nCET portion of the tumor is a predominantly low signal to normal tissue on ADC maps. Select cannot determine when findings are equivocal or mixed. (Select no ADC images if ADC images are not provided)
f18	Pial invasion	Enhancement of the overlying pia in continuity with enhancing or non-enhancing tumor
f19	Ependymal invasion	Tumor invasion of any adjacent ependymal surface in continuity with enhancing or non-enhancing tumor matrix
f20	Cortical involvement	Non-enhancing or enhancing tumor that extends to the cortical mantle, or if the cortex is no longer distinguishable relative to subjacent tumor

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

		Descriptions
VASARI	**Names**	**Definitions**
f21	Deep white matter invasion	Enhancing, or nCET tumor extending into the internal capsule, corpus callosum or brainstem
f22	nCET tumor Crosses Midline	Presence of any nCET tumor that extends into the contralateral hemisphere through white matter commissures (exclusive of herniated ipsilateral tissue)
f23	Enhancing tissue crosses midline	Presence of any CET that extends into the contralateral hemisphere through white matter commissures (exclusive of herniated ipsilateral tissue)
f24	Satellites	One or more areas of enhancement within the region of signal abnormality surrounding the dominant lesion but not contiguous in any part with the major tumor mass
f25	Calvarial remodeling	Visible erosion/remodelling of inner table of skull (possibly a secondary sign of slow growth)
f26	Extent of resection of enhancing tumor	Using the first postoperative scan (contrast-enhanced MR imaging) assessed for tumor residual. Estimate the proportion of enhancing tumors removed. Total resection of the component should be scored 100%
f27	Extent resection of nCET	Using the first postoperative scan (contrast-enhanced MR imaging) assessed for tumor residual. Estimate the proportion of non-enhancing tumors removed. Total resection of the component should be scored 100%
f28	Extent resection of vasogenic edema	Using the first postoperative scan (contrast-enhanced MR imaging) assessed for tumor residual. Estimate the proportion of edema removed. Total resection of the component should be scored 100%
f29&30	Lesion Size	Lesion size is defined as the largest perpendicular (x-y) cross-sectional diameter of the entire T2 signal abnormality (longest dimension x perpendicular dimension) measured in the single axial image that reveals the largest cross-sectional area of the lesion. (Can incorporate CET, nCET, necrosis, and edema)

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f1	Tumor location	0 = - 1 = Frontal 2 =Temporal 3=Insular 4=Parietal 5=Occipital 6=Brainstem 7=Cerebellum	1 = Frontal 2 =Temporal 3=Insular 4=Parietal 5=Occipital 6=Brainstem 7=Cerebellum 8=Basal ganglia 9=Thalamus 10-Corpus callosum
f2	Side of tumor epicenter	0= - 1=Right 2=Center/Bilateral 3=Left	1=Right 2=Center/Bilateral 3=Left
f3	Eloquent brain	0= - 1=None 2=Speech motor 3=Speech receptive 4=Motor 5=Vision	1=None 2=Speech motor 3=Speech receptive 4=Motor 5=Vision
f4	Enhancement Quality	0= - 1=None 2=Mild/Minimal 3=Marked/Avid	0= No contrast injected 1=None 2=Mild/Minimal 3=Marked/Avid

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f5	Proportion Enhancing	0= - 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	1=Minimal 2=<1/3 3= >1/3, <2/3 4=>2/3
f6	Proportion nCET	0=- 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	0=No nCET 1=Minimal 2=<1/3 3= >1/3, <2/3 4=>2/3

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f7	Proportion Necrosis	0=- 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	0=None 1=Minimal 2=<1/3 3= >1/3, <2/3 4=>2/3
f8	Cyst(s)	0=- 1= No 2= Yes	1=Absent 2=Present
f9	Multifocal or Multicentric	0 = - 1= n/a 2= Multifocal 3= Multicentric 4= Gliomatosis	1= Focal 2= Multifocal or Multicentric 3= Gliomatosis
f10	T1/FLAIR ratio	0= - 1= Expansive (T1~FLAIR) 2= Mixed (T1<FLAIR) 3= Infiltrative (T1<<FLAIR)	0=No FLAIR images 1= T1~FLAIR 2= T1<FLAIR 3= T1<<FLAIR

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f11	Thickness of enhancing margin	0=- 1= n/a 2= None 3= Thin 4= Thick/solid	1=Minimal 2= Thick/nodular 3=Solid
f12	Definition of the enhancing margin	0=- 1= n/a 2= Well-defined 3= Poorly-defined	1= Well-defined 2= Poorly-defined
f13	Definition of the non-enhancing margin (e.g. Grade III)	0=- 1= n/a 2= Smooth 3= Irregular	1= Well-defined 2= Poorly-defined
f14	Proportion of Edema	0= - 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	0=None 1=Minimal 2=<1/3 3= >1/3, <2/3 4=>2/3

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f15	Edema Crosses Midline	0=- 1=n/a 2= No 3= Yes	Not indicated
f16	Hemorrhage	0=- 1= No 2= Yes	1= No 2= Yes 3=Can not determine
f17	Diffusion	0=- 1= No image 2= Facilitated 3= Restricted 4=Neither/equivocal	1= Facilitated 2= Restricted 3=Can not determine 4=no ADC images
f18	Pial invasion	0=- 1= No 2= Yes	1=Absent 2=Present
f19	Ependymal invasion	0=- 1= No 2= Yes	1=Absent 2=Present
f20	Cortical involvement	0=- 1= No 2= Yes	1=Absent 2=Present

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f21	Deep white matter invasion	0=- 1= No 2= Yes	1=None 2=Corpus callosum 3=Internal capsule 4=Brainstem
f22	nCET tumor Crosses Midline	0=- 1= n/a (no nCET) 2= No 3= Yes	1= No 2= Yes
f23	Enhancing tissue crosses midline	0=- 1= n/a 2= No 3= Yes	1= No 2= Yes
f24	Satellites	0=- 1= No 2= Yes	1=Absent 2=Present
f25	Calvarial remodelling	0=- 1= No 2= Yes	1=Absent 2=Present

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f26	Extent of resection of enhancing tumor	0= - 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	Not indicated
f27	Extent resection of nCET	0= - 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	Not indicated

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f28	Extent resection of vasogenic edema	0= - 1= n/a 2=None (0%) 3= <5% 4= 6-33% 5= 34-67% 6= 68-95% 7= >95% 8=All (100%) 9= Indeterminate	Not indicated

Table S1. Descriptions and scoring system of VASARI features set *(Continued)*

Scoring system			
VASARI	**Names**	**Scoring: first version/round 1**	**Scoring: second version/round 2**
f29&30	Lesion Size	0=- 1= <0.5cm 2= 0.5 cm 3= 1.0 cm 4= 1.5 cm 5= 2.0 cm 6= 2.5 cm 7= 3.0 cm 8= 3.5 cm 9= 4.0 cm 10= 4.5 cm 11= 5.0 cm 12 = 5.5 cm 13= 6.0 cm 14= 6.5 cm 15= 7.0 cm 16= 7.5 cm 17= 8.0 cm 18= >8.0cm	Measure the lesion size

Abbreviations: CET = contrast-enhancing tumor, f = VASARI feature, nCET = non-contrast-enhancing tumor, VASARI = Visually Accessible Rembrandt [Repository for Molecular Brain Neoplasia Data] Images

Table S2. Recorded study data

- PubMed ID, title, and year of publication
- Study design and purpose (prediction of diagnosis, progression, overall survival, or progression-free survival)
- Timepoint related to surgery (pre/postoperative)
- Demographics (sex ratio, age)
- WHO classification edition and diagnosis
- Genetic feature selection (isocitrate dehydrogenase, 1p/19q co-deletion, telomerase reverse transcriptase, O(6)-methylguanine-DNA methyltransferase)
- Examined Visually Accessible Rembrandt [Repository for Molecular Brain Neoplasia Data] Images (VASARI) features
- Sample size
- Index test (VASARI feature set evaluation via human or machine learning approach)
- Statistical methods and metrics (area under the receiver operating characteristic curve, accuracy, sensitivity, specificity, negative predictive value, positive predictive value, hazard ratio, and their relevant confidence intervals, standard errors, and p-values)

Table S3. Details of the modified QUADAS-2 tool

Risk of bias assessment		
Domains	**Signalling questions with descriptions**	**Scoring with descriptions**
Patient selection	Is the study design clearly described?	**Low risk:** prospective (registered protocol, prospective data collection, primary/secondary goal) or retrospective **High risk:** Not mentioned
	Is the patient selection process clearly described?	**Low risk:** Consecutive sampling - all eligible patients from a well-defined period are included. Random sampling - randomly selected patients are included **High risk:** No consecutive or random sampling, or the description is unclear or missing important details
	Are the selection criteria clearly described?	**Low risk:** all relevant criteria are mentioned clearly **Medium risk:** key criteria are mentioned with some doubts about the details (inclusion criteria are mentioned, exclusion criteria are missing) **High risk:** key criteria are very vague or missing
	Is the information on participants complete?: demographics (age - mean+SD or range, male/female ratio, primary disease).	**Low risk:** information is complete for all patients **High risk:** important information is missing, unclear, or incomplete.

Table S3. Details of the modified QUADAS-2 tool *(Continued)*

Risk of bias assessment		
Domains	**Signalling questions with descriptions**	**Scoring with descriptions**
Index test	Are the index test results interpreted without knowledge of the results of the reference standard?	**Low risk:** The index test results are interpreted with blindness to the reference standard **Medium risk:** Relevant information is unclear or missing without suspicion of blindness **High risk:** Strong suspicion for reviewers not being blinded to the reference standard
	How many and which VASARI features are used as a part of the index test?	**Low risk:** All details about the VASARI features are given **Medium risk:** Information regarding the usage of VASARI features is incomplete: usage of which VASARI features are not given **High risk:** Information regarding the usage of VASARI features is unclear or missing
Reference standard	Is the tumor classification done according to the World Health Organization (WHO) classification 2021 of brain tumors?	**Low risk:** WHO CNS tumor classification, tumor type, and grade are reported, including IDH mutation and 1p/19q co-deletion status **High risk:** The tumor type is incorrectly named or not indicated, and/or tumor grade and mutation status are missing
	Is the response assessment done according to the state-of-the-art? (Response assessment: low/medium/high risk of bias/not available) The decision of progression (true progression vs. pseudoprogression) after brain tumor therapy should be done according to Response Assessment in Neuro-Oncology (RANO) criteria for gliomas and, ideally, with longer follow-up or performed histology to confirm the findings. Progression-free survival assessment does not involve response assessment.	**Low risk:** Longer follow-up and/or histology is performed to confirm the RANO findings on the follow-up at the question **Medium risk:** Only RANO or modified RANO criteria are used for treatment response evaluation **High risk:** No RANO criteria are used or reported **N/A:** Progression not evaluated

Table S3. Details of the modified QUADAS-2 tool *(Continued)*

Risk of bias assessment		
Domains	**Signalling questions with descriptions**	**Scoring with descriptions**
Flow and timing	Is the timing of the index test clearly described? The authors need to describe properly when the imaging data is acquired relative to the surgery (preoperative or postoperative).	**Low risk:** Relevant information with time intervals is stated clearly **Medium risk:** Relation to therapy stated, but the exact time interval is not specified **High risk:** Relevant information is missing or unclear
Data analysis	Are withdrawals from the study explained and uninterpretable results reported? Do the authors give reasons why subjects and scans are excluded from the study and properly list the numbers and reasons for exclusions? This question addresses both subject-level exclusions and also exclusions due to low data quality. Is the quality control procedure described?	**Low risk:** A mechanism for exclusion and quality control is clearly explained, and the number of excluded subjects/scans is given **High risk:** Data on non-participating subjects and excluded scans are incomplete or missing, and an unexplained mismatch between the number of recruited subjects examined scans and reported results.
	Which VASARI features are found statistically significant for the purpose of the study?	**Low risk:** Information regarding which VASARI features made the statistically significant difference is given **High risk:** Related information is unclear or missing
	Statistical tests: Are the statistical tests sufficient, reported, and correct? Are significant AND non-significant results stated with p-values? Are confidence intervals and/or effect sizes stated? Are all statistical tests used mentioned? In the case of voxel-wise statistics and multiple parameter comparison, is multiple testing correction used? Do the authors state the error rate if prediction model performance is reported?	**Low risk:** All of the above is mentioned and considered **Medium risk:** minor issues **High risk:** The statistics part has clear gaps of information hampering the reproduction of results OR used tests not fitting the research question OR missing key values like p-values or effect sizes

Table S3. Details of the modified QUADAS-2 tool *(Continued)*

Risk of bias assessment		
Domains	**Signalling questions with descriptions**	**Scoring with descriptions**
Calculation of Final Overall Risk of Bias Score		
Equation: [(count of category A x category point A) + (count of category B x category point B) + (count of category C x category point C)]/ total count of categories = Risk of bias score Z		
Category points: A (low risk of bias) = 0; B (medium risk of bias) = 1; C (high risk of bias) = 2		
Calculation of final score: Z=0-0.4 >>> Low; Z=0.5-1.4 >>> Medium; Z=1.5-2 >>>> High		
Applicability concerns assessment		
Domains	**Signalling questions**	**Scoring**
Patient selection	Are there concerns that the included patients do not match the review question?	Yes/no
Index test	Are there concerns that the index test, its conduct, or its interpretation differ from the review question?	Yes/no
Reference standard	Are there concerns that the target condition, as defined by the reference standard, does not match the review question?	Yes/no

Abbreviations: RANO = response assessment in neuro-oncology criteria, VASARI = Visually Accessible Rembrandt [Repository for Molecular Brain Neoplasia Data] Images, WHO-CNS tumor classification = The World Health Organization Classification of tumors of the Central Nervous System

Table S4. Key characteristics of included studies

Study	Sample size (n)	Female: Male	Mean age (SD) years	WHO-CNS system*	Entity	Time point	Objective is the prediction of?
Mazurowski et al. 2013	82	-	-	C	glioblastoma (TCIA data)	preop	OS
Gutman et al. 2013	75	24:51	58.2	C	glioblastoma (TCIA data)	preop	OS/IDH
Agarwal et al. 2013	46	-	56.7	C	grade 3, 4 glioma	postop	progression
Jain et al. 2014	45	15:30	-	C	glioblastoma (TCIA data)	preop	OS/PFS
Colen et al. 2014	104	38:66	58	C	glioblastoma (TCIA data)	preop	OS
Nicolasjilwan et al. 2015	102	36:66	57.7 (14.6)	C	glioblastoma (TCIA data)	preop	OS
Wangaryat-tawanich et al. 2015	94	31:63	57.5	C	glioblastoma (TCIA data)	preop	OS/PFS
Rao et al. 2016	92	32:60	-	C	glioblastoma (TCIA data)	preop	OS/PFS
Yu et al. 2016	126	51:75	-	C	grade 2-4 astrocytoma	preop	grade

Table S4. Key characteristics of included studies *(Continued)*

Study	Sample size (n)	Female: Male	Mean age (SD) years	WHO-CNS system*	Entity	Time point	Objective is the prediction of?
Zhou et al. 2017	165	-	-	C	grade 2, 3 glioma (TCIA data)	preop	OS/PFS/ grade/ progression/ IDH/1p/19q
Lasocki et al. 2017	153	64:89	IDH wild 64.9 (12) IDH mutant 50.6 (11.2)	C	glioblastoma	preop	IDH
Park et al. 2017	training 108 validation 40	training 54:54 validation -	-	C	glioblastoma	preop	OS
Ersoy et al. 2017	67	24:43	-	B	glioblastoma	preop	MGMT/ TERT
Park et al. 2018	training 175 validation 40	training 82:93 validation 19:21	training 44.6 (12.9) validation 46.9 (12.1)	C	grade 2, 3 glioma	preop	IDH/1p/19q
Peeken et al. 2018	189	75:114	-	C	glioblastoma	preop postop	OS/PFS
Su et al. 2019	40	15:25	52.9 (10.2)	B	grade 2-4 glioma	preop	grade

Table S4. Key characteristics of included studies *(Continued)*

Study	Sample size (n)	Female: Male	Mean age (SD) years	WHO-CNS system*	Entity	Time point	Objective is the prediction of?
Su et al. 2019	46	21:25	47.8 (12)	C	grade 3 glioma	preop	IDH
Peeken et al. 2019	189 (training 132 validation 57)	68: 121	-	C	glioblastoma	preop postop	OS/PFS
Lee et al. 2019	86	36:50	44.5	B	grade 3 glioma	preop	OS/PFS
Ivanidze et al. 2019	29	TERT wild 3:10 TERT mutant 4:12	-	C	glioblastoma	preop	TERT
Chen et al. 2019	training 85 validation 42	training 32:53 validation 13:29	training 57.8 (13.5) validation 60.0 (13.8)	C	glioblastoma (TCIA data)	preop	OS
Hyare et al. 2019	146 (grade 2,3 astrocytoma 120 glioblastoma 26)	IDH wild astro.18:34 IDH mutant astro. 30:38 glioblastoma 13:13	IDH wild astrocytoma 54 IDH mutant astrocytoma 37 glioblastoma 42	C	grade 2,3 astrocytoma, glioblastoma	preop	IDH

Table S4. Key characteristics of included studies *(Continued)*

Study	Sample size (n)	Female: Male	Mean age (SD) years	WHO-CNS system*	Entity	Time point	Objective is the prediction of?
Park et al. 2020	training 112 validation 46	training 61:51 validation 25:21	training 49.9 (16.2) validation 48 (15.3)	B	IDH wild grade 2, 3 glioma (TCIA data)	preop	OS
Lu et al. 2020	training 127 validation 54	68:113	58 (12.8)	B	glioblastoma	preop	OS
Cao et al. 2021	102	42:60	45.3 (16.3)	C	grade 2, 3 glioma	preop	IDH
Verduin et al. 2021	training 142 validation 46	training 57:85 validation 17:29	-	C	glioblastoma	preop	OS/IDH/ MGMT
Ahn et al. 2021	training 125 validation 51	training 62:63 validation 22:29	training 57.6 (13.7) validation 59.8 (11.2)	B	glioblastoma	preop	IDH/TERT
Wang et al. 2021	training 86 validation 36	50:72	-	C	glioblastoma	preop postop	progression
Ge et al. 2021	72	32:40	45.1 (17.2)	C	grade 2-4 glioma	preop	grade
Sun et al. 2022	335	146:189	44.9 (12.5)	C	grade 2, 3 glioma	preop	IDH/ 1p/19q

Table S4. Key characteristics of included studies *(Continued)*

Study	Sample size (n)	Female: Male	Mean age (SD) years	WHO-CNS system*	Entity	Time point	Objective is the prediction of?
Ruan et al. 2022	training 175 validation 25	75:125	55.1 (15.4)	C	glioblastoma (TCIA data)	preop	OS
Sahu et al. 2022	148	49:99	-	C	grade 4 glioma	preop	IDH
Sacli-Bilmez et al. 2023	98	32-66	54.4 (13.2)	C	glioblastoma	preop	OS
Gemini et al. 2023	126	51-75	55.3	C	grade 1-4 glioma	preop	IDH/grade
You et al. 2023	102	-	-	A	grade 1-4 glioma	preop	grade

Abbreviations: IDH = isocitrate dehydrogenase, MGMT = O(6) -methylguanine-DNA methyltransferase, OS = overall survival, preop = preoperative, PFS = progression-free survival, postop = postoperative, SD = standard deviation, TCIA = The Cancer Imaging Archive, TERT = telomerase reverse transcriptase, WHO-CNS = World Health Organization Classification of Central Nervous System tumors

*Meaning of ABC labels for WHO-CNS system column: A = current WHO-CNS classification system (5th edition, 2021) was used; B = an outdated WHO-CNS classification system was used (2007 or 2016); C= WHO-CNS classification system was not specified.

Note: The term "glioblastoma " covers the definitions of all respectively used WHO-CNS classifications, as most of the studies either used outdated classification systems or did not mention explicitly which classification system was used.

Table S5. Overview of quality assessment results of all included studies according to modified QUADAS-2

Study	Risk of Bias						Applicability concerns		
	Patient selection	Index test	Reference standard	Flow/ timing	Data analysis	Overall score	Patient selection	Index test	Reference standard
Mazurowski et al. 2013	M	M	H	H	L	M	Yes	No	No
Gutman et al. 2013	M	M	H	M	L	M	Yes	No	No
Agarwal et al. 2013	M	L	H	M	M	M	Yes	No	No
Jain et al. 2014	M	M	H	M	L	M	Yes	No	No
Colen et al. 2014	M	M	H	M	M	M	Yes	No	No
Nicolasjilwan et al. 2015	M	M	H	H	M	M	Yes	No	No
Wangaryattawanich et al. 2015	M	M	H	H	M	M	Yes	No	No
Rao et al. 2016	M	M	H	H	M	M	Yes	No	No
Yu et al. 2016	L	M	H	M	L	M	Yes	No	No
Lasocki et al. 2017	L	M	H	L	L	M	Yes	No	No
Zhou et al. 2017	M	L	H	M	L	M	Yes	No	No
Park et al. 2017	L	L	H	M	L	M	Yes	No	No
Ersoy et al. 2017	H	M	L	M	L	M	No	No	No
Peeken et al. 2018	L	L	H	M	M	M	Yes	No	No
Park et al. 2018	L	M	H	L	L	M	Yes	No	No
Su et al. 2019	M	L	H	M	L	M	Yes	No	No

Table S5. Overview of quality assessment results of all included studies according to modified QUADAS-2 *(Continued)*

Study	Risk of Bias						Applicability concerns		
	Patient selection	Index test	Reference standard	Flow/ timing	Data analysis	Overall score	Patient selection	Index test	Reference standard
Su et al. 2019	L	L	H	M	L	M	Yes	No	No
Peeken et al. 2019	L	M	H	M	M	M	Yes	No	No
Lee et al. 2019	L	M	H	M	L	M	Yes	No	No
Ivanidze et al. 2019	M	M	H	M	M	M	Yes	No	No
Chen et al. 2019	M	L	H	M	M	M	Yes	No	No
Hyare et al. 2019	L	L	H	L	M	M	Yes	No	No
Park et al. 2020	L	L	H	M	M	M	Yes	No	No
Lu et al. 2020	L	L	H	M	M	M	Yes	No	No
Cao et al. 2021	L	M	H	M	L	M	Yes	No	No
Verduin et al. 2021	L	L	H	M	M	M	Yes	No	No
Ahn et al. 2021	L	L	H	M	L	M	Yes	No	No
Wang et al. 2021	L	L	H	L	L	L	Yes	No	No
Ge et al. 2021	L	L	H	M	L	M	Yes	No	No
Sun et al. 2022	L	L	H	M	M	M	Yes	No	No
Ruan et al. 2022	L	L	H	M	L	M	Yes	No	No
Sahu et al. 2022	M	L	H	H	M	M	Yes	No	No

Table S5. Overview of quality assessment results of all included studies according to modified QUADAS-2 *(Continued)*

Study	Risk of Bias						Applicability concerns		
	Patient selection	Index test	Reference standard	Flow/ timing	Data analysis	Overall score	Patient selection	Index test	Reference standard
Sacli-Bilmez et al. 2023	M	L	H	L	M	M	Yes	No	No
Gemini et al. 2023	L	L	H	L	L	L	Yes	No	No
You et al. 2023	M	L	H	H	L	M	No	No	Yes

Abbreviations: QUADAS = Quality Assessment of Diagnostic Accuracy Studies, L = low risk of bias, M = medium risk of bias, H = high risk of bias

Table S6. Studies predicting IDH mutation status using VASARI features

Studies	Tumor types	Main findings
Gutman et al. 2013	glioblastoma	**Evaluated VASARI features:** f5-7, f14 and f29&30 **Significant VASARI features:** no significant association was found
Lasocki et al. 2017	glioblastoma	**Evaluated VASARI features**: f1-2, f5-12, f14, f16-25, f29&30 **Significant VASARI features**: f14 p=0.009 • All five IDH mutant patients had 5–33% edema, while the majority of IDH wild-type had >33% edema (44% of IDH wild-type patients with 34–67% edema, 14% with 68–95%) • Sensitivity, specificity, and overall accuracy for predicting IDH mutation based on different imaging features: • all tumors with a frontal lobe location (f1) (sensitivity 40%, specificity 64%, accuracy 63%) • all tumors with >33% non-enhancing tumor (f6) (sensitivity 60%, specificity 79%, accuracy 78%) • tumors with either a frontal lobe location (f1) or >33% non-enhancing tumor (f6) (sensitivity 100%, specificity 51%, accuracy 52%)
Verduin et al. 2021	glioblastoma	**Evaluated VASARI features:** f1-14, f16, f18-24, f29&30 **Significant VASARI features:** no significant association was reported **The selected features for the VASARI model:** f2, f3, f6, f10, f11, f12, f14, f18, f29&30 • In the test dataset, only radiomics features reached statistical significance with an AUC of 0.82 (95% CI 0.65–0.95), which improved upon combining with VASARI features • In the external validation set, neither VASARI nor radiomics features or the combination were able to predict the IDH status
Ahn et al. 2021	glioblastoma	**Evaluated VASARI features:** f1-14, f16, f18-25 **Significant VASARI features:** f4 p=0.012, f5 p=0.003, f7 p=0.009, f8 p=0.021, f12 p<0.001, f16 p=0.017, f19 p=0.048

Table S6. Studies predicting IDH mutation status using VASARI features *(Continued)*

Studies	Tumor types	Main findings
Sun et al. 2022	grade 2/3 glioma	**Evaluated VASARI features:** f1-25 **Significant VASARI features:** f1, f4-6, f11, f13, f20, f22, f23 were significantly associated with IDH status. However, no p values were reported • Nine VASARI features (f1, f4-6, f11, f13, f20, f22, f23) and T2-FLAIR mismatch sign were selected to build a qualitative model. • AUC levels for IDH wild-type tumors in the testing cohort: • VASARI+T2-FLAIR mismatch = 0.75, 95% CI: 0.60, 0.89 (the best single data type model) vs radiomics = 0.66, 95% CI: 0.51, 0.80 vs. clinical = 0.66, 95% CI: 0.48, 0.83 • Combined (VASARI+T-FLAIR mismatch + radiomics + clinical) = 0.86, 95% CI: 0.75, 0.98 (the highest performance)
Cao et al. 2021	grade 2/3 glioma	**Evaluated VASARI features:** f1-2, f4-5, f7-25, f29&30 **Significant VASARI features:** no significant association was reported • After feature selection, the top 5 VASARI features were f4 (AUC 0.75), f21 (AUC 0.73), f1 (AUC 0.68), f7 (AUC 0.68) and f10 (AUC 0.63) • Performance of VASARI model: training cohort (f1, f4, f7, f10, f21) AUC 0.83; validation cohort (only f4) AUC 0.78 • Performance of radiomics model with top 10 radiomic features in both training and validation cohorts: AUC 0.85 • Performance of combined model1 (f4+radiomics): AUC 0.88 • Performance of combined model2 (f4+radiomics+clinical): AUC 0.86
Zhou et al. 2017	grade 2/3 glioma	**Evaluated VASARI features:** f1-25, f29&30 **Significant VASARI features:** no significant association was reported • Prediction performance estimation in 100 bootstrap testing samples for optimal multivariable VASARI models, including f7 and f29&30: AUC 0.73 • Prediction performance estimation in 100 bootstrap testing samples for optimal multivariable texture models: AUC 0.86 • AUC results obtained on out-of-bag estimates in random forest analysis for different combinations of set variables: VASARI AUC 0.60 vs. clinical AUC 0.67 vs. texture AUC 0.79 vs. texture+VASARI AUC 0.83 vs. texture+VASARI+clinical AUC 0.86

Table S6. Studies predicting IDH mutation status using VASARI features *(Continued)*

Studies	Tumor types	Main findings
Park et al. 2018	grade 2/3 glioma	**Evaluated VASARI features:** f1-14, f16-25, f29&30 **Significant VASARI features for grade 2+3 gliomas**: f1 p<0.001, f2 p<0.001, f4 p=0.002, f5 p<0.001, f7 p=0.007, f8 p=0.046, f9 p<0.001, f10 p<0.001, f11 p<0.001, f12 p=0.001, f13 p<0.001, f14 p=0.007, f17 p=0.004, f18 p=0.015, f19 p=0.009, f20 p<0.001, f21 p=0.012, f29 p=0.002 • Independently associated features with IDH status by the LASSO procedure: f1 OR=2.38, f5 OR=1.66, f9 OR=2.93, f13 OR=1.30 • AUC for the optimal model: discovery set 0.86 (95% CI 0.78 –0.93), validation set 0.78 (95% CI, 0.62 –0.89) **Significant VASARI features for grade 2 gliomas**: f2 p=0.005, f9 p<0.001, f10 p=0.049, f11 p=0.018, f13 p=0.006, f18 p=0.048, f19 p=0.003, f20 p=0.043 • Independently associated features with IDH status by the LASSO procedure: f2 OR=1.36, f9 OR=7, f18 OR=1.48, f19 OR=1.62 • AUC for the optimal model: 0.83 (95% CI, 0.75–0.91) **Significant VASARI features for grade 3 gliomas**: f1 p=0.006, f2 p=0.005, f5 p=0.005, f9 p<0.001, f10 p<0.048, f11 p=0.022, f12 p=0.001, f13 p=0.014, f20 p=0.002 • Independently associated features with IDH status by the LASSO procedure: f1 OR=2.33, f5 OR=1.25, f9 OR=2.27, f20 OR=0.99 • AUC for the optimal model: 0.87 (95% CI, 0.79 –0.95)
Hyare et al. 2019	grade 2/3 astrocytoma	**Evaluated VASARI features:** f1-14, f16-f22, f24-25 **Significant VASARI features:** f1 p=0.001, f3 p=0.007, f9 p=0.005, f10 p<0.001, f13 p<0.001, f20 p=0.034, f21 p<0.001, f24 p=0.042 • The excellent performance (AUC = 0.92) was achieved with only the top five variables: age at presentation (OR = 0.94), tumor location within the thalamus (f1; OR = 0.15), involvement of speech receptive areas (f3; OR = 0.21), deep white matter invasion of the brainstem (f21; OR = 0.10), and T1/FLAIR signal ratio (f10; OR = 1.63)

Table S6. Studies predicting IDH mutation status using VASARI features *(Continued)*

Studies	Tumor types	Main findings
Su et al. 2019	grade 3 glioma	**Evaluated VASARI features:** f1-2, f4-8, f11-12, f14, f16, f17 **Significant VASARI features:** f4 p=0.003, f5 p=0.001, f6 p=0.001, f11 p<0.001, f12 p<0.001, f14 p=0.001 • Receiver operating characteristic results of different models: VASARI model (only f6) AUC 0.92 (95% CI: 0.80, 0.98) vs. texture analyses model (ADC entropy) AUC 0.72 (95% CI: 0.57, 0.85) vs. combined model (f6+ADC entropy) AUC 0.95 (95% CI: 0.85, 0.99)
Sahu et al. 2022	grade 4 glioma	**Evaluated VASARI features:** f1-14, f16-f25, f29&30 **Significant VASARI features:** f4 p<0.001, f7 p<0.001, f8 p=0.001, f14 p=0.025 • On multivariate analysis, using a stepwise logistic regression, f7 (OR 0.04, p<0.001) was independently correlated for IDH wild-type phenotype, while the other semantic features that were significant on the univariate analysis lost their significance on the multivariate analysis
Gemini et al. 2023	grade 1-4 glioma	**Evaluated VASARI features:** f1-25, f29&30 **Significant VASARI features:** f4 p< 0.05, f5 p=0.00034, f6 p=000124, f7 p=00339 • f4 model: accuracy 0.78, sensitivity of 0.64, specificity 0.81, AUC 0.73 • f5 model (cut-off level=6-33%): accuracy 0.81, sensitivity 0.48, specificity 0.89, AUC 0.73 • f6 model (cut-off level=68-95%): accuracy 0.84, sensitivity 0.48, specificity 0.93, AUC 0.76 • f7 model (cut-off level=<5%); accuracy 0.74, sensitivity 0.80, specificity 0.73, AUC 0.79

Abbreviations: AUC = area under the receiver operating characteristic curve, f = VASARI feature, f1 = location, f2 = side of lesion, f3 = eloquent brain, f4 = enhancement quality, f5 = proportion enhancing, f6 = proportion nonenhancing, f7 = proportion necrosis, f8 = cyst(s), f9 = multifocality, f10 = T1/FLAIR ratio, f11 = thickness of enhancing margin, f12 = definition of the enhancing margin, f13 = definition of the nonenhancing margin, f14 = proportion edema, f15 = edema crosses midline, f16 = hemorrhage, f17 = diffusion, f18 = pial invasion, f19 = ependymal invasion, f20 = cortical involvement, f21 = deep white matter invasion, f22 = non-enhancing tumor crosses midline, f23 = enhancing tumor crosses midline, f24 = satellites f25 = calvarial remodelling, f29&30 = lesion size, IDH = isocitrate dehydrogenase, OR= Odds ratio, VASARI = Visually Accessible Rembrandt [Repository for Molecular Brain Neoplasia Data] Images

Table S7. Studies predicting tumor grade, 1p/19q codeletion, MGMT methylation, TERT promoter status, and progression using VASARI features

Studies	Tumor types	Main findings
Prediction of histological grade		
Su et al. 2019	grade 2 vs. grade 3/4 glioma	**Evaluated VASARI features:** f1, f2, f4-8, f12, f14 and f17 **Significant VASARI features:** f4, f5, f6, f12, f14, f17 ($p<0.05$) and enhancement quality (f4) showed the best performance (AUC 0.87). • The combined model, including dynamic contrast-enhanced MRI and susceptibility-weighted imaging parameters, outperformed the VASARI (enhancement quality (f4)) model (AUC 0.99).
Ge et al. 2021	grade 2 vs. grade 3/4 glioma	**Evaluated VASARI features**: f4 and f5 **Significant VASARI features**: enhancement quality (f4; $p<0.001$) and enhancing tumor proportion (f5; $p=0.001$) • The combined model using synthetic relaxometry, arterial spin labelling, and diffusion-weighted imaging metrics (T1+PD+CBF+ADC: AUC 0.99) outperformed the VASARI model (AUC 0.84) with a significant difference between these models ($p=0.0016$).
Zhou et al. 2017	grade 2 vs. grade 3 glioma	**Evaluated VASARI features:** f1-25, f29, f30 **The selected features for the VASARI model:** enhancing tumor proportion (f5) + definition of nonenhancing margin (f13) + diffusion (f17) (AUC 0.78) • The VASARI model was the best-performing single-layer model (AUC 0.73). • The combined model (clinical, radiomics, and VASARI features) showed the highest performance (AUC 0.78).
Yu et al. 2016	grade 2-4 astrocytoma	**Evaluated VASARI features:** f1-16, f18-25, f29, f30 **Significant VASARI features for differentiation:** grade 2-4 >> f4-7, f11, f14, f1, f18, f19, f23, f24/ grade 2 vs. grade 3/4 >> f4-7, f11, f12, f14, f16, f18, f19, f23/ grade 2 vs. grade 3 >> f4-6, f11, f12, f14, f16/ grade 3 vs. grade 4 >> f4, f5, f6, f11, f16 • Enhancement quality (f4) was associated with a 31-fold increased risk of grade 3/4 astrocytoma compared to grade 2 astrocytoma (OR 31.05; 95% CI: 1.09, 897.69). • Edema proportion (f14) performed best in differentiating grade 2 from grade 3 gliomas ($p=0.015$). • Nonenhancing tumor proportion (f6) was the single significant feature that differentiated grade 3 from grade 4 gliomas ($p=0.011$).

Table S7. Studies predicting tumor grade, 1p/19q codeletion, MGMT methylation, TERT promoter status, and progression using VASARI features *(Continued)*

Studies	Tumor types	Main findings
Gemini et al. 2023	grade 1/2 vs. grade 3/4 glioma	**Evaluated VASARI features:** f1-25, f29, f30 **Significant VASARI features**: enhancement quality (f4; p=0.00028), enhancing tumor proportion (f5; p=0.00388), nonenhancing tumor proportion (f6; p=0.0032), necrosis proportion (f7; p=0.035) • VASARI model of f4: accuracy 0.78, sensitivity 0.80, AUC 0.73 • VASARI model of f5: accuracy 0.79, sensitivity 0.87, AUC 0.68 • VASARI model of f6: accuracy 0.74, sensitivity 0.75, AUC 0.76 • VASARI model of f7: accuracy 0.71, sensitivity 0.71, AUC 0.74
You et al. 2023	grade 1/2 vs. grade 3/4 glioma	**Evaluated VASARI features:** f1-25 **Significant VASARI features**: f8 (p=0.046), f14 (0.029), f21 (0.011), f24 (0.035) • Multivariate analysis: edema proportion (f14; p=0.036) and deep white matter invasion (f21; p=0.026) was chosen for VASARI model • Model 1 (f14+f21+Rad-score): accuracy 0.93, sensitivity 0.95, AUC 0.97 • Model 2 (Rad-score): accuracy 0.91, sensitivity 0.91, AUC 0.94 • Model 3 (f14+f21): accuracy 0.76, sensitivity 0.78, AUC 0.83
Prediction of 1p/19q codeletion status		
Sun et al. 2022	grade 2/3 glioma	**Evaluated VASARI features:** f1-25 **Significant VASARI features:** f1, f4-6, f11, f13, f20, f22, f23 • The imaging model (f1, f4-6, f11, f13, f20, f22, f23, and T2-FLAIR mismatch sign) was the best single-layer model in the testing set (AUC-IDH mutant-19/19q codeleted: 0.79, AUC-IDH mutant- 1p/19q non-codeleted: 0.76) • The combined model (imaging, clinical, and radiomics variables) performed better than all single models in training (AUC-IDH mutant-19/19q codeleted: 0.82, AUC-IDH mutant-19/19q non-codeleted: 0.82) and test sets (AUC-IDH mutant-19/19q codeleted: 0.80, AUC-IDH mutant-19/19q non-codeleted: 0.81).
Park et al. 2018	grade 2/3 glioma	**Evaluated VASARI features:** f1-14, f16-25, f29, f30 **Significant VASARI features:** pial invasion (f18; p=0.039) and diffusion (f17; p=0.020)

Table S7. Studies predicting tumor grade, 1p/19q codeletion, MGMT methylation, TERT promoter status, and progression using VASARI features *(Continued)*

Studies	Tumor types	Main findings
Zhou et al. 2017	grade 2/3 glioma	**Evaluated VASARI features:** f1-25, f29, f30 **Selected VASARI features for the VASARI model:** f1+f5+f9+f14 (AUC 0.78) • The VASARI model's performance (AUC 0.68) was not better than either radiomics (AUC 0.88) or the combined model (AUC 0.89).
Prediction of MGMT methylation status		
Verduin et al. 2021	glioblastoma	**Evaluated VASARI features:** f1-f14, f16, f18-f24, f29, f30 **Selected VASARI features for the VASARI model:** f6, f19, f21, f29, f30 + f29mean, f29median • The VASARI model showed moderate performance in the test (AUC 0.67) and validation sets (AUC 0.62), where radiomics features failed. • The combined model (VASARI and radiomics features) performed better than the VASARI model in the training set (AUC 0. 84) but not in the validation set (0.67).
Ersoy et al. 2017	glioblastoma	**Evaluated VASARI features:** f1-14, f18-f24 • No significant association between VASARI features and MGMT p>0.05
Prediction of TERT promoter status		
Ahn et al. 2021	glioblastoma	**Evaluated VASARI features:** f1-14, f16, f18-f25 **Significant VASARI features:** necrosis proportion (f7; p=0.013), definition of enhancing margin (f12; p=0.001), and enhancing tumor crossing midline (f23; p=0.012)
Ivanidze et al. 2019	glioblastoma	**Evaluated VASARI features:** f1-25 **Significant VASARI features:** nonenhancing tumor crossing midline (f22; p= 0.014)
Ersoy et al. 2017	glioblastoma	**Evaluated VASARI features:** f1-14, f18-f24 • No significant association between VASARI features and TERT p>0.05
Prediction of progression vs. non-progression		

Table S7. Studies predicting tumor grade, 1p/19q codeletion, MGMT methylation, TERT promoter status, and progression using VASARI features *(Continued)*

Studies	Tumor types	Main findings
Agarwal et al. 2013	grade 3/4 glioma	Examined differences between true progression (TP) and pseudoprogression (PsP) **Evaluated VASARI features:** f4-6, f9-12, f14-24, f29, f30 • Lesion size (f29&30): **single significant VASARI feature**, larger in the PsP group (p=0.002) • Definition of enhancing margin (f12): poorly defined in the PsP group (100 %) compared to the TP group (72 %) (p = 0.088, trend towards) • Enhancing tumor crossing midline (f23): 28 % of patients in the TP group compared to none in the PsP group (p = 0.088, trend towards)
Zhou et al. 2017	grade 2/3 glioma	**Evaluated VASARI features:** f1-25, f29, f30 **Selected VASARI features:** f1, f5, f7, f10, f11, AUC 0.58 • The radiomics model (texture analysis) was the best model (AUC 0.70) compared to other single-layer models, including VASARI (AUC 0.55) and combined models (AUC 0.69)
Wang et al. 2021	glioblastoma	Examined the association between early and later recurrence and VASARI features. **Evaluated VASARI features:** f1-29 **Significant VASARI features:** enhancing tumor proportion (f5; OR=1.61; p=0.038) and deep white matter invasion (f21; OR=3.45; p=0.002) were predictors of early recurrence, the extent of resection of enhancing tumor (f26; OR=0.51; p=0.007) was a predictor of later recurrence.

Abbreviations: AUC = area under the receiver operating characteristic curve, f = VASARI feature, f1 = location, f2 = side of lesion, f3 = eloquent brain, f4 = enhancement quality, f5 = proportion enhancing, f6 = proportion nonenhancing, f7 = proportion necrosis, f8 = cyst(s), f9 = multifocality, f10 = T1/FLAIR ratio, f11 = thickness of enhancing margin, f12 = definition of the enhancing margin, f13 = definition of the nonenhancing margin, f14 = proportion edema, f15 = edema crosses midline, f16 = hemorrhage, f17 = diffusion, f18 = pial invasion, f19 = ependymal invasion, f20 = cortical involvement, f21 = deep white matter invasion, f22 = non-enhancing tumor crosses midline, f23 = enhancing tumor crosses midline, f24 = satellites f25 = calvarial remodelling, f26 = extent of resection of enhancing tumor, f27 = extent of resection of non-enhancing tumor, f28 = extent of resection of vasogenic edema, f29&30 = lesion size, IDH = isocitrate dehydrogenase, MGMT = O(6)-methylguanine-DNA methyltransferase, OR= Odds ratio, TERT = telomerase reverse transcriptase, VASARI = Visually Accessible Rembrandt [Repository for Molecular Brain Neoplasia Data] Images

SUPPLEMENTAL FIGURES

Study	Wangaryattawanich et al. 2015	Peeken et al. 2018
Tumor type	glioblastoma	glioblastoma
Sample size	94	189
feature 1	1.83	frontal: 1.00, temporal: 0.78, insular: 1.65, parietal: 0.97, occipital: 1.36, brainstem: 1.08
feature 2	0.84	1.50
feature 3	1.13	speech motor: 0.77, speech: 1.11, receptive: 0.72, motor/vision: 0.79
feature 4	0.89	0.90
feature 5		0.93
feature 6		0.86
feature 7		0.95
feature 8	0.95	1.00
feature 9	2.26	**1.85**
feature 10	**1.79**	0.95
feature 11	1.19	1.28
feature 12	1.45	1.29
feature 13	1.47	**1.47**
feature 14		1.19
feature 15		1.21
feature 16	0.72	0.90
feature 17	0.69	facilitated: 0.78, restricted: 0.68, neither/equal: 0.93
feature 18	0.76	0.81
feature 19	**1.98**	**2.00**
feature 20	0.80	0.91
feature 21	**2.51**	**1.80**
feature 22	**2.18**	1.21
feature 23	**5.46**	1.23
feature 24	**2.87**	**1.61**
feature 25	0.92	0.98
feature 26		**0.52**
feature 27		**0.68**
feature 28		**0.66**
feature 29	**1.02**	1.01
feature 30	1.02	1.02
FWER correction	No	No

Figure S1: Color table describing HRs of single VASARI features in progression-free survival predicting studies using univariable Cox proportional models. Thresholds of single HR were defined arbitrarily, and each color represents a different range: blue, HR <0.75; green, HR 0.75 - <0.85; light red, HR 0.85 - 1.25; orange, HR >1.25 - 2.5; dark red, HR >2.5. Values highlighted in bold indicate HRs with statistical significance (*P* value < .05). The list of VASARI features and their detailed descriptions, along with respective scoring systems for each feature are provided in Online Supplemental Data.

Chapter 4

Human Performance in Predicting Enhancement Quality of Gliomas using Gadolinium-Free MRI Sequences

Azizova, A., *Wamelink, I.J.H.G., Prysiazhniuk, Y., Cakmak, M., Kaya, E., Petr, J., Barkhof, F., Keil, V.C.*

Published: ***Journal of Neuroimaging***
2024; vol. 34,6:673-693; doi:10.1111/jon.13233

ABSTRACT:

Background and Purpose: To develop and test a decision tree for predicting contrast enhancement quality and shape using pre-contrast MRI sequences in a large adult-type diffuse glioma cohort.

Methods: Preoperative MRI scans (development/optimization/test sets: n=31/38/303, male=17/22/189, mean age=52/59/56.7 years, high-grade glioma=22/33/249) were retrospectively evaluated, including pre-and post-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. Enhancement prediction decision tree (EPDT) was developed using development and optimization sets, incorporating four imaging features: necrosis, diffusion restriction, T2 inhomogeneity, and nonenhancing tumor margins. EPDT accuracy was assessed on a test set by three raters of variable experience. True enhancement features (gold standard) were evaluated using pre- and post-contrast T1-weighted images. Statistical analysis used confusion matrices, Cohen's/Fleiss' kappa, and Kendall's W. Significance threshold was $P < 0.05$.

Results: Raters 1, 2, and 3 achieved overall accuracies of 0.86 [95%-confidence interval (CI): 0.81-0.90], 0.89 (95%-CI: 0.85-0.92), and 0.92 (95%-CI: 0.89-0.95), respectively, in predicting enhancement quality (marked, mild, or no enhancement). Regarding shape, defined as the thickness of enhancing margin (solid, rim, or no enhancement), accuracies were 0.84 (95%-CI: 0.79-0.88), 0.88 (95%-CI: 0.84-0.92), and 0.89 (95%-CI: 0.85-0.92). Intra-rater inter-group agreement comparing predicted and true enhancement features consistently reached substantial levels [≥0.68 (95%-CI: 0.61-0.75)]. Inter-rater comparison showed at least moderate agreement [group: ≥0.42 (95%-CI: 0.36-0.48), pairwise: ≥0.61 (95%-CI: 0.50-0.72)]. Among the imaging features in the EPDT, necrosis assessment displayed the highest intra- and inter-rater consistency [≥0.80 (95%-CI: 0.73-0.88)].

Conclusion: The proposed enhancement prediction decision tree has high accuracy in predicting enhancement patterns of gliomas irrespective of rater experience.

INTRODUCTION

Neuro-oncological imaging is fundamentally linked to the use of gadolinium-based contrast agent (GBCA)-enhanced images, which are part of the recommended minimum standard MRI protocol for the imaging of brain tumors.[1] The utilization of GBCA is state-of-the-art for the diagnosis, preoperative evaluation, and response assessment.[2] The recently updated Response Assessment in Neuro-oncology criteria recognize GBCA-enhanced MRI as the most sensitive and reproducible way for assessing brain tumors while also highlighting the importance of GBCA-free sequences.[3] The presence of contrast enhancement serves as an indicator of an unfavorable prognosis[4,5] and is used to accurately define the resection margins of adult-type diffuse gliomas.[6] While supramarginal resection beyond GBCA-enhanced tumor margins using GBCA-free sequences, such as FLAIRectomy,[7,8] is potentially associated with better outcomes, GBCA-enhanced sequence-based evaluation remains the standard and is preferred by most neurosurgeons. Hence, the identification of contrast enhancement is essential for the effective management of these patients.

