CLINICAL INVESTIGATIVE STUDY

JOURNAL OF
NEUROIMAGING
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF NEUROIMAGING

WILEY

Human performance in predicting enhancement quality of gliomas using gadolinium-free MRI sequences

Aynur Azizova[1,2] | Ivar J. H. G. Wamelink[1,2] | Yeva Prysiazhniuk[3,4] | Marcus Cakmak[1,5] | Elif Kaya[6] | Jan Petr[1,7] | Frederik Barkhof[1,8,9] | Vera C. Keil[1,2,8] ⓘ

[1]Department of Radiology and Nuclear Medicine, Amsterdam UMC, location VUMC, Amsterdam, The Netherlands

[2]Imaging and Biomarkers, Cancer Center Amsterdam, Amsterdam, The Netherlands

[3]Second Faculty of Medicine, Department of Pathophysiology, Charles University, Prague, Czech Republic

[4]Motol University Hospital, Prague, Czech Republic

[5]University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[6]Faculty of Medicine, Ankara Yıldırım Beyazıt University, Ankara, Türkiye

[7]Helmholtz-Zentrum Dresden-Rossendorf, Institute of Radiopharmaceutical Cancer Research, Dresden, Germany

[8]Brain Imaging, Amsterdam Neuroscience, Amsterdam, The Netherlands

[9]Queen Square Institute of Neurology and Center for Medical Image Computing, University College London, London, UK

Correspondence
Vera C. Keil, Department of Radiology and Nuclear Medicine, Amsterdam UMC, location VUMC, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands.
Email: v.c.w.keil@amsterdamumc.nl

Abstract

Background and Purpose: To develop and test a decision tree for predicting contrast enhancement quality and shape using precontrast magnetic resonance imaging (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 postcontrast 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 postcontrast T1-weighted images. Statistical analysis used confusion matrices, Cohen's/Fleiss' kappa, and Kendall's W. Significance threshold was $p < .05$.

Results: Raters 1, 2, and 3 achieved overall accuracies of .86 (95% confidence interval [CI]: .81-.90), .89 (95% CI: .85-.92), and .92 (95% CI: .89-.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 .84 (95% CI: .79-.88), .88 (95% CI: .84-.92), and .89 (95% CI: .85-.92). Intrarater intergroup agreement comparing predicted and true enhancement features consistently reached substantial levels (≥.68 [95% CI: .61-.75]). Interrater comparison showed at least moderate agreement (group: ≥.42 [95% CI: .36-.48], pairwise: ≥.61 [95% CI: .50-.72]). Among the imaging features in the EPDT, necrosis assessment displayed the highest intra- and interrater consistency (≥.80 [95% CI: .73-.88]).

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

KEYWORDS
enhancement, gadolinium-based contrast agent, GBCA, glioma, MRI, VASARI

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 magnetic resonance imaging (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 noninsured 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 postcontrast 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 as follows: (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 1 month before surgery, and (iii) MRI scan including precontrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery (FLAIR), diffusion-weighted imaging, and postcontrast T1-weighted sequences. Exclusion criteria were as follows: (i) pediatric patients, (ii) patients who had explicitly refused
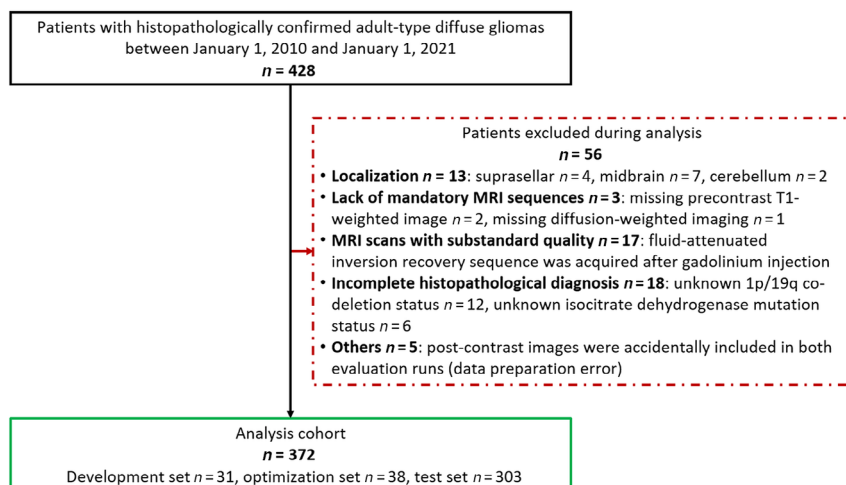
FIGURE 1 Flow chart describes the details of the patient enrollment. n = number.