There are multiple reasons why patients with brain tumors in practice may not receive GBCA during their MRI examination. GBCAs are associated with several putative health and environmental hazardous effects.[9] While nephrogenic systemic fibrosis is one of the well-recognized rare side effects and is mostly limited to linear GBCA types, the deposition of any GBCA in different body parts, including the brain, with currently unknown medical consequences[10] raises concerns in some patients. Special precautions always need to be taken when using GBCA in vulnerable patient groups, including children, as well as pregnant and breastfeeding women.[11,12] On a global scale, GBCA contributes to the already high healthcare costs in neuro-oncology, which can burden non-insured individuals and the healthcare sector, particularly in low- and middle-income countries.[13–15] GBCA availability can be problematic, too, in some areas.[16] For these reasons, radiologists face the burden of evaluating scans of neuro-oncological patients who did not undergo contrast-enhanced MRI without any knowledge of how far this affects their professional judgment. A consistent and valid methodology to predict contrast enhancement in brain tumors from nonenhanced sequences could improve decision-making. While artificial intelligence (AI)-derived synthetic post-contrast imaging may be a

possible future alternative,[17] it does not serve the needs of clinical radiologists today, and its approaches lack a comparison with human performance.

This study's purpose is, therefore, to develop a decision tree tool for radiologists to predict contrast enhancement intensity and shape using GBCA-free MRI sequences and to test its accuracy in comparison with contrast-enhanced T1-weighted images in a large adult-type diffuse glioma cohort. The results of this study will deliver valuable insights for science dedicated to the advancement of synthetic contrast enhancement, as no head-to-head comparative studies exist with human raters.

METHODS

Study Design

This is a retrospective study approved by the institutional medical ethics review committee (VUmc_2021-0437). Informed consent was waived.

Study sample

All eligible patients with preoperative MRI scans extracted from our in-house glioma database (IMAGO) between January 1, 2010, and January 1, 2021, were included. The inclusion criteria were: i) adult patients with grade 2-4 adult-type diffuse gliomas according to the 2021 World Health Organization CNS tumor classification system, ii) last preoperative brain MRI within one month before surgery, iii) MRI scan including pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging, and post-contrast T1-weighted sequences. Exclusion criteria were: i) pediatric patients, ii) patients who had explicitly refused consent for their data to be used for scientific research, iii) absence of mandatory MRI sequences, iv) MRI scans with inadequate quality such as motion artifacts, v) lack of a confirmed histopathological diagnosis, and vi) tumors localized in suprasellar, midline and cerebellar areas. Fifty-six patients were excluded (for details, refer to Figure 1).

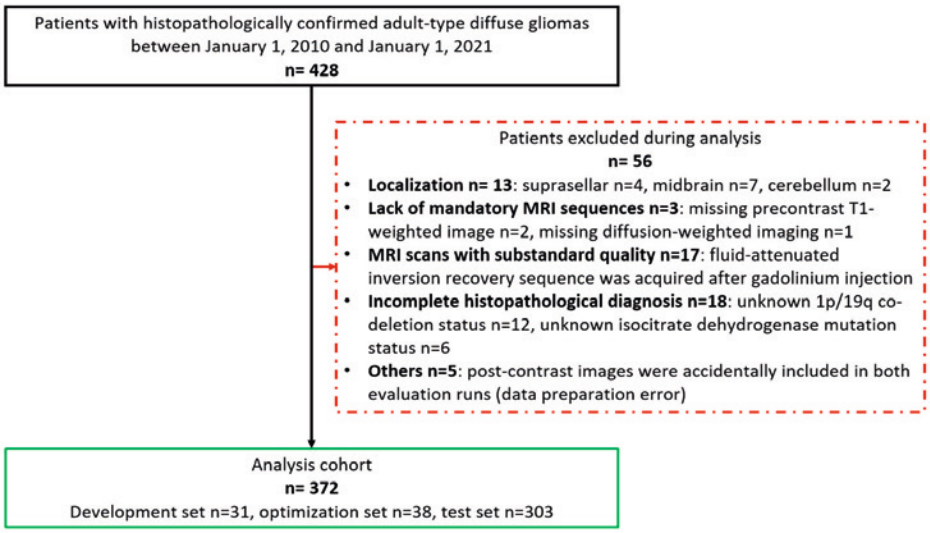


Figure 1: Flow chart describes the details of the patient enrollment. n = number.

MRI details

Pretherapy MRI scans had been acquired on seven MRI scanners according to standardized brain tumor imaging protocol,[1] including three 1.5-Tesla (T) MRI and four 3-T MRI machines (see Table 1).

Table 1. Overview of MRI parameters

Sequence type	Parameters			
	TE; ms	TR; ms	TI; ms	FA; ms
GE 1.5T Signa HDxt				
2D T1w	9-12	520-600	-	90
T2w	98-104	4,376-4,840	-	90
2D FLAIR	118-120	9,002-9,502	2250	90
3D FLAIR	96-122	6,000-6,500	1,925-1,987	90
DWI (b-1000)	81-105	8,000-8,500	-	90
2D CE-T1w	9-12	520-600	-	90
3D CE-T1w	3-5	8-13	0-450	12
GE 3T Discovery MR750				
2D T1w	7.9-9.7	600-731	-	90-125
3D T1w	2	4.7	650	15

Table 1. Overview of MRI parameters *(Continued)*

Sequence type	Parameters			
	TE; ms	TR; ms	TI; ms	FA; ms
T2w	82-88	4,889-6,872	-	90-111
3D FLAIR	126-135	8,000-8,002	2,331-2,347	90
DWI (b-1000)	62-87	4,000-7,200	-	90
2D CE-T1w	7.9-8.4	600-650	-	90-125
3D CE-T1w	2-3.2	4.6-8.3	450-650	15
Philips 1.5T Achieva				
3D T1w	3.3-4.6	6.5-8.7	-	8-10
T2w	100-110	3,404-5,251	-	90
2D FLAIR	140	9,000-11,000	2,800	90
3D FLAIR	286-306	4,800	1,660	90
DWI (b-1000)	72-119	2,674-6,448	-	90
3D CE-T1w	3.3-4.6	6.7-8.7	-	8-10
Philips 3T Ingenuity				
2D T1w	10	599	-	70
3D T1w	3	7	-	12
T2w	85	2,767-3,182	-	90
3D FLAIR	279	4,800	1,650	90
DWI (b-1000)	74-97	3,496-6,354	-	90
2D CE-T1w	10	599	-	70
3D CE-T1w	3	7	-	12
Siemens 1.5T Avanto				
2D T1w	7.8-17	500-718	-	90
3D T1w	4.5-11	700-2,700	0-950	8-120
T2w	93-104	2,830-5,562	-	150-180
2D FLAIR	88-109	8,870-9,000	2,500	150
3D FLAIR	334	6,500	2,200	120
DWI (b-1000)	90-122	3,400-10,500	-	90
2D CE-T1w	8.7-17	550-718	-	90
3D CE-T1w	2.9-4.5	1,900-2,700	950-1,100	8

Table 1. Overview of MRI parameters *(Continued)*

Sequence type	Parameters			
	TE; ms	TR; ms	TI; ms	FA; ms
Siemens 3T MAGNETOM Vida				
3D T1w	2.3	2,300	900	8
T2w	74	4,100-6,280	-	150
3D FLAIR	388-430	5,000-7,700	1,650-2,400	120
DWI (b-1000)	68	3,200	-	90
3D CE-T1w	2.3	2,300	900	8
Toshiba 3T Titan3T				
2D T1w	8	550	-	80
T2w	90	5,500-5,526	-	90
3D FLAIR	451	5,600	1,900	90
DWI (b-1000)	82	7,500	-	90
2D CE-T1w	8	550	-	80
3D CE-T1w	2.4	5.7	900	9

Table 1 describes MRI parameters for each of the seven scanners used in this study.
Note: All values are obtained from the Digital Imaging and Communications in Medicine headers. Evaluations were made in axial plane.
Abbreviations: 1.5T/3T = 1.5 tesla/3 tesla, 2D/3D = 2-/3-dimensional, CE-T1w = contrast-enhanced T1-weighted, DWI = diffusion weighted imaging, FA = flip angle, FLAIR = fluid-attenuated inversion recovery, ms = millisecond, TE = time of echo, TI = time of inversion, TR = time of repetition, w = weighted

Datasets

Datasets were prepared and pseudonymized by I.W., a fourth-year Ph.D. student in neuro-oncology. The included patients (n=372) were randomly distributed among three sets. Development (n=31) and optimization (n=38) sets were used to develop and improve the Enhancement Prediction Decision Tree (EPDT), respectively. A test set (n=303) was used to assess the accuracy of the EPDT in a larger cohort. RADIANT software (version:3.4.1.13367, Medixant, Poznan, Poland, https://www.radiantviewer.com/) was used to access the pseudonymized scans.

Enhancement prediction decision tree

The ratings were carried out independently by two raters (V.K., eleven years of neuroradiology experience; A.A., five years of neuroradiology experience).

In the initial stage, the raters assessed the GBCA-free sequences of the development set (n=31) consisting of pre-contrast sequences, blinded to diagnosis and contrast-enhanced T1-weighted scans. Based on previous clinical experience, the raters were tasked to predict the enhancement quality (pEQ), which is typically qualitatively assessed from pre- and post-contrast T1-weighted images. According to the standard definition of the Visually AcceSAble Rembrandt Images (VASARI) feature 4,[18] pEQ was categorized into three groups: marked-avid enhancement, mild-barely discernible enhancement, and no evident enhancement. The raters were also tasked to provide a justification and annotation of arguments for their decision for every case and answer the following questions: 1) Why does the rater anticipate marked/mild/no enhancement? and 2) Which imaging feature(s) were instrumental in the decision-making process? The raters needed to refer to the VASARI features set,[18] a standardized vocabulary for glioma imaging, to guide them in their decision-making process. The following VASARI features were rated for each case based on experience and the possibility of evaluation without post-contrast images: eloquent brain (feature 3), presence of necrosis (modified feature 7), multifocality (feature 9), T1/fluid-attenuated inversion recovery (FLAIR) ratio (feature 10), nonenhancing tumor margins (feature 13), substantial edema (modified feature 14), hemorrhage (feature 16), diffusion (feature 17), ependymal invasion (feature 19), cortical involvement (feature 20), and deep white matter invasion (feature 21). Furthermore, raters had the liberty to include any other non-VASARI features based on their clinical expertise that they deemed instrumental in their decision-making. One week later, the true enhancement quality (tEQ) was rated independently by the same two raters using pre-contrast and additional post-contrast T1-weighted images side-by-side. Subtraction images were not used as movement between the scans might reduce the quality of subtractions. The tEQ determined by each rater was considered the individual ground truth.

Upon completing the rating independently, both raters noted five VASARI features, including the presence of necrosis, multifocality, nonenhancing tumor margins, substantial edema, and diffusion as instrumental ones in their deci-

sion-making process. Among non-VASARI imaging features, the T2-FLAIR mismatch sign and T2 inhomogeneity were consistently preferred by each of the raters as helpful in their decision-making process. Subsequently, they reached a consensus on the primarily selected VASARI and non-VASARI imaging features by jointly reevaluating all cases to identify the most predictive features, assessing both the pre- and post-contrast images. Ultimately, they identified necrosis, nonenhancing tumor margins, diffusion, and T2 inhomogeneity as the most influential features in predicting enhancement quality, which correlated well with post-contrast images. Following this, raters proposed the preliminary version of EPDT (Figure 2) by determining the order of these identified four imaging features based on their subjective joint evaluation.

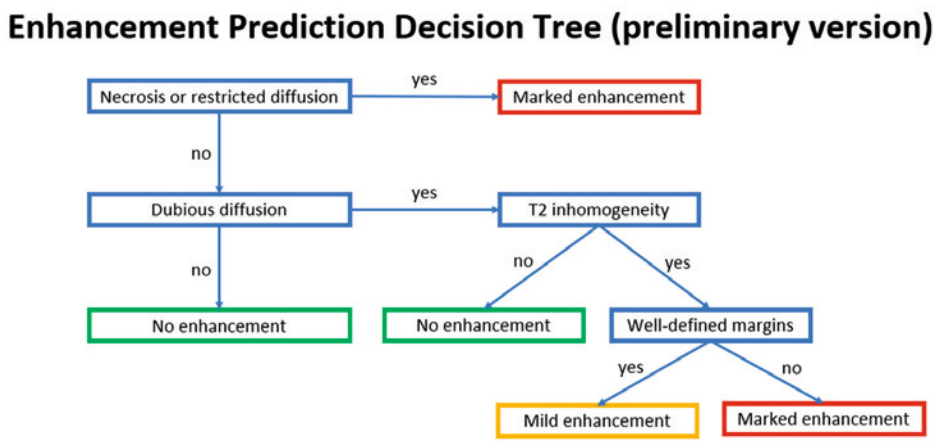


Figure 2: Flow chart describes the preliminary version of the enhancement prediction decision tree.

In a second step, an optimization set (n=38) was assessed to determine the precision of the EPDT and its potential utility in a larger test set. The same raters (V.K., A.A.) evaluated the pEQ by exclusively considering the imaging features included in the EPDT using GBCA-free sequences. Additionally, raters were tasked to predict the thickness of the enhancing margin (pTEM), given its apparent correlation with the presence of necrosis, based on their observation of the development set results. pTEM was classified into the categories "rim enhancement" and "solid enhancement." Tumors with predicted necrotic components using pre-contrast MRI sequences were classified into the "rim

pTEM" group, while others were placed in the "solid pTEM" group. One week later, raters assessed the tEQ and true thickness of enhancing margin (tTEM), the individual ground truth, using pre- and post-contrast T1-weighted sequences. Considering tTEM, if an enhancing area covered a central necrotic region, the enhancing margin was categorized as a "rim." Conversely, if there was only solid enhancement without a rim surrounding the necrotic area, the margin was classified as "solid." According to the optimization set outcomes, minor adaptations of the EPDT were allowed before starting the evaluation of the test set using the final version of EPDT (Figure 3). The only modification made was the incorporation of the pTEM into the decision tree.

Figures 4 and 5 show representative images depicting EPDT imaging features.

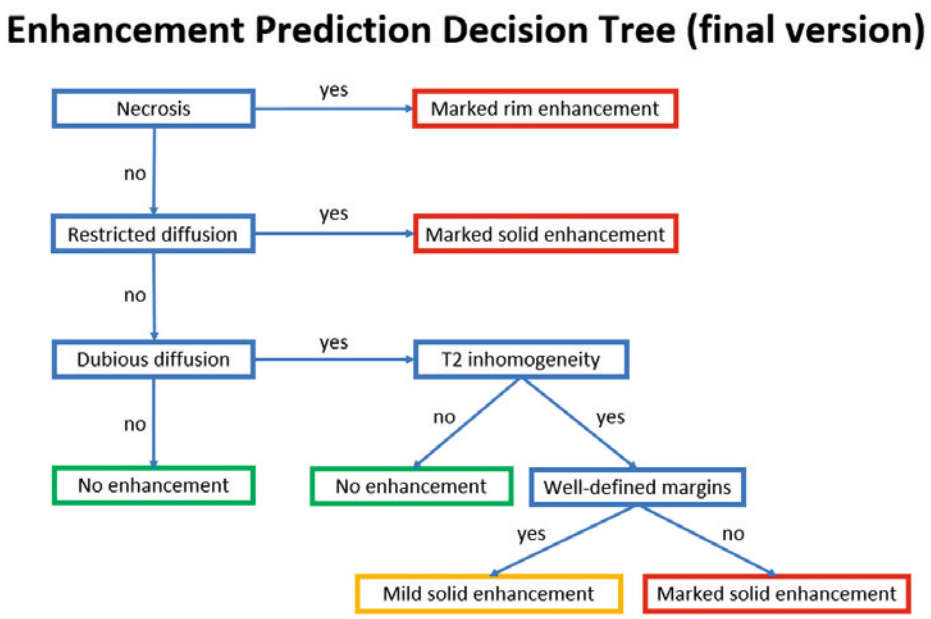


Figure 3: Flow chart describes the final version of the enhancement prediction decision tree.

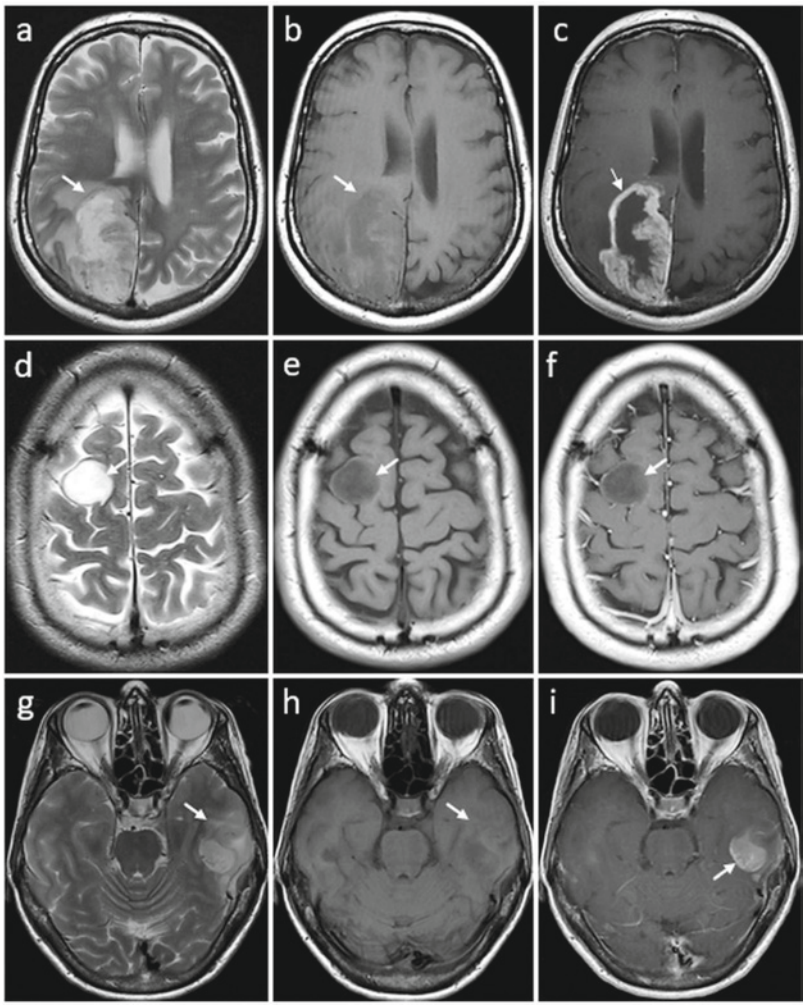


Figure 4: Demonstrative cases for the evaluation of necrosis, T2 inhomogeneity and nonenhancing tumor margins. Necrosis (a,b,c; glioblastoma, IDH-wildtype): right parietal lesion (white arrows) with central necrosis characterized by irregular and thick margins, and internal characteristics of T2 hyperintensity (a) and T1 hypointensity (b). Contrast-enhanced image (c) shows marked rim enhancement. T2 homogeneity and well-defined margins (d,e,f; glioblastoma, IDH-wildtype): right frontal lesion (white arrows) with homogeneous T2 hyperintense signal and well-defined margins (d). There is no signal difference on the contrast-enhanced image (f) compared to the pre-contrast T1-weighted image (e) compatible with nonenhancing glioma. T2 inhomogeneity and ill-defined margins (g,h,i; glioblastoma, IDH-wildtype): left temporal lesion (white arrows) with T2 heterogeneous signal and ill-defined margins (g). Contrast-enhanced image shows a significant signal increase in the posterior part of the lesion (i) compared to the pre-contrast T1-weighted image (h), compatible with marked solid enhancement.

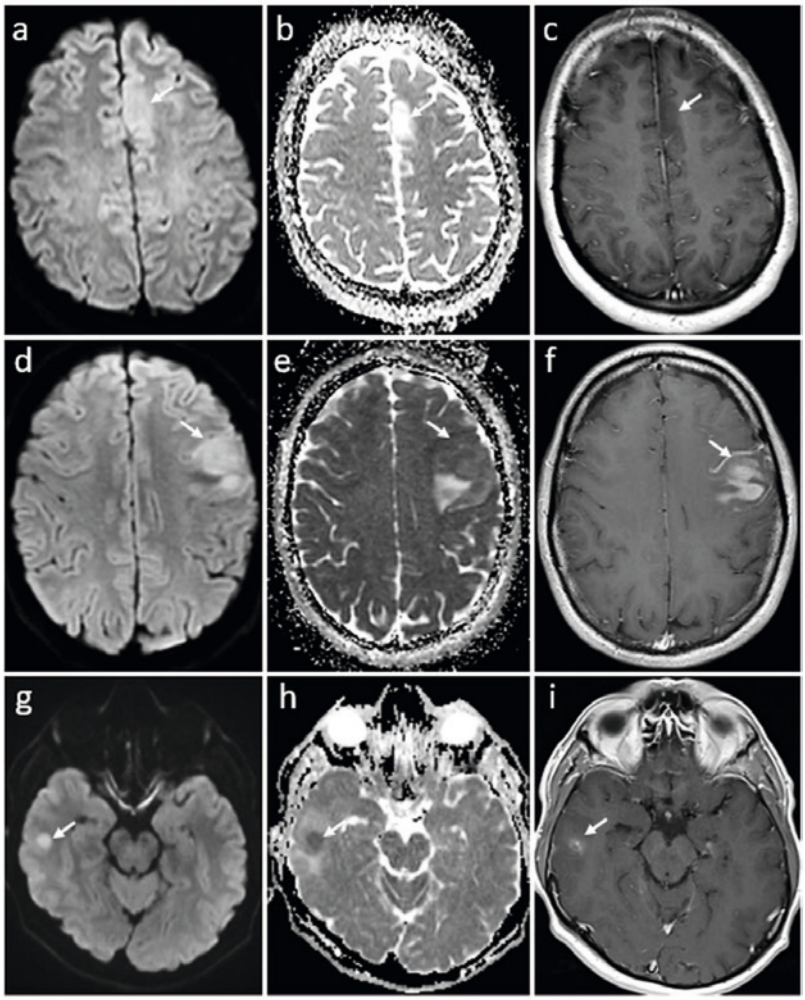


Figure 5: Demonstrative cases for the evaluation of diffusion. Facilitated diffusion (a,b,c; low-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted): Left frontal lesion (white arrows) with hyperintense signal on b-1000 map of diffusion-weighted imaging (DWI) (a) and higher apparent diffusion coefficient (ADC) signal (b) than the normal cortex. There is no enhancement on the contrast-enhanced image (c). Dubious diffusion (d,e,f; glioblastoma, IDH-wildtype): Left frontal lesion (white arrows) with hyperintense signal on DWI (d) and intermediate ADC signal (e) comparable to the normal cortex. Contrast-enhanced image shows marked solid enhancement (f). Restricted diffusion (g,h,i; glioblastoma, IDH-wildtype): Right temporal lesion (white arrows) with hyperintense signal on DWI (g) and lower ADC signal (h) than the normal cortex. Contrast-enhanced image shows marked solid enhancement in the corresponding area (i).

Test set

The performance of human raters predicting enhancement quality (EQ) and thickness of enhancing margin (TEM) was evaluated in a large cohort (n=303). The evaluation was carried out independently by three raters: the same raters who developed EPDT (V.K., AA.) plus a third rater M.C., a fourth-year medical student with no prior radiology experience. The third rater, M.C., underwent iterative EPDT training using the development and optimization set cohort until their rating was deemed adequate. All raters were provided with guide material for the rating, which included detailed definitions of the involved imaging features (Table 2) and the final EPDT flow chart (Figure 3).

Table 2. Definition of imaging features involved in enhancement prediction decision tree including enhancement quality and thickness of enhancing margin

Imaging features	Definitions
Necrosis, modified VASARI feature 7* (yes, no)	Region displaying irregular and/or thick margins, accompanied by imaging features of T1 hypointensity, T2 hyperintensity, and elevated ADC values resembling fluid Areas of a cyst, a cluster of microcysts, or a dilated perivascular space should be excluded
Diffusion, VASARI feature 17* (no/facilitated, dubious, yes/restricted)	No/facilitated- high or low signal intensity on b-1000 map of DWI with relevant high ADC values compared to the normal brain parenchyma Dubious- high signal intensity on b-1000 map of DWI with relevant normal brain parenchyma-like ADC values Yes/restricted- high signal intensity on b-1000 map of DWI with relevant low ADC values compared to the normal brain parenchyma Areas with low ADC signal intensity related to necrotic/hemorrhagic components should be excluded
T2 signal inhomogeneity (no/homogeneous, yes/heterogeneous)	Homogeneous- almost the same signal intensity throughout the tumor except for the lesion rim, vessels (dark dots or lines), cysts, perivascular spaces, and probably infiltrated but normal-appearing cortex compared to the other tumor parts Heterogeneous- mainly different signal intensity, including hypointense, isointense, and/or hyperintense signal compared to normal brain cortex, throughout the tumor
Nonenhancing tumor margins, VASARI feature 13* (well-defined, ill-defined)	Well defined- tumor margins should be considered well-defined if they can easily be traced throughout almost the entire tumor (>90% of the tumor volume) Ill-defined- fuzzy, blurred margins or margins following white matter tracts and difficult to differentiate from surrounding edema should be considered ill-defined

Table 2. Definition of imaging features involved in enhancement prediction decision tree including enhancement quality and thickness of enhancing margin *(Continued)*

Imaging features	Definitions
Enhancement quality, VASARI feature 4* (marked, mild, no)	Qualitative degree of contrast enhancement is defined as having all or portions of the tumor that demonstrate a higher signal on the postcontrast T1-weighted images compared to precontrast T1-weighted images Marked enhancement- obvious tissue enhancement characterized by the significantly higher signal on the postcontrast T1-weighted images compared to precontrast T1-weighted images Mild enhancement- when a barely discernible but unequivocal degree of enhancement is present relative to pre-contrast images No enhancement- no difference between pre-contrast and post-contrast images
Thickness of enhancing margin, modified VASARI feature 11* (rim, solid)	Rim- if there is an enhancing rim around central necrosis, the grade should be rim Solid- if there is only solid enhancement and no rim, the grade should be solid

Table 2 describes the definition of imaging features involved in enhancement prediction decision tree including enhancement quality and thickness of enhancing margin.
Abbreviations: ADC = apparent diffusion coefficient, DWI = diffusion-weighted imaging, VASARI = Visually AcceSAble Rembrandt Images
*https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project[18]

Raters used final EPDT, relying on GBCA-free sequences, to assess pEQ and pTEM, while tEQ and tTEM, representing the individual gold standard, were evaluated using pre- and post-contrast T1-weighted images. EPDT predictions and ground truth assessment of contrast-enhanced T1-weighted scans were performed in two runs. In each run, the subject order was randomized, and either GBCA-free or GBCA-enhanced scans were randomly presented for evaluation. The second run contained the same patients in a differently randomized order and their unseen respective GBCA-free or GBCA-enhanced datasets. This approach was chosen to reduce case recognition and mitigate confirmation bias, which would be caused by rating all GBCA-free scans consecutively.

Figure 6 describes the study pipeline.

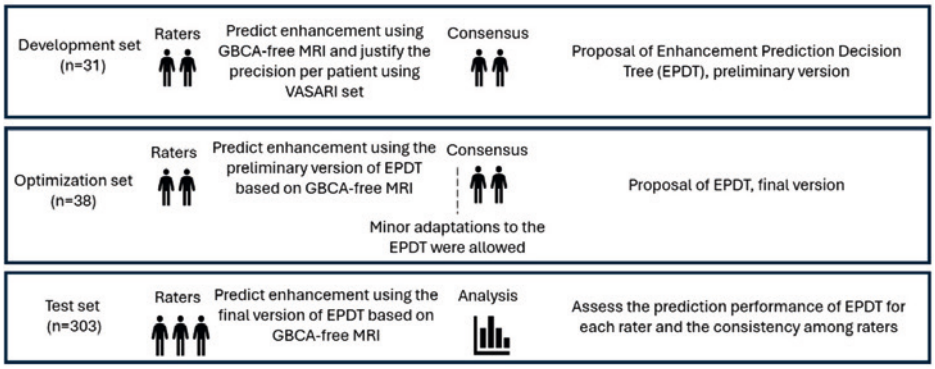


Figure 6: Flowchart demonstrates all steps of the study design. GBCA = gadolinium-based contrast agent, n = number, VASARI = Visually AcceSAble Rembrandt Images.

Histomolecular diagnosis

The histomolecular diagnosis followed the 2021 World Health Organization CNS tumor classification. Isocitrate dehydrogenase (IDH) status was determined through immunohistochemistry, next-generation sequencing, and/or methylation profiling, and 1p/19q-codeletion status was evaluated using loss of heterozygosity (LOH) analysis or methylation profiling. Gliomas are classified into low-grade if they are grade 2 and high-grade if they are grade 3 or 4. IDH-wild-type diffuse gliomas are considered high-grade regardless of histological grade due to their typically aggressive clinical behavior.

Statistical Analysis

The prediction performance of human raters was assessed using confusion matrices. Accuracy, sensitivity, specificity, and positive and negative predictive values were calculated accordingly. A subgroup analysis was conducted to investigate the potential relationship between histomolecular glioma diagnosis and failed enhancement predictions.

Inter-rater agreement was analyzed by comparing all three raters (group) and pairwise. Binary pEQ analysis was also done for the assessment of EQ by combining the categories “marked enhancement” and “mild enhancement” in the category “presence of enhancement,” as often the clinical consequence is linked to the presence, not the extent of enhancement.

Pairwise inter-rater agreement in unordered features was assessed with unweighted Cohen's kappa, complemented with prevalence-adjusted and bias-adjusted kappa (PABAK),[19,20] to show the potential impact of imbalance in the dataset. Agreement in the ordered features (diffusion restriction, enhancement quality) was assessed with linearly weighted Cohen's kappa. This methodology was also used for the intra-rater inter-group agreement analysis, which compares the predicted and true enhancement features per rater using GBCA-free and GBCA-enhanced datasets, respectively.

Group inter-rater agreement in unordered and ordered features was assessed with Fleiss' kappa and Kendall's W (coefficient of concordance), respectively.

Agreement values were interpreted as follows: 0.01-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, 0.81-0.99 almost perfect.[21] Hotelling's T2 test was used according to the study by Vanbelle[22] to compare the agreements between GBCA-free and GBCA-enhanced assessments. No statistical correction was applied to prevent artificial improvement of the results as this could obscure the significance of the difference between GBCA-free and GBCA-enhanced datasets. The threshold for significance was $P < 0.05$.

R package 4.3.0 was employed for the analyses. Additionally, the "multiagree" R[22] package was used for bootstrapping (1000 iterations) of metrics to estimate confidence intervals and for Hotelling's T2 test. The epiR package 2.0.68 (R Foundation for Statistical Computing, Vienna, Austria, https://CRAN.R-project.org/package=epiR) was used for PABAK.

RESULTS

Development and optimization sets

The application of the preliminary version of EPDT improved pEQ overall accuracy in the optimization set (n=38, male=22, mean age=59±15.9 years, high-grade (grade 3/4)=33, IDH-wildtype=32, 1p/19-codeleted=1) round compared with the development set (n=31, male=17, mean age=52±14.9 years, high-grade (grade 3/4)=22, IDH-wildtype=15, 1p/19-codeleted=7) round, from 0.68/0.64

to 0.95/0.97, respectively, for raters 1 and 2. The results for binary pEQ were 0.84/0.74 and 0.97/0.97, respectively.

Test set: Prediction performance of human raters

Table 3 describes the cohort characteristics of the test set.

Table 3. The main characteristics of the test set

Age, years±SD	56.7±14.2	
Female / Male	114 (38%) /189 (62%)	
Tumor location	frontal n=114, parietal n=62, temporal n=87, occipital n=16, insula n=14, thalamus n=8, corpus callosum n=2	
Tumor side	right n=153, middle n=7, left n=143	
Histological grade	LGG n=54	HGG n=249
IDH mutation status	IDHm n=82	IDHwt n=221
1p/19q co-deletion status	1p/19q-codeleted n=34	1p/19q-non-codeleted n=269 (IDHm n=48, IDHwt n=221)

Table 3 describes the primary characteristics of the test set (n=303), encompassing patient demographics, descriptive imaging features, and histopathological results of the included adult-type diffuse gliomas.
Abbreviations: HGG = high-grade (grade 3/4) glioma, IDHm/wt = isocitrate dehydrogenase mutant/wildtype, LGG = low-grade (grade 2) glioma, n = number, SD = standard deviation

The overall accuracy of the pEQ (marked, mild, or no enhancement) was 0.86 (95%-CI: 0.81-0.90), 0.89 (95%-CI: 0.85-0.92), and 0.92 (95%-CI: 0.89-0.95) for raters 1, 2, and 3, respectively. In particular, mild enhancement was often falsely classified (Table 4). The results improved when binary pEQ ("presence and absence of enhancement") were assessed: 0.89 (95%-CI: 0.85-0.92), 0.92 (95%-CI: 0.89-0.95), and 0.93 (95%-CI: 0.89-0.95). The overall accuracy of pTEM (solid, rim, or no enhancement) was 0.84 (95%-CI: 0.79-0.88), 0.88 (95%-CI: 0.84-0.92), and 0.89 (95%-CI: 0.85-0.92) for raters 1, 2, and 3, respectively. Table 4 lists rater-based confusion matrices and sensitivity, specificity, and positive and negative predictive values.

Table 4. Human prediction performance results for enhancement features

Results		Rater 1				Rater 2				Rater 3			
		EQ (enhancement categories: 0=no, 1=mild, 2=marked)											
Confusion matrix			Predicted				Predicted				Predicted		
			0	1	2		0	1	2		0	1	2
		True 0	53	1	7	True 0	46	0	18	True 0	60	2	8
		True 1	8	0	10	True 1	2	0	9	True 1	0	5	0
		True 2	17	0	207	True 2	3	2	223	True 2	12	1	215
Sensitivity	0	0.68				0.90				0.83			
	1	0.00				0.00				0.63			
	2	0.92				0.89				0.96			
Specificity	0	0.96				0.93				0.96			
	1	0.94				0.96				1.00			
	2	0.78				0.91				0.84			
PPV	0	0.87				0.72				0.86			
	1	0.00				0.00				1.00			
	2	0.92				0.98				0.94			
NPV	0	0.90				0.98				0.95			
	1	1.00				0.99				0.99			
	2	0.78				0.64				0.89			

Table 4. Human prediction performance results for enhancement features *(Continued)*

Results		Rater 1			Rater 2			Rater 3		
EQ binary (enhancement categories 0=absent, 1=present)										
Confusion matrix			Predicted			Predicted			Predicted	
			0	1		0	1		0	1
	True	0	53	8	True 0	46	18	True 0	60	10
		1	25	217	1	5	234	1	12	221
Sensitivity		0.68			0.90			0.83		
Specificity		0.96			0.93			0.96		
PPV		0.87			0.72			0.86		
NPV		0.90			0.98			0.95		
TEM (enhancement categories 0=no, 1=rim, 2=solid)										
Confusion matrix			Predicted			Predicted			Predicted	
			0	1	2	0	1	2	0	1	2
	True	0	53	3	5	True 0 46	1	17	True 0 60	6	4
		1	6	186	10	1 1	190	5	1 7	190	7
		2	19	6	15	2 4	8	31	2 5	4	20
Sensitivity	0	0.68			0.90			0.83		
	1	0.95			0.95			0.95		
	2	0.50			0.58			0.65		
Specificity	0	0.96			0.93			0.96		
	1	0.85			0.94			0.86		
	2	0.91			0.95			0.97		

Table 4. Human prediction performance results for enhancement features *(Continued)*

Results		Rater 1	Rater 2	Rater 3
PPV	0	0.87	0.72	0.86
	1	0.92	0.97	0.93
	2	0.38	0.72	0.69
NPV	0	0.90	0.98	0.95
	1	0.91	0.92	0.90
	2	0.94	0.92	0.96

Table 4 describes the prediction performance for each rater in evaluating glioma enhancement features.

Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using the gadolinium-based contrast agent (GBCA)-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences)

Abbreviations: EQ = enhancement quality, NPV = negative predictive value, PPV = positive predictive value, TEM = thickness of enhancing margin

Subgroup analysis exploring failed predictions

Subgroup analysis revealed that failed predictions of EQ and TEM were much more frequent in low-grade (grade 2) and IDH-mutant gliomas than in high-grade (grade 3/4) and IDH-wildtype counterparts. Table 5 provides the details for rater 1 as an example. Therefore, we reassessed the performance of the prediction accuracy of EPDT after excluding oligodendrogliomas, as these were most frequently falsely classified. The overall accuracy of pEQ improved to 0.89 (95%-CI: 0.85-0.93), 0.91 (95%-CI: 0.87-0.95), and 0.94 (95%-CI: 0.91-0.97) for raters 1, 2, and 3, respectively. Similarly, the results for binary pEQ [0.92 (95%-CI: 0.88-0.95), 0.95 (95%-CI: 0.91-0.97), and 0.94 (95%-CI: 0.91-0.97)] and pTEM [0.87 (95%-CI: 0.82-0.91), 0.91 (95%-CI: 0.87-0.94), and 0.91 (95%-CI: 0.87-0.94)] also increased (for details see Table 6).

Table 5. Subgroup analysis results exploring failed enhancement predictions based on histopathological diagnosis for Rater 1

Confusion matrix					Possible failed EPDT steps
EQ (enhancement categories: 0=no, 1=mild, 2=marked)					
		Predicted			0-1, 0-2, 1-0: dubious diffusion or T2 inhomogeneity 1-2: necrosis/diffusion restriction or well-defined margins 2-0: necrosis/diffusion restriction or well-defined margins
		0	1	2	
True	0	53	1 [LGG/HGG n=1 (2%), ODG n=(3%)]	7 [LGG/HGG n=4 (7%)/3 (1%), ODG/Astro n=4 (12%)/3 (6%)]	
	1	8 [LGG/HGG n=4 (7%)/4 (2%), ODG/Astro/GB n=3 (9%)/2 (4%)/3 (1%)]	0	10 [LGG n=4 (7%), HGG n=6 (2%), ODG n=2 (6%), Astro n=4 (8%), GB n=4 (2%)]	
	2	17 [LGG/HGG n=5 (9%)/12 (5%), ODG/Astro/GB n=4 (12%)/6 (13%)/7 (3%)]	0	207	
EQ binary (enhancement categories 0=absent, 1=present)					
		Predicted			0-1, 1-0: necrosis/diffusion restriction or dubious diffusion or T2 inhomogeneity
		0	1		
True	0	53	8 ([LGG/HGG n=5 (9%)/3 (1%), ODG/Astro n=5 (15%)/3 (6%)]		
	1	25 [LGG/HGG n=9 (17%)/16 (6%), ODG/Astro/GB n=7 (21%)/8 (17%)/10 (5%)]	217		

Table 5. Subgroup analysis results exploring failed enhancement predictions based on histopathological diagnosis for Rater 1 *(Continued)*

Confusion matrix					Possible failed EPDT steps
TEM (enhancement categories 0=no, 1=rim, 2=solid)					
		Predicted			0-1, 1-0, 1-2, 2-1: necrosis 0-2, 2-0: diffusion restriction or dubious ADC or T2 inhomogeneity
		0	1	2	
True	0	53	3 [LGG/HGG n=2 (4%)/1 (<1%), ODG/Astro n=2 (6%)/1 (2%)]	5 [LGG/HGG n=3 (6%)/2 (<1%), ODG/Astro n=3 (9%)/2 (4%)]	
	1	6 [LGG n=6 (11%), ODG/Astro/GB n=1 (3%)/3 (6%)/2 (1%)]	186	10 [LGG n=10 (19%), ODG/Astro/GB n=1 (3%)/1 (2%)/8 (4%)]	
	2	19 [LGG/HGG n=9 (17%)/10 (4%), ODG/Astro/GB n=6 (18%)/5 (10%)/8 (4%)]	6 [LGG/HGG n=2 (4%)/4 (2%), ODG/Astro/GB n=1 (3%)/1 (2%)/4 (2%)]	15	

Table 5 describes the subgroup analysis results exploring failed enhancement predictions based on histopathological diagnosis for Rater 1.
Total sample size: LGG/HGG n=54 (18%)/249 (82%), ODG/Astro/GB n=34 (11%)/48 (16%)/221 (73%)
Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using the gadolinium-based contrast agent (GBCA)-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences)
Abbreviations: Astro = astrocytoma, IDH-mutant, EQ = enhancement quality, EPDT = enhancement prediction decision tree, GB = glioblastoma, IDH-wildtype, HGG = high-grade (grade 3/4) glioma, ODG = oligodendroglioma, IDH-mutant and 1p/19q-codeleted, LGG = low-grade (grade 2) glioma, n= number, TEM = thickness of enhancing margin

Table 6. Human prediction performance results for enhancement features after excluding oligodendroglioma cases

Results		Rater 1				Rater 2				Rater 3			
EQ (enhancement categories: 0=no, 1=mild, 2=marked)													
Confusion matrix			Predicted				Predicted				Predicted		
			0	1	2		0	1	2		0	1	2
	True	0	39	0	3	True 0	34	0	10	True 0	40	2	6
		1	5	0	8	True 1	1	0	8	True 1	0	4	0
		2	13	0	201	2	3	1	212	2	7	1	209
Sensitivity	0		0.68				0.89				0.85		
	1		0.00				0.00				0.57		
	2		0.95				0.92				0.97		
Specificity	0		0.99				0.96				0.96		
	1		0.95				0.97				1.00		
	2		0.77				0.90				0.85		
PPV	0		0.93				0.77				0.83		
	1		0.00				0.00				1.00		
	2		0.94				0.98				0.96		

Table 6. Human prediction performance results for enhancement features after excluding oligodendroglioma cases *(Continued)*

Results		Rater 1	Rater 2	Rater 3
NPV	0	0.92	0.98	0.97
	1	0.00	1.00	0.99
	2	0.80	0.66	0.88

EQ binary (enhancement categories 0=absent, 1=present)

Confusion matrix			Predicted 0	Predicted 1			Predicted 0	Predicted 1			Predicted 0	Predicted 1
	True	0	39	3	True	0	34	10	True	0	40	8
		1	18	209		1	4	221		1	7	214
Sensitivity			0.68				0.89				0.85	
Specificity			0.99				0.96				0.96	
PPV			0.93				0.77				0.83	
NPV			0.92				0.98				0.97	

TEM (enhancement categories 0=no, 1=rim, 2=solid)

Confusion matrix			Predicted 0	Predicted 1	Predicted 2			Predicted 0	Predicted 1	Predicted 2			Predicted 0	Predicted 1	Predicted 2
	True	0	39	1	2	True	0	34	0	10	True	0	40	4	4
		1	5	183	9		1	1	188	5		1	6	188	7
		2	13	5	12		2	3	5	23		2	1	3	16
Sensitivity	0		0.68					0.89						0.85	
	1		0.97					0.97						0.96	
	2		0.52					0.60						0.59	

Table 6. Human prediction performance results for enhancement features after excluding oligodendroglioma cases *(Continued)*

Results		Rater 1	Rater 2	Rater 3
Specificity	0	0.99	0.96	0.96
	1	0.83	0.92	0.82
	2	0.93	0.97	0.98
PPV	0	0.93	0.77	0.83
	1	0.93	0.97	0.94
	2	0.40	0.74	0.80
NPV	0	0.92	0.98	0.97
	1	0.92	0.93	0.90
	2	0.95	0.94	0.96

Table 6 describes the prediction performance for each rater in evaluating glioma enhancement features after excluding oligodendroglioma cases.
Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using the gadolinium-based contrast agent (GBCA)-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences).

Abbreviations: EQ = enhancement quality, NPV = negative predictive value, PPV = positive predictive value, TEM = thickness of enhancing margin

Inter- and intra-rater agreement for EQ and TEM

Group inter-rater agreement analysis (Figure 7) revealed moderate agreement for both pEQ and tEQ [Kendall`s W 0.42 (95%-CI: 0.36-0.48) and 0.55 (95%-CI: 0.48-0.62)], substantial and almost perfect agreement for the pTEM and tTEM, respectively [Fleiss' kappa 0.66 (95%-CI: 0.60-0.71) and 0.83 (95%-CI: 0.79-0.88)]. The results were further improved for both pEQ and tEQ [Fleiss' kappa 0.65 (95%-CI: 0.57-0.73) and 0.87 (95%-CI: 0.82-0.93)] when binary analysis for enhancement quality (enhancing/nonenhancing) was applied. Pairwise inter-rater agreements were substantial [≥0.61 (95%-CI 0.50-0.72)] and almost perfect [≥0.82 (95%-CI: 0.75-0.89)] for the predicted and true features, respectively (see Table 7).

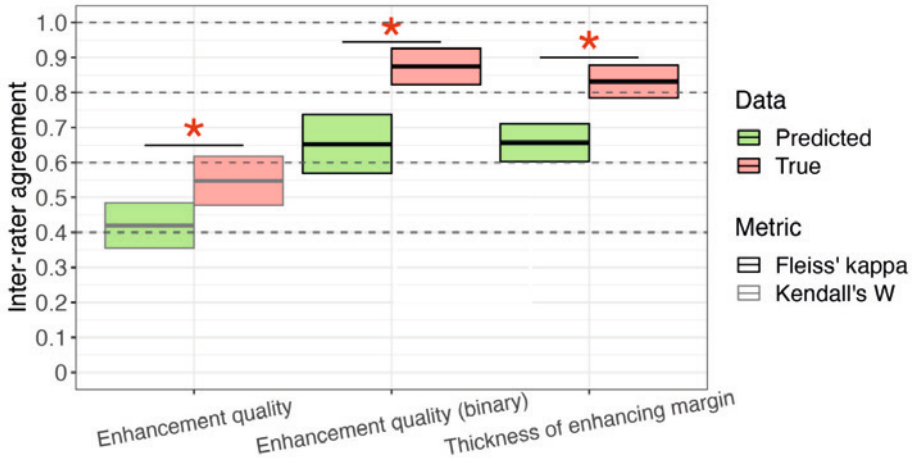


Figure 7: Group inter-rater agreement analysis in true (red color bars) and predicted (green color bars) enhancement quality and thickness of enhancing margin among all three raters. Red stars show significant differences between evaluation agreements of true and predicted datasets. Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using gadolinium-based contrast agent (GBCA)-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences).

Comparison analysis of agreements showed significant differences between the evaluation agreements of predicted and true features (compare Figure 7 and Table 7 for details). Intra-rater inter-group agreement analysis comparing predicted and true enhancement features per rater demonstrated substantial to almost perfect agreements [≥0.68 (95%-CI: 0.61-0.75)] for each rater without significant differences in agreement among all raters (*P*-values: EQ 0.10, binary EQ 0.22, TEM 0.07) (see Figure 8 and Table 8).

Table 7. Pairwise inter-rater agreement results for the assessment of the true/predicted enhancement

EQ/TEM		Raters 1 & 2	Raters 1 & 3	Raters 2 & 3
tEQ*		0.90; 95%CI 0.86-0.95	0.89; 95%CI 0.85-0.94	0.92; 95%CI 0.87-0.96
pEQ*		0.66; 95%CI 0.57-0.78	0.70; 95%CI 0.61-0.79	0.61; 95%CI 0.51-0.71
P-value^		<0.001	<0.001	<0.001
tEQ (binary)	**	0.87; 95%CI 0.79-0.94	0.87; 95%CI 0.80-0.94	0.88; 95%CI 0.82-0.95
	#	0.91	0.91	0.92
pEQ (binary)	**	0.68; 95%CI 0.58-0.78	0.68; 95%CI 0.58-0.77	0.61; 95%CI 0.50-0.72
	#	0.78	0.76	0.74
P-value^		<0.001	<0.001	<0.001
tTEM	**	0.84; 95%CI 0.78-0.91	0.82; 95%CI 0.75-0.89	0.84; 95%CI 0.78-0.89
	#	0.84	0.82	0.83
pTEM	**	0.69; 95%CI 0.62-0.76	0.64; 95%CI 0.57-0.71	0.64; 95%CI 0.57-0.72
	#	0.68	0.64	0.64
P-value^		<0.001	<0.001	<0.001

Table 7 describes the pairwise inter-rater agreement results for enhancement features based on assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets.
Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using GBCA-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences).
Abbreviations: CI = confidence interval, pEQ/tEQ = predicted/true enhancement quality, pTEM/tTEM = predicted/true thickness of enhancing margin,
*weighted Cohen's kappa
^comparison of the agreements in true- and predicted EQ/TEM
**unweighted Cohen`s kappa
prevalence-adjusted and bias-adjusted kappa

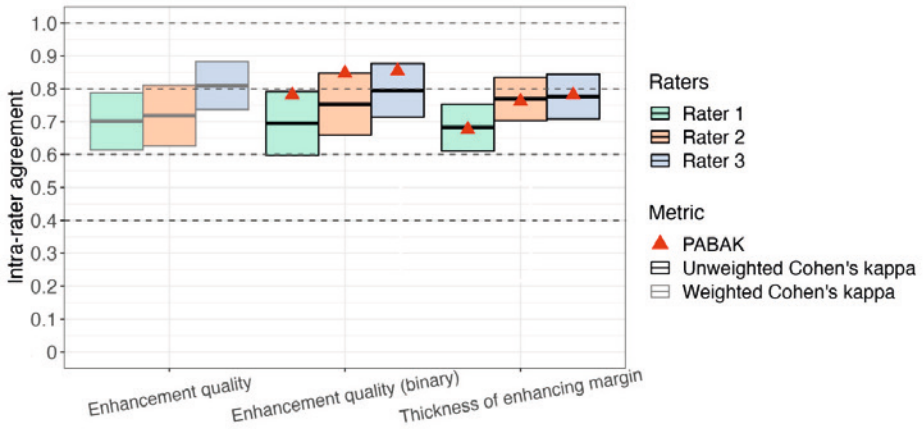


Figure 8: Intra-rater inter-group agreement analysis shows the comparison between true and predicted enhancement features per rater. Prevalence-adjusted and bias-adjusted kappa (PABAK) values (red triangles) are comparable with unweighted Cohen's kappa for thickness of enhancing margin, indicating a negligible impact of imbalance on the agreement metrics. However, there are increased PABAK values (red triangles) compared to Cohen's kappa for enhancement quality (binary), indicating the imbalance in the dataset. Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using gadolinium-based contrast agent (GBCA)-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences).

Table 8. Intra-rater inter-group agreement results for assessment of the true/predicted enhancement

tEQ/tTEM vs. pEQ/pTEM		Rater 1	Rater 2	Rater 3
EQ*		0.70; 95%CI 0.61-0.79	0.72; 95%CI 0.63-0.81	0.81; 95%CI 0.74-0.88
P-value^		0.10		
EQ (binary)	**	0.69; 95%CI 0.60-0.79	0.75; 95%CI 0.66-0.85	0.80; 95%CI 0.71-0.88
	#	0.78	0.85	0.85
P-value^		0.22		
TEM	**	0.68; 95%CI 0.61-0.75	0.77; 95%CI 0.70-0.83	0.78; 95%CI 0.71-0.85
	#	0.68	0.76	0.78
P-value^		0.07		

Table 8 describes the intra-rater inter-group agreement results comparing predicted and true enhancement features per rater.
Note: Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using gadolinium-based contrast agent (GBCA)-free dataset (pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, diffusion-weighted imaging). True enhancement features, representing the individual reference standard, are based on assessing the GBCA-enhanced dataset (GBCA-free and post-contrast T1-weighted sequences)
Abbreviations: CI = confidence interval, pEQ/tEQ- predicted/true enhancement quality, pTEM/tTEM-predicted/true thickness of enhancing margin
*weighted Cohen's kappa
^comparison of the agreements among all raters
**unweighted Cohen`s kappa
#prevalence-adjusted and bias-adjusted kappa

Agreement analysis for single imaging features involved in EPDT

Group inter-rater agreement analysis revealed almost perfect agreement for necrosis identification in both GBCA-enhanced and GBCA-free datasets [Fleiss' kappa 0.85 (95%-CI: 0.80-0.90) and 0.83 (95%-CI: 0.78-0.88)]. The agreement for other imaging features was moderate [≥0.43 (95%-CI: 0.28-0.57)]. Notably, the availability of post-contrast images did not significantly influence imaging feature agreement (*P*-values: necrosis 0.49, diffusion restriction 0.81, T2 inhomogeneity 0.63, nonenhancing tumor margins 0.65) (see Figure 9 and Table 9). A pairwise inter-rater agreement was almost perfect for necrosis [≥0.80 (95%-CI: 0.73-0.88)] and fair to moderate for other features [≥0.33 (95%-CI: 0.25-0.41)] in both GBCA-enhanced and GBCA-free datasets. However, applying PABAK analysis showed that the results for T2 inhomogeneity and nonenhancing tumor margins were substantial to almost perfect (≥0.71) level (Table 10). Intra-rater

agreement in the assessment of necrosis was almost perfect [≥0.82 (95%-CI: 0.75-0.89)] for all three raters. This analysis in the evaluation of other features showed substantial- almost perfect agreements [≥0.66 (95%-CI: 0.55-0.77)] for raters 2 and 3 while being fair to moderate [≥0.35 (95%-CI: 0.18-0.53)] for rater 1. Applying PABAK analysis revealed better agreement for T2 inhomogeneity and nonenhancing tumor margins (≥0.78; Figure 10 and Table 11).

Figure 11 shows case examples of successful and failed enhancement predictions based on necrosis assessment.

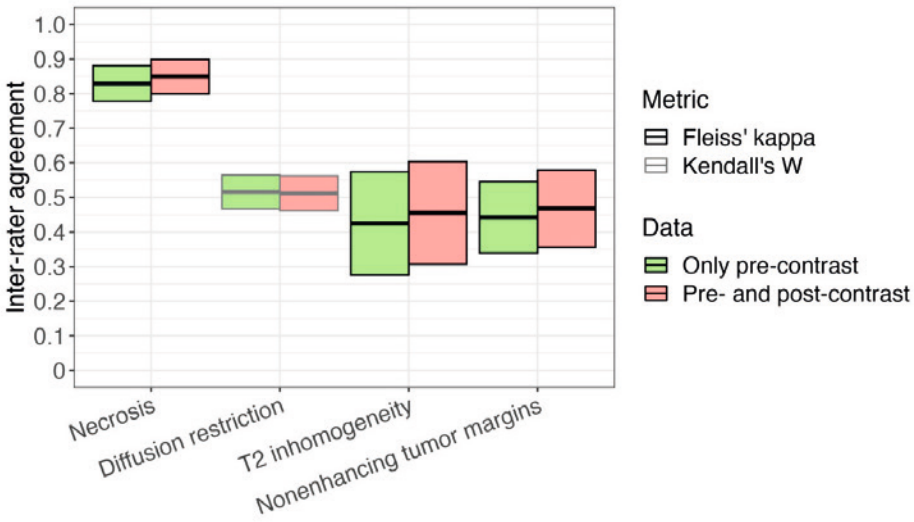


Figure 9: Group inter-rater agreement analysis in imaging features involved in enhancement prediction decision tree based on gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets among all three raters. There are no significant differences between evaluation agreements of GBCA-free and GBCA-enhanced datasets. Note: GBCA-free dataset includes pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and post-contrast T1-weighted sequences.

Table 9. Group inter-rater agreement results for assessment of single imaging features

Features involved in EPDT	GBCA-enhanced dataset	GBCA-free dataset
Necrosis*	0.85; 95%CI 0.80-0.90	0.83; 95%CI 0.78-0.88
P-value^	0.49	
Diffusion restriction**	0.52; 95%CI 0.46-0.56	0.52; 95%CI 0.47-0.56
P-value^	0.81	
T2 inhomogeneity*	0.46; 95%CI 0.31-0.60	0.43; 95%CI 0.28-0.57
P-value^	0.63	
Nonenhancing tumor margins*	0.47; 95%CI 0.36-0.58	0.44; 95%CI 0.34-0.56
P-value^	0.65	

Table 9 describes the inter-rater agreement results for single imaging features based on assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets among all raters.
Note: GBCA-free dataset includes pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and post-contrast T1-weighted sequences.
Abbreviations: CI = confidence interval, EPDT = enhancement prediction decision tree, GBCA = gadolinium-based contrast agent
*Fleiss` kappa,
^comparison of the agreements between the assessments of GBCA-enhanced and GBCA-free datasets
**Kendall`s W

Table 10. Pairwise inter-rater agreement results for assessment of single imaging features

EPDT imaging features		Raters 1 & 2	Raters 1 & 3	Raters 2 & 3
Necrosis_GBCA-enhanced	**	0.85; 95%CI 0.78-0.91	0.84; 95%CI 0.77-0.90	0.86; 95%CI 0.80-0.92
	#	0.87	0.86	0.87
Necrosis_GBCA-Free	**	0.87; 95%CI 0.81-0.93	0.80; 95%CI 0.73-0.88	0.82; 95%CI 0.75-0.88
	#	0.88	0.82	0.83
P-value^		0.63	0.42	0.26
Diffusion restriction_GBCA-enhanced*		0.42; 95%CI 0.34-0.50	0.33; 95%CI 0.25-0.41	0.42; 95%CI 0.33-0.52
Diffusion restriction_GBCA-Free*		0.42; 95%CI 0.34-0.49	0.39; 95%CI 0.31-0.47	0.37; 95%CI 0.28-0.47
P-value^		0.94	0.17	0.21
T2 inhomogeneity_GBCA-enhanced	**	0.41; 95%CI 0.24-0.59	0.39; 95%CI 0.22-0.57	0.58; 95%CI 0.40-0.76
	#	0.82	0.81	0.88
T2 inhomogeneity_GBCA-free	**	0.43; 95%CI 0.27-0.60	0.38; 95%CI 0.22-0.55	0.49; 95%CI 0.30-0.67
	#	0.80	0.77	0.87
P-value^		0.84	0.90	0.36
Nonenhancing tumor margins_GBCA-enhanced	**	0.43; 95%CI 0.28-0.57	0.47; 95%CI 0.34-0.61	0.49; 95%CI 0.37-0.62
	#	0.74	0.73	0.71
Nonenhancing tumor margins_GBCA-free	**	0.36; 95%CI 0.20-0.52	0.36; 95%CI 0.21-0.51	0.58; 95%CI 0.45-0.70
	#	0.74	0.71	0.78
P-value^		0.46	0.15	0.24

Table 10 describes the pairwise inter-rater agreement results for single imaging features based on assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets.
Note: GBCA-free dataset includes pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and post-contrast T1-weighted sequences.
Abbreviations: CI = confidence interval, EPDT = enhancement prediction decision tree, GBCA = gadolinium-based contrast agent
*weighted Cohen`s kappa
^comparison of the agreements between GBCA-enhanced and GBCA-free assessments
**unweighted Cohen`s kappa
prevalence-adjusted and bias-adjusted kappa

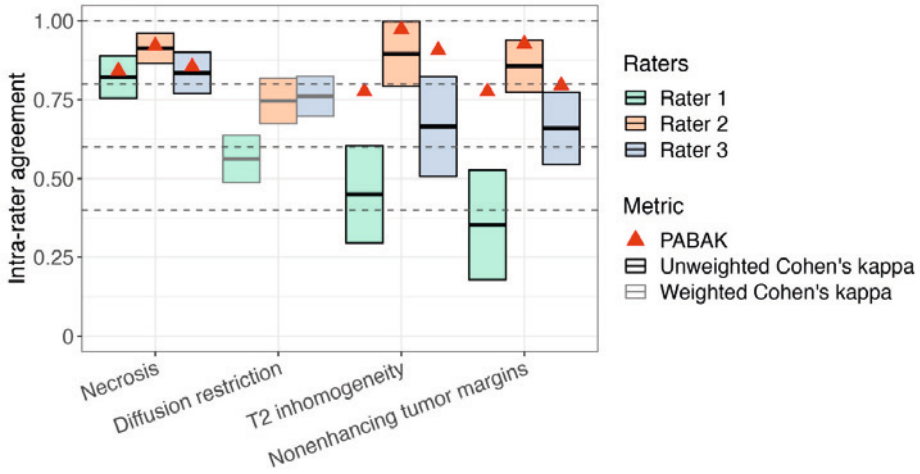


Figure 10: Intra-rater agreement analysis in the assessment of imaging features involved in enhancement prediction decision tree based on the evaluations of gadolinium-enhanced contrast agent (GBCA)-enhanced and GBCA-free datasets for each rater. Prevalence-adjusted and bias-adjusted kappa (PABAK) values (red triangles) are comparable with Cohen's kappa for the necrosis, indicating a negligible impact of imbalance on the agreement metrics. However, there are increased PABAK values (red triangles) compared to Cohen's kappa for the T2 inhomogeneity and nonenhancing tumor margins, indicating the imbalance in the dataset. None of the features show significant differences in agreement among the raters. Note: GBCA-free dataset includes pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and post-contrast T1-weighted sequences.

Table 11. Intra-rater agreement results for assessment of single imaging features

EPDT imaging features		Rater 1	Rater 2	Rater 3
Necrosis	**	0.82; 95%CI 0.75-0.89	0.91; 95%CI 0.87-0.96	0.83; 95%CI 0.77-0.90
	#	0.84	0.92	0.85
P-value^		0.10		
Diffusion restriction*		0.56; 95%CI 0.49-0.64	0.75; 95%CI 0.67-0.82	0.76; 95%CI 0.70-0.82
P-value^		0.10		
T2 inhomogeneity	**	0.45; 95%CI 0.30-0.60	0.90; 95%CI 0.79-0.99	0.67; 95%CI 0.51-0.82
	#	0.78	0.97	0.91
P-value^		0.22		
Nonenhancing tumor margin	**	0.35; 95%CI 0.18-0.53	0.86; 95%CI 0.77-0.93	0.66; 95%CI 0.55-0.77
	#	0.78	0.93	0.80
P-value^		0.07		

Table 11 describes the intra-rater agreement results for single imaging features between assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets per rater.
Note: GBCA-free dataset includes pre-contrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and post-contrast T1-weighted sequences.
Abbreviations: CI = confidence interval, EPDT = enhancement prediction decision tree
*weighted Cohen`s kappa
^comparison of the agreements among all raters
**unweighted Cohen`s kappa
prevalence-adjusted and bias-adjusted kappa

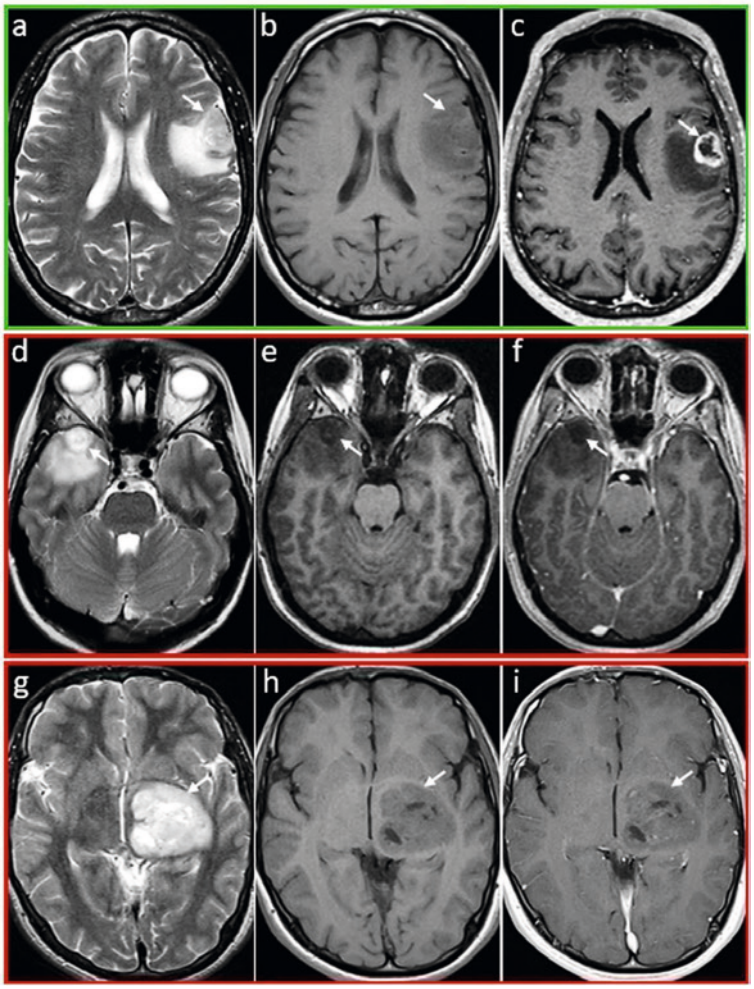


Figure 11: Demonstrative cases illustrating successful (green) versus failed (red) prediction of glioma enhancement based on the evaluation of necrosis using the gadolinium-based contrast-free dataset. Successful prediction of both enhancement quality (pEQ) and thickness of enhancing margin (pTEM; a,b,c; glioblastoma, IDH-wildtype): T2-weighted (a) and pre-contrast T1-weighted images (b) show signal characteristics of central necrosis (white arrows) within the left frontal lesion. The pEQ and pTEM are marked and rim, respectively, based on the enhancement prediction decision tree (EPDT). The contrast-enhanced T1-weighted image (c, white arrow) confirms the prediction results showing marked rim enhancement surrounding the necrotic part of the tumor. Failed prediction of enhancement patterns (d,e,f; high-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted): Right temporal lesion (white arrows) with T2 (d) and T1 (e) signal characteristics of small tumor necrosis. The pEQ and pTEM are "marked" and "rim" enhancement, respectively, based on EPDT. However, contrast-enhanced T1-weighted image (f) showed no increase in the T1 signal intensity compared to pre-contrast T1-weighted image (e)

compatible with nonenhancing glioma. Failed prediction of enhancement patterns (g,h,i; low-grade astrocytoma, IDH-mutant): Left thalamic lesion (white arrows) with T2 (g) and T1 (h) signal characteristics of small multiple necrotic areas in the tumor. The pEQ and pTEM are marked and rim, respectively, based on EPDT. However, the lesion shows a mild solid enhancement pattern on the contrast-enhanced T1-weighted image (i, white arrow).

DISCUSSION

MRI contrast enhancement presence and pattern are commonly used diagnostic and prognostic pillars in oncological neuroradiology. The EPDT algorithm presented here demonstrated that predicting EQ and TEM with GBCA-free MRI sequences is feasible with high accuracy regardless of the rater's experience level. This underscores its independent applicability in clinical settings, even for less-trained readers. Intra-rater inter-group agreement comparing predicted and true enhancement features was consistently substantial for all raters without improvement when using GBCA-enhanced images. It suggests that assessments do not significantly rely on GBCA-enhanced images to evaluate glioma enhancement.

There is hardly any literature studying enhancement prediction using human evaluation methods. Especially for unsupervised AI approaches, it is, however, relevant to know the comparative performance of humans in predicting contrast enhancement and to gain insight into the potential image features triggering decisions. One machine-learning radiomics model predicted glioma enhancement quality using T2-FLAIR images,[23] which demonstrated high accuracy levels [AUC of 0.81 (95%-CI: 0.71-0.90)] in an external validation cohort. Despite its high performance, this method might not outperform human raters using the EPDT algorithm presented here.

The method presented in this study demonstrates the predictive capabilities of human rating of GBCA-free data in a systematic way using a proposed EPDT algorithm and documents its performance for both enhancement features of pEQ and pTEM in a general adult-type diffuse glioma population and different subgroups. The outcomes showed that EPDT works better for high-grade and IDH-wildtype gliomas compared to low-grade and IDH mutant gliomas. Calabrese et al.,[24] who reported an AI-based approach for synthetic contrast

enhancement in low-grade and high-grade gliomas, also noted a histology-dependent factor in enhancement prediction, with lower Dice scores in low-grade gliomas [0.58 (95%-CI: 0.49-0.68) vs. 0.65 (95%-CI: 0.63-0.67)]. Generally, studies on synthetic GBCA-enhanced image generation have shown promising results.[24–26] Kleesiek et al.[25] demonstrated a sensitivity and specificity of over 90% in qualitative and quantitative evaluations of their model, comparing the generated enhancement maps with standard contrast-enhanced T1-weighted images. Interestingly, despite a totally different approach than ours, they also observed a failure of their model with wrongly predicted mildly enhancing tumors as well as the misshaping of the predicted enhancing margin, which we defined as pTEM.

In our proposed human-based EPDT algorithm, the predictive performance did not decrease as the experience level of the raters decreased, with the trained student rater (rater 3) showing at least non-inferior accuracy and consistency. This may reflect the structured approach facilitated by EPDT, which mainly utilizes VASARI features, a standardized glioma vocabulary.[18] Furthermore, raters received comprehensive guidance on all VASARI and non-VASARI features involved in the decision tree. The experience of the trained radiologists may have caused a less pertinent adherence to the decision tree definitions. The pertinent application of the standardized definitions likely reduced the impact of experience levels on decision-making. These results align with the literature,[27–29] highlighting the benefits of standardization in radiological image assessment, diminishing the reliance on experience level across diverse clinical settings. Furthermore, both group and pairwise inter-rater reliability for predicted enhancement were moderate or better, indicating the potential applicability and generalizability of the EPDT algorithm. The reproducible application of EPDT also by less specialized personnel is promising for its use in situations where GBCA administration is not available, e.g., in radiology units of low- and middle-income countries.

When introducing a new diagnostic tool, such as the EPDT, achieving high inter-rater agreement is crucial to guarantee consistency and ensure that decision criteria are easily reproducible. Various studies[30–33] found high inter-rater agreements when assessing the EPDT's most consistent feature, necrosis, with kappa values ranging from 0.71 to 0.96, thus supporting our findings. The re-

maining three features of the EPDT (diffusion restriction, T2 inhomogeneity, and nonenhancing tumor margin) yielded fair to moderate inter-rater agreement outcomes in regular tests. The PABAK analysis, which compensates for unbalanced group comparisons, showed that the results were, in fact, even better, with a level ranging from substantial to almost perfect, correcting the potential influence of dataset imbalance. Prior studies reported agreement values for these imaging features, ranging from 0.36 to 0.85 for diffusion[31–34] and 0.77 to 0.96 for nonenhancing tumor margin.[32,33,35]

The suggested EPDT algorithm could serve as a valuable tool in clinical settings for directing the pretreatment care of gliomas, particularly for patients at higher risk who cannot receive GBCAs due to various factors or who decline intravenous contrast administration. However, in its current version, it should not be used as a substitute for GBCA-enhanced imaging for all patients undergoing MRI examination. The next step is to study its applicability in an external validation cohort as well as its performance in more complex diagnostic scenarios, e.g., when applied to other types of brain lesions than adult-type diffuse gliomas, or when predicting enhancement characteristics in post-treatment scenarios, such as distinguishing treatment-related changes from tumor progression. The post-treatment situation can lead to confounders, especially e.g., from therapy-induced diffusion restriction, an imaging feature of the EPDT in its current version.

This study has several limitations. The retrospective study design involved a single center as the main inherent constraint, potentially impacting the study's overall generalizability and external validity, even though multiple scanners with variable imaging protocols were employed. The datasets were unbalanced between enhancing and nonenhancing tumors. However, their proportions reflect those encountered in real-life conditions within the epidemiological context. Additionally, the inclusion of only treatment-naïve adult-type diffuse gliomas in this study leaves unanswered questions regarding the performance of the proposed decision tree in post-treatment settings or for other tumor types, such as grade 1 and pediatric gliomas or infratentorial/suprasellar tumors. Perfusion weighted imaging, essential for glioma evaluation, was not assessed in this study due to inconsistent availability of arterial spin labeling (ASL), a GBCA-free alternative to the routinely applied GBCA-based perfusion tech-

nique of dynamic susceptibility contrast (DSC)-MRI. Considering the potential relation between increased perfusion and enhancement, perfusion imaging data, particularly ASL, should be evaluated for its contribution to the accuracy of the EPDT. Although EPDT aims to provide a reasonably easy-to-use clinical tool, it might not fully cover all variations in real-world imaging, leading to some differences between the raters or in repeated evaluations. Further validation studies encompassing more raters, different tumors, various therapy stages, and advanced imaging techniques, both GBCA-free and GBCA-enhanced, such as ASL, MR spectroscopy, DSC-MRI, or dynamic contrast-enhanced perfusion MR, are crucial to translating the findings of this study. Moreover, identifying the exact location of predicted enhancing regions, multifocality or satellites and the impact of T1 hyperintense or gradient-echo susceptibility regions on the evaluation was not the focus of this study, paving the way for potential new research directions.

In conclusion, this study proposes an enhancement intensity and shape prediction decision tree utilizing visual imaging features assessed through GBCA-free MRI sequences. The outcomes demonstrate robust and highly accurate predictive performance of enhancement features even in inexperienced raters. Furthermore, this study provides relevant insights for AI study designs about predicted post-GBCA imaging and opportunities for direct applications by radiologists.

REFERENCES

1. Ellingson BM, Bendszus M, Boxerman J, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol 2015;17:1188–98.
2. Weller M, van den Bent M, Preusser M, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. Nat Rev Clin Oncol 2021;18:170–86.
3. Wen PY, van den Bent M, Youssef G, et al. RANO 2.0: Update to the response assessment in neuro-oncology criteria for high- and low-grade gliomas in adults. J Clin Oncol 2023;41:5187–99.
4. Wang Y, Wang K, Wang J, et al. Identifying the association between contrast enhancement pattern, surgical resection, and prognosis in anaplastic glioma patients. Neuroradiology 2016;58:367–74.
5. Castet F, Alanya E, Vidal N, et al. Contrast-enhancement in supratentorial low-grade gliomas: a classic prognostic factor in the molecular age. J Neurooncol 2019;143:515–23.
6. Verburg N, de Witt Hamer PC. State-of-the-art imaging for glioma surgery. Neurosurg Rev 2021;44:1331–43.
7. Certo F, Altieri R, Maione M, et al. FLAIRectomy in supramarginal resection of glioblastoma correlates with clinical outcome and survival Analysis: s prospective, single institution, case series. Oper Neurosurg (Hagerstown) 2021;20:151–63.
8. Haddad AF, Young JS, Morshed RA, Berger MS. FLAIRectomy: Resecting beyond the contrast margin for glioblastoma. Brain Sci 2022;12:544.
9. Fraum TJ, Ludwig DR, Bashir MR, Fowler KJ. Gadolinium-based contrast agents: a comprehensive risk assessment. J Magn Reson Imaging 2017;46:338–53.
10. Blomqvist L, Nordberg GF, Nurchi VM, Aaseth JO. Gadolinium in medical imaging-usefulness, toxic reactions and possible countermeasures-a review. Biomolecules 2022;12:742.
11. Blumfield E, Moore MM, Drake MK, et al. Survey of gadolinium-based contrast agent utilization among the members of the society for pediatric radiology: a quality and safety committee report. Pediatr Radiol 2017;47:665–73.
12. Proença F, Guerreiro C, Sá G, Reimão S. Neuroimaging safety during pregnancy and lactation: a review. Neuroradiology 2021;63:837–45.
13. Crowson MG, Rocke DJ, Hoang JK, Weissman JL, Kaylie DM. Cost-effectiveness analysis of a non-contrast screening MRI protocol for vestibular schwannoma in patients with asymmetric sensorineural hearing loss. Neuroradiology 2017;59:727–36.
14. Raizer JJ, Fitzner KA, Jacobs DI, et al. Economics of malignant gliomas: a critical review. J Oncol Pract 2015;11:e59–65.
15. Ogbole GI, Adeyomoye AO, Badu-Peprah A, Mensah Y, Nzeh DA. Survey of magnetic resonance imaging availability in West Africa. Pan Afr Med J 2018;30:240.
16. Anazodo UC, Ng JJ, Ehiogu B, et al. A framework for advancing sustainable magnetic resonance imaging access in Africa. NMR Biomed 2023;36:e4846.
17. Pasquini L, Napolitano A, Pignatelli M, et al. Synthetic post-contrast imaging through artificial intelligence: clinical applications of virtual and augmented contrast media. Pharmaceutics 2022;14:2378.
18. The Cancer Genome Atlas (TCGA) Phenotype Research Group. VASARI research project. In: Cancer Imaging Archive. Available at https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project. Accessed December 19, 2023.
19. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.
20. Vanbelle S. A new interpretation of the weighted kappa coefficients. Psychometrika 2016;81:399–410.

21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

22. Vanbelle S. Comparing dependent kappa coefficients obtained on multilevel data. Biom J 2017;59:1016–34.

23. He H, Guo E, Meng W, et al. [Predicting cerebral glioma enhancement pattern using a machine learning-based magnetic resonance imaging radiomics model]. Nan Fang Yi Ke Da Xue Xue Bao 2024;44:194–200.

24. Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Cha S. Feasibility of simulated postcontrast MRI of glioblastomas and lower-grade gliomas by using three-dimensional fully convolutional neural networks. Radiol Artif Intell 2021;3:e200276.

25. Kleesiek J, Morshuis JN, Isensee F, et al. Can virtual contrast enhancement in brain MRI replace gadolinium?: a feasibility study. Invest Radiol 2019;54:653–60.

26. Wang Y, Wu W, Yang Y, et al. Deep learning-based 3D MRI contrast-enhanced synthesis from a 2D noncontrast T2Flair sequence. Med Phys 2022;49:4478–93.

27. Parillo M, Mallio CA, Pileri M, et al. Interrater reliability of brain tumor reporting and data system (BT-RADS) in the follow up of adult primary brain tumors: a single institution experience in Italy. Quant Imaging Med Surg 2023;13:7423–31.

28. Nam YK, Park JE, Park SY, et al. Reproducible imaging-based prediction of molecular subtype and risk stratification of gliomas across different experience levels using a structured reporting system. Eur Radiol 2021;31:7374–85.

29. Bellini D, Panvini N, Rengo M, et al. Diagnostic accuracy and interobserver variability of CO-RADS in patients with suspected coronavirus disease-2019: a multireader validation study. Eur Radiol 2021;31:1932–40.

30. Park M, Lee S-K, Chang JH, et al. Elderly patients with newly diagnosed glioblastoma: can preoperative imaging descriptors improve the predictive power of a survival model? J Neurooncol 2017;134:423–31.

31. Su C-Q, Lu S-S, Han Q-Y, Zhou M-D, Hong X-N. Intergrating conventional MRI, texture analysis of dynamic contrast-enhanced MRI, and susceptibility weighted imaging for glioma grading. Acta radiol 2019;60:777–87.

32. Zhou H, Vallières M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 2017;19:862–70.

33. Park YW, Han K, Ahn SS, et al. Prediction of IDH1-mutation and 1p/19q-codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. AJNR Am J Neuroradiol 2018;39:37–42.

34. Hyare H, Rice L, Thust S, et al. Modelling MR and clinical features in grade II/III astrocytomas to predict IDH mutation status. Eur J Radiol 2019;114:120–27.

35. Park CJ, Han K, Shin H, et al. MR image phenotypes may add prognostic value to clinical features in IDH wild-type lower-grade gliomas. Eur Radiol 2020;30:3035–45.

Chapter 5

Preoperative Prediction of Diffuse Glioma Type and Grade in Adults: a Gadolinium-Free MRI-Based Decision Tree

Azizova, A., *Prysiazhniuk, Y., Wamelink, I.J.H.G., Cakmak, M., Kaya, E., Wesseling, P., de Witt Hamer, P.C., Verburg, N., Petr, J., Barkhof, F., Keil, V.C.*

Published: ***European Radiology***
2025; vol. 35,3:1242-1254; doi: 10.1007/s00330-024-11140-5

ABSTRACT

Objectives: To develop a gadolinium-free MRI-based diagnosis prediction decision tree (DPDT) for adult-type diffuse gliomas and to assess the added value of a gadolinium-based contrast agent (GBCA) enhanced images.

Materials and Methods: This study included preoperative grade 2-4 adult-type diffuse gliomas (World Health Organization 2021) scanned from January 2010 to 2021. The DPDT, incorporating eleven GBCA-free MRI features, was developed using 18% of the dataset based on consensus readings. Diagnosis predictions involved grade (grade 2 vs. grade 3/4) and molecular status (IDH and 1p/19q). GBCA-free diagnosis was predicted using DPDT, while GBCA-enhanced diagnosis included post-contrast images. The accuracy of these predictions was assessed by three raters with varying experience levels in neuroradiology using the test dataset. Agreement analyses were applied to evaluate the prediction performance/reproducibility.

Results: The test dataset included 303 patients (age (SD): 56.7 (14.2) years, female/male: 114/189, low-grade/high-grade: 54/249, IDH-mutant/wildtype: 82/221, 1p/19q-codeleted/intact: 34/269). Per-rater GBCA-free predictions achieved ≥0.85 (95%-CI: 0.80-0.88) accuracy for grade and ≥0.75 (95%-CI: 0.70-0.80) for molecular status, while GBCA-enhanced predictions reached ≥0.87 (95%-CI: 0.82-0.90) and ≥0.77 (95%-CI: 0.71-0.81), respectively. No accuracy difference was observed between GBCA-free and GBCA-enhanced predictions. Group inter-rater agreement was moderate for GBCA-free (0.56 (95%-CI: 0.46-0.66)) and substantial for GBCA-enhanced grade prediction (0.68 (95%-CI: 0.58-0.78), p = 0.008), while substantial for both GBCA-free (0.75 (95%-CI: 0.69-0.80) and GBCA-enhanced (0.77 (95%-CI: 0.71-0.82), p = 0.51) molecular status prediction.

Conclusion: The proposed GBCA-free diagnosis prediction decision tree performed well, with GBCA-enhanced images adding little value to the preoperative diagnostic accuracy of adult-type diffuse gliomas.

KEY POINTS

Question Given health and environmental concerns, is there a gadolinium-free imaging protocol to preoperatively evaluate gliomas comparable to the gadolinium-enhanced standard practice?

Findings The proposed gadolinium-free diagnosis prediction decision tree for adult-type diffuse gliomas performed well, and gadolinium-enhanced MRI demonstrated only limited improvement in diagnostic accuracy.

Clinical relevance Even inexperienced raters effectively classified adult-type diffuse gliomas using the gadolinium-free diagnosis prediction decision tree, which, until further validation, can be used alongside gadolinium-enhanced images to respect standard practice, despite this study showing that gadolinium-enhanced images hardly improved diagnostic accuracy.

INTRODUCTION

Gadolinium-based contrast agent (GBCA)-enhanced MRI is the current standard imaging modality for managing brain tumors, including adult-type diffuse gliomas, aiding diagnosis and treatment decisions [1]. Nonetheless, enhancement is an imperfect measure for both tumor malignancy and resectability of tumor borders [2]. Tumors displaying enhancement may not always be high-grade gliomas [3]; conversely, high-grade gliomas may lack enhancement [4]. This conflict is acknowledged in the latest Response Assessment in Neuro-oncology criteria (RANO 2.0), which also stresses the diagnostic relevance of GBCA-free sequences [5].

While being a standard imaging practice [1], GBCA increasingly raises concerns about associated side effects, with safety recommendations relying solely on expert opinion rather than prospective experimental evidence [6]. Although certain linear GBCAs were restricted due to their link with nephrogenic systemic fibrosis, renal impairment remains the primary catalyst for this condition, with uncertainty about whether normal renal function excludes the risk [6]. With uncertain clinical implications, GBCA, mainly in linear forms, was also identified to accumulate in the body [7]. Furthermore, studies indicate anthropogenic medical gadolinium accumulation in ecosystems, raising concerns about aquatic life and urban water safety [8]. Beyond these challenges, longer examination times, increased financial costs [9], and limited availability in low-middle-income countries [10] are stimuli to the shift from GBCA-enhanced MRI to GBCA-free MRI. Additionally, vulnerable populations, such as pregnant or breastfeeding women [11] and children [12], necessitate careful consideration due to putative GBCA exposure risks.

Various artificial intelligence (AI) methods hold the potential for substituting GBCA with synthetic GBCA-enhanced images [13] or reducing their dosage through augmented GBCA-enhanced images [14]. Yet, their integration into clinical practice is lagging. While advanced imaging techniques like arterial spin labeling (ASL) [15] or amide proton transfer chemical exchange saturation transfer (APT-CEST) [16] introduce alternative GBCA-free parameters, their utilization is constrained by availability and variability in acquisition parameters. Conversely, conventional MRI sequences are a component of daily practice and

provide essential glioma imaging biomarkers, such as T2-FLAIR mismatch sign or cysts, many of which can be assessed without GBCA-enhanced images [17]. However, previous studies predominantly assessed these biomarkers, such as necrosis [18–20], with GBCA-enhanced MRI, as it has been the standard of care, leaving open questions about the predictive added value of GBCA-enhanced images. Further maturation of AI-based and advanced MRI methods and their clinical translation into glioma management will be a long process. Therefore, qualitative parameter evaluation of conventional GBCA-free MRI, combined with a simple decision tree, might be the near-future solution for phasing out GBCA use in glioma, as it is more time-efficient than quantitative approaches.

As a first step to develop and establish a general GBCA-free MRI-based diagnosis prediction decision tree for brain tumors, this study aims to assess the additive value of GBCA-enhanced imaging in predicting histomolecular diagnosis in adult-type diffuse gliomas.

MATERIALS AND METHODS

Study Sample

This retrospective single-center study received approval from the institutional medical ethics review board (Vumc_2021-0437). Informed consent was waived. Eligible patient cases from the hospital glioma database (IMAGO) registered from January 2010 to January 2021 were consecutively added to the trial database. The eligibility criteria are listed in Table 1.

Table 1. Eligibility criteria

Inclusion criteria	Exclusion criteria
(a) patients with grade 2-4 adult-type diffuse gliomas based on the 5th WHO-CNS tumor classification	(a) pediatric patients
(b) presence of IDH mutation and 1p/19q-codeletion status	(b) missing/incomplete histopathological diagnosis
(c) no more than one month gap between preoperative MRI and surgery	(c) suprasellar, midline, and cerebellar tumors as adult-type diffuse gliomas are rare and may have distinct radiological features in these locations

Table 1. Eligibility criteria *(Continued)*

Inclusion criteria	Exclusion criteria
(d) availability of the following mandatory MRI sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and post-contrast T1-weighted	(d) missing MRI sequences
	(e) MRI scans with suboptimal quality, including movement-related artifacts
	(f) patients who declined permission for their data to be used in research during their original stay at the institution (scientific use opt-out)
	(g) data preparation errors during the randomization of gadolinium-free and gadolinium-enhanced images of the same patients into separate evaluation sessions

Table 1 describes the inclusion and exclusion criteria for this study.
DWI/ADC = diffusion-weighted imaging/apparent diffusion coefficient, FLAIR = fluid-attenuated inversion recovery, IDH = isocitrate dehydrogenase, WHO-CNS = world health organization-central nervous system

MRI and datasets

Seven MRI scanners provided the images used in this study (Table S1). I.W., a fourth-year Ph.D. candidate in neuro-oncology, conducted data preparation, including pseudonymization. Eligible patients were randomly assigned into three subsets: development (n=38), optimization (n=31), and test (n=303). RADIANT software facilitated access to pseudonymized datasets (3.4.1.13367; https://www.radiantviewer.com/). Two raters (V.K., 11 years of neuroradiology experience; A.A., 5 years of neuroradiology experience) explored development and optimization datasets. The test dataset was independently assessed, blinded to the reference standard, by three raters (V.K., A.A., and M.C.). Rater 3 (M.C.) was a fourth-year medical student without prior radiology experience who underwent training using the optimization dataset.

Reference standard

Histomolecular diagnosis, based on the 2021 World Health Organization classification, served as the reference standard. Isocitrate dehydrogenase (IDH) status was determined via immunohistochemistry, next-generation sequencing, and/or methylation profiling, and 1p/19q-codeletion status was assessed using loss

of heterozygosity (LOH) analysis or methylation profiling. The final histomolecular diagnosis of glioblastoma in IDH-wildtype cases was determined based on additional molecular markers (e.g., TERT promoter mutation, EGFR amplification, and combination of chromosome 7 gain and chromosome 10 loss) and supporting histological features (e.g., necrosis, microvascular proliferation, and high mitotic index). A small subset of IDH-wildtype diffuse gliomas (n=16) that lacked molecular analysis (not otherwise specified) or had negative molecular markers (not elsewhere classified) were included in the study as glioblastoma, IDH-wildtype based on their final multidisciplinary team diagnosis indicating aggressive clinical behavior. Grade 2 gliomas were categorized as low-grade (LGG), while grade 3/4 as high-grade (HGG). IDH-wildtype diffuse gliomas were accepted as HGG regardless of their histological grade because of their generally aggressive clinical behavior.

Diagnosis Prediction Decision Tree (DPDT)

Two raters assessed a development dataset initially comprising only GBCA-free MRI scans (pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and SWI (if present)) to predict histomolecular diagnoses: (1) glioma grade (LGG vs. HGG) and (2) molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype). One week later, raters reexamined all cases, integrating the post-contrast T1-weighted sequence to evaluate the added value of the GBCA-enhanced scans using common clinical radiology practice. For instance, tumors with avid enhancement or rim enhancement were assigned to HGG and glioblastoma, IDH-wildtype groups, respectively. In both rating rounds, they justified their decisions on a case-by-case basis by identifying key imaging features, drawing from individual clinical experience, and utilizing literature-based biomarkers [18–25], as well as the Visually AcceSAble Rembrandt Images (VASARI), glioma imaging features set [26]. Subsequently, the raters collaboratively analyzed the results to identify helpful biomarkers for GBCA-free diagnosis prediction correlated with the reference standard. Following this consensus, a DPDT comprising seven VASARI (necrosis, diffusion, hemorrhage, non-enhancing tumor margin, calvarial remodeling, cysts, proportion of edema) and four non-VASARI (T2-FLAIR mismatch sign, T2 signal homogeneity, calcification, midline shift) imaging features, each previously linked to the respective histomolecular diagnosis [18–25],

was proposed; see Fig. 1, Table S2 and Supplementary material. Fig. S1-S5 depict case examples for DPDT imaging features.

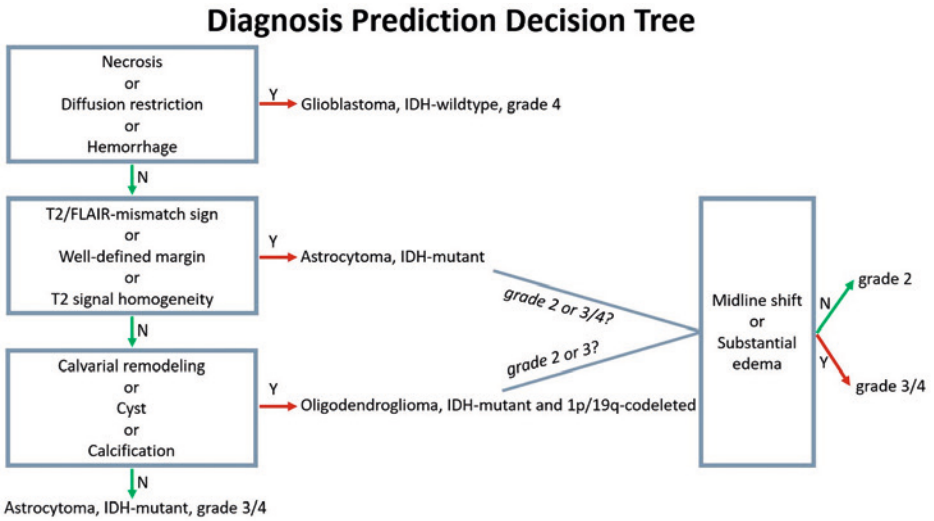


Figure 1: Diagnosis prediction decision tree (DPDT) based on GBCA-free MRI sequences. Flow chart describes DPDT for adult-type diffuse gliomas encompassing seven VASARI (necrosis, diffusion, hemorrhage, non-enhancing tumor margin, calvarial remodeling, cysts, proportion of edema) and four non-VASARI (T2-FLAIR mismatch sign, T2 signal homogeneity, calcification, midline shift) imaging features guiding the histomolecular diagnosis decision. *GBCA = Gadolinium-based contrast agent,* VASARI = Visually AcceSAble Rembrandt Images.

The optimization dataset was assessed, both with and without GBCA-enhanced images, at a one-week interval to gauge the effectiveness of DPDT using only the exclusive imaging features included in this tree, its impact on interrater agreement, and its potential applicability in a larger cohort. The GBCA-free diagnosis was evaluated using DPDT based on GBCA-free MRI sequences, while the GBCA-enhanced diagnosis included the post-contrast T1-weighted images alongside the DPDT.

Test dataset

Three raters (V.K., A.A., M.C.) assessed the diagnoses using the GBCA-free *and* GBCA-enhanced DPDT in a larger cohort to compare the predictive diagnostic accuracy using GBCA-free versus GBCA-enhanced scans. A DPDT guide including definitions of imaging features, was provided to raters (Supplementary

material). MRI scans were randomly distributed across two rating sessions. The first session assessed GBCA-free and GBCA-enhanced scans from different patients. In the second session, scans from the same patients, which had not yet been rated as GBCA-enhanced or GBCA-free, were evaluated in a differently randomized order. This approach aimed to mitigate confirmation bias by ensuring that GBCA-free or GBCA-enhanced scans were not exclusively assessed in the same session.

Statistical Analysis

Rater prediction performance was evaluated using overall accuracy for multiple classes (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype), along with accuracy, sensitivity, specificity, and negative and positive predictive values for binary classes (e.g., HGG vs. LGG or astrocytoma, IDH-mutant vs. others). The performance between GBCA-free and GBCA-enhanced datasets was compared using McNemar's test with Yates continuity correction [27].

The prediction performance was also assessed across different subgroups, including age, sex, tumor location, and tumor laterality (right/left/midline), to identify factors that might influence diagnostic accuracy. Logistic regression and Pearson's Chi-squared test were used for continuous and categorical variables, respectively. This subgroup analysis was conducted on a combined dataset of all raters, with separate evaluations for GBCA-free and GBCA-enhanced scans.

Inter-rater agreement was analyzed both collectively and pairwise among the raters. Kendall's W and Fleiss' kappa were used for group inter-rater agreement, and weighted and unweighted Cohen's kappa were used for pairwise inter-rater/intra-rater inter-group agreements in ordered and unordered features, respectively. Intra-rater inter-group agreements examined the consistency within raters by comparing the GBCA-free and GBCA-enhanced predictions. Unweighted Cohen's kappa was supplemented with prevalence-adjusted and bias-adjusted kappa (PABAK) for binary features and between two groups to compensate for a possible influence of dataset diagnosis imbalances due to naturally different tumor incidence rates [28].

The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect [29]. The comparison of agreements was conducted using the Z-test or Hotelling's T2 test [30] with the "multiagree" R package.

Data analysis was conducted by Y.P., a third-year Ph.D. student in neuroscience, using R package 4.3.0. Bootstrapping and PABAK calculations were performed using "multiagree" R and epiR (2.0.68 https://CRAN.R-project.org/package=epiR) packages, respectively. The significance threshold was $p < 0.05$.

RESULTS

Figure 2 illustrates the patient cohort. Table 2 lists the demographics of the study cohort.

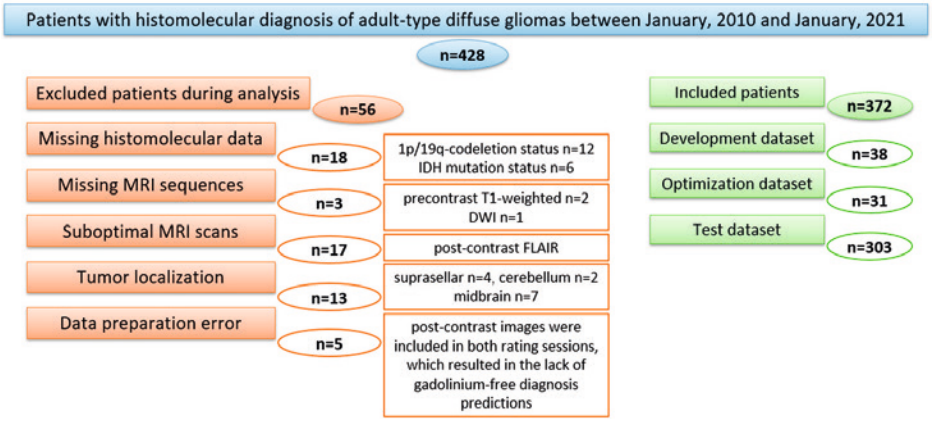


Figure 2: Patient enrollment diagram. The flow chart depicts the patients included and excluded in this study. IDH = isocitrate dehydrogenase.

Table 2. Patient demographics

Demographics	Datasets		
	Development	Optimization	Test
Sample size	n=38	n=31	n=303
Age (SD) (years)	59 (15.9)	52 (14.9)	56.7 (14.2)
Sex (female/male)	16/22	14/17	114/189
Histological grade (LGG/HGG)	5/33	9/22	54/249
IDH mutation status (IDH-mutant/IDH-wildtype)	6/32	16/15	82/221
1p/19q-codeletion status (codeleted/intact)	1/37	7/24	34/269

Table 2 describes the main demographics for the development, optimization, and test datasets.
IDH = isocitrate dehydrogenase, HGG = high-grade glioma (grade 3/4), LGG = low-grade glioma (grade 2), SD = standard deviation

Development and optimization datasets

The raters' GBCA-free prediction performance improved in the optimization dataset compared to the development dataset. For instance, the overall accuracy of GBCA-free molecular diagnosis prediction for raters 1 and 2 increased from 0.76 and 0.74 in the development dataset to 0.81 and 0.84 in the optimization dataset, respectively (Tables S3/S4 and Fig. S6/7). The comparison of GBCA-free and GBCA-enhanced histomolecular predictions per rater revealed no significant differences in both development and optimization datasets (all $p > 0.05$).

Following the implementation of DPDT, there was an improvement in inter-rater agreement in GBCA-free molecular diagnosis prediction in the development (0.45 (95%-CI: 0.18-0.71)) and optimization (0.78 (95%-CI: 0.57-0.98)) datasets, showing a trend towards significance ($p = 0.06$). Slight improvements were observed in the GBCA-free grade prediction (0.80 (95%-CI: 0.54-1.06) vs. 0.85 (95%-CI: 0.64-1.05), $p = 0.81$), as well as GBCA-enhanced molecular diagnosis (0.51 (95%-CI: 0.23-0.79) vs. 0.64 (95%-CI: 0.42-0.87), $p = 0.50$) and GBCA-enhanced grade (0.61 (95%-CI: 0.28-0.94) vs. 0.74 (95%-CI: 0.46-1.01), $p = 0.59$) prediction.

Test dataset

Predictive performance of the raters using GBCA-free vs. GBCA-enhanced scans

The accuracy in predicting tumor grade (LGG vs. HGG) using GBCA-free scans was 0.85 (95%-CI: 0.80-0.88), 0.88 (95%-CI: 0.84-0.92), and 0.86 (95%-CI: 0.82-0.90) for raters 1, 2, and 3, respectively. The corresponding accuracies for GBCA-enhanced scans were 0.88 (95%-CI: 0.84-0.91), 0.87 (95%-CI: 0.82-0.90), and 0.87 (95%-CI: 0.83-0.90). Regarding using GBCA-free scans for predicting the molecular status, the overall accuracies were 0.77 (95%-CI: 0.72-0.82), 0.76 (95%-CI: 0.71-0.81), and 0.75 (95%-CI: 0.70-0.80) for raters 1, 2, and 3, respectively. The corresponding overall accuracies for GBCA-enhanced scans were 0.77 (95%-CI: 0.71-0.81), 0.77 (95%-CI: 0.72-0.82), and 0.78 (95%-CI: 0.72-0.82). Comparing the GBCA-free and GBCA-enhanced outcomes revealed insignificant differences (all $p > 0.05$) except for rater 1's sensitivity in grade prediction (GBCA-free/GBCA-enhanced 0.90/0.95, $p = 0.006$); see Table 3, Fig. 3, and Fig. S8.

Table 3. Prediction performance of the raters in the test dataset

Results	Rater 1	Rater 2	Rater 3
Histopathological grade with GBCA-free MRI			
Accuracy	0.85	0.88	0.86
Sensitivity	0.90	0.95	0.95
Specificity	0.59	0.56	0.46
Positive predictive value	0.91	0.91	0.89
Negative predictive value	0.56	0.71	0.66
Histopathological grade with GBCA-enhanced MRI			
Accuracy	0.88	0.87	0.87
Sensitivity	0.95	0.94	0.96
Specificity	0.56	0.54	0.46
Positive predictive value	0.91	0.90	0.89
Negative predictive value	0.70	0.66	0.69

Table 3. Prediction performance of the raters in the test dataset *(Continued)*

Results		Rater 1	Rater 2	Rater 3
Molecular diagnosis with GBCA-free MRI				
Overall accuracy *(astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype)*		0.77	0.76	0.75
Accuracy	Astrocytoma, IDH-mutant	0.81	0.80	0.78
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.89	0.88	0.89
	Glioblastoma, IDH-wildtype	0.85	0.85	0.83
Sensitivity	Astrocytoma, IDH-mutant	0.54	0.58	0.63
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.42	0.35	0.29
	Glioblastoma, IDH-wildtype	0.88	0.86	0.85
Specificity	Astrocytoma, IDH-mutant	0.85	0.84	0.80
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.95	0.94	0.97
	Glioblastoma, IDH-wildtype	0.78	0.80	0.80
Positive predictive value	Astrocytoma, IDH-mutant	0.41	0.40	0.37
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.50	0.44	0.53
	Glioblastoma, IDH-wildtype	0.92	0.92	0.92
Negative predictive value	Astrocytoma, IDH-mutant	0.91	0.91	0.92
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.93	0.92	0.92
	Glioblastoma, IDH-wildtype	0.70	0.68	0.61
Molecular diagnosis with GBCA-enhanced MRI				
Overall accuracy *(astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype)*		0.77	0.77	0.78
Accuracy	Astrocytoma, IDH-mutant	0.80	0.80	0.79
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.89	0.88	0.89
	Glioblastoma, IDH-wildtype	0.84	0.85	0.86

Table 3. Prediction performance of the raters in the test dataset *(Continued)*

Results		Rater 1	Rater 2	Rater 3
Sensitivity	Astrocytoma, IDH-mutant	0.50	0.63	0.67
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.35	0.38	0.26
	Glioblastoma, IDH-wildtype	0.89	0.86	0.88
Specificity	Astrocytoma, IDH-mutant	0.86	0.84	0.82
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.96	0.95	0.97
	Glioblastoma, IDH-wildtype	0.72	0.83	0.83
Positive predictive value	Astrocytoma, IDH-mutant	0.40	0.42	0.41
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.50	0.48	0.56
	Glioblastoma, IDH-wildtype	0.90	0.93	0.93
Negative predictive value	Astrocytoma, IDH-mutant	0.90	0.92	0.93
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.92	0.92	0.91
	Glioblastoma, IDH-wildtype	0.70	0.69	0.72

Table 3 describes the diagnostic prediction performance, with and without GBCA-enhanced scans, per rater in the test dataset (n=303).
Evaluations with GBCA-free MRI were based on pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and SWI (if present) sequences using the Diagnosis Prediction Decision Tree (DPDT). Evaluations with GBCA-enhanced MRI included post-contrast T1-weighted images in addition to the assessment conducted with GBCA-free MRI.
IDH = isocitrate dehydrogenase, GBCA = gadolinium-based contrast agent

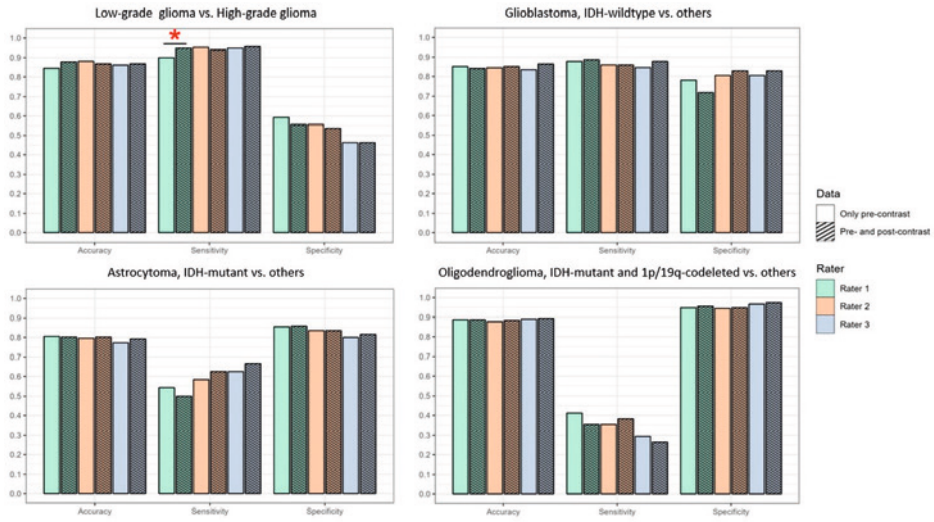


Figure 3: Per-rater prediction performance of histomolecular diagnosis of adult-type diffuse gliomas using GBCA-free vs. GBCA-enhanced scans. Color bar charts show the prediction performance, including accuracy, sensitivity, and specificity levels in predicting tumor grade, IDH mutation, and 1p/19q-codeletion status using GBCA-free and GBCA-enhanced scans per rater (rater 1 = green bar, rater 2 = orange bar, rater 3 = blue bar). Comparison of GBCA-free and GBCA-enhanced predictions revealed no significant differences ($p > 0.05$) except for rater 1's sensitivity in histopathological grade prediction (red star, $p = 0.006$). GBCA = Gadolinium-based contrast agent.

Predictive performance in different subgroups using GBCA-free and GBCA-enhanced scans

Subgroup analysis showed a significant correlation between the prediction of tumor grade or molecular status and patient age. The diagnostic accuracy improved with increasing patient age, regardless of whether GBCA-free or GBCA-enhanced scans were used (all $p < 0.001$). Similarly, a significant correlation was observed between GBCA-free or GBCA-enhanced diagnosis predictions and tumor location (all $p < 0.001$). Tumors in the thalamus were more frequently misclassified than those in other locations, particularly the parietal and temporal lobes; see Table S5. Predictions did not vary between patient sex or tumor laterality subgroups (all $p > 0.05$).

Agreement analysis for histomolecular diagnosis

Group inter-rater agreement for tumor grade prediction was higher using GBCA-enhanced scans (0.68 (95%-CI: 0.58-0.78)) than using GBCA-free scans (0.56 (95%-CI: 0.46-0.66,) p = 0.008). Outcomes for molecular status prediction were substantial for both GBCA-free (0.75 (95%-CI: 0.69-0.80)) and GBCA-enhanced scans (0.77 (95%-CI: 0.71-0.82), p = 0.51); see Fig. 4.

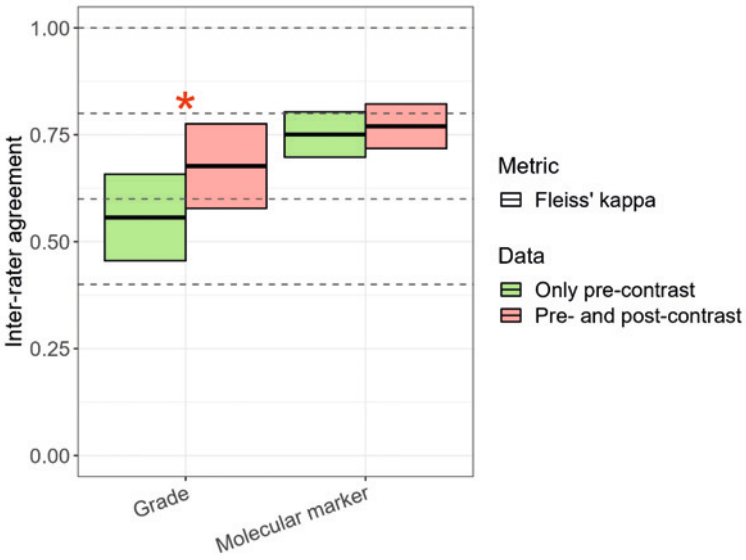


Figure 4: Group inter-rater agreement in histomolecular diagnosis prediction of adult-type diffuse gliomas using GBCA-free vs. GBCA-enhanced scans. The color box plot shows inter-rater agreement in predicting tumor grade (low-grade: grade 2 vs. high-grade: grade 3/4) and molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype) among all raters. Green bars depict the results based on the evaluation of GBCA-free scans (only pre-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and SWI (if present)) and red bars show the results of the evaluation using GBCA-enhanced scans (pre- and post-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted). Comparison of agreements between GBCA-free and GBCA-enhanced predictions was significant for tumor grade (red star, p = 0.008) while it was insignificant for molecular marker (p > 0.05). Note: The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect. GBCA = Gadolinium-based contrast agent, IDH = isocitrate dehydrogenase.

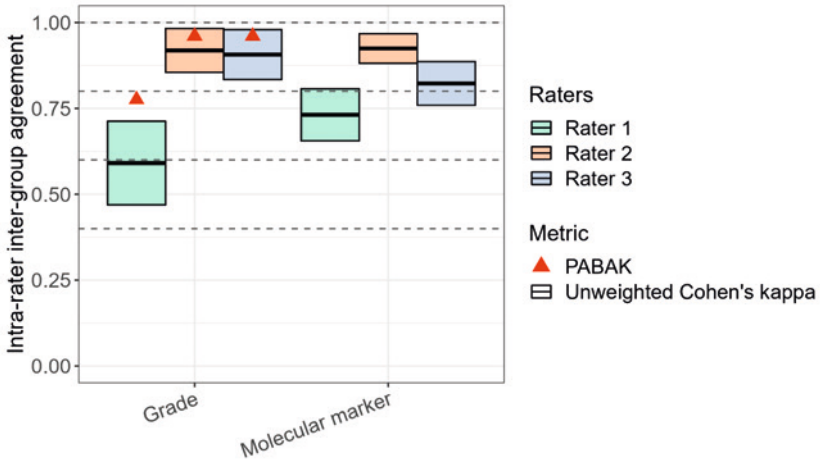


Figure 5: Intra-rater inter-group agreement in histomolecular diagnosis prediction of adult-type diffuse gliomas. The color box plot describes intra-rater inter-group agreement comparing GBCA-free and GBCA-enhanced histomolecular diagnosis predictions of each rater (rater 1 = green bar, rater 2 = orange bar, rater 3 = blue bar). Diagnosis predictions include tumor grade (low-grade: grade 2 vs. high-grade: grade 3/4) and molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype) evaluations. Red triangles indicate prevalence-adjusted and bias-adjusted kappa (PABAK) values that compensate for the possible influence of dataset diagnosis imbalances. Comparison of agreements among all raters revealed significant differences for both tumor grade and molecular marker ($p < 0.001$). Note: The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect. GBCA = Gadolinium-based contrast agent, IDH = isocitrate dehydrogenase.

Pairwise inter-rater agreements for tumor grade prediction were moderate or better for both GBCA-free (≥0.52 (95%-CI 0.38-0.65)) and GBCA-enhanced (≥0.59 (95%-CI 0.50-0.73)) scans. The results increased to substantial or almost perfect levels (≥0.74) after applying PABAK analysis, accounting for dataset imbalance. Outcomes for the molecular status prediction were substantial for both GBCA-free (≥0.73 (95%-CI 0.65-0.80)) and GBCA-enhanced (≥0.74 (95%-CI 0.66-0.81)) scans. Comparison analysis between GBCA-free and GBCA-enhanced agreements showed no significant differences (all $p > 0.05$) except for agreements in grade prediction for raters 1&2 (GBCA-free/GBCA-enhanced 0.53 (95%-CI: 0.40-0.66)/0.74 (95%-CI: 0.63-0.86), $p = 0.003$) and raters 1&3

(GBCA-free/GBCA-enhanced 0.52 (95%-CI: 0.38-0.65)/0.69 (95%-CI: 0.57-0.81), p = 0.02); see Table S6 and Fig. S9.

Intra-rater inter-group agreements (Fig. 5) for grade prediction were moderate 0.59 (95%-CI: 0.47-0.71) for rater 1 and almost perfect for rater 2 (0.92 (95%-CI: 0.85-0.98)) and rater 3 (0.91 (95%-CI: 0.83-0.98)). Applying PABAK to account for diagnosis incidence imbalance further improved the results (raters 1/2/3: 0.78/0.96/0.96). The corresponding results for molecular status prediction were substantial for rater 1 (0.73 (95%-CI: 0.66-0.81)) and almost perfect for rater 2 (0.92 (95%-CI: 0.88-0.97)) and rater 3 (0.82 (95%-CI: 0.76-0.89)). The comparison of intra-rater agreements among all raters showed significant differences for both tumor grade (p < 0.001) and molecular status prediction ($p < 0.001$).

Agreement analysis for DPDT imaging features

Group inter-rater agreements were consistent (all p > 0.05) between the evaluation of GBCA-free and GBCA-enhanced scans except for hemorrhage and midline shift, showing significant differences (p = 0.02 and p = 0.04, respectively). The robust feature with almost perfect agreement was necrosis (≥0.83 (95%-CI: 0.78-0.88)). Calcification and midline shift showed substantial agreements (≥0.61 (95%-CI: 0.31-0.90)) while other features reached fair to moderate levels (0.35 (95%-CI: 0.19-0.51) - 0.58 (95%-CI: 0.40-0.76)); see Table 4 and Fig. 6.

Table 4. Group inter-rater agreement in the evaluation of imaging features included in DPDT

Imaging features	GBCA-free scans	GBCA-enhanced scans
Necrosis*	0.83 (95%-CI: 0.78-0.88)	0.85 (95%-CI: 0.80-0.90)
Diffusion restriction**	0.52 (95%-CI: 0.47-0.56)	0.51 (95%-CI: 0.46-0.56)
Hemorrhage*^	0.48 (95%-CI: 0.39-0.57)	0.40 (95%-CI: 0.31-0.49)
T2-FLAIR mismatch sign*	0.57 (95%-CI: 0.37-0.77)	0.58 (95%-CI: 0.40-0.76)
Nonenhancing tumor margin*	0.44 (95%-CI: 0.33-0.54)	0.47 (95%-CI: 0.36-0.58)
T2 homogeneity*	0.43 (95%-CI: 0.28-0.57)	0.46 (95%-CI: 0.31-0.61)
Calvarial remodeling*	0.49 (95%-CI: 0.33-0.65)	0.56 (95%-CI: 0.39-0.73)
Cyst*	0.40 (95%-CI: 0.24-0.57)	0.35 (95%-CI: 0.19-0.51)
Calcification*	0.61 (95%-CI: 0.31-0.90)	0.63 (95%-CI: 0.31-0.95)
Midline shift*^	0.68 (95%-CI: 0.61-0.75)	0.73 (95%-CI: 0.67-0.80)

Table 4. Group inter-rater agreement in the evaluation of imaging features included in DPDT *(Continued)*

Imaging features	GBCA-free scans	GBCA-enhanced scans
Substantial edema*	0.39 (95%-CI: 0.30-0.47)	0.38 (95%-CI: 0.30-0.46)

Table 4 shows the group inter-rater agreement results for imaging features involved in DPDT using both GBCA-free and GBCA-enhanced scans.
* Fleiss` kappa, ** Kendall`s W, ^ Features with significant difference (P < .05) between GBCA-free and GBCA-enhanced agreement values
DPDT = diagnosis prediction decision tree, GBCA= gadolinium-based contrast agent, GBCA-free scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) sequences, GBCA-enhanced scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted sequences
Interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect

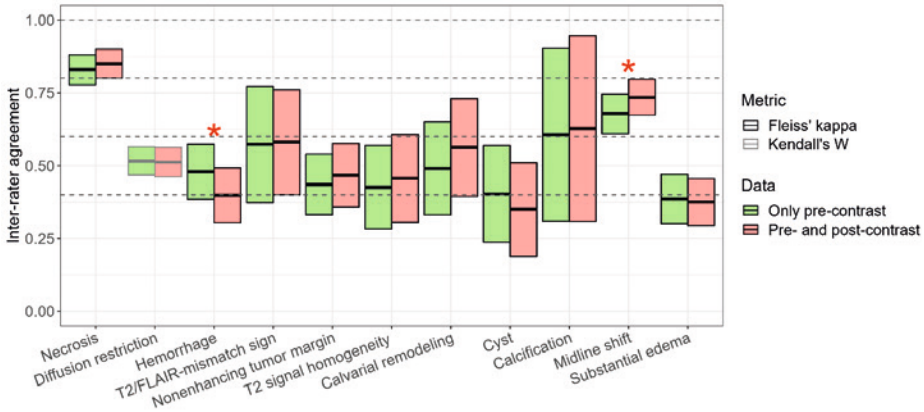


Figure 6: Group inter-rater agreement in evaluating imaging features included in the Diagnosis Prediction Decision tree (DPDT) for adult-type diffuse gliomas. The color box plot shows inter-rater agreement in evaluating single DPDT imaging features using GBCA-free or GBCA-enhanced scans among all raters. Green bars depict the results based on the assessment of GBCA-free scans (only pre-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC and SWI (if present)) and red bars show the results of the evaluation using GBCA-enhanced scans (pre- and post-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted). Comparison of agreements between GBCA-free and GBCA-enhanced assessments was insignificant (p > 0.05) except for hemorrhage and midline shift (red stars, p = 0.02 and p = 0.04, respectively). Note: The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect. GBCA = Gadolinium-based contrast agent.

Table 5. Intra-rater inter-group agreement in the evaluation of imaging features included in DPDT

Imaging features	Rater 1	Rater 2	Rater 3
Necrosis* *PABAK*	0.82 (95%-CI: 0.75-0.89) *0.84*	0.91 (95%-CI: 0.87-0.96) *0.92*	0.83 (95%-CI: 0.77-0.90) *0.85*
Diffusion restriction**	0.56 (95%-CI: 0.49-0.64)	0.75 (95%-CI: 0.67-0.82)	0.76 (95%-CI: 0.70-0.82)
Hemorrhage* *PABAK*	0.69 (95%-CI: 0.59-0.80) *0.80*	0.87 (95%-CI: 0.81-0.93) *0.89*	0.87 (95%-CI: 0.77-0.97) *0.96*
T2-FLAIR mismatch sign* *PABAK*	0.76 (95%-CI: 0.52-1.00) *0.97*	1.00 (95%-CI: 1.00-1.00) *1.00*	0.68 (95%-CI: 0.45-0.91) *0.95*
Nonenhancing tumor margin* *PABAK*	0.35 (95%-CI: 0.18-0.52) *0.78*	0.85 (95%-CI: 0.77-0.94) *0.93*	0.66 (95%-CI: 0.54-0.77) *0.80*
T2 homogeneity* *PABAK*	0.45 (95%-CI: 0.30-0.60) *0.78*	0.90 (95%-CI: 0.79-1.00) *0.97*	0.66 (95%-CI: 0.50-0.83) *0.91*
Calvarial remodeling* *PABAK*	0.64 (95%-CI: 0.44-0.84) *0.92*	0.95 (95%-CI: 0.88-1.00) *0.99*	0.72 (95%-CI: 0.51-0.93) *0.95*
Cyst* *PABAK*	0.40 (95%-CI: 0.21-0.59) *0.81*	0.74 (95%-CI: 0.55-0.93) *0.95*	0.55 (95%-CI: 0.28-0.82) *0.94*
Calcification* *PABAK*	0.75 (95%-CI: 0.48-01.00) *0.90*	1.00 (95%-CI: 1.00-1.00) *1.00*	0.87 (95%-CI: 0.61-1.00) *0.97*
Midline shift* *PABAK*	0.81 (95%-CI: 0.73-0.89) *0.84*	0.91 (95%-CI: 0.86-0.96) *0.91*	0.83 (95%-CI: 0.77-0.89) *0.83*
Substantial edema* *PABAK*	0.65 (95%-CI: 0.55-0.74) *0.70*	0.78 (95%-CI: 0.68-0.89) *0.89*	0.84 (95%-CI: 0.77-0.90) *0.83*

Table 5 shows the intra-rater inter-group agreement results for imaging features involved in DPDT, comparing the evaluation of GBCA-free scans with GBCA-enhanced scans for each rater.

*unweighted Cohen`s kappa, **weighted Cohen's kappa

DPDT = diagnosis prediction decision tree, GBCA= gadolinium-based contrast agent, GBCA-free scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) sequences, GBCA-enhanced scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted sequences

The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect

Pairwise inter-rater agreements were at least substantial for necrosis (≥0.80 (95%-CI: 0.73-0.87)) with further improvement after applying PABAK analysis (≥0.82). The outcomes for other features varied between 0.13 (95%-CI: -0.04-0.31) and 0.80 (95%-CI: 0.52-1.07). However, the results increased with PABAK analysis (range: 0.50-94), showing the impact of dataset imbalance. There were no significant differences (all $p > 0.05$) between GBCA-free and GBCA-enhanced agreements with a few exceptions (rater 1&3: cyst $p = 0.04$, midline shift $p = 0.03$, rater 1&2: midline shift $p = 0.02$); see Table S7 and Fig. S10.

Intra-rater agreements were almost perfect for necrosis (≥0.82 (95%-CI: 0.75-0.89)) and midline shift (≥0.81 (95%-CI 0.73-0.89)), substantial or better for hemorrhage, T2-FLAIR mismatch sign, calvarial remodeling, calcification and substantial edema (≥0.64 (95%-CI: 0.44-0.84)), fair or better for other features (≥0.35 (95%-CI: 0.18-0.52)). The comparison of intra-rater agreements among all raters showed significant differences (all $p < 0.05$) except for the T2-FLAIR mismatch sign ($p = 0.93$) and calcification ($p = 0.96$); see Table 5 and Fig. S11.

DISCUSSION

In this study, we developed a DPDT, incorporating eleven imaging features from conventional GBCA-free MRI for adult-type diffuse gliomas. DPDT, assessed by three raters with variable levels of experience, demonstrated high predictive performance for the classification of both tumor grade (accuracy ≥0.85 (95%-CI: 0.80-0.88)) and molecular status (overall accuracy ≥0.75 (95%-CI: 0.70-0.80). Adding GBCA-enhanced images to the evaluation showed comparable results (accuracy ≥ 0.87 (95%-CI: 0.82-0.90) and overall accuracy ≥0.77 (95%-CI: 0.71-0.81), respectively). Comparison of GBCA-free and GBCA-enhanced outcomes (accuracy, sensitivity, and specificity) revealed insignificant differences except for rater 1's sensitivity in grade prediction (GBCA-free/GBCA-enhanced = 0.90/0.95, $p = 0.006$).

Our study suggests that the proposed DPDT using GBCA-free MRI could be as reliable as standard GBCA-enhanced MRI in preoperative diagnostic glioma assessment. Previous studies often evaluated the diagnostic efficacy of conventional MRI, including GBCA-enhanced images, making it challenging to

determine the complementary role of GBCA. For instance, Du et al. [31] and Setyawan et al. [20] proposed preoperative glioma grading (AUC: 0.93 and 1.00, respectively) and IDH genotyping (AUC: 0.86 and 0.93, respectively) models incorporating enhancement features alongside other GBCA-free imaging features such as hemorrhage or cysts. Although a recent study [32] proposed an MRI scoring system utilizing GBCA-free features, it specifically assessed non-enhancing gliomas. However, focusing exclusively on GBCA-free imaging features without excluding enhancing gliomas could help to better comprehend the additional benefit of GBCA in decision-making. Our study addresses this gap by assessing a large glioma cohort through a head-to-head comparison of GBCA-free and GBCA-enhanced MRI, ultimately refuting the additive value of GBCA. The evaluation across different patient subgroups, such as age or tumor location, was similar for GBCA-free and GBCA-enhanced evaluations, stressing the limited added value of GBCA-enhanced images. The subgroup analysis revealed that DPDT performs better in older age groups, likely due to the high prevalence of glioblastoma in this demographic, which often exhibits typical imaging features such as necrosis, facilitating identification. The performance was less accurate for tumors in the thalamus, possibly because thalamic lesions display more distinct imaging features than hemispheric tumors. Importantly, inter-rater agreement regarding histomolecular diagnosis was not improved by GBCA use, highlighting the potential value of GBCA-free DPDT in real-world clinical settings, while the IDH prediction accuracy of only 77% even with GBCA indicates limitations for radiology in general.

Our DPDT algorithm comprises eleven imaging features, each correlated with the respective histomolecular glioma diagnosis in previous studies [18–25]. Among these, necrosis, a glioblastoma biomarker in DPDT, was the robust feature, demonstrating at least substantial inter-/intra-rater agreement (≥0.80 (95%-CI: 0.73-0.87)). A recent study [33] investigating the reliability of imaging-based necrosis found a strong agreement between this and pathological necrosis (0.77 (95%-CI: 0.64-0.90)). That study observed a significant correlation between imaging-based necrosis and tumor grade as well as IDH status (p < 0.001), alongside substantial inter-rater agreement ((0.67 (95%-CI: 0.49-0.85)), comparable to our study. However, their necrosis assessment relied on GBCA-enhanced MRI, similar to most previous studies [18–20]. Conversely, our evaluation of necrosis utilized both GBCA-free and GBCA-enhanced MRI, re-

vealing almost perfect intra-rater agreements (≥0.82 (95%-CI: 0.75-0.89)), thus underscoring the efficacy of GBCA-free MRI in this context. Among other DPDT features, varying levels of inter- and intra-rater agreement (range: 0.13 (95%-CI: -0.04-0.31)- 1.00 (95%-CI: 1.00-1.00)) were observed. T2-FLAIR mismatch sign (≥0.51 (95%-CI: 0.23-0.80)) and calcification (≥0.49 (95%-CI: 0.08-0.89)) emerged as the most consistent imaging features associated with astrocytoma and oligodendroglioma, respectively, in DPDT. Group inter-rater agreements for hemorrhage and midline shift differed significantly between GBCA-free and GBCA-enhanced evaluations. The discrepancy in hemorrhage assessment may stem from inconsistent SWI availability and varying degrees of pre-contrast T1 hyperintensity, rather than from the availability of GBCA-enhanced images. Similarly, for the midline shift, the disagreement could result from the prevalence of cases with a minimal shift around the 5 mm threshold, leading to measurement variations between evaluations and raters. Utilizing the decision tree enables a systematic approach to imaging assessment, potentially improving diagnostic thoroughness, as less experienced raters demonstrated comparable or better performance to experienced radiologists. Beyond assessment by radiologists, the present study may deliver insights for researchers focusing on AI-based algorithms, as DPDT provides imaging characteristics relevant to algorithm decision-making.

This study has several limitations demanding future research. The study focuses on a specific scenario, and a relatively small sample size rated by only two observers was used to develop DPDT. Besides the scientific outcomes, the study's clinical value is potentially limited to situations where tissue diagnosis is not feasible, e.g., due to poor clinical condition or tumor location or when GBCA administration is contraindicated or not desired. Moreover, the omission of clinical factors such as age and the lack of a longitudinal evaluation in DPDT will have impacted diagnostic predictions. Additionally, despite IDH-wildtype diffuse gliomas being classified as HGG in this study regardless of histological grade, recent studies suggest that histologically grade 2 IDH-wildtype diffuse gliomas with isolated telomerase reverse transcriptase (TERT) promoter mutation may exhibit a more favorable outcome than glioblastoma, IDH-wildtype [34]. Further work is needed to incorporate clinical factors and other imaging biomarkers, including temporal imaging evaluation, and extend DPDT to other brain tumor differentials, such as metastasis or lymphoma, and non-tumor le-

sions, such as demyelination or infection. Another significant limitation is the exclusion of perfusion data due to the inconsistent availability of ASL data for comparison with the standard method of dynamic susceptibility contrast (DSC)-MRI. Provided a head-to-head comparison can be guaranteed, perfusion should be examined in a future DPDT study. Moreover, an extended consecutive study should incorporate time-tracking of the DPDT usage to evaluate its clinical utility compared to standard evaluation.

In conclusion, the proposed decision tree enables non-invasive preoperative diagnosis of adult-type diffuse gliomas using only GBCA-free MRI, regardless of the rater's experience level. Future research should develop a generalized decision tree with diverse brain mass lesions and advanced imaging techniques and test it with additional raters and new data.

REFERENCES

1. Ellingson BM, Bendszus M, Boxerman J et al (2015) Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol 17:1188–1198
2. Haddad AF, Young JS, Morshed RA, Berger MS (2022) FLAIRectomy: Resecting beyond the Contrast Margin for Glioblastoma. Brain Sci 12:544
3. Castet F, Alanya E, Vidal N et al (2019) Contrast-enhancement in supratentorial low-grade gliomas: a classic prognostic factor in the molecular age. J Neurooncol 143:515–523
4. Eichberg DG, Di L, Morell AA et al (2020) Incidence of high grade gliomas presenting as radiographically non-enhancing lesions: experience in 111 surgically treated non-enhancing gliomas with tissue diagnosis. J Neurooncol 147:671–679
5. Wen PY, van den Bent M, Youssef G et al (2023) RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults. J Clin Oncol 41:5187–5199
6. Leyba K, Wagner B (2019) Gadolinium-based contrast agents: why nephrologists need to be concerned. Curr Opin Nephrol Hypertens 28:154–162
7. van der Molen AJ, Quattrocchi CC, Mallio CA et al (2024) Ten years of gadolinium retention and deposition: ESMRMB-GREC looks backward and forward. Eur Radiol 34:600–611
8. Chaban YV, Vosshenrich J, McKee H et al (2024) Environmental Sustainability and MRI: Challenges, Opportunities, and a Call for Action. J Magn Reson Imaging 59:1149–1167
9. Crowson MG, Rocke DJ, Hoang JK et al (2017) Cost-effectiveness analysis of a non-contrast screening MRI protocol for vestibular schwannoma in patients with asymmetric sensorineural hearing loss. Neuroradiology 59:727–736
10. Ogbole GI, Adeyomoye AO, Badu-Peprah A et al (2018) Survey of magnetic resonance imaging availability in West Africa. Pan Afr Med J 30:240
11. Proença F, Guerreiro C, Sá G, Reimão S (2021) Neuroimaging safety during pregnancy and lactation: a review. Neuroradiology 63:837–845
12. Blumfield E, Moore MM, Drake MK et al (2017) Survey of gadolinium-based contrast agent utilization among the members of the Society for Pediatric Radiology: a Quality and Safety Committee report. Pediatr Radiol 47:665–673
13. Xie H, Lei Y, Wang T et al (2022) Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision. Med Phys 49:3278–3287
14. Pasumarthi S, Tamir JI, Christensen S et al (2021) A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain MRI. Magn Reson Med 86:1687–1700
15. Qu Y, Kong D, Wen H et al (2022) Perfusion measurement in brain gliomas using velocity-selective arterial spin labeling: comparison with pseudo-continuous arterial spin labeling and dynamic susceptibility contrast MRI. Eur Radiol 32:2976–2987
16. Togao O, Hiwatashi A, Yamashita K et al (2017) Grading diffuse gliomas without intense contrast enhancement by amide proton transfer MR imaging: comparisons with diffusion- and perfusion-weighted imaging. Eur Radiol 27:578–588
17. Lasocki A, Anjari M, Örs Kokurcan S, Thust SC (2021) Conventional MRI features of adult diffuse glioma molecular subtypes: a systematic review. Neuroradiology 63:353–362
18. Lasocki A, Buckland ME, Drummond KJ et al (2022) Conventional MRI features can predict the molecular subtype of adult grade 2-3 intracranial diffuse gliomas. Neuroradiology 64:2295–2305

19. Nam YK, Park JE, Park SY et al (2021) Reproducible imaging-based prediction of molecular subtype and risk stratification of gliomas across different experience levels using a structured reporting system. Eur Radiol 31:7374–7385
20. Setyawan NH, Choridah L, Nugroho HA et al (2024) Beyond invasive biopsies: using VASARI MRI features to predict grade and molecular parameters in gliomas. Cancer Imaging 24:3
21. Çelik S, Öven BB, Demir MK et al (2021) Magnetic resonance imaging criteria for prediction of isocitrate dehydrogenase (IDH) mutation status in patients with grade II-III astrocytoma and oligodendroglioma. Clin Neurol Neurosurg 207:106745
22. Corell A, Ferreyra Vega S, Hoefling N et al (2020) The clinical significance of the T2-FLAIR mismatch sign in grade II and III gliomas: a population-based study. BMC Cancer 20:450
23. Han Z, Chen Q, Zhang L et al (2022) Radiogenomic association between the T2-FLAIR mismatch sign and IDH mutation status in adult patients with lower-grade gliomas: an updated systematic review and meta-analysis. Eur Radiol 32:5339–5352
24. Smits M (2016) Imaging of oligodendroglioma. Br J Radiol 89:20150857
25. Johnson DR, Diehn FE, Giannini C et al (2017) Genetically Defined Oligodendroglioma Is Characterized by Indistinct Tumor Borders at MRI. AJNR Am J Neuroradiol 38:678–684
26. VASARI research project - the cancer imaging archive (TCIA) public access - cancer imaging archive wiki Available via https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project. Accessed 19 Aug 2024
27. Kim S, Lee W (2017) Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? Stat Methods Med Res 26:142–154
28. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 85:257–268
29. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
30. Vanbelle S (2017) Comparing dependent kappa coefficients obtained on multilevel data. Biom J 59:1016–1034
31. Du N, Zhou X, Mao R et al (2022) Preoperative and Noninvasive Prediction of Gliomas Histopathological Grades and IDH Molecular Types Using Multiple MRI Characteristics. Front Oncol 12:873839
32. Kang KM, Song J, Choi Y et al (2024) MRI Scoring Systems for Predicting Isocitrate Dehydrogenase Mutation and Chromosome 1p/19q Codeletion in Adult-type Diffuse Glioma Lacking Contrast Enhancement. Radiology 311:e233120
33. Ma H, Zeng S, Xie D et al (2024) Looking through the imaging perspective: the importance of imaging necrosis in glioma diagnosis and prognostic prediction - single centre experience. Radiol Oncol 58:23–32
34. Berzero G, Di Stefano AL, Ronchi S et al (2021) IDH-wildtype lower-grade diffuse gliomas: the importance of histological grade and molecular assessment for prognostic stratification. Neuro Oncol 23:955–966

ELECTRONIC SUPPLEMENTARY MATERIAL

Rationale for the selection of DPDT imaging features

The selection of specific imaging features for DPDT was primarily guided by the clinical experience of the raters, who also referenced established imaging biomarkers [1-8] commonly used in clinical practice.

Glioblastoma, IDH-wildtype:

Necrosis and hemorrhage are well-recognized imaging biomarkers of glioblastoma, IDH-wildtype. However, their diffusion properties can still identify tumors without these characteristic features. Glioblastomas, particularly those imaged before necrosis occurs, often show diffusion restriction. As a result, these three features were considered independent imaging biomarkers of glioblastoma, IDH-wildtype, in DPDT.

Astrocytoma, IDH-mutant:

Astrocytomas exhibit a range of imaging features. T2-FLAIR mismatch sign is considered the most reliable biomarker, with studies showing up to 100% specificity. Tumors displaying this sign typically have a homogeneous T2 signal and well-defined tumor margins. However, not all astrocytomas with homogeneous T2 signal or well-defined margins exhibit the T2-FLAIR mismatch sign. Thus, these three features were treated as independent biomarkers for astrocytomas. Although astrocytomas with these features are often low-grade, the presence of a midline shift or substantial edema may suggest an aggressive, high-grade tumor. High-grade astrocytomas may also display heterogeneous T2 signal and ill-defined margins. In such cases, after excluding the imaging features of oligodendrogliomas, as discussed below, a diagnosis of high-grade astrocytoma should be considered.

Oligodendroglioma, IDH-mutant, and 1p/19q-codeleted:

Oligodendrogliomas typically have a better prognosis but present with a heterogeneous appearance on MRI. Tumors with T2 heterogeneity or ill-defined borders should be evaluated for the possibility of 1p/19q codeletion. Supporting imaging features such as calvarial remodeling, cyst, or calcification are well-documented in the literature and commonly used in clinical practice. Thus, these three imaging features were considered as a oligodendroglioma

biomarkers in tumors with T2 heterogeneity and ill-defined borders in DPDT. While determining the grade of oligodendrogliomas can be challenging, the presence of a midline shift or substantial edema may suggest a higher grade.

1. Lasocki A, Buckland ME, Drummond KJ et al (2022) Conventional MRI features can predict the molecular subtype of adult grade 2-3 intracranial diffuse gliomas. Neuroradiology 64:2295–2305
2. Nam YK, Park JE, Park SY et al (2021) Reproducible imaging-based prediction of molecular subtype and risk stratification of gliomas across different experience levels using a structured reporting system. Eur Radiol 31:7374–7385
3. Setyawan NH, Choridah L, Nugroho HA et al (2024) Beyond invasive biopsies: using VASARI MRI features to predict grade and molecular parameters in gliomas. Cancer Imaging 24:3
4. Çelik S, Öven BB, Demir MK et al (2021) Magnetic resonance imaging criteria for prediction of isocitrate dehydrogenase (IDH) mutation status in patients with grade II-III astrocytoma and oligodendroglioma. Clin Neurol Neurosurg 207:106745
5. Corell A, Ferreyra Vega S, Hoefling N et al (2020) The clinical significance of the T2-FLAIR mismatch sign in grade II and III gliomas: a population-based study. BMC Cancer 20:450
6. Han Z, Chen Q, Zhang L et al (2022) Radiogenomic association between the T2-FLAIR mismatch sign and IDH mutation status in adult patients with lower-grade gliomas: an updated systematic review and meta-analysis. Eur Radiol 32:5339–5352
7. Smits M (2016) Imaging of oligodendroglioma. Br J Radiol 89:20150857
8. Johnson DR, Diehn FE, Giannini C et al (2017) Genetically Defined Oligodendroglioma Is Characterized by Indistinct Tumor Borders at MRI. AJNR Am J Neuroradiol 38:678–684

Explanation of the Diagnosis Prediction Decision Tree (DPDT)

DPDT incorporates seven VASARI (necrosis, diffusion restriction, hemorrhage, non-enhancing tumor margins, calvarial remodeling, cysts, and proportion of edema) and four non-VASARI imaging features (T2-FLAIR mismatch sign, T2 signal homogeneity, calcification, and midline shift). Initially, necrosis, diffusion restriction (visually assessed on both b-1000 and ADC maps), or hemorrhage serve as independent imaging biomarkers for glioblastoma, IDH-wildtype. In cases where these biomarkers are absent, the subsequent evaluation focuses on the T2/FLAIR mismatch sign, well-defined non-enhancing tumor margin, or T2 signal homogeneity. The presence of any of these features independently marks astrocytoma, IDH-mutant, while the absence prompts assessment of calvarial remodeling, cysts, or calcification. The presence of these latter features suggests oligodendroglioma, IDH-mutant and 1p/19q-codeleted, whereas their absence indicates high-grade astrocytoma, IDH-mutant. The midline shift or substantial edema is the chosen imaging feature for determining the

histological grade of astrocytomas or oligodendrogliomas, as these features correlate with high-grade gliomas. Refer to Figure 1 for an illustration of DPDT.

Guide material for the raters

Definition of single imaging features used in Diagnosis Prediction Decision Tree (DPDT)

- **Necrosis, modified VASARI feature 7* (yes, no); see Figure S1 (a-c)**
  - area with irregular and/or thick margins and the following internal characteristics: T1 hypointensity, T2 hyperintensity, and high ADC values like fluid
  - should not be cysts, clusters of microcysts, or enlarged perivascular space
- **Diffusion restriction, VASARI feature 17* (yes: restricted or no: dubious/facilitated); see Figure S1 (d-i)**
  - restricted- characterized by high signal intensity on TRACE/DWI and notably low ADC values compared to the normal brain parenchyma
  - dubious/facilitated- characterized by high signal intensity on TRACE/DWI with corresponding ADC values similar to the normal brain parenchyma (dubious) or high/low signal intensity on TRACE/DWI with ADC values notably higher than the typical brain parenchyma (facilitated)
  - the areas featuring reduced ADC signal intensity due to necrotic or hemorrhagic tumor components should be disregarded

- **Hemorrhage, VASARI feature 16* (yes, no); see Figure S2**
  - yes- (a) the presence of drop-out or blooming on post-processed SWI images exhibiting identical signal intensity to internal cerebral veins/superior sagittal sinus on filtered phase images or (b) any intrinsic hyperintense foci in proximity to or within the necrotic region of the tumor on pre-contrast T1-weighted images should also be recognized as hemorrhage
  - no- (a) the absence of drop-out or blooming on post-processed SWI images or the presence of drop-out or blooming with a signal intensity opposite to internal cerebral veins/superior sagittal sinus, indicative of calcification on filtered phase images, or (b) when observations are unclear on pre-contrast T1-weighted images or may potentially signify mineral presence rather than hemorrhage
- **T2-Flair mismatch sign (yes, no); see Figure S3 (a,b)**
  - yes- the presence of complete/near complete homogeneous hyperintense signal on T2-weighted images with relatively homogeneous or heterogeneous hypointense signal in most of these regions (> 90%) on T2-FLAIR, except for a hyperintense complete or incomplete peripheral rim
  - no- the presence of heterogeneous signal intensity on T2-weighted images and/or the lack of suppressed T2 signal on FLAIR, including the absence of a hyperintense peripheral rim
- **Non-enhancing tumor margin, VASARI feature 13* (well-defined, ill-defined); see Figure S3 (c,d)**
  - well-defined- tumor margins should be deemed well-defined when they can be easily followed consistently across nearly the entire tumor, encompassing more than 90% of its volume.
  - ill-defined- indistinct or blurred margins, or margins that align with white matter tracts and are challenging to distinguish from surrounding edema across most of the tumor volume.
- **T2 signal homogeneity (yes/homogeneous, no/heterogeneous); see Figure S3 (c,d)**
  - homogeneous- almost the same signal intensity throughout the tumor except for the lesion rim, vessels (dark dots or lines), cysts, perivascular spaces, and probably infiltrated but normal-appearing cortex compared to the other tumor parts.

  - heterogeneous- mainly different signal intensity, including hypointense, isointense, and/or hyperintense signal compared to normal brain cortex, throughout the tumor
- **Cyst, VASARI feature 8* (yes, no); see Figure S4 (a-c)**
  - well-defined, rounded, often eccentric regions of high T2-weighted signal and low T1-weighted signal essentially resembling the signal intensity of cerebrospinal fluid, with very thin, uniform, smooth, non-enhancing or regularly enhancing walls, possibly with thin, regular, internal septations
- **Calcification (yes, no); see Figure S4 (d-h)**
  - yes- the presence of drop-out or blooming on post-processed SWI images with a signal intensity opposite to that of internal cerebral veins/superior sagittal sinus on filtered phase images. When present, the calcified choroid plexus or pineal gland serves as an internal reference for calcification; a similar signal intensity to these structures suggests calcification rather than hemorrhage.
  - no- the absence of drop-out or blooming on post-processed SWI images or the presence of drop-out or blooming with the same signal intensity as internal cerebral veins/superior sagittal sinus, indicative of hemorrhage on filtered phase images.
- **Calvarial remodeling, VASARI feature 25* (yes, no); see Figure S5 (a,b)**
  - yes- the presence of the evident erosion or remodeling of the inner table of the skull, possibly indicative of the gradual growth of the tumor.
  - no – the lack of observable bone erosion or remodeling near the tumor or the tumor is far from the calvarial bones.
- **Midline shift (yes, no); see Figure S5 (c)**
  - yes- the presence of more than 5 mm shift of the midline structures to the left or right according to the line drawn coplanar with falx connecting the anterior and posterior superior sagittal sinus attachments
  - no- the absence of a midline shift or the presence of a midline shift measuring 5 mm or less, based on the line drawn coplanar with the falx connecting the anterior and posterior superior sagittal sinus attachments.
- **Substantial edema, modified VASARI feature 14 (yes, no); see Figure S5 (d)**
  - yes- significant edema is defined as equal to or exceeding half of the tumor volume (≥ 50%). The signal intensity of edema surrounding the tumor should surpass that of the non-enhancing tumor and be lower than that of cerebrospinal fluid. Pseudopods are indicative of edema.

- no- the absence of edema surrounding the tumor or the presence of edema constituting < 50% of the tumor volume.

** VASARI features according to VASARI Research Project https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project.*

SUPPLEMENTARY TABLES

Table S1. MRI scan parameters

Sequence parameters	Scanners						
	GE 1.5T Signa HDxt	***GE 3T DISCOVERY MR750***	***Philips 1.5T Achieva***	***Philips 3T Ingenuity***	***Siemens 1.5T Avanto***	***Siemens 3T MAGNETOM Vida***	***Toshiba 3T Titan3T***
2D T1-weighted							
plane	axial	axial	-	axial	axial	-	axial
TE; ms, min-max	9-12	7.9-9.7	-	10	7.8-17	-	8
TR; ms, min-max	520-600	600-731	-	599	500-718	-	550
FA; min-max	90	90-125	-	70	90	-	80
3D T1-weighted							
plane	-	axial	axial	axial	axial	sagittal	-
TE; ms, min-max	-	2	3.3-4.6	3	4.5-11	2.3	-
TR; ms, min-max	-	4.7	6.5-8.7	7	700-2,700	2,300	-
TI; ms, min-max	-	650	-	-	0-950	900	-
FA; min-max	-	15	8-10	12	8-120	8	-
T2-weighted							
plane	axial	axial	axial	axial	axial	axial	axial
TE; ms, min-max	98-104	82-88	100-110	85	93-104	74	90
TR; ms, min-max	4,376-4,840	4,889-6,872	3,404-5,251	2,767-3,182	2,830-5,562	4,100-6,280	5,500-5,526
FA; min-max	90	90-111	90	90	150-180	150	90

Table S1. MRI scan parameters *(Continued)*

Sequence parameters	**Scanners**						
	GE 1.5T Signa HDxt	***GE 3T DISCOVERY MR750***	***Philips 1.5T Achieva***	***Philips 3T Ingenuity***	***Siemens 1.5T Avanto***	***Siemens 3T MAGNETOM Vida***	***Toshiba 3T Titan3T***
2D FLAIR							
plane	axial	-	axial	-	axial	-	-
TE; ms, min-max	118-120	-	140	-	88-109	-	-
TR; ms, min-max	9,002-9,502	-	9,000-11,000	-	8,870-9,000	-	-
TI; ms, min-max	2,250	-	2,800	-	2,500	-	-
FA; min-max	90	-	90	-	150	-	-
3D FLAIR							
plane	sagittal	sagittal	axial	sagittal	sagittal	sagittal	axial
TE; ms, min-max	96-122	126-135	286-306	279	334	388-430	451
TR; ms, min-max	6,000-6,500	8,000-8,002	4,800	4,800	6,500	5,000-7,700	5,600
TI; ms, min-max	1,925-1,987	2,331-2,347	1,660	1,650	2,200	1,650-2,400	1,900
FA; min-max	90	90	90	90	120	120	90
DWI							
plane	axial	axial	axial	axial	axial	axial	axial
TE; ms, min-max	81-105	62-87	72-119	74-97	90-122	68	82
TR; ms, min-max	8,000-8,500	4,000-7,200	2,674-6,448	3,496-6,354	3,400-10,500	3,200	7,500
FA; min-max	90	90	90	90	90	90	90

Table S1. MRI scan parameters *(Continued)*

Sequence parameters	Scanners						
	GE 1.5T Signa HDxt	***GE 3T DISCOVERY MR750***	***Philips 1.5T Achieva***	***Philips 3T Ingenuity***	***Siemens 1.5T Avanto***	***Siemens 3T MAGNETOM Vida***	***Toshiba 3T Titan3T***
b value	1,000	1,000	1,000	1,000	1,000	1,000	1,000
SWI							
plane	-	axial	-	axial	axial	axial	-
TE; ms, min-max	-	25	-	25-30	40	25	-
TR; ms, min-max	-	31	-	18-21	49	31	-
FA; min-max	-	15	-	10	15	15	-
2D post-contrast T1-weighted							
plane	axial	axial	-	axial	axial	-	axial
TE; ms, min-max	9-12	7.9-8.4	-	10	8.7-17	-	8
TR; ms, min-max	520-600	600-650	-	599	550-718	-	550
FA; min-max	90	90-125	-	70	90	-	80
3D post-contrast T1-weighted							
plane	axial	axial	axial	axial	sagittal	axial	axial
TE; ms, min-max	3-5	2-3.2	3.3-4.6	3	2.9-4.5	2.3	2.4
TR; ms, min-max	8-13	4.6-8.3	6.7-8.7	7	1,900-2,700	2,300	5.7
TI; ms, min-max	0-450	450-650	-	-	950-1,100	900	900
FA; min-max	12	15	8-10	12	8	8	9

Table S1 describes the MRI acquisition parameters for all seven scanners.
2D/3D = 2/3-dimensional, DWI = diffusion-weighted imaging, FA = flip angle, FLAIR = fluid-attenuated inversion recovery, max = maximum value, min = minimum value, ms = millisecond, 1.5T/3T = 1.5 tesla/3 tesla, TE = time of echo, TI = time of inversion, TR = time of repetition, SWI = susceptibility-weighted imaging
The evaluation of MRI sequences while assigning ratings was done in the axial plane. When necessary, 3D sagittal scans were reformatted to the axial plane.
All values are extracted from the DICOM headers

Table S2. Imaging features involved in the Diagnosis Prediction Decision Tree

Seven VASARI features (Vf)		**Four non-VASARI features**
Presence of necrosis	modified Vf 7	T2-FLAIR mismatch sign
Diffusion restriction	Vf 17	T2 signal homogeneity
Hemorrhage	Vf 16	Calcification
Non-enhancing tumor margin	Vf 13	Midline shift
Calvarial remodeling	Vf 25	
Cyst	Vf 8	
Substantial edema	modified Vf 14	

Table S2 describes seven VASARI and four non-VASARI imaging features involved in the Diagnosis Prediction Decision Tree (DPDT) for grade 2-4 adult-type diffuse gliomas

Table S3. Prediction performance of the raters in the development dataset

Results		**Rater 1**	**Rater 2**
Histopathological grade with GBCA-free MRI			
Accuracy		0.95	0.89
Sensitivity		0.97	0.91
Specificity		0.80	0.80
Positive predictive value		0.97	0.97
Negative predictive value		0.80	0.57
Histopathological grade with GBCA-enhanced MRI			
Accuracy		0.97	0.87
Sensitivity		1.00	0.88
Specificity		0.80	0.80
Positive predictive value		0.97	0.97
Negative predictive value		1.00	0.50
Molecular diagnosis with GBCA-free MRI			
Overall accuracy *(astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype)*		0.76	0.74
Accuracy	Astrocytoma, IDH-mutant	0.76	0.82
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.95	0.89
	Glioblastoma, IDH-wildtype	0.82	0.76

Table S3. Prediction performance of the raters in the development dataset *(Continued)*

Results		**Rater 1**	**Rater 2**
Sensitivity	Astrocytoma, IDH-mutant	0.60	0.80
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	1.00	1.00
	Glioblastoma, IDH-wildtype	0.78	0.72
Specificity	Astrocytoma, IDH-mutant	0.79	0.82
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.95	0.89
	Glioblastoma, IDH-wildtype	1.00	1.00
Positive predictive value	Astrocytoma, IDH-mutant	0.30	0.40
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.33	0.20
	Glioblastoma, IDH-wildtype	1.00	1.00
Negative predictive value	Astrocytoma, IDH-mutant	0.93	0.97
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	1.00	1.00
	Glioblastoma, IDH-wildtype	0.46	0.40
Molecular diagnosis with GBCA-enhanced MRI			
Overall accuracy *(astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype)*		0.82	0.84
Accuracy	Astrocytoma, IDH-mutant	0.82	0.87
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.97	0.95
	Glioblastoma, IDH-wildtype	0.84	0.87
Sensitivity	Astrocytoma, IDH-mutant	0.60	0.80
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	1.00	1.00
	Glioblastoma, IDH-wildtype	0.84	0.84
Specificity	Astrocytoma, IDH-mutant	0.85	0.88
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.97	0.95
	Glioblastoma, IDH-wildtype	0.83	1.00
Positive predictive value	Astrocytoma, IDH-mutant	0.38	0.50
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.50	0.33
	Glioblastoma, IDH-wildtype	0.96	1.00

Table S3. Prediction performance of the raters in the development dataset *(Continued)*

Results		Rater 1	Rater 2
Negative predictive value	Astrocytoma, IDH-mutant	0.93	0.97
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	1.00	1.00
	Glioblastoma, IDH-wildtype	0.50	0.55

Table S3 shows the diagnostic prediction performance, with and without GBCA-enhanced scans, per rater in the development dataset (n=38).
Evaluations with GBCA-free MRI were based on pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and SWI (if present) sequences. Evaluations with GBCA-enhanced MRI included post-contrast T1-weighted images in addition to GBCA-free MRI sequences.

IDH = isocitrate dehydrogenase, GBCA = gadolinium-based contrast agent

Table S4. Prediction performance of the raters in the optimization dataset

Results		Rater 1	Rater 2
Histopathological grade with GBCA-free MRI			
Accuracy		0.97	0.90
Sensitivity		1.00	0.91
Specificity		0.89	0.89
Positive predictive value		0.96	0.95
Negative predictive value		1.00	0.80
Histopathological grade with GBCA-enhanced MRI			
Accuracy		0.87	0.97
Sensitivity		0.67	1.00
Specificity		0.95	0.89
Positive predictive value		0.86	0.96
Negative predictive value		0.88	1.00
Molecular diagnosis with GBCA-free MRI			
Overall accuracy *(astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype)*		0.81	0.84
Accuracy	Astrocytoma, IDH-mutant	0.84	0.94
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.94	0.90
	Glioblastoma, IDH-wildtype	0.84	0.84
Sensitivity	Astrocytoma, IDH-mutant	0.84	0.78
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.94	0.71
	Glioblastoma, IDH-wildtype	0.84	0.93

Table S4. Prediction performance of the raters in the optimization dataset *(Continued)*

Results		**Rater 1**	**Rater 2**
Specificity	Astrocytoma, IDH-mutant	0.67	1.00
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.71	0.96
	Glioblastoma, IDH-wildtype	0.93	0.75
Positive predictive value	Astrocytoma, IDH-mutant	0.75	1.00
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	1.00	0.83
	Glioblastoma, IDH-wildtype	0.78	0.78
Negative predictive value	Astrocytoma, IDH-mutant	0.87	0.92
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.92	0.92
	Glioblastoma, IDH-wildtype	0.92	0.92
Molecular diagnosis with GBCA-enhanced MRI			
Overall accuracy *(astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype)*		0.90	0.87
Accuracy	Astrocytoma, IDH-mutant	0.90	0.94
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.94	0.90
	Glioblastoma, IDH-wildtype	0.97	0.90
Sensitivity	Astrocytoma, IDH-mutant	0.90	0.88
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.94	0.71
	Glioblastoma, IDH-wildtype	0.97	0.93
Specificity	Astrocytoma, IDH-mutant	0.89	0.95
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.86	0.96
	Glioblastoma, IDH-wildtype	0.93	0.88
Positive predictive value	Astrocytoma, IDH-mutant	0.80	0.89
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.86	0.83
	Glioblastoma, IDH-wildtype	1.00	0.88
Negative predictive value	Astrocytoma, IDH-mutant	0.95	0.95
	Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	0.96	0.92
	Glioblastoma, IDH-wildtype	0.94	0.93

Table S4 shows the diagnostic prediction performance, with and without GBCA-enhanced scans, per rater in the optimization dataset (n=31).
Evaluations with GBCA-free MRI were based on pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and SWI (if present) sequences using the Diagnosis Prediction Decision Tree (DPDT). Evaluations with GBCA-enhanced MRI included post-contrast T1-weighted images in addition to the assessment conducted with GBCA-free MRI.
IDH = isocitrate dehydrogenase, GBCA = gadolinium-based contrast agent

Table S5. Failed GBCA-free and GBCE-enhanced diagnosis predictions across various tumor locations

Tumor locations	Failed GBCA-free predictions (%)		Failed GBCA-enhanced predictions (%)	
	Tumor grade	Molecular diagnosis	Tumor grade	Molecular diagnosis
Frontal	21%	32%	20%	31%
Parietal	4%	15%	3%	15%
Insula	19%	17%	19%	17%
Temporal	8%	21%	9%	19%
Thalamus	25%	46%	25%	42%
Occipital	15%	19%	10%	17%
Corpus callosum	0%	17%	0%	17%

Table S5 shows failed diagnosis predictions, including tumor grade and molecular diagnosis, across different tumor locations using GBCA-free and GBCA-enhanced scans based on a combined dataset of all raters.

GBCA= gadolinium-based contrast agent, GBCA-free scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) sequences, GBCA-enhanced scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted sequences

Tumor grade prediction includes low-grade (grade 2) vs. high-grade (grade 3/4) gliomas. Molecular diagnosis prediction includes astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant, and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype

Table S6. Pairwise inter-rater agreement in the prediction of the histomolecular diagnosis

Histomolecular diagnosis prediction	Raters 1&2	Raters 1&3	Raters 2&3
Tumor grade $_{\text{GBCA-free}}$*	0.53 (95%-CI: 0.40-0.66)	0.52 (95%-CI: 0.38-0.65)	0.65 (95%-CI: 0.52-0.78)
PABAK	*0.74*	*0.74*	*0.84*
Tumor grade $_{\text{GBCA-enhanced}}$*	0.74 (95%-CI: 0.63-0.86)	0.69 (95%-CI: 0.57-0.81)	0.59 (95%-CI: 0.46-0.73)
PABAK	*0.87*	*0.86*	*0.82*
Molecular diagnosis $_{\text{GBCA-free}}$*	0.75 (95%-CI: 0.68-0.82)	0.73 (95%-CI: 0.65-0.80)	0.77 (95%-CI: 0.70-0.84)
Molecular diagnosis $_{\text{GBCA-enhanced}}$*	0.78 (95%-CI: 0.71-0.85)	0.74 (95%-CI: 0.66-0.81)	0.79 (95%-CI: 0.72-0.86)

Table S6 shows the pairwise inter-rater agreement results for the prediction of histomolecular diagnosis using GBCA-free and GBCA-enhanced scans.
*unweighted Cohen`s kappa
GBCA= gadolinium-based contrast agent, GBCA-free scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) sequences, GBCA-enhanced scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted sequences, PABAK = Prevalence-adjusted and bias-adjusted kappa
Tumor grade prediction includes low-grade (grade 2) vs. high-grade (grade 3/4) gliomas. Molecular diagnosis prediction includes astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype. Interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect

Table S7. Pairwise inter-rater agreement in the evaluation of imaging features included in DPDT

Imaging features	Raters 1&2	Raters 1&3	Raters 2&3
Necrosis $_{\text{GBCA-free}}$*	0.87 (95%-CI: 0.81-0.93)	0.80 (95%-CI: 0.73-0.87)	0.82 (95%-CI: 0.75-0.88)
PABAK	*0.88*	*0.82*	*0.83*
Necrosis $_{\text{GBCA-enhanced}}$*	0.85 (95%-CI: 0.79-0.92)	0.84 (95%-CI: 0.77-0.91)	0.86 (95%-CI: 0.80-0.92)
PABAK	*0.87*	*0.86*	*0.87*
Diffusion restriction $_{\text{GBCA-free}}$**	0.42 (95%-CI: 0.34-0.49)	0.39 (95%-CI: 0.31-0.47)	0.37 (95%-CI: 0.28-0.46)
Diffusion restriction $_{\text{GBCA-enhanced}}$**	0.42 (95%-CI: 0.34-0.50)	0.34 (95%-CI: 0.26-0.42)	0.43 (95%-CI: 0.33-0.52)
Hemorrhage $_{\text{GBCA-free}}$*	0.65 (95%-CI: 0.55-0.74)	0.48 (95%-CI: 0.35-0.61)	0.33 (95%-CI: 0.22-0.43)
PABAK	*0.72*	*0.73*	*0.55*
Hemorrhage $_{\text{GBCA-enhanced}}$*	0.54 (95%-CI: 0.44-0.64)	0.40 (95%-CI: 0.26-0.53)	0.29 (95%-CI: 0.19-0.39)
PABAK	*0.64*	*0.70*	*0.50*
T2-FLAIR mismatch sign $_{\text{GBCA-free}}$*	0.52 (95%-CI: 0.27-0.78)	0.60 (95%-CI: 0.31-0.89)	0.59 (95%-CI: 0.36-0.82)
PABAK	*0.93*	*0.95*	*0.93*
T2-FLAIR mismatch sign $_{\text{GBCA-enhanced}}$*	0.53 (95%-CI: 0.27-0.78)	0.51 (95%-CI: 0.23-0.80)	0.60 (95%-CI: 0.45-0.88)
PABAK	*0.93*	*0.93*	*0.94*
Nonenhancing tumor margin $_{\text{GBCA-free}}$*	0.35 (95%-CI: 0.20-0.50)	0.36 (95%-CI: 0.22-0.51)	0.58 (95%-CI: 0.45-0.70)
PABAK	*0.74*	*0.71*	*0.78*
Nonenhancing tumor margin $_{\text{GBCA-enhanced}}$*	0.43 (95%-CI: 0.28-0.57)	0.47 (95%-CI: 0.34-0.61)	0.49 (95%-CI: 0.37-0.62)
PABAK	*0.74*	*0.73*	*0.71*
T2 homogeneity $_{\text{GBCA-free}}$*	0.44 (95%-CI: 0.28-0.60)	0.39 (95%-CI: 0.22-0.55)	0.49 (95%-CI: 0.28-0.69)
PABAK	*0.80*	*0.77*	*0.87*
T2 homogeneity $_{\text{GBCA-enhanced}}$*	0.42 (95%-CI: 0.23-0.60)	0.40 (95%-CI: 0.22-0.58)	0.57 (95%-CI: 0.39-0.75)
PABAK	*0.82*	*0.81*	*0.88*
Calvarial remodeling $_{\text{GBCA-free}}$*	0.54 (95%-CI: 0.35-0.74)	0.35 (95%-CI: 0.13-0.58)	0.56 (95%-CI: 0.36-0.76)
PABAK	*0.89*	*0.87*	*0.90*

Table S7. Pairwise inter-rater agreement in the evaluation of imaging features included in DPDT *(Continued)*

Imaging features	Raters 1&2	Raters 1&3	Raters 2&3
Calvarial remodeling $_{GBCA\text{-}enhanced}$*	0.58 (95%-CI: 0.40-0.77)	0.53 (95%-CI: 0.31-0.75)	0.56 (95%-CI: 0.35-0.76)
PABAK	*0.89*	*0.91*	*0.91*
Cyst $_{GBCA\text{-}free}$*	0.42 (95%-CI: 0.23-0.62)	0.36 (95%-CI: 0.17-0.55)	0.46 (95%-CI: 0.19-0.72)
PABAK	*0.85*	*0.84*	*0.91*
Cyst $_{GBCA\text{-}enhanced}$*	0.47 (95%-CI: 0.27-0.68)	0.13 (95%-CI: -0.04-0.31)	0.45 (95%-CI: 0.19-0.71)
PABAK	*0.87*	*0.82*	*0.91*
Calcification $_{GBCA\text{-}free}$*	0.73 (95%-CI: 0.45-1.02)	0.55 (95%-CI: 0.21-0.88)	0.54 (95%-CI: 0.13-0.95)
PABAK	*0.90*	*0.84*	*0.87*
Calcification $_{GGBCA\text{-}enhanced}$*	0.80 (95%-CI: 0.52-1.07)	0.49 (95%-CI: 0.08-0.89)	0.62 (95%-CI: 0.21-1.04)
PABAK	*0.94*	*0.84*	*0.90*
Midline shift $_{GBCA\text{-}free}$*	0.64 (95%-CI: 0.56-0.73)	0.60 (95%-CI: 0.51-0.69)	0.80 (95%-CI: 0.73-0.86)
PABAK	*0.68*	*0.62*	*0.80*
Midline shift $_{GBCA\text{-}enhanced}$*	0.73 (95%-CI: 0.65-0.82)	0.69 (95%-CI: 0.61-0.77)	0.78 (95%-CI: 0.71-0.85)
PABAK	*0.76*	*0.71*	*0.79*
Substantial edema $_{GBCA\text{-}free}$*	0.53 (95%-CI: 0.43-0.64)	0.51 (95%-CI: 0.42-0.60)	0.25 (95%-CI: 0.18-0.32)
PABAK	*0.66*	*0.50*	*0.21*
Substantial edema $_{GBCA\text{-}enhanced}$*	0.45 (95%-CI: 0.34-0.57)	0.50 (95%-CI: 0.41-0.59)	0.28 (95%-CI: 0.20-0.35)
PABAK	*0.62*	*0.50*	*0.27*

Table S7 shows the pairwise inter-rater agreement results for imaging features involved in DPDT using both GBCA-free and GBCA-enhanced scans.
*unweighted Cohen`s kappa, **weighted Cohen's kappa
DPDT = diagnosis prediction decision tree, GBCA= gadolinium-based contrast agent, GBCA-free scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) sequences, GBCA-enhanced scans = pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted sequences, PABAK = Prevalence-adjusted and bias-adjusted kappa
The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect

SUPPLEMENTARY FIGURES AND FIGURE CAPTIONS

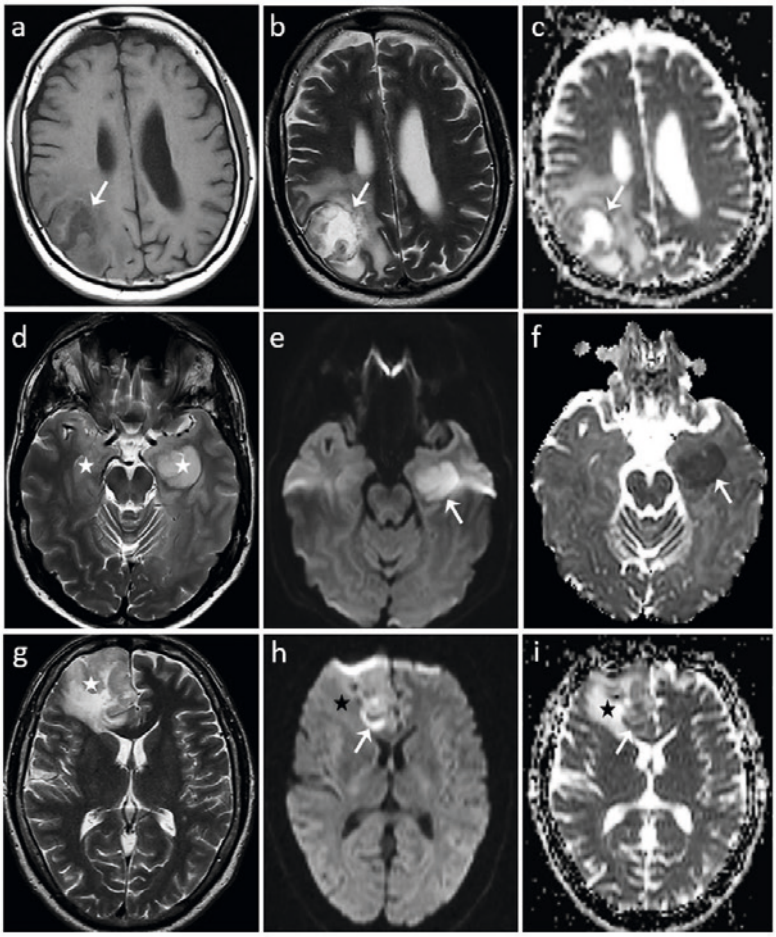


Figure S1: Case examples for the evaluation of necrosis and diffusion.

Necrosis (a,b,c; glioblastoma, IDH-wildtype): Axial MR images show a right parietal lesion with typical features of central necrosis characterized by T1 hypointense (a), T2 hyperintense (b) and high fluid-like ADC signal (c) covered by irregular and thick margins (white arrows). *Restricted diffusion (d,e,f; glioblastoma, IDH-wildtype):* Axial MR images show left and right mesial temporal/hippocampal infiltrative T2 hyperintense lesions (d, white stars) with restricted diffusion on the left side characterized by high TRACE/DWI signal (e; white arrow) and corresponding low ADC signal (f, white arrow). *Dubious and facilitated diffusion (g,h,i; high-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted):* Axial MR images demonstrate right frontal T2 hyperintense lesion (g, white star) with dubious diffusion characterized by high TRACE/DWI signal (h, white arrow) plus corresponding intermediate healthy cortex-like ADC signal (i, white arrow), and facilitated diffusion characterized by low TRACE/DWI signal (h, black star) plus corresponding high ADC signal (i, black star).

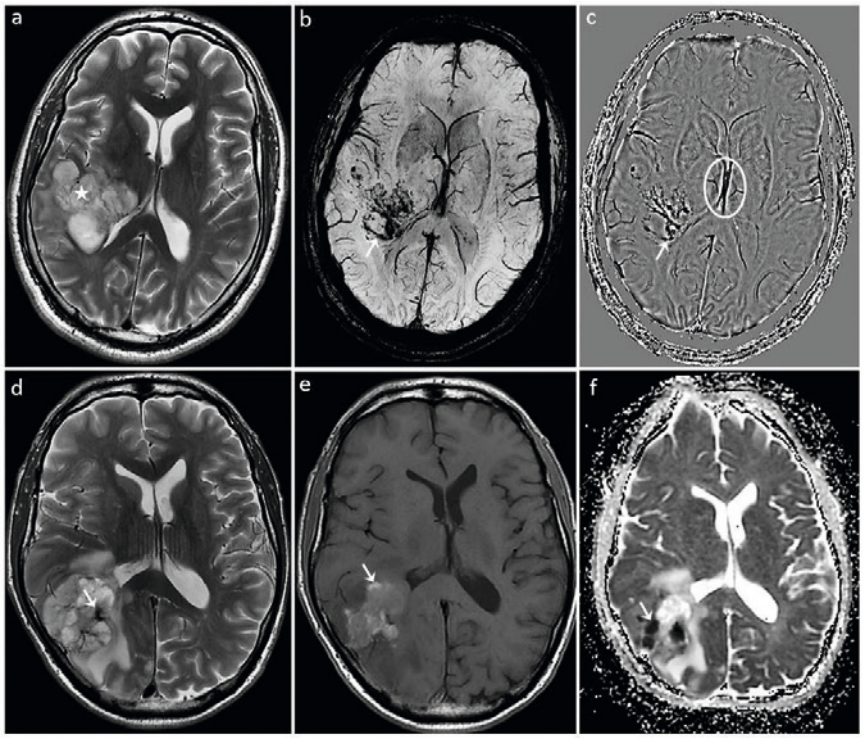


Figure S2: Case examples for the evaluation of hemorrhage.

Hemorrhage on SWI (a,b,c; glioblastoma, IDH-wildtype): Axial MR images show a right temporal T2 hyperintense lesion (a, white star) with hemorrhage characterized by the blooming on post-processed SWI image (b, white arrow) with the same signal intensity as internal cerebral veins on filtered phase images (c, white arrow pointing out hemorrhage and white circle highlighting internal cerebral vein). *Hemorrhage on pre-contrast T1-weighted image (d,e,f; glioblastoma, IDH-wildtype):* Axial MR images demonstrate right temporoparietal lesion with hemorrhagic necrosis characterized by a hyperintense signal within the necrotic part of the tumor on pre-contrast T1-weighted image (e, white arrow). There are also T2 hypointensity (d, white arrow) and low ADC signal (f, white arrow) related to hemorrhage. SWI images were not acquired for this case.

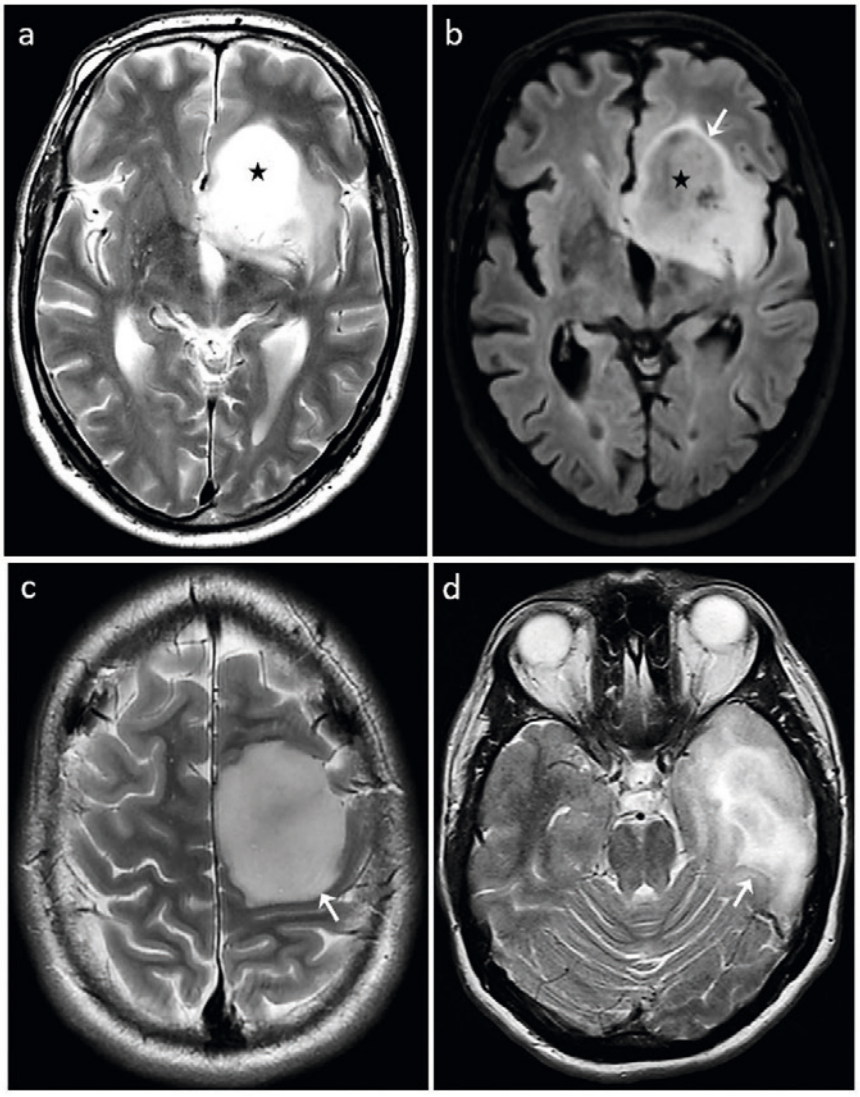


Figure S3: Case examples for the evaluation of T2-FLAIR mismatch sign, non-enhancing tumor margin, and T2 signal homogeneity.

T2-FLAIR mismatch sign (a,b; low-grade astrocytoma, IDH-mutant): Axial T2 weighted image shows left fronto-insular hyperintense nearly-homogeneous hyperintense lesion (a, black star) and FLAIR image demonstrates near-complete signal drop-out in the lesion (b, black star) except for a hyperintense peripheral rim (b, white arrow). *Well-defined margin and T2 homogeneity (c; low-grade astrocytoma, IDH-mutant):* Axial T2 weighted image shows left frontal lesion with homogeneous hyperintense signal and well-defined margins (white arrow). *Ill-defined margin and T2 inhomogeneity (d; high-grade astrocytoma, IDH-mutant):* Axial T2 weighted image shows left temporal lesion with inhomogeneous hyperintense signal and ill-defined margins (white arrow).

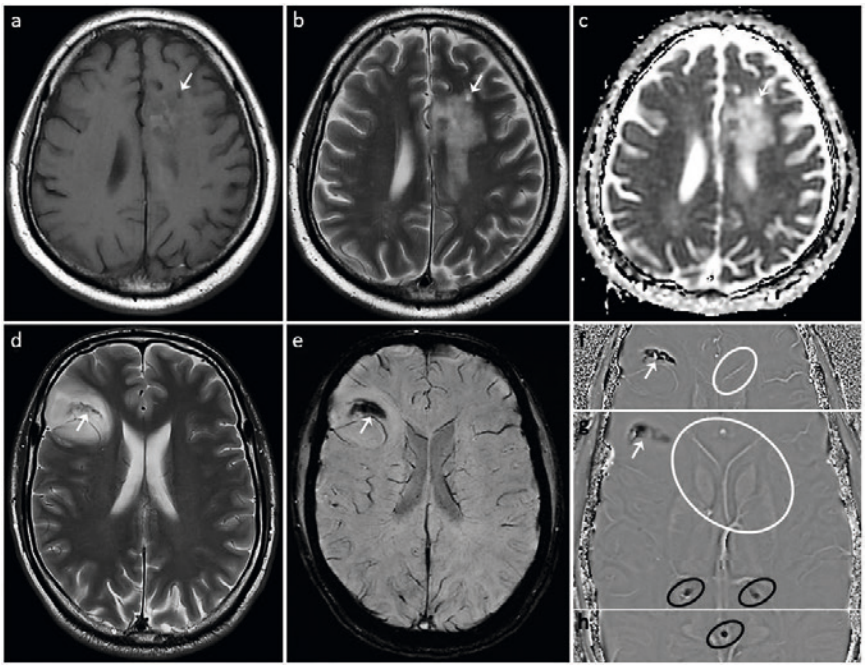


Figure S4: Case examples for the evaluation of cyst and calcification.

Cyst (a,b,c; high-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted): Axial MR images show a left frontal lesion consisting of eccentrically located small cyst with regular and smooth margins characterized by a well-defined rounded area of low T1-weighted (a, white arrow), bright T2-weighted (b, white arrow) and high ADC signal (c, white arrow) matching cerebrospinal fluid signal intensity. *Calcification (d,e,f,g,h; low-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted):* Axial MR images show right frontal T2 hyperintense heterogeneous lesion with internal calcification (d, white arrow) characterized by the blooming on post-processed SWI image (e, white arrow). This area displays a mainly hypointense signal on filtered phase image (f (the same level as in image e) and g (the choroid plexus level, white arrows), similar to calcified choroid plexus (g, dark circles) and pineal gland (h, dark circle), and opposite to the internal cerebral veins (f and g, white circles). These signal features suggest calcification rather than hemorrhage, using the internal cerebral veins as a reference for hemorrhage and the calcified choroid plexus/pineal gland as a reference for calcification.

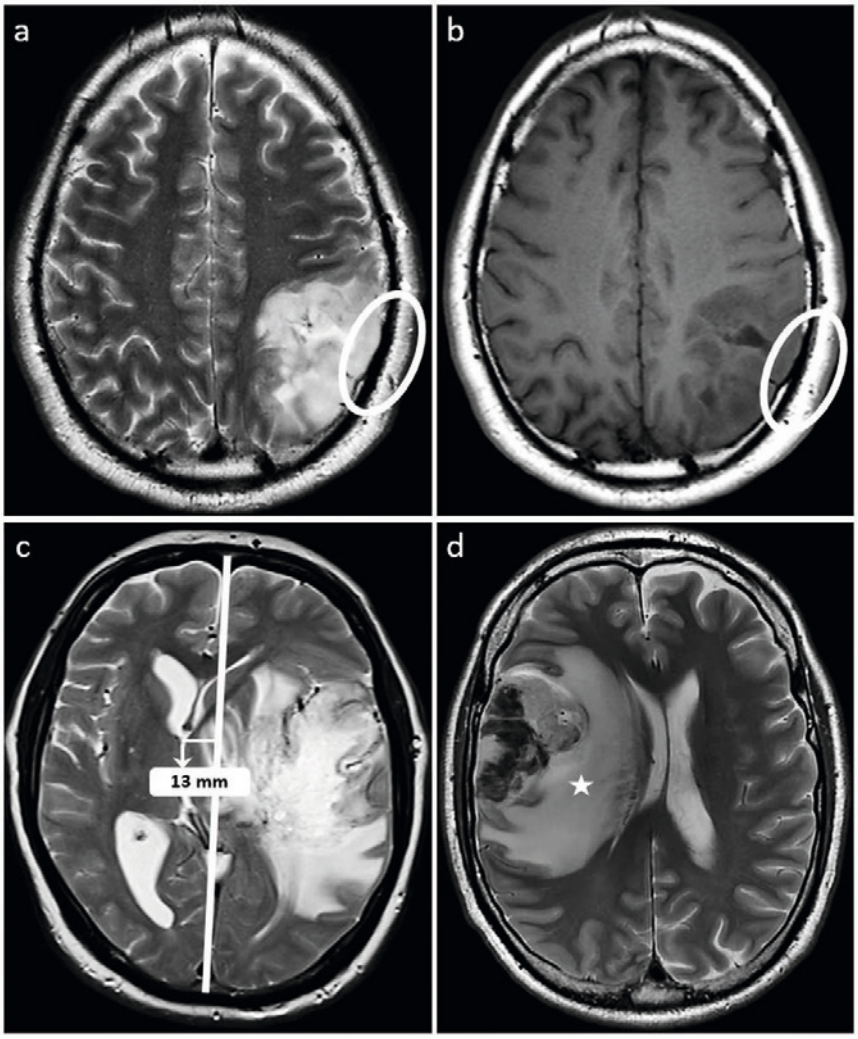


Figure S5: Case examples for the evaluation of calvarial remodeling, midline shift, and substantial edema.

Calvarial remodeling (a,b; low-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted): Axial T2-weighted (a) and pre-contrast T1-weighted (b) images show left parietal heterogeneous lesion causing discernible thinning/ remodeling of the inner table of parietal bone (white circles). *Midline shift (c; glioblastoma, IDH-wildtype):* Axial T2 weighted image shows left temporo-insular heterogeneous lesion causing more than 5 mm shift of the midline structures (horizontal line and text box indicating the midline shift degree of 13 mm) to the right according to the line drawn coplanar with falx connecting the anterior and posterior superior sagittal sinus attachments (white vertical line). *Substantial edema (d; glioblastoma, IDH-wildtype):* Axial T2-weighted image shows a right frontal heterogeneous lesion with hemorrhagic central necrosis lesion causing edema ≥50% of the tumor volume (white star).

Development dataset

Rater 1

True: reference standard	Predicted GBCA-free grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	4	1
High-grade glioma	1	32

True: reference standard	Predicted GBCA-enhanced grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	4	1
High-grade glioma	0	33

True: reference standard	Predicted GBCA-free molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	1	0	0
Astro, IDH-mut	2	3	0
GB, IDH-wt	0	7	25

True: reference standard	Predicted GBCA-enhanced molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	1	0	0
Astro, IDH-mut	1	3	1
GB, IDH-wt	0	5	27

Rater 2

True: reference standard	Predicted GBCA-free grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	4	1
High-grade glioma	3	30

True: reference standard	Predicted GBCA-enhanced grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	4	1
High-grade glioma	4	29

True: reference standard	Predicted GBCA-free molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	1	0	0
Astro, IDH-mut	1	4	0
GB, IDH-wt	3	6	23

True: reference standard	Predicted GBCA-enhanced molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	1	0	0
Astro, IDH-mut	1	4	0
GB, IDH-wt	1	4	27

Figure S6: Per-rater confusion matrices for predicting histomolecular diagnosis of adult-type diffuse gliomas using GBCA-free and GBCA-enhanced scans in the development dataset.

Diagnosis predictions include tumor grade (low-grade glioma: grade 2 vs. high-grade glioma: grade 3/4) and molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype) evaluations. The reference standard is the histomolecular diagnosis. Green and red color boxes show correct and false predictions, respectively. *GBCA = Gadolinium-based contrast agent; IDH = isocitrate dehydrogenase;* Oligo, IDH-mut & 1p/19-codel. = oligodendroglioma; IDH-mutant and 1p/19q-codeleted; Astro, IDH-mut= astrocytoma, IDH-mutant; GB, IDH-wt = glioblastoma, IDH-wildtype.

Optimization dataset

Rater 1

True: reference standard	Predicted GBCA-free grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	8	1
High-grade glioma	0	22

True: reference standard	Predicted GBCA-enhanced grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	6	3
High-grade glioma	1	21

True: reference standard	Predicted GBCA-free molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	5	1	1
Astro, IDH-mut	0	6	3
GB, IDH-wt	0	1	14

True: reference standard	Predicted GBCA-enhanced molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	6	1	0
Astro, IDH-mut	1	8	0
GB, IDH-wt	0	1	14

Rater 2

True: reference standard	Predicted GBCA-free grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	8	1
High-grade glioma	2	20

True: reference standard	Predicted GBCA-enhanced grade	
	Low-grade glioma	High-grade glioma
Low-grade glioma	8	1
High-grade glioma	0	22

True: reference standard	Predicted GBCA-free molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	5	0	2
Astro, IDH-mut	0	7	2
GB, IDH-wt	1	0	14

True: reference standard	Predicted GBCA-enhanced molecular diagnosis		
	Oligo, IDH-mut & 1p/19q-codel.	Astro, IDH-mut	GB, IDH-wt
Oligo, IDH-mut & 1p/19q-codel.	5	1	1
Astro, IDH-mut	0	8	1
GB, IDH-wt	1	0	14

Figure S7: Per-rater confusion matrices for predicting histomolecular diagnosis of adult-type diffuse gliomas using GBCA-free and GBCA-enhanced scans in the optimization dataset.

Diagnosis predictions include tumor grade (low-grade glioma: grade 2 vs. high-grade glioma: grade 3/4) and molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype) evaluations. The reference standard is the histomolecular diagnosis. Green and red color boxes show correct and false predictions, respectively. *GBCA = Gadolinium-based contrast agent; IDH = isocitrate dehydrogenase;* Oligo, IDH-mut & 1p/19-codel. = oligodendroglioma; IDH-mutant and 1p/19q-codeleted; Astro, IDH-mut= astrocytoma, IDH-mutant; GB, IDH-wt = glioblastoma, IDH-wildtype.

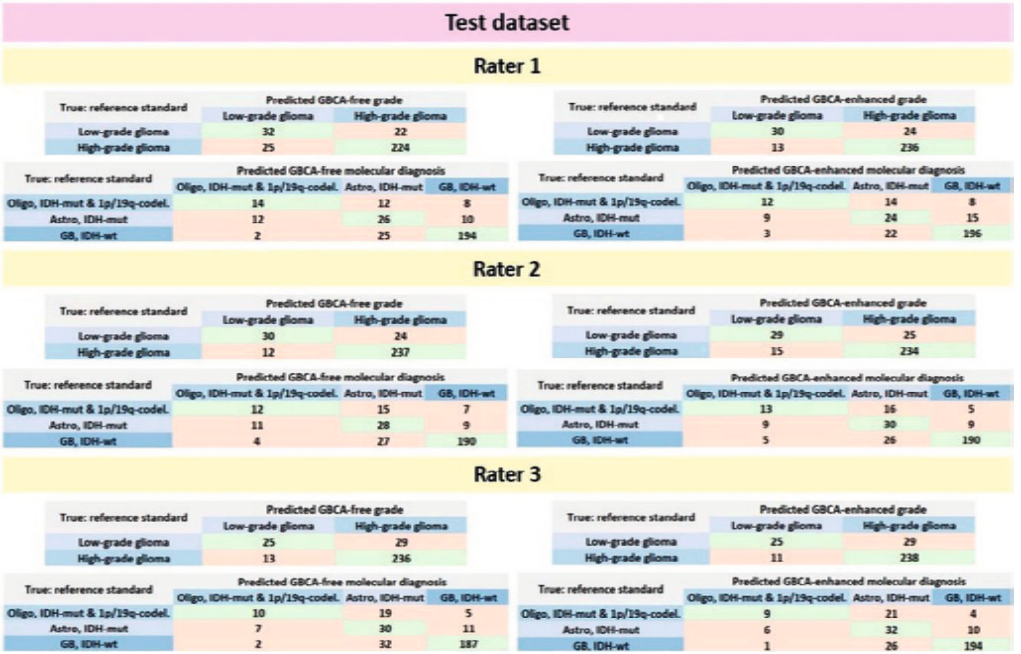


Figure S8: Per-rater confusion matrices for predicting histomolecular diagnosis of adult-type diffuse gliomas using GBCA-free and GBCA-enhanced scans in the test dataset.

Diagnosis predictions include tumor grade (low-grade glioma: grade 2 vs. high-grade glioma: grade 3/4) and molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype) evaluations. The reference standard is the histomolecular diagnosis. Green and red color boxes show correct and false predictions, respectively. *GBCA = Gadolinium-based contrast agent; IDH = isocitrate dehydrogenase;* Oligo, IDH-mut & 1p/19-codel. = oligodendroglioma; IDH-mutant and 1p/19q-codeleted; Astro, IDH-mut= astrocytoma, IDH-mutant; GB, IDH-wt = glioblastoma, IDH-wildtype.

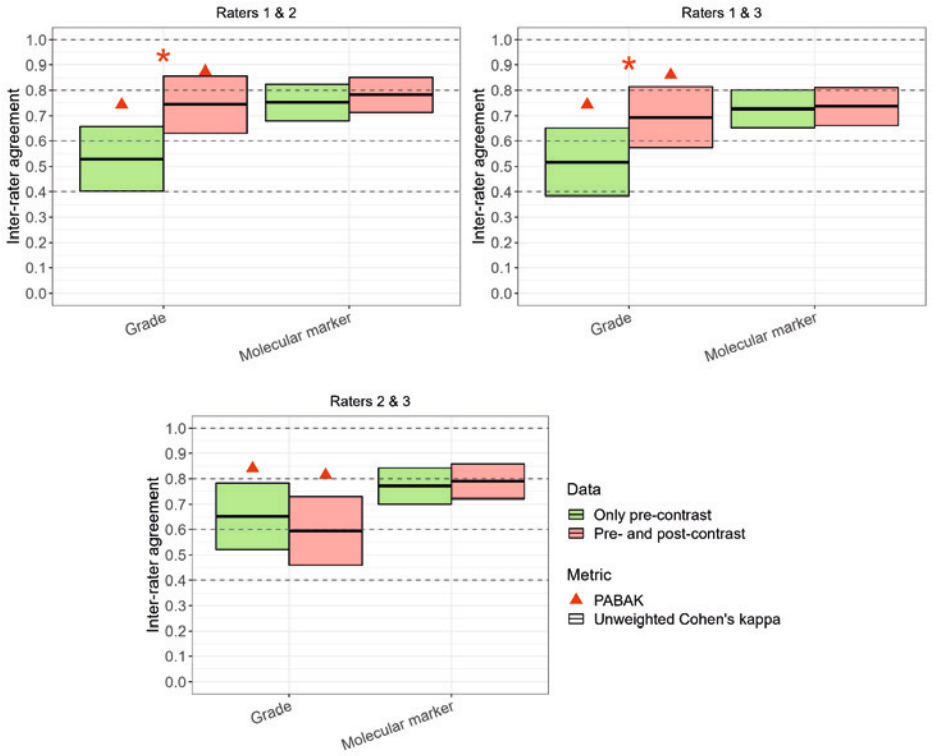


Figure S9: Pairwise inter-rater agreement in histomolecular diagnosis prediction of adult-type diffuse gliomas using GBCA-free vs. GBCA-enhanced scans.

Color box plots show inter-rater agreement in predicting tumor grade (low-grade: grade 2 vs. high-grade: grade 3/4) and molecular status (astrocytoma, IDH-mutant vs. oligodendroglioma, IDH-mutant and 1p/19q-codeleted vs. glioblastoma, IDH-wildtype) among various two raters (rater 1&2, rater 1&3, rater 2&3). Green bars depict the results based on the evaluation of GBCA-free scans (only pre-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, and SWI (if present)) and red bars show the results of the evaluation using GBCA-enhanced scans (pre- and post-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted). Red triangles indicate prevalence-adjusted and bias-adjusted kappa (PABAK) values that compensate for the possible influence of dataset diagnosis imbalances. Comparison of agreements between GBCA-free and GBCA-enhanced predictions was insignificant ($p > 0.05$) except for agreements in grade prediction for raters 1&2 and raters 1&3 (red stars, $p = 0.003$ and $p = 0.02$, respectively). *Note: The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect. GBCA = Gadolinium-based contrast agent, IDH = isocitrate dehydrogenase.*

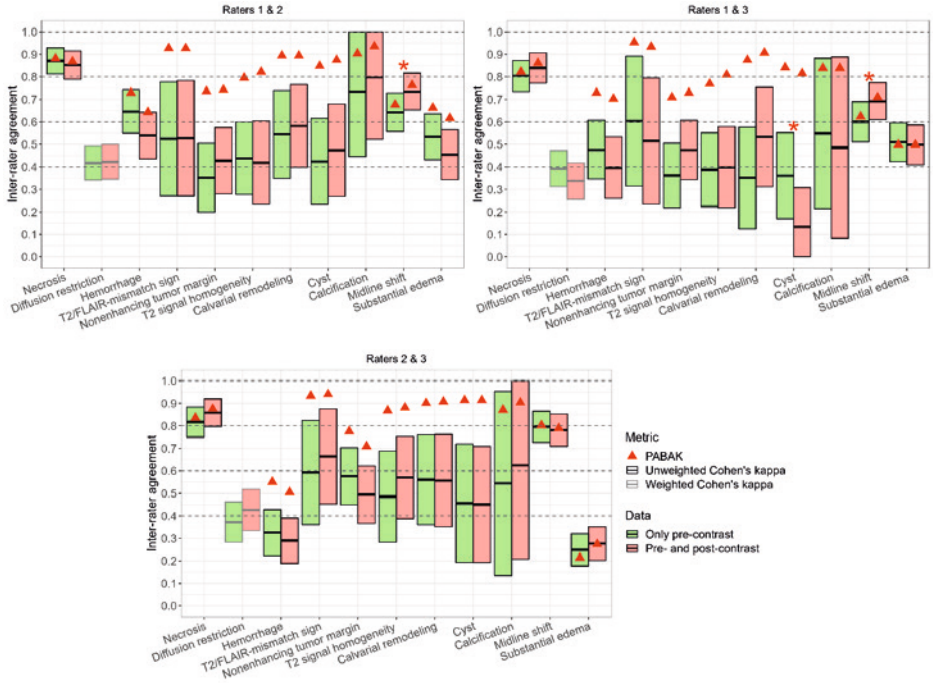


Figure S10: Pairwise inter-rater agreement in the evaluation of imaging features included in the Diagnosis Prediction Decision tree (DPDT) for adult-type diffuse gliomas using GBCA-free vs. GBCA-enhanced scans.

Color box plots show inter-rater agreement in the evaluation of single DPDT imaging features using either GBCA-free or GBCA-enhanced scans among various two raters (rater 1&2, rater 1&3, rater 2&3). Green bars depict the results based on the evaluation of GBCA-free scans (only pre-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC and SWI (if present)) and red bars show the results of the evaluation using GBCA-enhanced scans (pre- and post-contrast sequences: pre-contrast T1-weighted, T2-weighted, FLAIR, DWI/ADC, SWI (if present) + post-contrast T1-weighted). Red triangles indicate prevalence-adjusted and bias-adjusted kappa (PABAK) values that compensate for the possible influence of dataset diagnosis imbalances. Comparison of agreements between GBCA-free and GBCA-enhanced assessments was insignificant ($p > 0.05$) except for rater 1&2 midline shift (red star, $p = 0.02$) and rater 1&3 cyst (red star, $p = 0.04$) and midline shift (red star, $p = 0.03$). *Note: The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect. GBCA = Gadolinium-based contrast agent.*

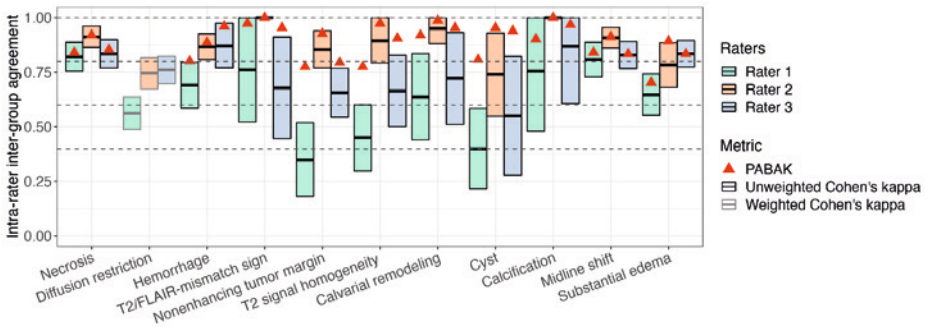


Figure S11: Intra-rater inter-group agreement in the evaluation of imaging features included in the Diagnosis Prediction Decision tree (DPDT) for adult-type diffuse gliomas.

Color box plot describes intra-rater inter-group agreement comparing GBCA-free and GBCA-enhanced evaluations of single DPDT imaging features by each rater (rater 1 = green bar, rater 2 = orange bar, rater 3 = blue bar). Red triangles indicate prevalence-adjusted and bias-adjusted kappa (PABAK) values that compensate for the possible influence of dataset diagnosis imbalances. Comparison of agreements among all raters revealed significant differences ($p < 0.05$) except for T2-FLAIR mismatch sign and calcification ($p > 0.05$). *Note: The interpretation of agreement values was as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect. GBCA = Gadolinium-based contrast agent.*

Chapter 6

Visual versus Region-of-Interest Based Diffusion Evaluation and their Diagnostic Impact in Adult-Type Diffuse Gliomas

Azizova, A., Prysiazhniuk, Y., Cakmak, M., Kaya, E., Petr, J., Barkhof, F., Wamelink, I.J.H.G., Keil, V.C.

Submitted: ***Neuroradiology***

ABSTRACT

Purpose: To evaluate the comparability and reproducibility of standardized visual versus region-of-interest (ROI)-based diffusion assessment and their prediction capacity for isocitrate dehydrogenase (IDH) mutation status in adult gliomas.

Methods: Preoperative MRI scans, including diffusion-weighted imaging (DWI), of grade 2-4 adult-type diffuse gliomas (n=303) were evaluated by three raters and repeated after one month. Visual assessment used the categorization of the Visually AcceSAble Rembrandt Images-feature 17 classes (facilitated, dubious, restricted). ROI-based assessment placed circular ROI on the visually perceived lowest apparent diffusion coefficient (ADC) areas (absolute/aADC) and contralateral normal-appearing white matter (normalized/nADC). Agreement and correlation analysis between visual and ROI-based assessments were performed. Logistic regression was conducted for IDH prediction in non-necrotic and non-hemorrhagic cases.

Results: ROI-based assessment demonstrated superior inter- and intra-rater agreement (≥0.56 (95%-CI: 0.48-0.63)) than visual assessment (≥0.34 (95%-CI: 0.26-0.42)). Thresholds of 1,090 and 623 x 10^{-6} mm^2/s for aADC, and 1.38 and 0.80 for nADC, distinguishing facilitated, dubious, and restricted diffusion, significantly correlated with visual assessments (P < .001). IDH classification accuracy of visual assessment was comparable to that of the ROI-based method using thresholds of aADC 1,048 x 10^{-6} mm^2/sn and nADC 1.38 (visual vs. aADC/nADC: 69% vs. 73%/70%). However, neither method achieved a balanced performance between specificity (99% vs. 81%/75%) and sensitivity (14% vs. 57%/61%).

Conclusion: ROI-based diffusion assessment guided by visual input showed superior reproducibility than visual assessment alone. However, visual assessment strongly correlated with ADC thresholds and demonstrated comparable IDH prediction accuracy, suggesting potentially equivalent clinical utility to ROI-based assessment.

INTRODUCTION

MRI is the primary modality for monitoring adult-type diffuse gliomas, the most prevalent malignant brain tumors in adults, providing essential diagnostic and prognostic information [1–3]. Conventionally, radiologists interpret MRI data visually, and several visually assessed MRI biomarkers are now part of routine clinical practice [4]. Quantitative MRI sequences and their standardized or quantitative evaluation are gaining attention, with growing evidence demonstrating their efficacy in distinguishing tumor characteristics [5–7].

Diffusion-weighted imaging (DWI) is the most established quantitative MRI technique and the only one recommended as a standard component of brain tumor imaging protocol [1]. It measures the Brownian motion of water molecules in tissues, providing insights into cellular density [8]. Commonly, DWI is analyzed by visually comparing the b-1000 isotropic map juxtaposed to an apparent diffusion coefficient (ADC) map. A European glioma imaging survey reported that 78% of surveyed neuroradiologists prefer visual assessment of potential diffusion restriction [9]. While visual analysis is quick and widely adopted, it is inherently subjective. To address this, the Visually AcceSAble Rembrandt Images (VASARI) glioma imaging features set [10] introduced a standardized approach, categorizing diffusion assessment (feature 17) into three classes: restricted, dubious, and facilitated. In clinical practice, radiologists can opt against this plain visual analysis and choose a region-of-interest (ROI)-based assessment of diffusivity that produces absolute but normalizable values. Absolute ADC values directly reflect the diffusion properties within the ROI but are affected by technical factors. Normalized ADC values, the ratio of the absolute ADC to that of normal-appearing white matter, reduce variability across sequences and scanners but depend on accurate reference region selection and may obscure direct comparability of absolute values.

Diffusion assessment is critical in radiological decision-making, often alongside other MRI sequences [11, 12]. However, its independent diagnostic relevance has also been evaluated to better understand its clinical impact [13, 14]. One important application is isocitrate dehydrogenase (IDH) status prediction, a key biomarker for classifying adult-type diffuse gliomas. While visual assessment has been explored for its IDH genotyping, a recent meta-analysis [15]

found limited evidence of its significance compared to other visual imaging features. In contrast, ROI-based assessments show strong potential for distinguishing IDH-wildtype tumors from IDH-mutant tumors [16–18], although the lack of consensus on ADC thresholds remains a significant limitation [19]. Almost all of these studies included tumors with necrosis and hemorrhage, defining features of IDH-wildtype tumors [4, 12, 20–23], which degrade ADC map quality and limit the reliability of diffusion assessments. Excluding such tumors could enhance the predictive accuracy and clinical applicability of DWI. Moreover, there is scant evidence in the literature regarding the correlation between visual and ROI-based approaches and the comparative diagnostic accuracy of these methods.

This study aims to evaluate if the visual assessment of diffusion - represented by VASARI feature 17 - is comparable with an ROI-based assessment and similarly reproducible. To explore a possible diagnostic impact of the diffusion evaluation method in daily practice, we assess the methods' capacity to predict IDH status in adult-type diffuse gliomas, excluding tumors with necrosis and hemorrhage.

METHODS

Study cohort

The medical ethics review committee (VUmc_2021-0437) approved this retrospective single-center study and waived informed consent. Eligible patients between January 2010 and January 2021 were taken from a cohort presented in previous publications [12, 24]. The study cohort was sourced from the pseudonymized hospital glioma database (IMAGO) by I.W., a fourth-year neuro-oncology Ph.D. student. Inclusion criteria were treatment-naïve patients with grades 2-4 adult-type diffuse gliomas according to the 5th World Health Organization Central Nervous System Classification. Patients with preoperative MRI data consisting of pre-contrast T1-weighted, T2-weighted, T2-FLAIR, post-contrast T1-weighted images, and DWI b-0 and b-1000 images with ADC maps generated automatically on the scanner were analyzed. Exclusion criteria were incomplete histomolecular diagnosis (e.g., missing IDH status), incomplete or suboptimal preoperative MRI (e.g., motion artifacts), a more than one-month interval

between preoperative MRI and surgery, suprasellar, midbrain, and cerebellar tumors due to their distinct radiophenotype, and pediatric age group.

MRI data and analysis

MR images of all patients (n=303) were acquired on three 1.5T (n=120) and four 3T scanners (n=183); see Supplementary Table 1 for details. Three raters with different levels of radiology experience (V.C.K., eleven years of neuroradiology experience; A.A., five years of neuroradiology experience; M.C., a fourth-year medical student with one month of specialized radiology training for this study using a different small cohort (n=69) from the hospital glioma database) independently conducted imaging evaluations using RADIANT software (version 3.4.1.13367; https://www.radiantviewer.com/). The raters evaluated the visual and ROI-based methods, focusing on the solid tumor parts. Hemorrhage, necrosis, cysts, and peritumoral edema identified from the evaluation of pre- and post-contrast T1-weighted, T2-weighted, and T2-FLAIR images were excluded from the assessments. In multifocal/multicentric glioma cases, the most aggressive-looking lesion defined by the visually lowest ADC signal was considered. Visual and ROI-based evaluations were performed twice (measurements 1 and 2) by the same raters at a one-month interval for all enrolled patients. Raters were blinded to the histomolecular diagnosis during evaluation.

Visual and ROI-based DWI assessments

Visual evaluations were conducted using the VASARI feature 17: facilitated, dubious, and restricted diffusion (Supplementary Fig.1). Facilitated diffusion is marked by a high or low b-1000 signal with a corresponding ADC signal higher than normal brain tissue. Dubious diffusion is identified by a high signal on b-1000 images with a corresponding ADC signal resembling normal brain tissue. Restricted diffusion is characterized by a high DWI signal intensity on b-1000 images with a corresponding lower signal on ADC maps than normal brain tissue. If the lesion showed a heterogeneous diffusion pattern, the lowest diffusion score was recorded, irrespective of the relative size of the area.

ROI-based assessments included placing circular ROI on areas on the ADC map that visually appeared to have the lowest ADC (absolute ADC; aADC, mm^2/s); see Supplementary Fig.1. The slice with the largest area of this visually lowest ADC region was exclusively considered. The mean value of the measured area

was recorded. The circular ROI size was 20-40 mm^2 to standardize the measurements across the raters. The raters were instructed to cover the region with the lowest ADC as completely as possible without extending into areas with a visually different ADC. A same-size circular ROI was also positioned on the contralateral normal-appearing white matter (ADC_{NAWM}) for normalization ($aADC/ADC_{NAWM}$ = normalized ADC; nADC).

Statistical analysis

Analysis was conducted using R package 4.3.0 by Y.P., a third-year Ph.D. student in neuroscience. The significance threshold was $P < .05$.

Descriptive analysis

Visual assessment was expressed as percentages per category. ROI-based measurements were summarized with the mean and standard deviation.

Rater agreement

Consistency among raters was evaluated separately for the two assessment rounds through group and pairwise inter-rater agreements. In contrast, per-rater consistency between the first and second assessments was measured using intra-rater agreement analysis. Bootstrapping with 1000 iterations was used for all agreement analyses to calculate the confidence intervals. ROI-based measurements were analyzed using an intraclass correlation coefficient with two-way random-effects and mixed-effects models for inter-rater and intra-rater agreements. In visual assessment, Kendall's W and Cohen's weighted kappa were used for group inter-rater and pairwise inter-rater/intra-rater agreements. Agreement values were interpreted as follows: 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect [25].

Correlation between visual and ROI-based DWI assessments

The overall distribution of ROI-based measurements (mean, standard deviation) within visual assessment classes was calculated using all six measurements. Logistic regression analysis was then used to determine aADC and nADC thresholds for distinguishing different visual assessment classes (facilitated vs. dubious, dubious vs. restricted). Spearman rank correlation analysis was conducted to identify the relationship between visual assessment and thresholded ROI-based measurements. The interpretation of the correlation

coefficient (ρ) was as follows: 0.00–0.09, negligible; 0.10–0.39, weak; 0.40–0.69, moderate; 0.70–0.89, strong; and 0.90–1.00, very strong [26].

Inter-method IDH classification prediction performance

Considering the IDH status of cases, descriptive analyses were repeated for both visual and ROI-based measurements. Confusion matrices were used to assess the classification performance, treating IDH-wildtype gliomas as the positive class. The radiological IDH prediction was conducted only on cases without visual hemorrhage or necrosis, as these imaging features are primarily associated with IDH-wildtype tumors [4, 12, 20–23] and provide evident descriptive characteristics, making diffusion status assessment less relevant. Hemorrhage and necrosis were evaluated using contrast-enhanced MRI as part of a previously published study [12], and cases with these imaging features identified by at least two raters were excluded from the analysis.

For visual assessments, cases rated as *restricted diffusion* were classified as IDH-wildtype, while a rating of *dubious plus facilitated diffusion* was classified as IDH-mutant. The rationale is the results of a previous study showing respective assumptions to be predictive [12].

Logistic regression analysis determined the IDH classification thresholds for the ROI-based values of both aADC and nADC. To ensure robustness and prevent data leakage, measurements across all raters were included and divided into training (70%) and test (30%) sets at the patient level. This approach ensured that all measurements from the same patient were assigned exclusively to either the training or test set, preserving the independence of the datasets. The Random OverSampling Examples method [27] addressed a class imbalance regarding IDH status in the training set. Subsequently, the diagnostic performance of these thresholded ROI-based measurements was calculated.

RESULTS

Descriptive analysis

Table 1 shows the study cohort demographics.

Table 1. Study cohort demographics

Sample size: number	303
Age: years (standard deviation)	56.7 (14.2)
Sex: female/male	114/189
Isocitrate dehydrogenase status: mutant (codeleted/intact)/wildtype	82 (34/48)/221
Histological World Health Organization grade: grade 2/grade 3 or 4	54/249

Caption: Table 1 describes the main characteristics of the study cohort

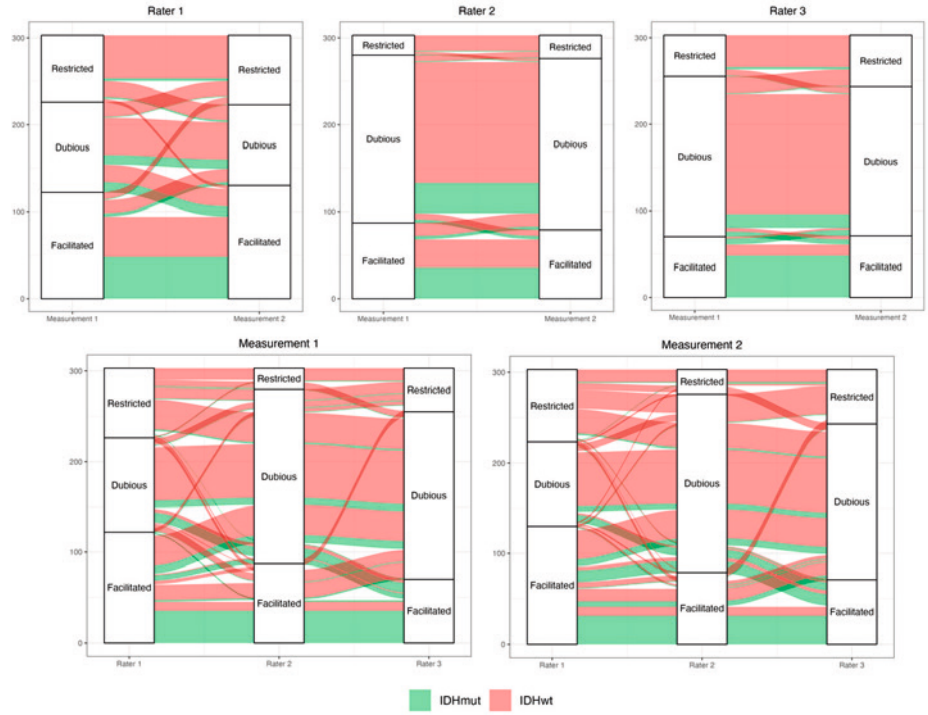


Figure 1: Alluvial plots show the distribution of visual assessment classes per rater between measurements and per measurement between raters. Green and red colors represent isocitrate dehydrogenase-mutant (IDHmut) and -wildtype (IDHwt) gliomas, respectively.

The distribution of visual assessment classes per rater and measurement is shown in Fig.1 and Supplementary Table 2. There was variability among evaluations by different raters, particularly in the *restricted diffusion class*. Overall, *restricted diffusion class* was the least chosen category (restricted vs. facilitated and dubious: 7-26% vs. 74-93%), with rater 2 assigning only 7% and 9% of cases to this category in assessments 1 and 2, respectively. These percentages were higher for rater 1 (25-26%) and rater 3 (16-20%). The most frequently chosen class was the *dubious diffusion class* for rater 2 (assessment 1/2: 64%/65%) and 3 (assessment 1/2: 61%/57%), while rater 1 primarily assigned cases to the *facilitated diffusion class* (assessment 1/2: 40%/43%).

The mean and standard deviation of aADC ranged between 864±323 and 1,000±391 x 10^{-6} mm^2/s across all measurements. The mean and standard deviation of nADC ranged between 1.13±0.42 and 1.30±0.53; see Fig.2 and Supplementary Table 3.

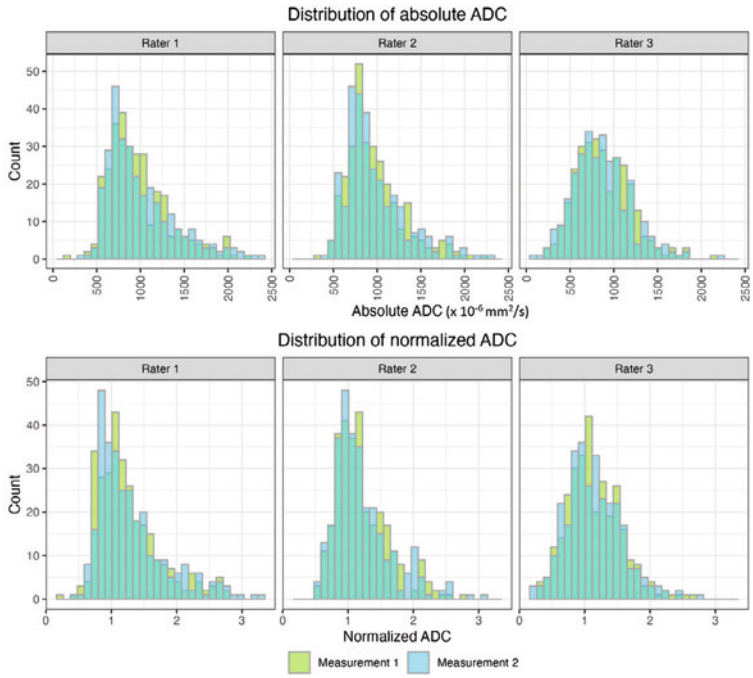


Figure 2: Stacked histograms show the distribution of absolute and normalized ADC values per rater and measurement. Green bars represent the overlap between the first (yellow bars) and second (blue bars) measurements.

Rater agreement

Group inter-rater agreements in measurements 1 and 2 were moderate for visual assessment (0.51 (95%-CI: 0.46-0.56) and 0.52 (95%-CI: 0.47-0.56)) and substantial for both aADC (0.66 (95%-CI: 0.60-0.72) and 0.64 (95%-CI: 0.56-0.70)) and nADC (0.62 (95%-CI: 0.56-0.68) and 0.62 (95%-CI: 0.55-0.68)). Pairwise inter-rater agreements were fair-moderate (≥0.34 (95%-CI: 0.26-0.42) and moderate-substantial (≥0.56 (95%-CI: 0.48-0.63) for visual and ROI-based assessments, respectively; see Table 2 and Supplementary Fig.2. Intra-rater inter-measurement agreements were moderate-substantial (≥0.56 (95%-CI: 0.49-0.64)) for visual assessment and substantial-almost perfect for ROI-based assessments (≥0.73 (95%-CI: 0.67-0.77)); see Table 3 and Supplementary Fig.3.

Table 2. Pairwise inter-rater agreement in visual and ROI-based diffusion assessments

Rater pairs	Measurement 1	Measurement 2
Visual assessment*		
Rater 1 & 2	0.42 (95%-CI: 0.34-0.50)	0.42 (95%-CI: 0.34-0.49)
Rater 2 & 3	0.43 (95%-CI: 0.33-0.51)	0.37 (95%-CI: 0.28-0.46)
Rater 1 & 3	0.34 (95%-CI: 0.26-0.42)	0.39 (95%-CI: 0.31-0.47)
Absolute ADC mm^2/s**		
Rater 1 & 2	0.74 (95%-CI: 0.69-0.79)	0.69 (95%-CI: 0.63-0.74)
Rater 2 & 3	0.58 (95%-CI: 0.48-0.67)	0.58 (95%-CI: 0.47-0.67)
Rater 1 & 3	0.65 (95%-CI: 0.54-0.73)	0.63 (95%-CI: 0.47-0.74)
Normalized ADC**		
Rater 1 & 2	0.69 (95%-CI: 0.63-0.75)	0.66 (95%-CI: 0.58-0.72)
Rater 2 & 3	0.56 (95%-CI: 0.48-0.63)	0.58 (95%-CI: 0.50-0.65)
Rater 1 & 3	0.60 (95%-CI: 0.51-0.68)	0.62 (95%-CI: 0.48-0.72)

Caption: Table 2 demonstrates the pairwise inter-rater agreement results in visual and ROI-based assessments of diffusion.
*Cohen's weighted kappa
**Intraclass correlation coefficient

Table 3. Intra-rater inter-measurement agreement in visual and ROI-based diffusion assessments

Visual assessment*	
Rater 1	0.56 (95%-CI: 0.49-0.64)
Rater 2	0.75 (95%-CI: 0.68-0.82)
Rater 3	0.76 (95%-CI: 0.70-0.83)
Absolute ADC mm^2/s**	
Rater 1	0.76 (95%-CI: 0.70-0.80)
Rater 2	0.86 (95%-CI: 0.83-0.89)
Rater 3	0.79 (95%-CI: 0.75-0.83)
Normalized ADC**	
Rater 1	0.73 (95%-CI: 0.67-0.77)
Rater 2	0.85 (95%-CI: 0.81-0.87)
Rater 3	0.77 (95%-CI: 0.72-0.82)

Caption: Table 3 demonstrates intra-rater inter-measurement agreement results in visual and ROI-based assessments of diffusion.
*Cohen's weighted kappa
**Intraclass correlation coefficient

Correlation between visual and ROI-based DWI assessments

The mean and standard deviation of aADC within visual assessment classes was as follows: facilitated 1,298±314 x 10^{-6} mm^2/s, dubious 853±185 x 10^{-6} mm^2/s, restricted 583±172 x 10^{-6} mm^2/s. Logistic regression analysis yielded optimal aADC thresholds of 1,090 x 10^{-6} mm^2/s for facilitated vs. dubious diffusion and 623 x 10^{-6} mm^2/s for dubious vs. restricted diffusion. The means and standard deviations of nADC within facilitated, dubious, and restricted visual assessment classes were 1.69±0.42, 1.08±0.21, and 0.73±0.18, respectively. The nADC thresholds for facilitated vs. dubious and dubious vs. restricted diffusion were 1.38 and 0.80, respectively.

Subsequent analysis using the calculated thresholds revealed a strong correlation between visual and ROI-based assessments, with an overall correlation coefficient of ρ=0.79 (P < .001) for visual vs. aADC and ρ=0.81 (P < .001) for visual vs. nADC (Fig.3). The results per rater and measurement also revealed a consistently strong correlation (all P < .001); for details, see Table 4 and Supplementary Fig.4.

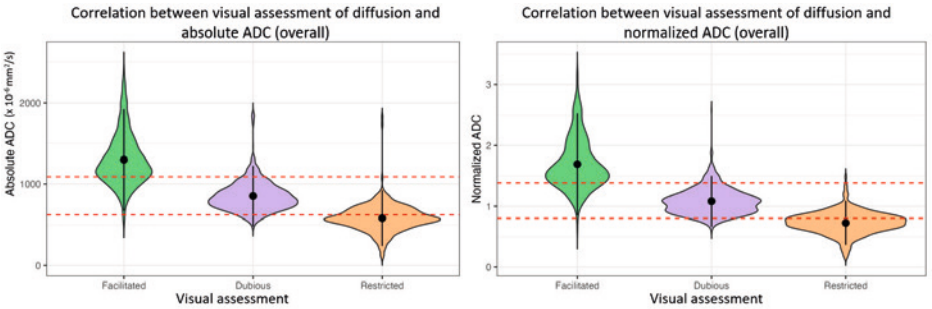


Figure 3: Violin plots show the overall correlation between visual and ROI-based assessments of diffusion, including absolute and normalized ADC, across all measurements. Red dashed lines represent the absolute/normalized ADC distribution thresholds within visual assessment classes (facilitated, dubious, and restricted) derived from logistic regression analysis (absolute ADC: 1,090 mm^2/s for facilitated vs. dubious diffusion and 623 mm^2/s for dubious vs restricted diffusion; normalized ADC: 1.38 for facilitated vs. dubious diffusion and 0.80 for dubious vs. restricted diffusion).

Table 4. The results of correlation analysis between visual and ROI-based diffusion assessments

Per rater visual vs. ROI-based assessments ρ* (P-value)		Measurement 1	Measurement 2
		ρ* (P-value)	
Rater 1	Visual vs. absolute ADC mm^2/s	0.78 (< .001)	0.77 (< .001)
	Visual vs. normalized ADC	0.78 (< .001)	0.82 (< .001)
Rater 2	Visual vs. absolute ADC mm^2/s	0.79 (< .001)	0.80 (< .001)
	Visual vs. normalized ADC	0.81 (< .001)	0.81 (< .001)
Rater 3	Visual vs. absolute ADC mm^2/s	0.80 (< .001)	0.83 (< .001)
	Visual vs. normalized ADC	0.82 (< .001)	0.86 (< .001)

Caption: Table 4 shows Spearman's rank correlation analysis results for each rater and measurement.
*Spearman's rank correlation coefficient

Inter-method IDH classification prediction performance

Visual assessments

In IDH-mutant gliomas, facilitated diffusion was the primary assessment class (52-80%) across all six measurements. For IDH-wildtype gliomas, facilitated (38-58%) and dubious (20-51%) classes were selected at similar rates. Restricted diffusion, the least common class overall, was more prevalent in IDH-wildtype gliomas (7-22%) compared to IDH-mutant gliomas (0-3%); see Fig.4 and Table 5.

When cases with visually restricted diffusion were accepted as IDH-wildtype and the remaining as IDH-mutant, the visual assessment achieved 69% accuracy, 14% sensitivity, 99% specificity, 89% positive predictive value, and 68% negative predictive value.

ROI-based assessments

The mean and standard deviation of aADC ranged from 979±264 to 1082±367 x 10^{-6} mm^2/s for IDH-wildtype and from 1,161±280 to 1,406±447 x 10^{-6} mm^2/s for IDH-mutant gliomas across all six measurements. For nADC, The mean and standard deviation ranged from 1.25±0.35 to 1.40±0.54 for IDH-wildtype and from 1.54±0.42 to 1.91±0.55 for IDH-mutant gliomas; see Fig.5 and Table 5.

Optimal IDH classification thresholds were 1,048 x 10^{-6} mm^2/s and 1.38 for aADC and nADC, respectively; see Fig.5. The subsequent classification accuracy, sensitivity, specificity, and positive and negative predictive values for aADC were 73%, 57%, 81%, 58%, and 80%, respectively. The corresponding results for nADC were 70% accuracy, 61% sensitivity, 75% specificity, 53% positive predictive value, and 80% negative predictive value.

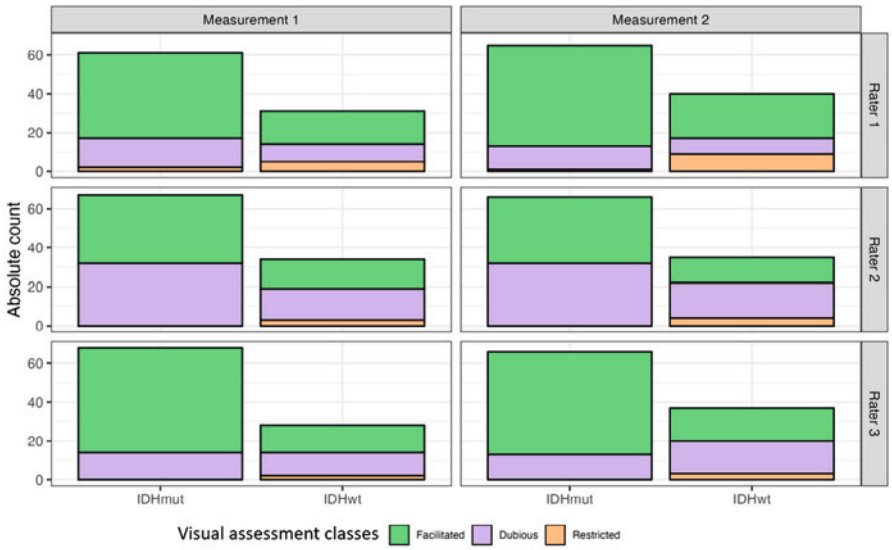


Figure 4: Stacked bar plots show the distribution of visual diffusion assessment classes per rater and measurement in isocitrate dehydrogenase (IDH)-mutant and -wildtype gliomas using non-necrotic and non-hemorrhagic adult-type glioma cases. IDHmut = IDH-mutant, IDHwt = IDH-wildtype.

Table 5. Distribution of visual and ROI-based assessment of diffusion in IDH-mutant and IDH-wildtype gliomas

		Rater 1		Rater 2		Rater 3	
		run 1	run 2	run 1	run 2	run 1	run 2
Distribution of visual assessment classes %							
IDH-mutant	Facilitated	72%	80%	52%	52%	79%	80%
	Dubious	25%	18%	48%	48%	21%	20%
	Restricted	3%	2%	0%	0%	0%	0%
IDH-wildtype	Facilitated	55%	58%	44%	38%	50%	46%
	Dubious	29%	20%	47%	51%	43%	46%
	Restricted	16%	22%	9%	11%	7%	8%
Distribution of absolute ADC mean±SD x 10^{-6} mm^2/s							
IDH-mutant		1,328±388	1,406±447	1,217±359	1,234±403	1,187±282	1,161±280
IDH-wildtype		1,066±372	1,082±367	1013±308	1,022±337	979±264	1,024±301
Distribution of normalized ADC mean±SD							
IDH-mutant		1.76±0.53	1.91±0.55	1.54±0.42	1.59±0.51	1.60±0.37	1.57±0.35
IDH-wildtype		1.38±0.54	1.40±0.54	1.27±0.39	1.25±0.41	1.25±0.35	1.40±0.40

Caption: Table 5 describes the distribution of visual and ROI-based assessment of diffusion in IDH-mutant and IDH-wildtype gliomas using non-necrotic and non-hemorrhagic adult-type diffuse gliomas.
Abbreviation: run 1/2 = first/second assessments, IDH = isocitrate dehydrogenase

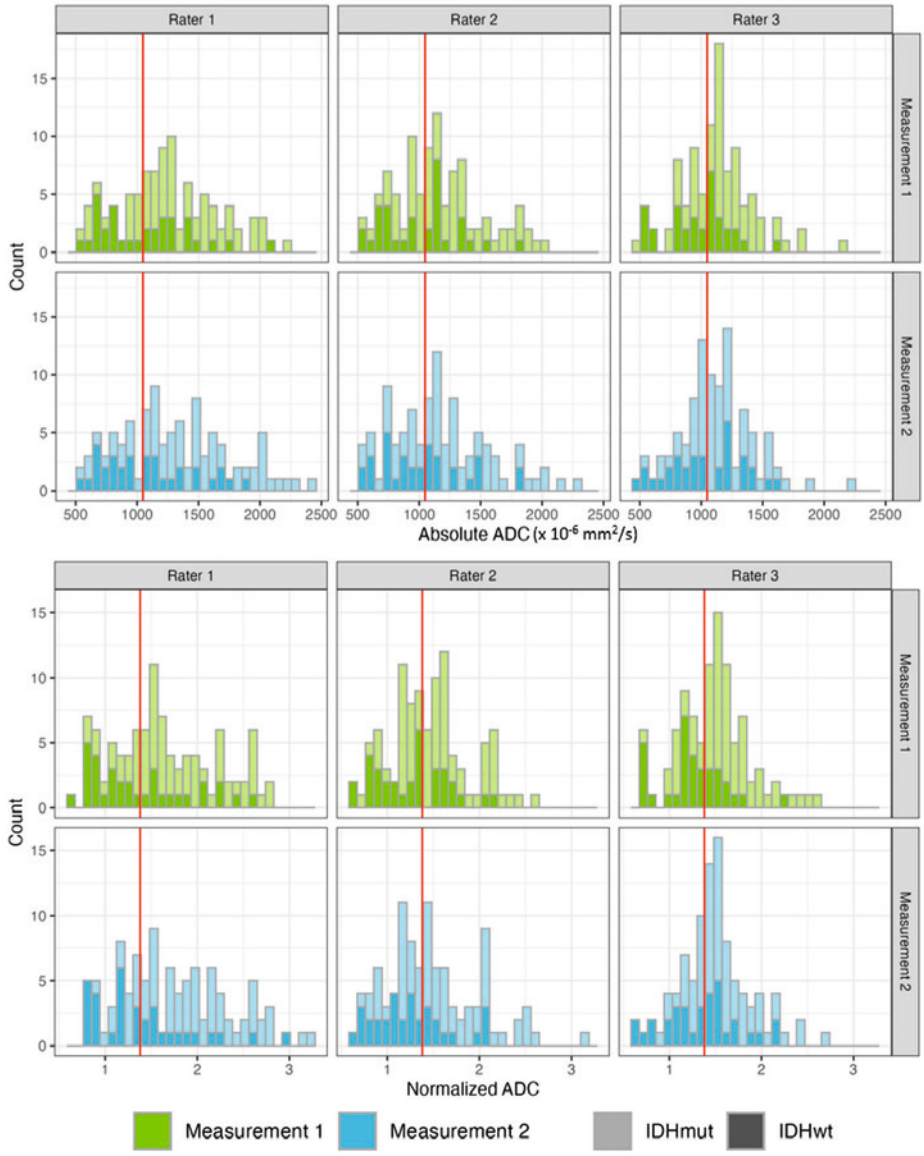


Figure 5: Stacked histograms show the absolute and normalized ADC distribution per rater and measurement in isocitrate dehydrogenase (IDH)-mutant and -wildtype gliomas using non-necrotic and non-hemorrhagic adult-type glioma cases. Red horizontal lines represent the optimal IDH classification thresholds of absolute/normalized ADC, derived from logistic regression analysis (absolute ADC threshold: 1,048 mm^2/s; normalized ADC threshold: 1.38).

DISCUSSION

This study evaluated the reproducibility, correlation, and IDH categorization performance of visual vs. ROI-based diffusion assessment in adult-type diffuse gliomas. ROI-based assessment demonstrated superior rater reproducibility, with moderate-almost perfect inter-/intra-rater agreements (≥0.56 (95%-CI: 0.48-0.63)) compared to fair-substantial agreements for visual assessment (≥0.34 (95%-CI: 0.26-0.42)). ADC thresholds of 1,090 and 623 x 10^{-6} mm^2/s for aADC and 1.38 and 0.80 for nADC, distinguishing facilitated, dubious, and restricted diffusion, however, significantly correlated well with visual assessments. For the clinical use case of IDH classification, visual assessment, when compared to the ROI-based method at a threshold of 1,048 x 10^{-6} mm^2/s for aADC and 1.38 for nADC, achieved the highest specificity (visual vs. aADC/nADC: 99% vs. 81%/75%), but had very low sensitivity (visual vs. aADC/nADC: 14% vs. 57%/61%). Regarding accuracy, the visual assessment showed comparable performance to the ROI-based method (69% vs. 73%/70%). This combination of results bears challenging implications for clinical practice.

Reproducibility is crucial when evaluating the consistency and reliability of imaging methods. Studies using VASARI criteria for visual assessment reported agreements ranging from fair to almost perfect (kappa 0.36-0.85), the spread highlighting the subjective nature of visual assessments [28–31]. ROI-based methods, being possibly more impartial, demonstrated superior reproducibility with intraclass correlation coefficient agreements ranging from 0.84 to 0.99 [16, 32, 33]. Our study further showed that an ROI-based evaluation offers a superior consistency (intraclass correlation coefficient 0.56-0.86) than the visual assessment (kappa 0.34-0.76). The lower agreement values for the ROI-based method in our study compared to previous studies may be attributed to methodological differences. For instance, two studies [16, 32] measured three visually defined lowest ADC areas and used the mean value of these measurements, while another study [33] included all solid tumor components with low ADC signal, potentially leading to more reliable evaluations. In contrast, our study focused on measuring only one area representing the visually lowest ADC signal to reflect the day-to-day practice and accurately correlate these two methods.

The literature lacked direct comparisons between visual and ROI-based diffusion assessments until now. A related study [34] compared both approaches but focused on glioma grading using a five-scale visual system, making direct comparison with our study challenging. Aligning with our results, they found a higher specificity (aADC 89% vs. visual 100%) and lower sensitivity (aADC 90% vs. visual 50%) for a visual evaluation. Our study established ADC thresholds for visual assessment classes in gliomas to facilitate method correlation, especially for hard-to-classify diffusion cases (aADC range: 623-1,090 x 10^{-6} mm^2/s and nADC range: 0.80-1.30), which are very common in gliomas and challenging for radiologists. These thresholds are presented to the community to streamline decision-making in clinical and research settings by integrating the reproducibility of ROI-based methods with the time efficiency of the visual method. Moreover, the results of this study could potentially be utilized to guide the application of diffusion data in advanced predictive models that incorporate artificial intelligence, which is currently hardly established.

On the other hand, the clinical use case we applied for this study may suggest that despite lower reproducibility, the visual assessment of diffusivity in glioma and possibly other brain tumors is sufficient and diagnostically comparable to the ROI-based method, thus clinically equally powerful. Obviously, IDH status is not exclusively based on DWI in clinical practice. The IDH analysis of this study is a means to demonstrate the relevance of the choice of diffusion analysis practice. Visual diffusion assessment for IDH classification was explored in several studies using the VASARI glioma imaging set [22, 23, 28–30, 35, 36]. Except for one study [28], none reported a significant predictive value for feature 17, visual assessment of diffusion, excluding it from multivariable models. Our study assessed the performance of different diffusion assessment approaches and found that the accuracy of visual assessment was comparable to that of aADC or nADC (69% vs. 73/70%). Nonetheless, both approaches revealed limitations, with the visual approach achieving high specificity but at the cost of low sensitivity, while the ROI-based method improved sensitivity but had a lower specificity. This imbalance highlights the challenge of reliably distinguishing glioma IDH characteristics using either method in isolation.

Studies using ROI-based diffusion assessment for IDH subtyping reported aADC thresholds between 900 and 1,200 x 10^{-6} mm^2/s [11, 13, 16, 37–39] and

nADC thresholds between 1.28-1.60 [13, 16, 39]. Similarly, our study identified thresholds within this range, with an aADC threshold of 1,048 x 10^{-6} mm^2/s and an nADC threshold of 1.38. The variation in reported thresholds likely reflects methodological heterogeneity across different studies; for example, Ma et al. [16] used the average of three visually identified lowest ADC areas for the tumor and a single ROI for normal-appearing white matter, reporting 65/92% sensitivity/specificity for aADC at 930 x 10^{-6} mm^2/s and 69/93% for nADC at 1.28. Another study [39] averaged four non-overlapping ROIs for the tumor and two for normal-appearing white matter, showing 84/68% sensitivity/specificity at 1,200 x 10^{-6} mm^2/s for aADC and 82/80% for nADC at 1.60. Thust et al. [13] used regional and volumetric ADC assessments with ADC_{min} measurements based on visually identified lowest ADC areas. They gathered three ROIs and took into account the mean value of the numerically lowest ADC measurement. Although cases with hemorrhage and necrosis were not excluded, the study focused on grade 2 and 3 gliomas, the majority of which were IDH-mutant (204 IDH-mutant vs. 79 IDH-wildtype), thereby reducing the likelihood of encountering these imaging features, thus closely matching our cohort. Their classification thresholds were $aADC_{min}$ at 1,070 x 10^{-6} mm^2/s, and $nADC_{min}$ at 1.40 with sensitivity/specificity values of 82/61% for $aADC_{min}$ and 86/62% for $nADC_{min}$.

This study has several limitations. First, this is a single-center study; however, it includes MRI data from multiple scanners and field strengths, reflecting normal clinical variability. Second, automated or volumetric ADC measurements were not included; instead, ROI placement was guided by visual evaluations, potentially introducing a collinearity bias. This method choice intentionally reflects real-world workflow but may limit the detection of visually subtle lowest ADC areas. However, full automation of ADC readings is unlikely to be implemented soon, particularly when excluding biasing hemorrhagic or necrotic areas is necessary. Additionally, the interrater variability in ADC measurements may have hindered the determination of a single optimal threshold for IDH classification. Future studies should aim to harmonize distributions across raters to mitigate interrater effects, incorporate external datasets with varied ADC quantification methods, and include other pathologies to increase the clinical impact of this study by validating and refining its findings.

CONCLUSIONS

ROI-based diffusion assessment with visual guidance in adult-type diffuse gliomas provided more reproducible results than visual assessment alone, but both techniques rendered results with a high degree of correlation and comparable accuracies to predict IDH status. Clinicians can, therefore, rely on their visual assessment of DWI only, but should consider confirming their visual assessment through ROI measurements when repetitive measurements are planned.

REFERENCES

1. Ellingson BM, Bendszus M, Boxerman J, et al (2015) Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. Neuro Oncol 17:1188–1198. https://doi.org/10.1093/neuonc/nov095
2. Wen PY, van den Bent M, Youssef G, et al (2023) RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults. J Clin Oncol 41:5187–5199. https://doi.org/10.1200/JCO.23.01059
3. Whitfield BT, Huse JT (2022) Classification of adult-type diffuse gliomas: Impact of the World Health Organization 2021 update. Brain Pathol 32:e13062. https://doi.org/10.1111/bpa.13062
4. Feraco P, Franciosi R, Picori L, et al (2022) Conventional MRI-Derived Biomarkers of Adult-Type Diffuse Glioma Molecular Subtypes: A Comprehensive Review. Biomedicines 10.: https://doi.org/10.3390/biomedicines10102490. https://doi.org/10.3390/biomedicines10102490
5. Hangel G, Schmitz-Abecassis B, Sollmann N, et al (2023) Advanced MR Techniques for Preoperative Glioma Characterization: Part 2. J Magn Reson Imaging 57:1676–1695. https://doi.org/10.1002/jmri.28663
6. Hirschler L, Sollmann N, Schmitz-Abecassis B, et al (2023) Advanced MR Techniques for Preoperative Glioma Characterization: Part 1. J Magn Reson Imaging 57:1655–1675. https://doi.org/10.1002/jmri.28662
7. Prysiazhniuk Y, Server A, Leske H, et al (2024) Diffuse glioma molecular profiling with Arterial Spin Labeling and Dynamic Susceptibility Contrast perfusion MRI: a comparative study. Neuro Oncol Adv vdae113. https://doi.org/10.1093/noajnl/vdae113
8. Huisman TAGM (2010) Diffusion-weighted and diffusion tensor imaging of the brain, made easy. Cancer Imaging 10 Spec no A:S163–71. https://doi.org/10.1102/1470-7330.2010.9023
9. Thust SC, Heiland S, Falini A, et al (2018) Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice. Eur Radiol 28:3306–3317. https://doi.org/10.1007/s00330-018-5314-5
10. VASARI research project - the cancer imaging archive (TCIA) public access - cancer imaging archive wiki. https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project. Accessed 19 Dec 2023
11. Xing Z, Yang X, She D, et al (2017) Noninvasive assessment of IDH mutational status in World Health Organization grade II and III astrocytomas using DWI and DSC-PWI combined with conventional MR imaging. AJNR Am J Neuroradiol 38:1138–1144. https://doi.org/10.3174/ajnr.A5171
12. Azizova, A., Prysiazhniuk, Y., Wamelink, I. J. H. G., et al (2024) Preoperative prediction of diffuse glioma type and grade in adults: a gadolinium-free MRI-based decision tree. European radiology, 10.1007/s00330-024-11140-5. Advance online publication. https://doi.org/10.1007/s00330-024-11140-5
13. Thust SC, Maynard JA, Benenati M, et al (2021) Regional and volumetric parameters for diffusion-weighted WHO grade II and III glioma genotyping: A method comparison. AJNR Am J Neuroradiol 42:441–447. https://doi.org/10.3174/ajnr.A6965
14. Zhang L, Min Z, Tang M, et al (2017) The utility of diffusion MRI with quantitative ADC measurements for differentiating high-grade from low-grade cerebral gliomas: Evidence from a meta-analysis. J Neurol Sci 373:9–15. https://doi.org/10.1016/j.jns.2016.12.008
15. Azizova A, Prysiazhniuk Y, Wamelink IJHG, et al (2024) Ten Years of VASARI Glioma Features: Systematic Review and Meta-Analysis of Their Impact and Performance. AJNR Am J Neuroradiol. https://doi.org/10.3174/ajnr.A8274. https://doi.org/10.3174/ajnr.A8274
16. Ma X, Cheng K, Cheng G, et al (2023) Apparent Diffusion Coefficient as Imaging Biomarker for Identifying IDH Mutation, 1p19q Codeletion, and MGMT Promoter Methylation Status in Patients With Glioma. J Magn Reson Imaging 58:732–738. https://doi.org/10.1002/jmri.28589

17. Wu J, Su R, Qiu D, et al (2022) Analysis of DWI in the classification of glioma pathology and its therapeutic application in clinical surgery: a case-control study. Transl Cancer Res 11:805–812. https://doi.org/10.21037/tcr-22-114

18. Cindil E, Sendur HN, Cerit MN, et al (2022) Prediction of IDH Mutation Status in High-grade Gliomas Using DWI and High T1-weight DSC-MRI. Acad Radiol 29 Suppl 3:S52–S62. https://doi.org/10.1016/j.acra.2021.02.002

19. Smits M (2021) MRI biomarkers in neuro-oncology. Nat Rev Neurol 17:486–500. https://doi.org/10.1038/s41582-021-00510-y

20. Guo X, Gu L, Li Y, et al (2023) Histological and molecular glioblastoma, IDH-wildtype: a real-world landscape using the 2021 WHO classification of central nervous system tumors. Front Oncol 13:1200815. https://doi.org/10.3389/fonc.2023.1200815

21. Ahn SS, An C, Park YW, et al (2021) Identification of magnetic resonance imaging features for the prediction of molecular profiles of newly diagnosed glioblastoma. J Neurooncol 154:83–92. https://doi.org/10.1007/s11060-021-03801-y

22. Sahu A, Patnam NG, Goda JS, et al (2022) Multiparametric Magnetic Resonance Imaging Correlates of Isocitrate Dehydrogenase Mutation in WHO high-Grade Astrocytomas. J Pers Med 13.: https://doi.org/10.3390/jpm13010072. https://doi.org/10.3390/jpm13010072

23. Gemini L, Tortora M, Giordano P, et al (2023) Vasari Scoring System in Discerning between Different Degrees of Glioma and IDH Status Prediction: A Possible Machine Learning Application? J Imaging Sci Technol 9.: https://doi.org/10.3390/jimaging9040075. https://doi.org/10.3390/jimaging9040075

24. Azizova A, Wamelink IJHG, Prysiazhniuk Y, et al. Human performance in predicting enhancement quality of gliomas using gadolinium-free MRI sequences. J Neuroimaging. 2024; doi: 10.1111/jon.13233.

25. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

26. Schober P, Boer C, Schwarte LA (2018) Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg 126:1763–1768. https://doi.org/10.1213/ANE.0000000000002864

27. Lunardon N, Menardi G, Torelli N (2014) ROSE: A package for binary imbalanced learning. R J 6:79

28. Park YW, Han K, Ahn SS, et al (2018) Prediction of IDH1-Mutation and 1p/19q-Codeletion Status Using Preoperative MR Imaging Phenotypes in Lower Grade Gliomas. AJNR Am J Neuroradiol 39:37–42. https://doi.org/10.3174/ajnr.A5421

29. Zhou H, Vallières M, Bai HX, et al (2017) MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 19:862–870. https://doi.org/10.1093/neuonc/now256

30. Hyare H, Rice L, Thust S, et al (2019) Modelling MR and clinical features in grade II/III astrocytomas to predict IDH mutation status. Eur J Radiol 114:120–127. https://doi.org/10.1016/j.ejrad.2019.03.003

31. Su C-Q, Lu S-S, Han Q-Y, et al (2019) Intergrating conventional MRI, texture analysis of dynamic contrast-enhanced MRI, and susceptibility weighted imaging for glioma grading. Acta radiol 60:777–787. https://doi.org/10.1177/0284185118801127

32. Maynard J, Okuchi S, Wastling S, et al (2020) World Health Organization grade II/III glioma molecular status: Prediction by MRI morphologic features and apparent diffusion coefficient. Radiology 296:111–121. https://doi.org/10.1148/radiol.2020191832

33. Hu Y-C, Yan L-F, Sun Q, et al (2017) Comparison between ultra-high and conventional mono b-value DWI for preoperative glioma grading. Oncotarget 8:37884–37895. https://doi.org/10.18632/oncotarget.14180

34. Phuttharak W, Thammaroj J, Wara-Asawapati S, Panpeng K (2020) Grading Gliomas Capability: Comparison between Visual Assessment and Apparent Diffusion Coefficient (ADC) Value Measurement on Diffusion-Weighted Imaging (DWI). Asian Pac J Cancer Prev 21:385–390. https://doi.org/10.31557/APJCP.2020.21.2.385

35. Su C-Q, Lu S-S, Zhou M-D, et al (2019) Combined texture analysis of diffusion-weighted imaging with conventional MRI for non-invasive assessment of IDH1 mutation in anaplastic gliomas. Clin Radiol 74:154–160. https://doi.org/10.1016/j.crad.2018.10.002

36. Cao M, Suo S, Zhang X, et al (2021) Qualitative and quantitative MRI analysis in IDH1 genotype prediction of lower-grade gliomas: A machine learning approach. Biomed Res Int 2021:1235314. https://doi.org/10.1155/2021/1235314

37. Villanueva-Meyer JE, Wood MD, Choi BS, et al (2018) MRI features and IDH mutational status of grade II diffuse gliomas: Impact on diagnosis and prognosis. AJR Am J Roentgenol 210:621–628. https://doi.org/10.2214/AJR.17.18457

38. Wasserman JK, Nicholas G, Yaworski R, et al (2015) Radiological and pathological features associated with IDH1-R132H mutation status and early mortality in newly diagnosed anaplastic astrocytic tumours. PLoS One 10:e0123890. https://doi.org/10.1371/journal.pone.0123890

39. Liu S, Zhang Y, Kong Z, et al (2022) Feasibility of evaluating the histologic and genetic subtypes of WHO grade II-IV gliomas by diffusion-weighted imaging. BMC Neurosci 23:72. https://doi.org/10.1186/s12868-022-00750-8

ONLINE SUPPLEMENT

Supplementary Tables

Supplementary Table 1. MRI sequence parameters per scanner

Scanner (patient number)	TE (ms)	TR (ms)	TI (ms)	FA (°)
GE 1.5T Signa HDxt (n=20)				
2D T1w	9-12	520-600	-	90
T2w	98-104	4,376-4,840	-	90
2D T2-FLAIR	118-120	9,002-9,502	2250	90
3D T2-FLAIR	96-122	6,000-6,500	1,925-1,987	90
DWI (b-1,000)	81-105	8,000-8,500	-	90
2D CE-T1w	9-12	520-600	-	90
3D CE-T1w	3-5	8-13	0-450	12
GE 3T DISCOVERY MR750 (n=97)				
2D T1w	7.9-9.7	600-731	-	90-125
3D T1w	2	4.7	650	15
T2w	82-88	4,889-6,872	-	90-111
3D T2-FLAIR	126-135	8,000-8,002	2,331-2,347	90
DWI (b-1,000)	62-87	4,000-7,200	-	90
2D CE-T1w	7.9-8.4	600-650	-	90-125
3D CE-T1w	2-3.2	4.6-8.3	450-650	15

Supplementary Table 1. MRI sequence parameters per scanner *(Continued)*

Scanner (patient number)	TE (ms)	TR (ms)	TI (ms)	FA (°)
Philips 1.5T Achieva (n=15)				
3D T1w	3.3-4.6	6.5-8.7	-	8-10
T2w	100-110	3,404-5,251	-	90
2D T2-FLAIR	140	9,000-11,000	2,800	90
3D T2-FLAIR	286-306	4,800	1,660	90
DWI (b-1,000)	72-119	2,674-6,448	-	90
3D CE-T1w	3.3-4.6	6.7-8.7	-	8-10
Philips 3T Ingenuity (n=64)				
2D T1w	10	599	-	70
3D T1w	3	7	-	12
T2w	85	2,767-3,182	-	90
3D T2-FLAIR	279	4,800	1,650	90
DWI (b-1,000)	74-97	3,496-6,354	-	90
2D CE-T1w	10	599	-	70
3D CE-T1w	3	7	-	12

Supplementary Table 1. MRI sequence parameters per scanner *(Continued)*

Scanner (patient number)	TE (ms)	TR (ms)	TI (ms)	FA (°)
Siemens 1.5T Avanto (n=85)				
2D T1w	7.8-17	500-718	-	90
3D T1w	4.5-11	700-2,700	0-950	8-120
T2w	93-104	2,830-5,562	-	150-180
2D T2-FLAIR	88-109	8,870-9,000	2,500	150
3D T2-FLAIR	334	6,500	2,200	120
DWI (b-1,000)	90-122	3,400-10,500	-	90
2D CE-T1w	8.7-17	550-718	-	90
3D CE-T1w	2.9-4.5	1,900-2,700	950-1,100	8
Siemens 3T MAGNETOM Vida (n=18)				
3D T1w	2.3	2,300	900	8
T2w	74	4,100-6,280	-	150
3D T2-FLAIR	388-430	5,000-7,700	1,650-2,400	120
DWI (b-1,000)	68	3,200	-	90
3D CE-T1w	2.3	2,300	900	8

Supplementary Table 1. MRI sequence parameters per scanner *(Continued)*

Scanner (patient number)	TE (ms)	TR (ms)	TI (ms)	FA (°)
Toshiba 3T Titan3T (n=4)				
2D T1w	8	550	-	80
T2w	90	5,500-5,526	-	90
3D T2-FLAIR	451	5,600	1,900	90
DWI (b-1,000)	82	7,500	-	90
2D CE-T1w	8	550	-	80
3D CE-T1w	2.4	5.7	900	9

Caption: Supplementary Table 1 lists the MRI sequence parameters for three 1.5T and four 3T MRI scanners.
Abbreviations: CE = contrast enhanced, FA = flip angle, ms = millisecond, n = number, 1.5T/3T = 1.5 tesla/3 tesla, TE = time of echo, TI = time of inversion, TR = time of repetition, 2D = 2-dimensional, 3D = 3-dimensional, ° = degrees

Supplementary Table 2. Distribution of visual assessment class of diffusion per rater and measurement

Raters	measurement 1			measurement 2		
	Restricted	Dubious	Facilitated	Restricted	Dubious	Facilitated
Rater 1	25%	35%	40%	26%	31%	43%
Rater 2	7%	64%	29%	9%	65%	26%
Rater 3	16%	61%	23%	20%	57%	23%

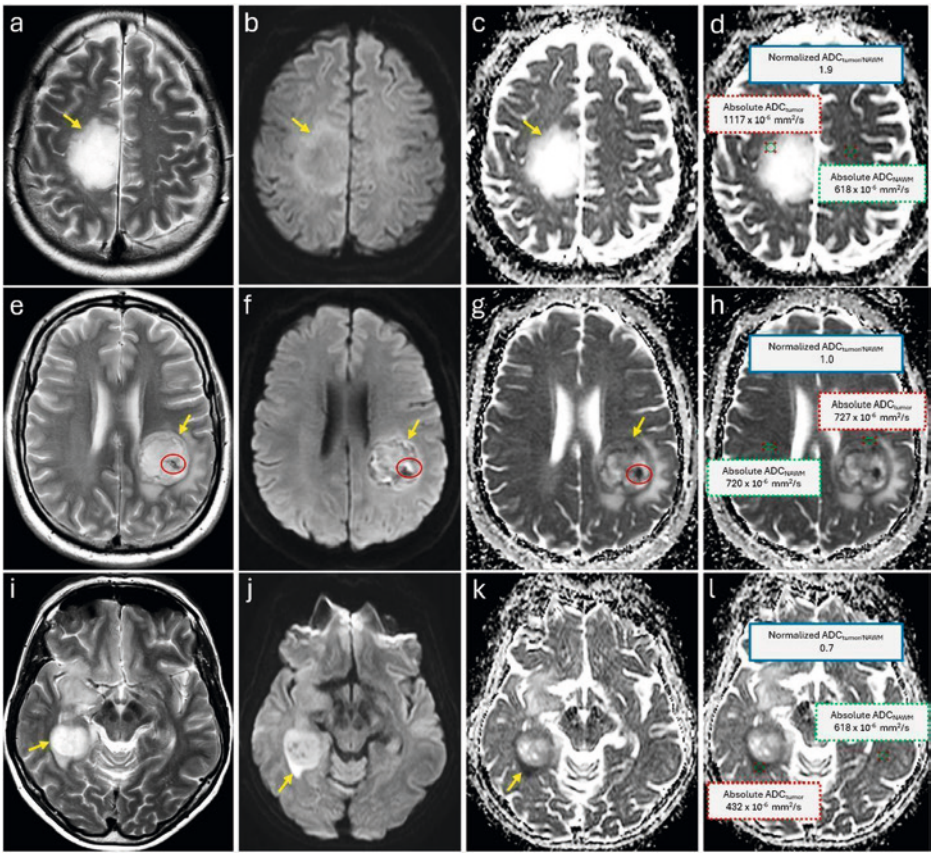
Caption: Supplementary Table 2 demonstrates the distribution of visual diffusion assessment classes for each rater across the first and second measurements

Supplementary Table 3. Distribution of absolute and normalized ADC per rater and measurement

Ratings per rater and measurements		Absolute ADC mean±SD x 10^{-6} mm^2/s	Normalized ADC mean±SD
Rater 1	measurement 1	977±353	1.26±0.48
	measurement 2	1,000±391	1.30±0.53
Rater 2	measurement 1	970±322	1.21±0.40
	measurement 2	965±339	1.21±0.43
Rater 3	measurement 1	882±312	1.15±0.41
	measurement 2	864±323	1.13±0.42

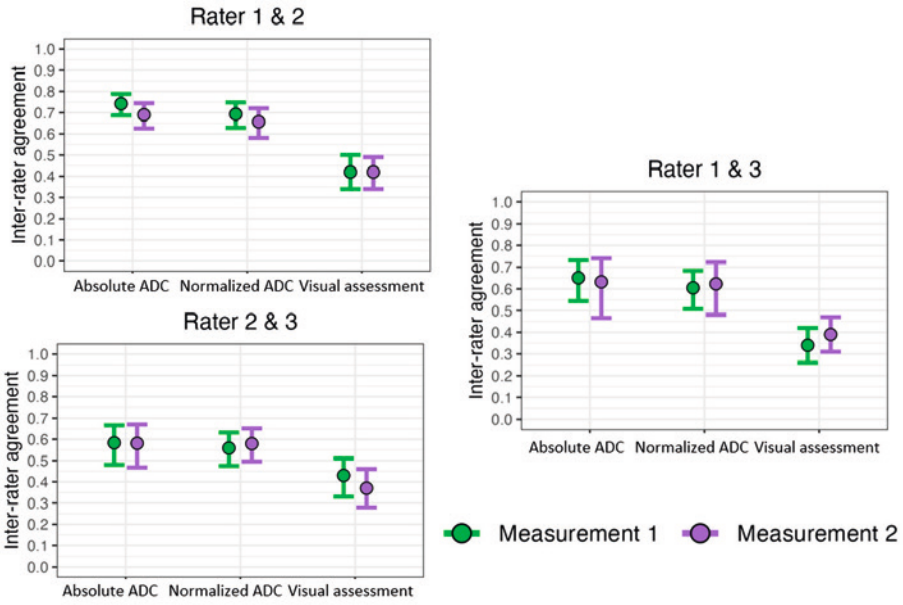
Caption: Supplementary Table 3 describes each rater's absolute and normalized ADC distribution across the first and second measurements.
Abbreviations: IDH = isocitrate dehydrogenase, SD = standard deviation

Supplementary Figures and Legends

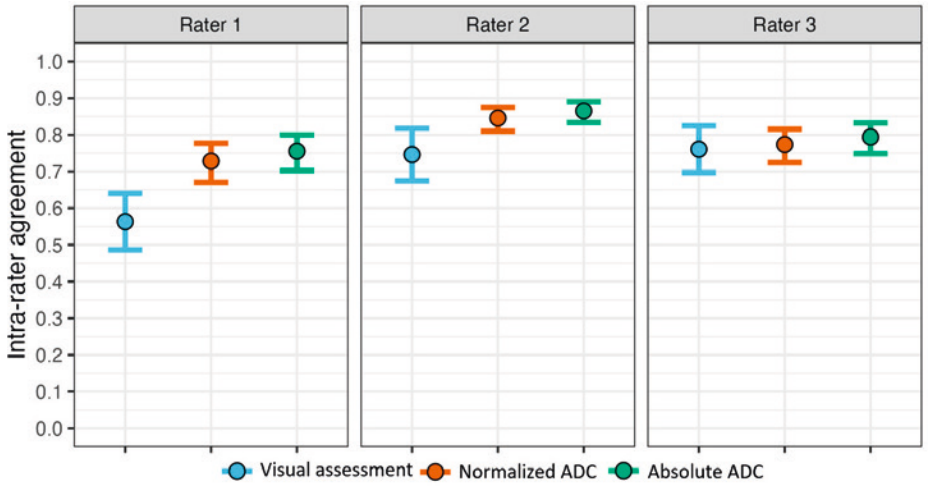


Supplementary Figure 1: Demonstrative cases depicting visual and ROI-based evaluation of diffusion in adult-type diffuse gliomas. Facilitated diffusion (a-d, isocitrate dehydrogenase (IDH)-wildtype glioblastoma): Right frontal T2-hyperintense lesion (a, yellow arrow) shows a peripheral hyperintense signal on b-1000 map of DWI (b, yellow arrow), which correlates with hyperintense ADC signal (c, yellow arrow) compared to normal-appearing brain tissue. Image d shows the ROI-based measurement of this visually defined area of the tumor (red text box, absolute ADC_{tumor} mm^2/s) and contralateral normal-appearing white matter (green text box, absolute ADC_{NAWM} mm^2/s) together with normalized $ADC_{tumor/NAWM}$ (blue text box). Dubious diffusion (e-h, IDH-wildtype, glioblastoma): Left parietal T2-hyperintense necrotic lesion (e, yellow arrow) shows a peripheral hyperintense signal on b-1000 map of DWI (f, yellow arrow), which correlates with isointense ADC signal (g, yellow arrow) resembling normal-appearing brain tissue. Red circles (e-g) show the area with restricted diffusion within the necrotic-hemorrhagic part of the tumor, which is excluded from the evaluation. Image h shows ROI-based measurements (text boxes). Restricted diffusion (i-l, IDH-wildtype, glioblastoma): Right temporal T2-hyperintense lesion (i,

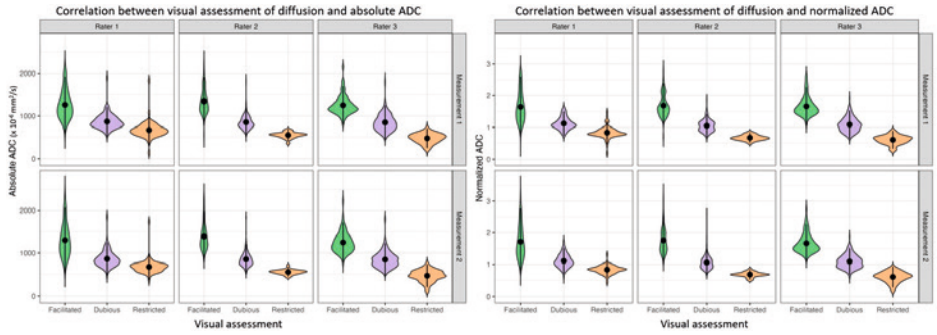
yellow arrow) shows a posterolateral hyperintense signal on b-1000 map of DWI (j, yellow arrow), which correlates with hypointense ADC signal (k, yellow arrow) compared to normal-appearing brain tissue. Image l shows ROI-based measurements (text boxes).



Supplementary Figure 2: Error bars show pairwise inter-rater agreements among different rater pairs across both measurements in evaluating visual and region-of-interest-based assessments of diffusion, including absolute and normalized ADC.



Supplementary Figure 3: Error bars show per-rater intra-rater inter-measurement agreements in evaluating visual and region-of-interest-based assessments of diffusion, including absolute and normalized ADC.



Supplementary Figure 4: Violin plots show per-rater and measurement correlation between visual and region-of-interest-based assessments of diffusion, including absolute and normalized ADC.

Chapter 7

Summary and Discussion

SUMMARY

Contrast-enhanced MRI using GBCAs is the standard practice for imaging brain tumors, including gliomas, due to its effectiveness in improving lesion visibility, aiding tumor characterization, and guiding therapeutic decisions. GBCAs accumulate in vital organs with unclear health consequences and contaminate water, as was shown for rivers and drinking water in the Netherlands, Germany and Poland, with a negative impact on aquatic life. Furthermore, GBCAs are substantial drivers of healthcare costs. Despite these concerns, no definitive evidence has demonstrated that GBCA-enhanced imaging is superior to GBCA-free approaches. Alternatively formulated, it is also unknown to what extent GBCA-free imaging protocols can be used to evaluate gliomas. Therefore, this thesis explores whether GBCA-free MRI protocols can reliably answer clinical questions for adult-type diffuse gliomas while maintaining diagnostic quality.

In Chapter 2, I assessed the current state and feasibility of GBCA-free or reduced-dose GBCA-enhanced imaging for gliomas and meningiomas through an extensive non-systematic literature review. This review showed that while most current guidelines recommend GBCAs at all stages of brain tumor management, some support GBCA-free monitoring, including Response Assessment in Pediatric Neuro-Oncology guidelines for nonenhancing pediatric low-grade gliomas and European Association of Neuro-Oncology and Danish guidelines for small, asymptomatic meningiomas. Evidence from two prospective trials suggests that reducing GBCA doses to 50-75% does not compromise diagnostic accuracy for gliomas and meningiomas. Over the last decade, interest in advanced GBCA-free techniques, including ASL, APT-CEST, and MRS, has grown, with increasing evidence supporting their comparable diagnostic performance to GBCA-enhanced methods for tumor grading and recurrence monitoring. For therapy planning, T2-FLAIR-guided approaches for glioblastoma resection and radiation therapy offer survival benefits without increasing neurologic risks. Artificial intelligence technologies, including deep learning models to generate synthetic contrast-enhanced images and radiomics from GBCA-free sequences, show great potential for replacing GBCAs, demonstrating equivalent performance in tumor grading and treatment response assessment. Overall, GBCA-free imaging is increasingly feasible for specific cases, particularly in pediatric low-grade tumors and small meningiomas. However, further studies

are essential to support its broader clinical adoption, especially for diffuse gliomas, the most common and malignant brain tumors in adults.

In Chapter 3, I conducted a systematic review and meta-analysis of Visually Accessible Rembrandt (Repository for Molecular Brain Neoplasia Data) Images (VASARI) features set, which consists of 30 glioma imaging descriptors. My motivation for this review was to evaluate the general potential of this standardized glioma imaging vocabulary, as the VASARI set primarily includes GBCA-free MRI features, which formed the foundation for the GBCA-free prediction algorithms described in Chapters 4 and 5. Analyzing data from 35 studies and 3,304 patients, I found that VASARI features were mainly used to predict overall survival and IDH mutation status. Multifocality, ependymal invasion, and enhancing tumor crossing midline were strong predictors of overall survival, with pooled hazard ratios of 1.80, 1.73, and 2.08, respectively. For IDH mutation prediction, models combining VASARI features achieved a pooled AUC of 0.82, although substantial heterogeneity across studies (I^2=100%) was noted. Enhancement quality and the proportion of enhancing tumor, necrosis, and edema emerged as key indicators of IDH status. Combined models integrating VASARI features with clinical, genomics, or radiomics data outperformed those relying solely on the VASARI set. However, variability in VASARI feature selection and modifications to scoring systems across studies limited the generalizability of the findings. These findings showed the need to develop automatic extraction methods to maximize the clinical utility of VASARI.

In Chapter 4, given the reliance on GBCA-enhanced MRI in neuroimaging, I aimed to determine whether key features associated with blood-brain barrier disruption, typically detected in contrast-enhanced imaging, could be identified from pre-contrast MRI using human visual analysis. To achieve this, I developed and tested an enhancement prediction decision tree (EPDT) to predict contrast enhancement quality and shape in adult-type diffuse gliomas using GBCA-free MRI sequences, including pre-contrast T1-weighted, T2-weighted, FLAIR, and DWI. The EPDT incorporated three VASARI (necrosis, diffusion restriction, non-enhancing tumor margins) and one non-VASARI (T2 inhomogeneity) glioma imaging features. Tested on 303 cases by three raters with varying experience levels, the EPDT achieved high accuracy for enhancement quality (marked, mild, no enhancement) with per-rater accuracies of 86%, 89%, and 92%, and for

enhancement shape (solid, rim, or no enhancement) with per-rater accuracies of 84%, 88%, and 89%. Substantial agreement was observed between predicted and true enhancement features within raters (≥0.68), while inter-rater reliability was moderate or better (group ≥0.42, pairwise ≥0.61). Necrosis was the most reproducible EPDT feature, achieving the highest intra- and inter-rater agreement (≥0.80). Agreement for other features was fair-to-moderate (≥0.33) and improved notably (≥0.71) after correcting for dataset imbalance. These findings demonstrate the EPDT's potential to accurately predict glioma enhancement features without relying on GBCAs or advanced expertise, which could support the advancement of AI-driven enhancement prediction techniques.

In Chapter 5, I designed a GBCA-free MRI-based diagnosis prediction decision tree (DPDT) to evaluate its effectiveness in preoperative glioma diagnosis and to compare it with GBCA-enhanced imaging. The DPDT incorporated seven VASARI (necrosis, diffusion, hemorrhage, non-enhancing tumor margin, calvarial remodeling, cysts, proportion of edema) and four non-VASARI (T2-FLAIR mismatch sign, T2 signal homogeneity, calcification, midline shift) imaging features. Three raters used a test dataset of 303 cases to predict tumor grade (grade 2 vs. grade 3/4) and molecular status (IDH and 1p/19q) from GBCA-free and GBCA-enhanced MRI. Per-rater GBCA-free predictions achieved ≥85% accuracy for tumor grade and ≥75% for molecular status, with no significant difference from GBCA-enhanced predictions, which achieved ≥87% and ≥77%, respectively. Group inter-rater agreements were moderate (0.56) and substantial (0.68) for GBCA-free and GBCA-enhanced predictions, respectively. These findings highlight the potential of GBCA-free imaging as a reliable alternative for glioma diagnosis, demonstrating that the added value of GBCA-enhanced MRI is limited and supporting the feasibility of a GBCA-free imaging approach in neuro-oncology.

In Chapter 6, I focused on two different assessment approaches of DWI (visual assessment versus region-of-interest (ROI)-based method), the most widely used advanced GBCA-free imaging technique and a crucial element in neuro-oncological MRI protocols. I evaluated the correlation between these approaches, their reproducibility, and their predictive diagnostic performance. For visual assessment, diffusion patterns were categorized using the VASARI-feature 17 classes (facilitated, dubious, restricted). ROI-based assessment placed

circular ROIs on the visually perceived areas with the lowest ADC, measuring absolute and normalized ADC (aADC/nADC). Using pre-operative MRI scans from 303 patients with grades 2-4 adult-type diffuse gliomas, ROI-based assessment demonstrated superior reproducibility, with higher inter- and intra-rater agreement (ROI-based ≥0.56 vs. visual ≥0.34). Despite this, a strong correlation was observed between ROI-based ADC thresholds (1,090 and 623 x 10^{-6} mm^2/s for aADC and 1.38 and 0.80 for nADC distinguishing facilitated, dubious and restricted visual diffusion categories) and visual assessment categories ($P < .001$). For IDH mutation prediction, cases without necrosis or hemorrhage were analyzed. Restricted diffusion in visual assessment was classified as IDH-wildtype, while dubious and facilitated diffusion indicated IDH-mutant. For the ROI-based method, optimal IDH classification thresholds were 1,048 x 10^{-6} mm^2/s and 1.38 for aADC and nADC, respectively. Visual assessment achieved comparable accuracy to the ROI-based method (visual vs. aADC/nADC: 69% vs. 73%/70%), and both methods showed an imbalance between specificity (99% vs. 81%/75%) and sensitivity (14% vs. 57%/61%). These findings show the equivalent clinical utility and strong correlation between visual and ROI-based assessments of DWI. Additionally, the results may guide the use of diffusion data in advanced AI-based predictive models, a field still underdeveloped.

DISCUSSION

This thesis is driven by growing concerns surrounding GBCAs, essential for contrast-enhanced MRI, as endorsed by current neuro-oncology guidelines[1,2]. While these contrast agents improve tumor visibility and guide clinical decisions, their health, environmental, and economic impacts cannot be overlooked. Gadolinium deposition in tissues, including the brain's deep grey matter, highlights the trade-offs of their widespread use, with unknown long-term consequences[3]. A shift from isolated cases to a pressing global issue, GBCA residues contaminate aquatic ecosystems, jeopardizing biodiversity and disrupting ecological balance[4–6]. Prolonged patient survival due to advancement in cancer care[7] further increases lifetime GBCA exposure, adding financial strain from the high costs of GBCAs, particularly in resource-limited settings[8]. Despite decades of use, a clear safety profile for vulnerable groups, such as pregnant women and children, remains lacking[9–12]. Moreover, GBCA injections cause discomfort for

patients, with a recent study by our research group showing that patients prefer avoiding GBCA injections if diagnostic accuracy is maintained[13].

While the use of GBCAs is well-established and critical in specific clinical scenarios, such as the detection of leptomeningeal disease or small metastasis, their universal necessity remains debatable. Evidence supporting their absolute requirement in all contexts is limited, and there is little to indicate that GBCA-free imaging compromises clinical decisions or that GBCA-enhanced image provides a significant advantage over alternatives in many routine neuroimaging applications. This thesis addresses this gap as part of the GLIOCARE project by evaluating a GBCA-free MRI approach for pre-operative glioma diagnosis. It explores whether conventional pre-contrast MRI, including DWI, can reliably answer key clinical questions, aiming to reshape neuro-oncology practice and promote sustainable, equitable care.

Standardizing glioma imaging features: the role of VASARI

Radiological evaluations to date rely on a visual analysis, referred to as "eye-balling" in clinical jargon. While standard practice, this approach is inherently subjective and requires standardized definitions and assessment criteria to ensure reproducibility and generalizability. Frameworks like BI-RADS for breast imaging[14] or PI-RADS for prostate imaging[15] exemplify the value of such standardization in improving diagnostic accuracy, guiding clinical management, and ultimately improving patient outcomes. Similarly, The Cancer Genome Atlas introduced the VASARI features set[16] nearly a decade ago to bring consistency and comparability to glioma imaging assessments.

The VASARI features set[16] comprises 30 imaging features, 25 of which are derived from GBCA-free MRI sequences, highlighting the wealth of information available from pre-contrast imaging. Initially designed for glial tumors, it has been used for other tumor types, including pediatric brain tumors[17] and ependymomas[18]. The VASARI set has demonstrated strong reproducibility, with high inter-rater agreement levels reported across studies. For example, Setyawan et al.[19] used the VASARI framework to predict glioma grade, IDH mutation, and MGMT methylation status, finding substantial agreement for all features and significant correlations with histomolecular characteristics. Their prediction models incorporated a range of VASARI features, with many of these being

GBCA-free imaging features, such as tumor location, eloquent brain, non-enhancing tumor margins, cyst, hemorrhage, T1/FLAIR ratio or diffusion characteristics. These qualities make the VASARI set an invaluable foundation for the GBCA-free MRI-based analysis described in Chapters 4, 5, and 6 of this thesis.

In Chapter 3, I conducted a meta-analysis of the VASARI features to identify their performance and clinical impact on glioma patients since their introduction a decade ago. The systematic search revealed its primary applications as pre-operative prognostication and IDH genotyping. Key prognostic features with the highest hazard ratios included multifocality (1.80), ependymal invasion (1.73), and enhancing tumor crossing midline (2.08). Features such as enhancement quality and the proportion of enhancing tumor, necrosis, and edema were commonly associated with IDH mutation status. However, methodological differences between studies prevented quantitative single-feature-based analysis for IDH genotyping. While a pooled analysis of multiple feature-based prediction models showed promise with an AUC of 0.82, it also showed considerable variability between studies, compromising comparability (I^2=100%).

Heterogeneity across studies, primarily due to some not evaluating all features and others modifying the scoring system, hindered reliable meta-analysis of the VASARI set. This limitation was likely caused by the time-consuming nature of extracting the complete VASARI set. Addressing this challenge, recent advances in AI offer promising solutions. A notable example is VASARI-auto, a machine learning-based tool developed in a recent study by Ruffle et al. [20]. This tool automatically extracted 15 VASARI features in 100 glioblastoma test cases, reducing analysis time from 317 seconds per case (manual extraction) to just 3 seconds, with inter-rater agreement comparable to that of human raters. A simulated economic analysis further stressed its impact, projecting that manual VASARI evaluations over three years would require 29,777 hours and cost over £1.5 million, whereas VASARI-auto could reduce these figures to 332 hours and £146. These findings establish VASARI-auto as a transformative tool for glioma imaging standardization, offering efficiency and cost-effectiveness. Future research should focus on extending the tool to cover the entire VASARI features set, enabling broader adoption in both clinical and research contexts.

In summary, the VASARI set appears to be a promising framework for standardizing glioma imaging, with growing research adoption and applications in clinical decision-making. However, in the context of this thesis's focus on GBCA-free imaging, the studies identified in this meta-analysis consistently combined GBCA-free and GBCA-enhanced imaging features, making it difficult to isolate the specific contributions of GBCA-enhanced imaging or to determine whether GBCA-free sequences alone are sufficient. I addressed this gap by evaluating the diagnostic accuracy and predictive power of GBCA-free imaging compared to GBCA-enhanced imaging in the following chapters.

Artificial intelligence or human intelligence for predicting glioma enhancement?

The shift from conventional practices in medical imaging necessitates demonstrating that alternative methods offer equivalent or superior diagnostic value while minimizing associated risks. As awareness grows around the side effects of GBCAs, the imaging field is actively pursuing viable alternatives. While existing research predominantly explores AI-based approaches to address this challenge, this thesis shifts the focus to human visual analysis- referred to here as "human intelligence"- the traditional method of evaluating radiological images, which remains underexplored in the context of glioma enhancement prediction. Such insights could not only validate the potential of human intelligence in predicting previously unseen tumor features but also contribute to the refinement of AI-based models by identifying specific GBCA-free imaging features that guide the decision-making process, eventually improving the performance of these models.

Deep learning-powered synthetic post-contrast imaging is among the most promising advancements in the AI field, as discussed in Chapter 2, which utilizes GBCA-free or low-dose GBCA-enhanced images as input[21]. For instance, Kleesiek et al.[22] reported excellent quantitative performance of synthetic contrast-enhanced maps generated from GBCA-free inputs, achieving sensitivity, specificity, and structural similarity index of 92%, 91%, and 87%, respectively, compared to standard post-contrast T1-weighted images. Another study[23] used 25%-dose contrast-enhanced images as input and demonstrated 96% sensitivity, 95% specificity, and 87% structural similarity index. Despite their high-quality outputs, implementing AI-generated synthetic contrast-enhanced imaging in

clinical settings presents significant challenges. These include limited access to annotated medical datasets, risks of algorithmic bias, ethical concerns about data privacy, and substantial infrastructural, legal, and financial investments required for broader adoption.[24] One potential solution to these challenges is the collaborative creation of larger, more diverse, and open-access annotated medical datasets. As part of the GLIOCARE project, we are preparing the in-house IMAGO glioma dataset for public release this year, with a cohort from this dataset used in my thesis.

In contrast, the potential of human raters to predict contrast enhancement using widely available non-contrast MRI sequences has received little attention. This straightforward and accessible approach could serve as a human benchmark for AI models, which remains essential to validate and improve these innovations. On the other hand, AI models are unlikely to receive regulatory approval to be used widely in clinical practice in the foreseeable future, making alternative methods using currently available resources necessary. In Chapter 4, I introduced an enhancement prediction decision tree for adult-type diffuse gliomas, offering a simple, easily implantable GBCA-free imaging alternative. This decision tree relies on four glioma imaging features, including necrosis, diffusion restriction, T2 inhomogeneity, and nonenhancing tumor margins (well-defined or ill-defined), to predict enhancement quality (marked, mild, or nonenhancing) and shape (rim, solid, and unenhanced). Developed using 69 glioma cases, the decision tree was tested on 303 cases to assess its robustness and generalizability, with predictions made by three raters of varying expertise. The raters achieved at least 86% accuracy for enhancement quality and 84% for enhancement shape, demonstrating the model's clinical applicability, even for less-trained readers. Furthermore, intra-rater agreements between predicted and true enhancement features were substantial, suggesting that glioma enhancement assessments mainly reflecting blood-brain barrier disruption do not rely heavily on post-contrast images. A comparative study[25] using a machine learning radiomic model based on T2-FLAIR images reported 98% sensitivity and 61% specificity for predicting binarized glioma enhancement quality (enhancing vs. nonenhancing). In comparison, when I applied the decision tree to binarized enhancement quality cases, sensitivity ranged from 68% to 90%, while specificity varied between 93% and 96% across raters, demonstrating comparable results despite the use of entirely different approaches.

The decision tree demonstrated varying performance across tumor subtypes, performing better for high-grade and IDH-wildtype gliomas compared to low-grade and IDH-mutant tumors. Similar findings were reported by Calabrese et al.,[26] who noted histology-dependent variations in synthetic contrast enhancement predictions with lower Dice scores for low-grade gliomas. Additionally, the decision tree often falsely classified the mild enhancement patterns, a challenge reported by another study[22] using deep learning-generated virtual contrast enhancement images. These limitations highlight the imperfections of enhancement prediction models, whether human or AI-driven, and show the need for further refinement. Despite these challenges, the decision tree proposed in this thesis offers an easily implementable solution for clinical use, especially when post-contrast image acquisition is impossible for various reasons. It may also accelerate advancements in virtual post-contrast image generation by defining the most informative imaging features for model input, laying the groundwork for future improvements in glioma enhancement predictions.

Predicting pre-operative glioma diagnosis using GBCA-free MRI

The diagnosis and management of brain tumors begin with MRI, which plays a pivotal role in distinguishing tumor types, including glioma subtypes, and guiding subsequent therapeutic decisions. Imaging findings help determine whether immediate intervention, such as biopsy and aggressive treatment, is required or if a less invasive, even watch-and-waiting strategy is more appropriate[27]. Contrast enhancement after GBCA administration is often associated with high-grade gliomas and poor prognosis. However, growing evidence suggests that reliance on enhancement features may lead to misclassification of gliomas, as enhancing tumors are not always high-grade[28], and non-enhancing tumors are not always low-grade[29]. This raises the critical question: how essential are GBCAs for accurate diagnosis? To address this, in Chapter 5, I evaluated the diagnostic power of GBCA-free MRI by directly comparing it to GBCA-enhanced MRI.

Using the same cohort described as in Chapter 4, I developed a diagnosis prediction decision tree based solely on pre-contrast MRI sequences (pre-contrast T1-weighted, T2-weighted, FLAIR, DWI). The decision tree incorporated seven VASARI (necrosis, diffusion, hemorrhage, non-enhancing tumor margin, calvarial remodeling, cysts, proportion of edema) and four non-VASARI (T2-

FLAIR mismatch sign, T2 signal homogeneity, calcification, midline shift) imaging features, all linked to histomolecular glioma diagnosis (IDH mutation, 1p/19q-codeletion status, and tumor grade). Three raters, with varying levels of experience, evaluated the cases twice in a randomized order: once using DPDT without post-contrast T1-weighted images and once with them. Across all raters, results showed no significant difference between GBCA-free and GBCA-enhanced predictions: GBCA-free predictions achieved ≥85% accuracy for the grade and ≥75% for molecular status, while GBCA-enhanced predictions reached ≥87% and ≥77%, respectively. These findings suggest that the added value of GBCAs may be less critical for accurate diagnosis than previously assumed. Furthermore, the implementation of the decision tree enables a systematic and standardized approach to imaging assessment, potentially improving diagnostic consistency, as less experienced raters achieved comparable performance to that of experienced radiologists.

A recent study by Yuan et al.[30] further supports these findings by demonstrating the potential of GBCA-free MRI in preoperative glioma characterization. Using a deep radiomics approach that combines deep learning with machine learning, they analyzed 206 treatment-naive cases of grade 2-4 adult gliomas to predict IDH mutation status without post-contrast images. Their model incorporated DTI, instead of DWI, alongside FLAIR, T2-weighted, and pre-contrast T1-weighted images, achieving an AUC of 0.85, with a sensitivity of 93% and a specificity of 57%. When DTI metrics were excluded, performance dropped to an AUC of 0.73, highlighting the importance of diffusion-related information. Although direct comparisons with GBCA-enhanced MRI were not provided, these findings, consistent with our results, reinforce the viability of GBCA-free MRI for glioma diagnosis. Moreover, the imaging features defined in our diagnosis prediction decision tree could provide valuable insights for researchers developing AI-based algorithms, with the potential to improve both their interpretability and diagnostic accuracy.

While this thesis focuses on conventional GBCA-free MRI for preoperative glioma assessments, advanced GBCA-free alternatives are being actively explored, showing promising diagnostic potential. Studies have demonstrated that these techniques can offer similar effectiveness to their GBCA-enhanced counterparts in preoperative glioma diagnosis. For example, a study[31] com-

paring GBCA-free ASL and GBCA-enhanced DSC perfusion MRI for molecular glioma diagnosis found comparable performance for detecting IDH (AUC 0.82 and 0.83) and TERT (AUC 0.70 and 0.81) mutations, though neither could differentiate 1p/19q-codeletion or MGMT promoter methylation. ASL also performed similarly to DSC-MRI in differentiating grade 2 gliomas from grade 3 and 4 gliomas (AUC 0.90)[32]. Another study[33] applying machine learning to ASL perfusion data achieved high performance in differentiating glioma subtypes, correctly identifying 80% of glioblastomas and 83% of astrocytomas.

APT-CEST, a relatively new advanced MRI technique providing insight into the tumor microenvironment, is another promising GBCA-free imaging alternative[34,35]. Studies have demonstrated its potential in glioma grading[36–39] and genotyping[40–42]. For example, Jiang et al.[43] found that IDH-wildtype grade 2 gliomas exhibited significantly higher APT-weighted signal intensity than IDH-mutant gliomas (AUC 0.89). Wu et al.[44] showed that APT-CEST effectively predicts IDH mutation status and glioma grade, with AUCs of 0.87 for IDH status and 0.86 for grade. A meta-analysis of 23 studies[45] found that APT-CEST reliably distinguishes low-grade from high-grade gliomas (pooled AUC 0.84). Furthermore, Hou et al.[39] found that combining APT-CEST with ASL improved diagnostic accuracy for glioma grading (AUC from 0.90 to 0.96) and IDH genotyping (AUC from 0.92 to 0.96).

Proton MR spectroscopy (MRS) also holds promise for pre-operative glioma assessment, particularly in identifying IDH-mutant gliomas by detecting 2-hydroxyglutarate (2HG) oncometabolite[34,46]. A meta-analysis by Suh et al.[47] reported that 2HG MRS could identify IDH-mutant gliomas with 95% pooled sensitivity and 91% pooled specificity. Despite its high accuracy, 2HG MRS has limited clinical application due to its technical requirements and the need for specialized expertise. As an alternative, detecting the choline/creatine ratio is a more practical option, being easier to implement in standard setups. For example, Zhao et al.[48] found that the choline/creatine ratio was significantly higher in IDH-wildtype gliomas. However, its sensitivity (67%) and specificity (58%) were lower than those of 2HG MRS, as reported in previous studies[49,50].

Despite these advances, challenges such as the need for standardized diagnostic criteria, cross-vendor reproducibility, and specialized training and ex-

pertise hinder widespread clinical adoption of the abovementioned advanced GBCA-free MRI alternatives[34,51]. However, the decision tree proposed in this thesis offers a simple GBCA-free alternative that relies on standardized MRI and does not demand advanced expertise. While further validation is necessary to establish its reliability as a clinical assessment tool, it holds potential for use in scenarios where patients cannot receive GBCA, providing radiologists with a practical diagnostic option in such cases and offering a human benchmark for AI-based models.

Diffusion-weighted imaging: visual versus ROI-based assessment

Among advanced GBCA-free imaging techniques, DWI stands out for its wider availability and integration into standardized brain tumor imaging protocols. DWI assesses the motion of water molecules within tissues, providing information about cellular density and structural organization. Two approaches are commonly used to analyze DWI ADC maps: the ROI-based method, which includes absolute and normalized ADC measurements, and visual assessment. Absolute ADC values reflect the diffusion properties within an ROI but are influenced by technical factors such as inter-scanner variability. Normalized ADC values, which ratio a tumor's absolute ADC to normal-appearing white matter, reduce variability across sequences and scanners. However, they depend on the accurate selection of the reference region and may limit direct comparison of absolute values. In contrast, visual assessment involves the subjective evaluation of ADC maps without performing any measurements. While subjective, visual assessment remains a standard clinical practice, as reflected in the VASARI framework, which categorizes diffusion into three classes: facilitated, dubious, and restricted. Despite being applied to various clinical questions, the relative merits of visual versus ROI-based methods remain unclear, as there is limited evidence comparing these approaches or demonstrating a clear correlation between them.

In Chapter 6, I compared visual and ROI-based DWI assessment methods to examine their correlation, reproducibility, and diagnostic prediction capacity in adult-type diffuse gliomas. Using a cohort of 303 cases described in Chapters 4 and 5, three raters evaluated DWI at two time points. Visual assessment was based on the VASARI-feature 17 criteria, and absolute and normalized ADC values were measured in the visually defined areas. The results provided

a nuanced perspective on both methods' reproducibility and clinical utility. The ROI-based method demonstrated superior reproducibility with moderate-to-almost-perfect inter- and intra-rater agreement (≥0.56 for absolute and normalized ADC vs. ≥0.34 for visual assessment), aligning with prior studies highlighting ROI-based methods' consistency[52–54] over visual assessments[55–58]. However, there was a strong correlation between visual assessment and ROI-based methods using ADC thresholds of 1,090 and 623 x 10^{-6} mm^2/s for absolute ADC and 1.38 and 0.80 for normalized ADC distinguishing visual classes of facilitated, dubious, and restricted diffusion. These thresholds offer guidance for clinical and research users, particularly in cases where visual assessment alone may be uncertain. Combining the reproducibility of ROI-based methods with the visual approach's time efficiency can help streamline decision-making. Moreover, these thresholds could play a role in automating VASARI feature 17 in future algorithms, advancing efforts to automate VASARI set extraction fully.

For IDH classification, I excluded cases with hemorrhage and necrosis, as these features are typically associated with IDH-wildtype tumors, reducing the clinical relevance and practicality of DWI assessment in such instances. In visual assessment, restricted diffusion was classified as IDH-wildtype, while dubious and facilitated diffusion as IDH-mutant, aligning with findings from Chapter 5. For the ROI-based method, optimal thresholds for IDH classification were determined as 1,048 x 10^{-6} mm^2/s and 1.38 for absolute ADC and normalized ADC. The visual assessment showed comparable (69%) accuracy to absolute ADC (73%) and normalized ADC (70%). Given the widespread use of visual assessment in daily clinical practice, our findings demonstrate its non-inferior accuracy for preoperative glioma evaluation. However, neither approach achieved a balance between specificity (visual vs. absolute ADC/normalized ADC: 99% vs. 81%/75%) and sensitivity (14% vs. 57%/61%). These findings are consistent with a related study [59], which also reported higher specificity for visual assessment (100 vs. 89%) but lower sensitivity (50% vs. 90%) compared to absolute ADC. While that study[59] focused on glioma grading rather than IDH classification and used a different method for visual and ADC assessments, the overall trend aligns. Additionally, the thresholds identified in our study align with previously reported ranges in the literature (900-1,200 x 10-6 mm^2/s for absolute ADC[52,60–64] and 1.28-1.60 for normalized ADC[52,63,64]) despite methodological differences across studies. This alignment reinforces the practical

value of these thresholds, particularly as a reference point for future clinical and research applications.

In addition to standard DWI and ADC mapping, there are alternative advanced diffusion techniques such as diffusion tensor imaging (DTI), diffusion kurtosis imaging (DKI), and intravoxel coherent motion (IVIM), providing deeper insights into tumor microstructure and physiology. DTI assesses the directionality of water diffusion, aiding preoperative planning by mapping white matter tracts to evaluate tumor infiltration[65]. It also provides quantitative measures, like fractional anisotropy, reflecting diffusion directionality and mean diffusivity, quantifying the overall diffusion, both associated with glioma grade and IDH mutation status [30,51,66,67]. For example, White et al.[66] found that fractional anisotropy metrics were significantly lower in low-grade gliomas compared to high-grade gliomas. DKI captures non-Gaussian diffusion behavior, offering insight into tissue complexity and heterogeneity, which can help differentiate tumor grades or molecular suptypes[68–72]. For instance, Bisdas et al.[70] used machine learning analyses of DKI, achieving 84% accuracy in predicting IDH status and 78% in glioma grading. IVIM separates diffusion and perfusion components, enabling simultaneous assessment of cellularity and microvascular density. A recent study by Yu et al. [73] explored the use of IVIM to predict IDH status, finding significant differences in IVIM parameters between IDH-wildtype and IDH-mutant tumors. Although these techniques hold promise for preoperative glioma characterization, their clinical adoption is hindered by longer acquisition times, motion sensitivity, and the need for standardized protocols.

In summary, this thesis highlights the importance of DWI, a GBCA-free advanced MRI technique, in glioma preoperative evaluation and demonstrates the need to explore and standardize the various assessment approaches. Visual assessment, a routine aspect of daily clinical practice, has been integral to both enhancement and diagnosis prediction decision trees described in Chapters 4 and 5. Chapter 6 showed that both visual and ROI-based assessment methods provide comparable diagnostic accuracy when considered in isolation, with a strong correlation between the two. However, the ROI-based method, which was found to be more reproducible, needs standardized thresholds for ADC values to be effectively integrated into clinical practice. Furthermore, future studies should investigate how these methods perform in conjunction with

other imaging features to better assess their overall comparability and diagnostic capability in clinical settings.

GBCA-free brain tumor imaging: feasible or fantasy?

My focus thus far has been on the preoperative applications of GBCA-free MRI, as detailed in Chapters 4, 5, and 6, which demonstrated promising findings. To fully understand its clinical impact, however, it is important to consider the potential of GBCA-free imaging across all stages of brain tumor management. In Chapter 2, I conducted a nonsystematic review of the current state-of-the-art in GBCA-free imaging, demonstrating that it is not only feasible for preoperative diagnosis but also holds promise for therapy planning and post-treatment follow-up.

GBCA-enhanced T1-weighted imaging remains the standard for preoperative therapy planning in enhancing gliomas, with resections often guided by the outer contrast-enhanced margin and a safety margin extending to T2-FLAIR hyperintense regions. T2-FLAIR imaging is typically used for nonenhancing gliomas. However, there is a tendency to include nonenhancing tumor and/or edema areas identified by T2-FLAIR imaging in enhancing gliomas, a strategy known as supramarginal resection, which has been associated with improved survival without compromising neurological outcomes. For instance, a study of 1229 patients found that including more FLAIR abnormalities in resection increased median survival from 15.5 months to 20.7 months[74]. FLAIR-guided radiation therapy has also shown promise, with one study reporting a median overall survival of 23 months using this method[75] versus 15 months in previous methods. Verburg et al.[76] studied glioma infiltration using multi-region biopsies from areas with and without imaging abnormalities to guide surgery and radiotherapy. Their findings revealed that, for enhancing gliomas, [18F]FET PET (AUC 0.76) or a combination of ADC and [18F]FET PET (AUC 0.89) outperformed GBCA-enhanced T1-weighted imaging (AUC 0.56). For non-enhancing gliomas, no imaging combination was superior to T2-FLAIR imaging (AUC 0.81 T2-FLAIR vs. 0.69 [18F]FET PET). These results, derived from limited sources, show the potential of GBCA-free therapy planning in gliomas, highlighting the need for further research to confirm its effectiveness and broaden its clinical application. Our ongoing research project, NEURAL-MRI, aims to identify the contrast-enhanced resectable tumor margins using GBCA-free MRI, comparing

the results to those of standard GBCA-enhanced T1-weighted imaging in 116 glioblastoma patients, with the hypothesis that GBCA-free MRI can accurately predict the surgical resection margin.

Regarding glioma follow-up, my literature review found no studies directly comparing conventional GBCA-free imaging with standard GBCA-enhanced imaging. However, a meta-analysis of 17 glioma studies indicated that DWI could accurately distinguish recurrent tumors from therapy-related changes (AUC 0.90), achieving a sensitivity of 82% and a specificity of 83%[77]. Other advanced GBCA-free alternatives, such as ASL[78–80] and APT-CEST[81,82], have also shown potential. For instance, Wang et al.[78] found that ASL performed comparably to GBCA-enhanced DSC perfusion MRI (accuracy: ASL 80%, DSC 83%) in distinguishing tumor recurrence from radiation-induced injury. Hou et al.[81] demonstrated that ASL (AUC 0.85) and APT-CEST (AUC 0.91) significantly differentiated tumor progression from treatment-related changes. Additionally, we recently conducted a meta-analysis of APT-CEST imaging to differentiate therapy-related changes from tumor progression or recurrence in brain tumors, including gliomas and brain metastasis[83]. This meta-analysis of 12 studies with 500 patients showed that APT-CEST performs well in gliomas, with a pooled sensitivity of 88% and specificity of 84%. However, we found significant inter-study heterogeneity in reported metrics, demonstrating the need for further validation. Building on these findings, our other ongoing research project aims to evaluate the role of conventional GBCA-free MRI in post-treatment follow-up, specifically comparing T2-FLAIR imaging to the standard post-GBCA T1-weighted imaging in identifying tumor progression and therapy-related changes.

Artificial intelligence (AI), including machine learning and deep learning, has been increasingly applied to glioma imaging to extract quantitative features beyond the scope of visual interpretation. However, limited evidence exists regarding the use of AI-powered algorithms for postoperative glioma assessments. Mammadov et al.[84] applied radiomics analysis to predict pseudoprogression in high-grade gliomas, with post-contrast T1-weighted images performing better (AUC 0.82) than pre-contrast T1-weighted images (AUC 0.65). However, another AI study using deep learning to differentiate tumor progression from pseudoprogression found that a GBCA-free model combining T2-FLAIR and DWI outperformed the post-contrast T1-weighted imaging-only model (AUC

0.57 vs. 0.80)[85]. Additionally, Jiang et al.[86] evaluated radiomic features from various pre- and post-contrast MRI sequences to differentiate treatment effects from tumor recurrence. They found that APT-CEST performed the best (accuracy 0.86), while among conventional MRI sequences, T2-FLAIR outperformed post-contrast T1-weighted images (accuracy 0.77 vs. 0.84), and T2-weighted images performed similarly to post-contrast T1-weighted images.

Finally, the effectiveness of fixed-interval imaging in glioma follow-up remains a topic of debate. As discussed in Chapter 2, several national and international guidelines, including the National Institutes of Health and Care Excellence (NICE), European Association of Neuro-Oncology (EANO), and Spanish guidelines, acknowledge insufficient evidence to determine optimal follow-up intervals. The NICE guidelines also highlight the potential drawbacks of frequent imaging, such as patient anxiety and costs. EANO recommends less frequent follow-up for stable low-grade gliomas, with additional MRI examinations triggered by new symptoms. Danish guidelines propose skipping early postoperative imaging (<48 hours) for nonenhancing gliomas due to the challenges of evaluating nonenhancing residual tumors and instead recommend assessing resection completeness only after 12 weeks. A meta-analysis by Thompson et al.[87] evaluated various imaging strategies for adult gliomas, including pre-specified interval imaging and symptomatic/triggered imaging. Still, the authors noted a need for more high-quality studies. They included only one study by Mrowczynski et al.[88], which assessed the impact of early postoperative MRI on overall survival in 125 glioblastoma patients and found that early postoperative MRI did not significantly influence survival outcomes.

The findings of this chapter highlight the potential of GBCA-free MRI as a feasible alternative for glioma management, with applications extending beyond preoperative diagnosis to therapy planning and post-treatment follow-up. GBCA-free techniques, such as T2-FLAIR imaging, ASL, and APT-CEST, show promise in identifying tumor margins, guiding resection strategies, and distinguishing tumor progression from treatment-related changes. AI-powered approaches are also being explored in this context, with promising results that could further improve diagnostic accuracy. However, significant gaps remain, primarily due to the limited and heterogeneous resources available, showing the need for further research. As discussed earlier, my ongoing projects focus

on evaluating the role of GBCA-free MRI in identifying resectable tumor margins and detecting tumor recurrence, with the goal of advancing the field.

Future perspectives

My thesis demonstrates that GBCA-free MRI techniques provide comparable performance to GBCA-enhanced methods for glioma imaging, presenting a viable alternative. However, to build on these findings, future research should address several key limitations and refine the proposed prediction models. The presented studies were retrospectively designed using a single-center dataset from Amsterdam UMC, limiting the findings' generalizability. Therefore, external validation across multiple centers with diverse patient populations and a range of raters is essential to improve the robustness and clinical applicability. To facilitate this, we are engaged in collaborative efforts with Nigerian researchers of the CAMERA network to explore the potential utilization of the Sub-Saharan glioma database. Moreover, patients enrolled prospectively in the GLIOCARE project will be assessed by both internal and external raters as a part of this comprehensive validation framework.

The decision tree models developed in this thesis were based on standardized brain MRI protocol that did not incorporate advanced imaging techniques such as perfusion-weighted imaging. Given the promising potential of ASL as a GBCA-free alternative to DSC-MRI, future studies should integrate ASL to refine predictive capabilities. Further incorporation of advanced imaging modalities, such as APT-CEST, MRS, and ROI-based assessments of DWI, will also improve model accuracy. The prospective phase of the GLIOCARE project, which includes ASL and APT-CEST imaging, will be a viable next step in addressing this limitation. Moreover, clinical factors such as patient age and Karnofsky performance score should also be integrated into the decision tree models for better diagnostic predictions.

In Chapter 6, the ROI-based assessment of DWI was guided by visual evaluation to reflect clinical practice, where radiologists often rely on visual cues to identify abnormal areas and complement them with quantitative measures. However, this introduces a potential collinearity bias regarding comparative assessments of these approaches. Future studies should incorporate automated or volumetric ADC map assessments and investigate their correlation with

visual evaluations to address this concern and better understand the value of different assessment approaches of glioma diffusion characteristics.

While the proposed models demonstrated potential in preoperative settings, their performance in more complex postoperative settings requires further investigation. Implementing these decision tree models in post-treatment situations is crucial for assessing their utility in managing glioma. However, post-treatment scenarios often introduce confounding factors, such as therapy-induced diffusion restriction or hemorrhage, imaging features of the prediction models in their current version. Such factors will necessitate algorithm modifications for accurate assessment.

Finally, refining the diagnostic workflow to distinguish gliomas from other tumors, such as lymphoma and metastasis, and tumor-like conditions, such as tumefactive demyelinating lesions, is critical for future research to optimize layered diagnostic decision-making. GBCA-free imaging techniques, such as DWI[89–92] and ASL[79,93,94], have shown promise in addressing these challenges. Integrating such techniques into a structured diagnostic framework can potentially improve GBCA-free diagnostic workflows.

Conclusions

This thesis provides evidence that a transition to GBCA-free MRI can be a viable alternative for managing patients with glial tumors. The proposed models based on GBCA-free imaging performed comparably to the current standard practice of GBCA-enhanced MRI in predicting glioma enhancement characteristics and histomolecular diagnosis. These results highlight the potential of GBCA-free approaches to improve patient safety and reduce costs while maintaining diagnostic integrity. Furthermore, the insights gained from this thesis could contribute to the future integration of AI-driven solutions into clinical practice by improving prediction accuracy through the systematic and structured introduction of key imaging features.

REFERENCES

1. Sanvito F, Kaufmann TJ, Cloughesy TF, et al. Standardized brain tumor imaging protocols for clinical trials: current recommendations and tips for integration. *Front Radiol* 2023;3:1267615.
2. Wen PY, van den Bent M, Youssef G, et al. RANO 2.0: Update to the Response Assessment in Neuro-Oncology Criteria for High- and Low-Grade Gliomas in Adults. *J Clin Oncol* 2023;41:5187–99.
3. Starekova J, Pirasteh A, Reeder SB. Update on Gadolinium Based Contrast Agent Safety, From the AJR Special Series on Contrast Media. *AJR Am J Roentgenol* https://doi.org/10.2214/AJR.23.30036.
4. Zhang Z, Jiang W, Gu T, et al. Anthropogenic gadolinium contaminations in the marine environment and its ecological implications. *Environ Pollut* 2024;359:124740.
5. Lenkinski RE, Rofsky NM. Contrast media-driven anthropogenic gadolinium: Knowns and unknowns. *Radiology* 2024;311:e240020.
6. Piarulli S, Riedel JA, Fossum FN, et al. Effects of gadolinium (Gd) and a Gd-based contrast agent (GBCA) on early life stages of zebrafish (Danio rerio). *Chemosphere* 2024;350:140950.
7. Samet JM, Bradley CJ. Long-term cancer survival: New insights from health professional cohorts. *JNCI Cancer Spectr* 2022;6.
8. Anazodo UC, Ng JJ, Ehiogu B, et al. A framework for advancing sustainable magnetic resonance imaging access in Africa. *NMR Biomed* 2023;36:e4846.
9. Proença F, Guerreiro C, Sá G, et al. Neuroimaging safety during pregnancy and lactation: a review. *Neuroradiology* 2021;63:837–45.
10. van Westrhenen A, Senders JT, Martin E, et al. Clinical challenges of glioma and pregnancy: a systematic review. *J Neurooncol* 2018;139:1–11.
11. Blumfield E, Moore MM, Drake MK, et al. Survey of gadolinium-based contrast agent utilization among the members of the Society for Pediatric Radiology: a Quality and Safety Committee report. *Pediatr Radiol* 2017;47:665–73.
12. Ouyang M, Bao L. Gadolinium contrast agent deposition in children. *J Magn Reson Imaging* https://doi.org/10.1002/jmri.29389.
13. Wamelink IJHG, Hempel HL, van de Giessen E, et al. The patients' experience of neuroimaging of primary brain tumors: a cross-sectional survey study. *J Neurooncol* 2023;162:307–15.
14. Breast Imaging Reporting & Data System.

https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

15. PI-RADS. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS
16. VASARI research project - the cancer imaging archive (TCIA) public access - cancer imaging archive wiki.
17. Biswas A, Amirabadi A, Wagner MW, et al. Features of Visually AcceSAble Rembrandt Images: Interrater reliability in pediatric brain tumors. *AJNR Am J Neuroradiol* 2022;43:304–8.
18. Li L, Fu Y, Zhang Y, et al. Magnetic resonance imaging findings of intracranial extraventricular ependymoma: A retrospective multi-center cohort study of 114 cases. *Cancer Med* 2023;12:16195–206.
19. Setyawan NH, Choridah L, Nugroho HA, et al. Beyond invasive biopsies: using VASARI MRI features to predict grade and molecular parameters in gliomas. *Cancer Imaging* 2024;24:3.
20. Ruffle JK, Mohinta S, Baruteau KP, et al. VASARI-auto: Equitable, efficient, and economical featurisation of glioma MRI. *NeuroImage Clin* 2024;44:103668.
21. Moya-Sáez E, de Luis-García R, Alberola-López C. Toward deep learning replacement of gadolinium in neuro-oncology: A review of contrast-enhanced synthetic MRI. *Front Neuroimaging* 2023;2:1055463.

22. Kleesiek J, Morshuis JN, Isensee F, et al. Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?: A Feasibility Study. *Invest Radiol* 2019;54:653–60.

23. Ammari S, Bône A, Balleyguier C, et al. Can deep learning replace gadolinium in neuro-oncology?: A reader study. *Invest Radiol* 2022;57:99–107.

24. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ* 2023;23:689.

25. He H, Guo E, Meng W, et al. [Predicting cerebral glioma enhancement pattern using a machine learning-based magnetic resonance imaging radiomics model]. *Nan Fang Yi Ke Da Xue Xue Bao* 2024;44:194–200.

26. Calabrese E, Rudie JD, Rauschecker AM, et al. Feasibility of Simulated Postcontrast MRI of Glioblastomas and Lower-Grade Gliomas by Using Three-dimensional Fully Convolutional Neural Networks. *Radiol Artif Intell* 2021;3:e200276.

27. Weller M, van den Bent M, Preusser M, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol* 2021;18:170–86.

28. Eichberg DG, Di L, Morell AA, et al. Incidence of high grade gliomas presenting as radiographically non-enhancing lesions: experience in 111 surgically treated non-enhancing gliomas with tissue diagnosis. *J Neurooncol* 2020;147:671–9.

29. Castet F, Alanya E, Vidal N. Contrast-enhancement in supra- tentorial low-grade gliomas: a classic prognostic factor in the molecular age. *J Neurooncol* 2019;143:515–23.

30. Yuan J, Siakallis L, Li HB, et al. Structural- and DTI- MRI enable automated prediction of IDH Mutation Status in CNS WHO Grade 2-4 glioma patients: a deep Radiomics Approach. *BMC Med Imaging* 2024;24:104.

31. Prysiazhniuk Y, Server A, Leske H, et al. Diffuse glioma molecular profiling with arterial spin labeling and dynamic susceptibility contrast perfusion MRI: A comparative study. *Neurooncol Adv* 2024;6:vdae113.

32. Qu Y, Kong D, Wen H, 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. *Eur Radiol* 2022;32:2976–87.

33. Amador K, Kniep H, Fiehler J, et al. Evaluation of an image-based classification model to identify glioma subtypes using Arterial Spin Labeling perfusion MRI on the publicly available UCSF glioma dataset. *Clin Neuroradiol* https://doi.org/10.1007/s00062-024-01465-5.

34. Hangel G, Schmitz-Abecassis B, Sollmann N, et al. Advanced MR Techniques for Preoperative Glioma Characterization: Part 2. *J Magn Reson Imaging* 2023;57:1676–95.

35. Deng H-Z, Zhang H-W, Huang B, et al. Advances in diffuse glioma assessment: preoperative and postoperative applications of chemical exchange saturation transfer. *Front Neurosci* 2024;18:1424316.

36. Zhang J, Zhu W, Tain R, et al. Improved differentiation of low-grade and high-grade gliomas and detection of tumor proliferation using APT contrast fitted from Z-spectrum. *Mol Imaging Biol* 2018;20:623–31.

37. Suh CH, Park JE, Jung SC, et al. Amide proton transfer-weighted MRI in distinguishing high- and low-grade gliomas: a systematic review and meta-analysis. *Neuroradiology* 2019;61:525–34.

38. Kang X-W, Xi Y-B, Liu T-T, et al. Grading of Glioma: combined diagnostic value of amide proton transfer weighted, arterial spin labeling and diffusion weighted magnetic resonance imaging. *BMC Med Imaging* 2020;20:50.

39. Hou H, Chen W, Diao Y, et al. 3D amide proton transfer-weighted imaging for grading glioma and correlating IDH mutation status: Added value to 3D pseudocontinuous arterial spin labelling perfusion. *Mol Imaging Biol* 2023;25:343–52.

40. Jiang S, Wen Z, Ahn SS, et al. Applications of chemical exchange saturation transfer magnetic resonance imaging in identifying genetic markers in gliomas. *NMR Biomed* 2023;36:e4731.

41. Mancini L, Casagranda S, Gautier G, et al. CEST MRI provides amide/amine surrogate biomarkers for treatment-naïve glioma sub-typing. *Eur J Nucl Med Mol Imaging* 2022;49:2377–91.

42. Joo B, Han K, Ahn SS, et al. Amide proton transfer imaging might predict survival and IDH mutation status in high-grade glioma. *Eur Radiol* 2019;29:6643–52.

43. Jiang S, Zou T, Eberhart CG, et al. Predicting IDH mutation status in grade II gliomas using amide proton transfer-weighted (APTw) MRI. *Magn Reson Med* 2017;78:1100–9.

44. Wu M, Jiang T, Guo M, et al. Amide proton transfer-weighted imaging and derived radiomics in the classification of adult-type diffuse gliomas. *Eur Radiol* 2024;34:2986–96.

45. Sotirios B, Demetriou E, Topriceanu CC, et al. The role of APT imaging in gliomas grading: A systematic review and meta-analysis. *Eur J Radiol* 2020;133:109353.

46. Di Ieva A, Magnussen JS, McIntosh J, et al. Magnetic resonance spectroscopic assessment of isocitrate dehydrogenase status in gliomas: The new frontiers of spectrobiopsy in neurodiagnostics. *World Neurosurg* 2020;133:e421–7.

47. Suh CH, Kim HS, Jung SC, et al. 2-Hydroxyglutarate MR spectroscopy for prediction of isocitrate dehydrogenase mutant glioma: a systemic review and meta-analysis using individual patient data. *Neuro Oncol* 2018;20:1573–83.

48. Zhao D, Duan L, Juratli TA, et al. Comprehensive multicentre retrospective analysis for predicting isocitrate dehydrogenase-mutant lower-grade gliomas. *Ann Clin Transl Neurol* https://doi.org/10.1002/acn3.52251.

49. Choi C, Raisanen JM, Ganji SK, et al. Prospective longitudinal analysis of 2-hydroxyglutarate magnetic resonance spectroscopy identifies broad clinical utility for the management of patients with IDH-mutant glioma. *J Clin Oncol* 2016;34:4030–9.

50. Tietze A, Choi C, Mickey B, et al. Noninvasive assessment of isocitrate dehydrogenase mutation status in cerebral gliomas by magnetic resonance spectroscopy in a clinical setting. *J Neurosurg* 2018;128:391–8.

51. Hirschler L, Sollmann N, Schmitz-Abecassis B, et al. Advanced MR Techniques for Preoperative Glioma Characterization: Part 1. *J Magn Reson Imaging* 2023;57:1655–75.

52. Ma X, Cheng K, Cheng G, et al. Apparent Diffusion Coefficient as Imaging Biomarker for Identifying IDH Mutation, 1p19q Codeletion, and MGMT Promoter Methylation Status in Patients With Glioma. *J Magn Reson Imaging* 2023;58:732–8.

53. Maynard J, Okuchi S, Wastling S, et al. World Health Organization grade II/III glioma molecular status: Prediction by MRI morphologic features and apparent diffusion coefficient. *Radiology* 2020;296:111–21.

54. Hu Y-C, Yan L-F, Sun Q, et al. Comparison between ultra-high and conventional mono b-value DWI for preoperative glioma grading. *Oncotarget* 2017;8:37884–95.

55. Park YW, Han K, Ahn SS, et al. Prediction of IDH1-Mutation and 1p/19q-Codeletion Status Using Preoperative MR Imaging Phenotypes in Lower Grade Gliomas. *AJNR Am J Neuroradiol* 2018;39:37–42.

56. Zhou H, Vallières M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol* 2017;19:862–70.

57. Hyare H, Rice L, Thust S, et al. Modelling MR and clinical features in grade II/III astrocytomas to predict IDH mutation status. *Eur J Radiol* 2019;114:120–7.

58. Su C-Q, Lu S-S, Han Q-Y, et al. Intergrating conventional MRI, texture analysis of dynamic contrast-enhanced MRI, and susceptibility weighted imaging for glioma grading. *Acta radiol* 2019;60:777–87.

59. Phuttharak W, Thammaroj J, Wara-Asawapati S, et al. Grading Gliomas Capability: Comparison between Visual Assessment and Apparent Diffusion Coefficient (ADC) Value Measurement on Diffusion-Weighted Imaging (DWI). *Asian Pac J Cancer Prev* 2020;21:385–90.

60. Villanueva-Meyer JE, Wood MD, Choi BS, et al. MRI features and IDH mutational status of grade II diffuse gliomas: Impact on diagnosis and prognosis. *AJR Am J Roentgenol* 2018;210:621–8.

61. Xing Z, Yang X, She D, et al. Noninvasive assessment of IDH mutational status in World Health Organization grade II and III astrocytomas using DWI and DSC-PWI combined with conventional MR imaging. *AJNR Am J Neuroradiol* 2017;38:1138–44.
62. Wasserman JK, Nicholas G, Yaworski R, et al. Radiological and pathological features associated with IDH1-R132H mutation status and early mortality in newly diagnosed anaplastic astrocytic tumours. *PLoS One* 2015;10:e0123890.
63. Thust SC, Maynard JA, Benenati M, et al. Regional and volumetric parameters for diffusion-weighted WHO grade II and III glioma genotyping: A method comparison. *AJNR Am J Neuroradiol* 2021;42:441–7.
64. Liu S, Zhang Y, Kong Z, et al. Feasibility of evaluating the histologic and genetic subtypes of WHO grade II-IV gliomas by diffusion-weighted imaging. *BMC Neurosci* 2022;23:72.
65. Li Y, Guo J, Zhang K, et al. Diffusion tensor imaging versus intraoperative subcortical mapping for glioma resection: a systematic review and meta-analysis. *Neurosurg Rev* 2023;46:154.
66. White ML, Zhang Y, Yu F, et al. Diffusion tensor MR imaging of cerebral gliomas: evaluating fractional anisotropy characteristics. *AJNR Am J Neuroradiol* 2011;32:374–81.
67. Luks TL, McKnight TR, Jalbert LE, et al. Relationship of in vivo MR parameters to histopathological and molecular characteristics of newly diagnosed, nonenhancing lower-grade gliomas. *Transl Oncol* 2018;11:941–9.
68. Raab P, Hattingen E, Franz K, et al. Cerebral gliomas: diffusional kurtosis imaging analysis of microstructural differences. *Radiology* 2010;254:876–81.
69. Van Cauter S, De Keyzer F, Sima DM, et al. Integrating diffusion kurtosis imaging, dynamic susceptibility-weighted contrast-enhanced MRI, and short echo time chemical shift imaging for grading gliomas. *Neuro Oncol* 2014;16:1010–21.
70. Bisdas S, Shen H, Thust S, et al. Texture analysis- and support vector machine-assisted diffusional kurtosis imaging may allow in vivo gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci Rep* 2018;8:6108.
71. Zerweck L, Hauser T-K, Klose U, et al. Glioma type prediction with dynamic contrast-enhanced MR imaging and diffusion kurtosis imaging-A standardized multicenter study. *Cancers (Basel)* 2024;16:2644.
72. Pang H, Dang X, Ren Y, et al. DKI can distinguish high-grade gliomas from IDH1-mutant low-grade gliomas and correlate with their different nuclear-to-cytoplasm ratio: a localized biopsy-based study. *Eur Radiol* 2024;34:7539–51.
73. Yu M, Ge Y, Wang Z, et al. The diagnostic efficiency of integration of 2HG MRS and IVIM versus individual parameters for predicting IDH mutation status in gliomas in clinical scenarios: A retrospective study. *J Neurooncol* 2024;167:305–13.
74. Li YM, Suki D, Hess K, et al. The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection? *J Neurosurg* 2016;124:977–88.
75. Duma CM, Kim BS, Chen PV, 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. *J Neurosurg* 2016;125:40–9.
76. Verburg N, Koopman T, Yaqub MM, et al. Improved detection of diffuse glioma infiltration with imaging combinations: a diagnostic accuracy study. *Neuro Oncol* 2020;22:412–22.
77. Du X, He Q, Zhang B, et al. Diagnostic accuracy of diffusion-weighted imaging in differentiating glioma recurrence from posttreatment-related changes: a meta-analysis. *Expert Rev Anticancer Ther* 2022;22:123–30.
78. Wang Y-L, Chen S, Xiao H-F, et al. Differentiation between radiation-induced brain injury and glioma recurrence using 3D pCASL and dynamic susceptibility contrast-enhanced perfusion-weighted imaging. *Radiother Oncol* 2018;129:68–74.

79. Lavrova A, Teunissen WHT, Warnert EAH, et al. Diagnostic accuracy of arterial spin labeling in comparison with dynamic susceptibility contrast-enhanced perfusion for brain tumor surveillance at 3T MRI. *Front Oncol* 2022;12:849657.

80. Seeger A, Braun C, Skardelly M, et al. Comparison of three different MR perfusion techniques and MR spectroscopy for multiparametric assessment in distinguishing recurrent high-grade gliomas from stable disease. *Acad Radiol* 2013;20:1557–65.

81. Hou H, Diao Y, Yu J, 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. *NMR Biomed* 2023;36:e4821.

82. Karimian-Jazi K, Enbergs N, Golubtsov E, et al. Differentiating glioma recurrence and pseudoprogression by APTw CEST MRI. *Invest Radiol* https://doi.org/10.1097/RLI.0000000000001145.

83. Essed RA, Prysiazhniuk Y, Wamelink IJ, et al. Performance of amide proton transfer imaging to differentiate true progression from therapy-related changes in gliomas and metastases. *Eur Radiol* https://doi.org/10.1007/s00330-024-11004-y.

84. Mammadov O, Akkurt BH, Musigmann M, et al. Radiomics for pseudoprogression prediction in high grade gliomas: added value of MR contrast agent. *Heliyon* 2022;8:e10023.

85. Bacchi S, Zerner T, Dongas J, et al. Deep learning in the detection of high-grade glioma recurrence using multiple MRI sequences: A pilot study. *J Clin Neurosci* 2019;70:11–3.

86. Jiang S, Guo P, Heo H-Y, et al. Radiomics analysis of amide proton transfer-weighted and structural MR images for treatment response assessment in malignant gliomas. *NMR Biomed* 2023;36:e4824.

87. Thompson G, Lawrie TA, Kernohan A, et al. Interval brain imaging for adults with cerebral glioma. *Cochrane Database Syst Rev* 2019;12:CD013137.

88. Mrowczynski OD, Zammar S, Bourcier AJ, et al. Utility of early postoperative magnetic resonance imaging after glioblastoma resection: Implications on patient survival. *World Neurosurg* 2018;120:e1171–4.

89. Eisenhut F, Schmidt MA, Putz F, et al. Classification of primary cerebral lymphoma and glioblastoma featuring dynamic susceptibility contrast and Apparent diffusion coefficient. *Brain Sci* 2020;10:886.

90. Wen J-B, Huang W-Y, Xu W-X-Z, et al. Differentiating primary central nervous system lymphomas from glioblastomas and inflammatory demyelinating pseudotumor using relative minimum apparent diffusion coefficients. *J Comput Assist Tomogr* 2017;41:904–9.

91. Doskaliyev A, Yamasaki F, Ohtaki M, et al. Lymphomas and glioblastomas: differences in the apparent diffusion coefficient evaluated with high b-value diffusion-weighted magnetic resonance imaging at 3T. *Eur J Radiol* 2012;81:339–44.

92. Ribas GA, de Mori LH, Freddi T de AL, et al. Primary central nervous system lymphoma: Imaging features and differential diagnosis. *Neuroradiol J* 2024;37:705–22.

93. Yu X, Hong W, Ye M, et al. Atypical primary central nervous system lymphoma and glioblastoma: multiparametric differentiation based on non-enhancing volume, apparent diffusion coefficient, and arterial spin labeling. *Eur Radiol* 2023;33:5357–67.

94. You S-H, Yun TJ, Choi HJ, et al. Differentiation between primary CNS lymphoma and glioblastoma: qualitative and quantitative analysis using arterial spin labeling MR imaging. *Eur Radiol* 2018;28:3801–10.

APPENDICES

NEDERLANDSE SAMENVATTING

Voor MRI beeldvorming van gliomen (een type hersentumoren) worden gadolinium houdende contrastmiddelen (GHCs) standaard in de praktijk gebruikt. GHCs verbeteren de zichtbaarheid van laesies, helpen bij het karakteriseren van tumoren en bepalen therapeutische beslissingen. Echter, stapelen GHCs zich in vitale organen op met onduidelijke gevolgen voor de gezondheid. Daarnaast vervuilen ze water, zoals is aangetoond in rivieren en drinkwater in Nederland. Het waterleven wordt hierdoor negatief beïnvloed. Bovendien zijn GHC's een aanzienlijke kostenveroorzaker in de gezondheidszorg. Ondanks deze zorgen is er geen absoluut bewijs dat beeldvorming met GHC betere resultaten heeft dan GHC-vrije alternatieven. Anders geformuleerd, is het ook onbekend in hoeverre GHC-vrije beeldvormingsprotocollen kunnen worden gebruikt om gliomen te evalueren. Daarom wordt in dit proefschrift onderzocht of GHC-vrije MRI-protocollen op betrouwbare wijze klinische vragen kunnen beantwoorden bij adult-type diffuse gliomen met behoud van diagnostische kwaliteit.

Door middel van een uitgebreid niet-systematisch literatuuronderzoek heb ik in hoofdstuk 2 de huidige stand van zaken en de haalbaarheid van GHC-vrije of gereduceerde GHC beeldvorming onderzocht voor gliomen en meningiomen. Uit dit overzicht bleek dat hoewel de meeste richtlijnen GHCs aanbevelen in alle stadia van de behandeling van hersentumoren, sommige richtlijnen GHC-vrije monitoring ondersteunen. Onder andere de Response Assessment in Pediatric Neuro-Oncology-richtlijnen voor niet-aankleurende pediatrische laaggradige gliomen, de richtlijnen van de European Association of Neuro-Oncology en 'de Danish richtlijnen voor kleine, asymptomatische meningiomen, ondersteunen GHC-vrije monitoring. Bewijs uit twee prospectieve onderzoeken suggereert dat het verlagen van de GHC-doses tot 50-75% de diagnostische nauwkeurigheid voor gliomen en meningiomen niet verandert. In de afgelopen tien jaar is de belangstelling voor geavanceerde GHC-vrije technieken, zoals ASL, APT-CEST en MRS gegroeid. Steeds meer bewijs laat zien dat de diagnostische prestaties van deze geavanceerde MRI technieken (op het gebied van tumor stadiëring en recidief monitoring) vergelijkbaar zijn met die van GHC-verrijkte methoden. T2-FLAIR-geleide benaderingen bieden overlevingsvoordelen bij het plannen van het chirurgisch verwijderen van glioblastomen en bij bestralingstherapie. Hierbij is er geen grotere kans op oncologische risico's dan wanneer planning

gedaan wordt met GHC beeldvorming. Kunstmatige intelligentie, zoals deep learning modellen die synthetische contrast verrijkte beelden kunnen genereren en radiomic features die uit GHC-vrije sequenties gehaald kunnen worden, hebben veel potentie om GHC's te vervangen. Deze technieken laten gelijkwaardige prestaties zien in tumorgradering en beoordeling van de behandelrespons. Over het geheel genomen is GHC-vrije beeldvorming steeds beter haalbaar voor specifieke gevallen, met name bij pediatrische laaggradige tumoren en kleine meningiomen. Verder onderzoek is echter essentieel om een bredere klinische toepasbaarheid te ondersteunen, met name voor diffuse gliomen, de meest voorkomende en kwaadaardigste hersentumoren bij volwassenen.

In hoofdstuk 3 heb ik een systematische review en meta-analyse uitgevoerd van de Visually Accessible Rembrandt (Repository for Molecular Brain Neoplasia Data) Images (VASARI) feature set, welke bestaat uit 30 glioom beeld-eigenschappen (image features). Mijn motivatie voor deze review was om de potentie van deze gestandaardiseerde glioom-beeldvormingsvocabulaire te evalueren. De VASARI-set bevat namelijk voornamelijk GHC-vrije MRI-kenmerken, die ook de basis zullen vormen voor de GHC-vrije voorspellingsalgoritmen die in hoofdstuk 4 en 5 worden beschreven. Bij het analyseren van gegevens van 35 onderzoeken en 3.304 patiënten vond ik dat VASARI-kenmerken voornamelijk werden gebruikt om de algehele overleving en IDH-mutatiestatus te voorspellen. Multifocaliteit, ependymale invasie en of een aankleurende tumor de de middellijn overschrijdt waren sterke voorspellers van algehele overleving, met gepoolde hazard ratio's van respectievelijk 1,80, 1,73 en 2,08. Voor de voorspelling van IDH-mutatie bereikten modellen die VASARI-kenmerken combineerden een gepoolde AUC van 0,82, hoewel er aanzienlijke heterogeniteit tussen onderzoeken (I^2=100%) werd waargenomen. De kwaliteit van aankleuring en de verhouding tussen aankleurende tumor, necrose en oedeem kwamen naar voren als de belangrijkste indicatoren van IDH-status. Gecombineerde modellen die VASARI kenmerken integreerden met klinische, genetische of radiomics eigenschappen presteerden beter dan modellen die alleen op de VASARI set vertrouwden. De generaliseerbaarheid van de bevindingen werd echter beperkt door de variabiliteit in VASARI feature selectie en wijzigingen in scoringsystemen tussen de verschillende onderzoeken. Deze bevindingen toonden de noodzaak om automatische extractiemethoden te ontwikkelen om de klinische bruikbaarheid van VASARI te maximaliseren.

In hoofdstuk 4 wilde ik, gezien de klinische afhankelijkheid van GHC-verrijkte MRI in de neurobeeldvorming, bepalen of visuele kenmerken die geassocieerd worden met een verstoring van de bloed-hersenbarrière geïdentificeerd kunnen worden uit pre-contrast MRI. Om dit te bereiken heb ik een enhancement prediction decision tree (EPDT) ontwikkeld en getest. Deze beslisboom kan door middel van GHC-vrije MRI sequenties contrast aankleuring en vorm voorspellen in adult-type diffuse gliomen. De MRI sequenties die de beslisboom gebruikt zijn pre-contrast T1, T2, FLAIR en DWI. De EPDT bevatte drie VASARI (necrose, diffusiebeperking, niet-aankleurende tumorranden) en één niet-VASARI (T2 inhomogeniteit) glioom MRI eigenschappen. Het EPDT, dat door 3 beoordelaars met verschillende ervaringsniveaus was getest op 303 patiënten, behaalde een hoge nauwkeurigheid voor kwaliteit van aankleuring (duidelijk aanwezig, gemiddeld, geen aankleuring) met een nauwkeurigheid per beoordelaar van 86%, 89% en 92%, en voor de vorm van aankleuring (vast, rand, of geen aankleuring) met een nauwkeurigheid per beoordelaar van 84%, 88% en 89%. Er werd aanzienlijke overeenstemming gevonden tussen voorspelde en werkelijke aankleuringseigenschappen binnen beoordelaars (≥0,68), terwijl de betrouwbaarheid tussen beoordelaars matig of beter was (groep ≥0,42, paarsgewijs ≥0,61). Necrose was de meest reproduceerbare EPDT eigenschap, met de hoogste overeenkomst binnen en tussen beoordelaars. (≥0,80). Overeenstemming voor andere beeld eigenschappen was redelijk tot matig (≥0,33) en verbeterde aanzienlijk (≥0,71) na correctie voor een onevenredig verdeelde dataset. Deze bevindingen tonen aan dat de EPDT in staat is om nauwkeurig glioma aankleuring te voorspellen zonder afhankelijk te zijn van GHCs of geavanceerde expertise, wat de AI-gedreven aankleuringsvoorspel modellen mogelijk zou kunnen ondersteunen.

In hoofdstuk 5 heb ik een beslisboom voor diagnosevoorspelling (DPDT) ontworpen dat preoperatieve gliomen kan diagnosticeren op basis van GHC-vrije MRI. Deze beslisboom is vervolgens ook vergeleken met diagnosticeren op basis van GHC-verrijkte beeldvorming. De DPDT bevatte zeven VASARI-beeldvormingskenmerken (necrose, diffusie, bloeding, niet-aankleurende tumormarge, calvariale remodeling, cysten, verhouding oedeem) en vier niet-VASARI-beeldvormingskenmerken (T2-FLAIR mismatch sign, T2-signaalhomogeniteit, calcificatie, midline shift). Drie beoordelaars gebruikten een testdataset van 303 patiënten om tumorgraad (graad 2 vs. graad 3/4) en moleculaire status (IDH en 1p/19q) te voorspellen op basis van GHC-vrije en GHC-verrijkte MRI.

Onafhankelijk behaalden de beoordelaars een nauwkeurigheid van ≥85% voor het voorspellen van tumorgraad en ≥75% voor het voorspellen van moleculaire status zonder gebruik te maken van contrastmiddel houdende MRI. Er waren geen significante verschillen met GHC-verrijkte voorspellingen, die respectievelijk een nauwkeurigheid van ≥87% en ≥77% bereikten. De groepsovereenstemming tussen beoordelaars was matig (0,56) en substantieel (0,68) voor respectievelijk GHC-vrije en GHC-verrijkte voorspellingen. Deze bevindingen benadrukken het potentieel van GHC-vrije beeldvorming als een betrouwbaar alternatief voor glioomdiagnose, tonen aan dat de toegevoegde waarde van GHC-verrijkte MRI beperkt is en ondersteunen de haalbaarheid van een GHC-vrije beeldvormingsaanpak in de neuro-oncologie.

In hoofdstuk 6 richtte ik me op twee verschillende beoordelingsbenaderingen van DWI (een visuele beoordeling versus een op region-of-interest (ROI)-gebaseerde methode). DWI is de meest gebruikte geavanceerde GHC-vrije beeldvormingstechniek en een cruciaal element in neuro-oncologische MRI-protocollen. Ik evalueerde de correlatie tussen deze methodes, hun reproduceerbaarheid en hun voorspellende diagnostische prestaties. Voor visuele beoordeling werden diffusiepatronen gecategoriseerd op basis van VASARI-feature 17 (gefaciliteerd, twijfelachtig, beperkt). Bij ROI-gebaseerde beoordeling werden cirkelvormige ROI's geplaatst op de visueel waargenomen gebieden met de laagste ADC, waarbij absolute en genormaliseerde ADC (aADC/nADC) werden gemeten. Bij gebruik van preoperatieve MRI-scans van 303 patiënten met graad 2-4 adult-type diffuse gliomen, liet de op ROI-gebaseerde beoordeling een superieure reproduceerbaarheid zien, met een hogere overeenkomst tussen en binnen beoordelaars. (ROI-gebaseerd 0,56 vs. visueel 0,34). Desondanks werd er een sterke correlatie gezien tussen ROI-gebaseerde ADC-drempelwaarden (1.090 en 623 x 10-6 mm2/s voor aADC en 1,38 en 0,80 voor nADC, waarbij onderscheid wordt gemaakt tussen gefaciliteerde, twijfelachtige en beperkte visuele diffusiecategorieën) en visuele beoordelingscategorieën ($P < .001$). Voor de voorspelling van IDH-mutatie werden alleen patiënten geanalyseerd waarbij geen necrose of bloedingen waren waargenomen. Beperkte diffusie bij visuele beoordeling werd geclassificeerd als IDH-wildtype, terwijl twijfelachtige en gefaciliteerde diffusie werden aangeduid als IDH-mutant. Voor de ROI-gebaseerde methode waren de optimale IDH-classificatiedrempels 1,048 x 10-6 mm2/s en 1,38 voor respectievelijk aADC en nADC. Visuele beoordeling behaalde een vergelijkbare

nauwkeurigheid (visueel vs. aADC/nADC: 69% vs. 73%/70%) en beide methoden vertoonden een onbalans tussen specificiteit (99% vs. 81%/75%) en gevoeligheid (14% vs. 57%/61%). Deze bevindingen tonen de gelijkwaardige klinische bruikbaarheid en sterke correlatie aan tussen visuele en ROI-gebaseerde beoordelingen van DWI. Daarnaast kunnen de resultaten een leidraad zijn voor het gebruik van diffusiegegevens in geavanceerde AI-gebaseerde modellen, een gebied dat nog steeds onderontwikkeld is.

LIST OF PUBLICATIONS

This thesis

Wamelink, I. J. H. G., **Azizova, A.**, Booth, T. C., Mutsaerts, H. J. M. M., Ogunleye, A., Mankad, K., Petr, J., Barkhof, F., & Keil, V. C. (2024). Brain Tumor Imaging without Gadolinium-based Contrast Agents: Feasible or Fantasy?. *Radiology*, 310(2), e230793. https://doi.org/10.1148/radiol.230793

Azizova, A., Prysiazhniuk, Y., Wamelink, I. J. H. G., Petr, J., Barkhof, F., & Keil, V. C. (2024). Ten Years of VASARI Glioma Features: Systematic Review and Meta-Analysis of Their Impact and Performance. *AJNR. American journal of neuroradiology*, 45(8), 1053–1062. https://doi.org/10.3174/ajnr.A8274

Azizova, A., Wamelink, I. J. H. G., Prysiazhniuk, Y., Cakmak, M., Kaya, E., Petr, J., Barkhof, F., & Keil, V. C. (2024). Human performance in predicting enhancement quality of gliomas using gadolinium-free MRI sequences. *Journal of neuroimaging*: official journal of the American Society of Neuroimaging, 34(6), 673–693. https://doi.org/10.1111/jon.13233

Azizova, A., Prysiazhniuk, Y., Wamelink, I. J. H. G., Cakmak, M., Kaya, E., Wesseling, P., de Witt Hamer, P. C., Verburg, N., Petr, J., Barkhof, F., & Keil, V. C. (2025). Preoperative prediction of diffuse glioma type and grade in adults: a gadolinium-free MRI-based decision tree. *European radiology*, 35(3), 1242-1254. https://doi.org/10.1007/s00330-024-11140-5

Azizova, A., Prysiazhniuk, Y., Cakmak, M., Kaya, E., Petr, J., Barkhof, F., Wamelink, I.J.H.G., Keil, V.C. Visual versus Region-of-Interest Based Diffusion Evaluation and their Diagnostic Impact in Adult-Type Diffuse Gliomas. *Neuroradiology. Under review*

Other publications

Essed, R. A., Prysiazhniuk, Y., Wamelink, I. J., **Azizova, A.**, & Keil, V. C. (2024). Performance of amide proton transfer imaging to differentiate true progression from therapy-related changes in gliomas and metastases. *European radiology*, 35(2), 580-591. https://doi.org/10.1007/s00330-024-11004-y

PORTFOLIO

Title	Location	Year	ECTS
Courses			
Writing a Data Management Plan	Online	2023	1.00
Basic Oncology	Local	2023	2.00
BioBusiness	Local	2023	3.00
Scientific Integrity	Online	2023	2.00
Entrepreneurship in Health & Life Sciences	Local	2023	1.50
BROK	Online	2023	1.50
Design your thesis with InDesign	Local	2024	0.10
Amsterdam UMC World of Science	Local	2024	0.50
Practical Biostatistics	Online	2024	1.40
International and National conferences			
4th Glioma MR Imaging 2.0 COST Action meeting	Porto/Portugal	2023	2.00
ESNR Annual Meeting 2023	Vienna/Austria	2023	2.00
CCA retreat	Local	2023	2.00
OOA retreat	Local	2023	2.00
Teaching			
Training students	Local	2023	5.71
Other			
RNG research day	Local	2023	0.50
EORTC ECI Mentoring & Career Development Programme Sessions	Online	2023	0.64
2023 CPTAC Scientific Symposium	Online	2023	0.18
ASAP Symposium	Local	2023	0.11
Introduction to CCA	Local	2023	0.10
Journal Club	Local	2023-2024	1.50
Internal scientific/multidisciplinary meetings and peer-review activities	Local/Online	2023-2024	3.75
Preparing a publication	Local/Online	2023-2024	2.00
GliMR 2.0 COST Action Training School	Padova/Italy	2024	1.50
Organization of Science and Awards Day, 2nd edition, March 8, 2024	Local	2024	3.00
Total ECTS: 39.99			

ACKNOWLEDGMENTS

As I sit down to write this, I find myself reflecting on the journey that brought me here. This PhD journey has been one of my life's most intense yet rewarding experiences. It challenged me in ways I never expected, but it also helped me grow- not just as a researcher but as a person. I learned to navigate the complexities of scientific research, express myself more clearly, and push through moments of doubt. However, this journey was never a solo one. Behind every achievement and hardship, there were people who supported, inspired, and uplifted me when I most needed it.

This is my opportunity **to thank** my supervisors who guided me, my colleagues who stood by me, my friends who cheered me on, and my family who gave me unconditional support.

My supervisors

Prof. Frederik Barkhof - the moment I decided to pursue research abroad and reached out to you, I felt your continuous support. Your extensive knowledge, breadth of experience, and capacity to simplify even the most complex matters have greatly influenced me. Despite the ups and downs of the PhD path, your guidance made it easier and was extremely motivating. You are among the role models that I try to emulate in my professional life. I appreciate your strong leadership and your constant inspiration to reach new heights.

Vera Keil - I want to thank you for so many things that I am not even sure where to start. From the very start, you were more than just a supervisor - you were a mentor, a motivator, and, at times, even a mind coach. Whenever I felt stuck and overwhelmed, a quick meeting with you was enough to reset my energy and give me a fresh perspective. I greatly respect your ability to break down problems, provide perceptive answers, and make everything seem manageable. I have learned so much from you - not just about research but also about resilience, patience, and kindness. You have made this journey a truly enjoyable experience, and I cannot thank you enough for that.

Jan Petr - if I had to describe you in one phrase, it would be "problem solver." No matter what the issue was, I always knew I could count on you for clear, constructive, and practical advice. Even though we didn't get many chances to meet in person, you were always incredibly accessible, making sure that I never felt lost. Your technical know-how, methodical approach, and productivity are very impressive. Many people who have worked with you would undoubtedly agree that having you as a supervisor is a privilege. Thank you for your continuous assistance and guidance.

My fellow researchers

Ivar Wamelink - you were my first go-to person whenever I needed help - whether it was for our research or simply for advice on adjusting to life in a new city. It's never easy to move to a new country and adjust to a new atmosphere, but your generosity, tolerance, and readiness to lend a hand were really helpful. I appreciate your being there for me no matter what, answering my questions, and making me feel welcomed and supported.

Yeva Prysiazhniuk - after meeting you, my research suddenly became so much easier. Your timely responses, responsible approach, and willingness to help whenever needed have been absolutely priceless. I truly appreciate all the effort and dedication you have put into supporting me throughout this journey. Thank you for everything you have done.

My team members, journal club colleagues, and everyone I shared discussions and meetings with - there were days when research felt overwhelming, but hearing your experiences and struggles reminded me that I was not alone. The simple act of sharing ideas, discussing challenges, and pushing through moments together made this process much more meaningful. I am thankful to have been part of such an inspiring group of researchers.

My friends

To my dear friends, **Nigar, Fatime, Natavan, Cavid, Omer, and Yasin** - you each hold a very special place in my heart. No matter how busy, stressed, and overwhelmed I was, I always knew that I had you by my side. Your support, your

laughter, and your presence made even the toughest day more bearable. There are countless things I could say, but I will keep it simple: thank you for always being there in ways I needed more than I ever realized.

My family

It is difficult for me to express how grateful I am to **my beloved family**. The past two years have been especially difficult because, in addition to completing my PhD, I also had to study for the highly competitive radiology exams. At times, when exhaustion overtook me and I began to doubt myself, it all seemed too much. And you were there each and every time.

My mother, my sisters Elnare and Gunel - these achievements would not have been possible without your invaluable support. No matter how far apart we were, I always felt your encouragement, your belief in me, and your unconditional love. But in chasing these goals, I also missed so much - precious time that I could have spent with you, important days we should have celebrated together, and simple everyday joy we should have shared. Yet, you always showed such profound kindness and understanding, continuously lifting me up and reminding me the true purpose behind this journey. Your support has meant everything to me, and I am endlessly grateful for that!

This has been a journey of challenges, growth, and discovery. I am deeply thankful to everyone who has been a part of it. Each of you has contributed in your own way, and this achievement belongs to all of us!

CURRICULUM VITAE

Aynur Azizova was born on May 28, 1991, in Azerbaijan. She completed secondary school in 2008 and earned her Doctor of Medicine degree from Azerbaijan Medical University in 2014. Following graduation, she began a Pediatrics Residency but realized within a few months that pediatrics was not the right fit for her. She then decided to pursue a career in radiology and completed her Radiology Residency in 2021 at Hacettepe University, Turkiye. During her residency, she discovered her passion for neuroradiology.

In 2021, she joined a one-year neuro-oncology research program at Hacettepe University. She also successfully applied for the European Society of Neuroradiology research fellowship program and started her work at Amsterdam UMC, The Netherlands, in 2022 under this program. Within six months of her fellowship, an opportunity emerged for a PhD position, and she began her PhD journey in March 2023 at the same university, resulting in this thesis (promotor: prof.dr. F. Barkhof and co-promotors: dr. V.C.W. Keil and dr. J. Petr).