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	4376-4840	-	90
2D FLAIR	118-120	9002-9502	2250	90
3D FLAIR	96-122	6000-6500	1925-1987	90
DWI (b-1000)	81-105	8000-8500	-	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
T2w	82-88	4889-6872	-	90-111
3D FLAIR	126-135	8000-8002	2331-2347	90
DWI (b-1000)	62-87	4000-7200	-	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	3404-5251	-	90
2D FLAIR	140	9000-11,000	2800	90
3D FLAIR	286-306	4800	1660	90
DWI (b-1000)	72-119	2674-6448	-	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	2767-3182	-	90
3D FLAIR	279	4800	1650	90
DWI (b-1000)	74-97	3496-6354	-	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	700-2700	0-950	8-120
T2w	93-104	2830-5562	-	150-180
2D FLAIR	88-109	8870-9000	2500	150
3D FLAIR	334	6500	2200	120
DWI (b-1000)	90-122	3400-10,500	-	90
2D CE-T1w	8.7-17	550-718	-	90
3D CE-T1w	2.9-4.5	1900-2700	950-1100	8

(Continues)

TABLE 1 (Continued)

Sequence type	Parameters			
	TE; ms	TR; ms	TI; ms	FA; ms
Siemens 3T MAGNETOM Vida				
3D T1w	2.3	2300	900	8
T2w	74	4100-6280	-	150
3D FLAIR	388-430	5000-7700	1650-2400	120
DWI (b-1000)	68	3200	-	90
3D CE-T1w	2.3	2300	900	8
Toshiba 3T Titan3T				
2D T1w	8	550	-	80
T2w	90	5500-5526	-	90
3D FLAIR	451	5600	1900	90
DWI (b-1000)	82	7500	-	90
2D CE-T1w	8	550	-	80
3D CE-T1w	2.4	5.7	900	9

Note: The table describes MRI parameters for each of the seven scanners used in this study. 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; TE, time of echo; TI, time of inversion; TR, time of repetition; w, weighted.

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).

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).

Datasets

Datasets were prepared and pseudonymized by I.W., a fourth-year PhD 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 (preliminary version)
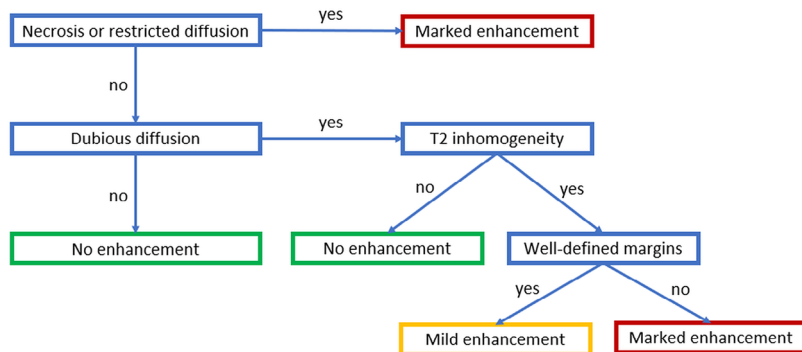
FIGURE 2   Flow chart describes the preliminary version of the enhancement prediction decision tree.

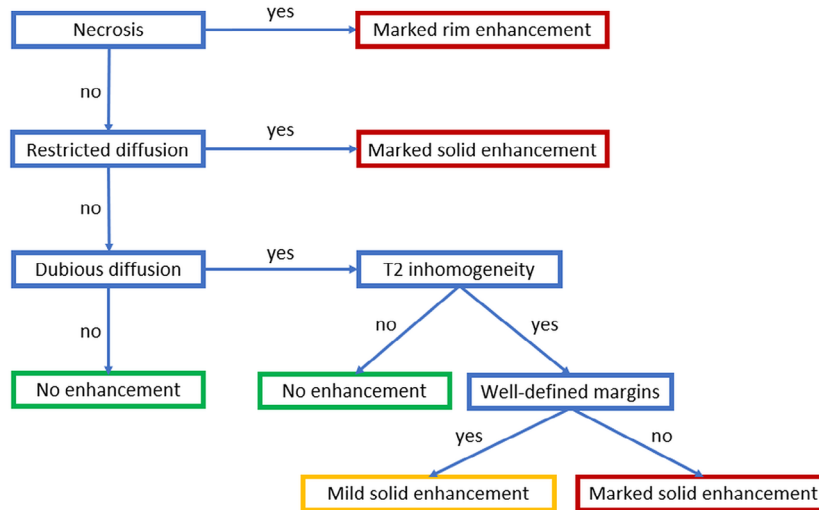### Enhancement prediction decision tree

The ratings were carried out independently by two raters (V.K., 11 years of neuroradiology experience; A.A., 5 years of neuroradiology experience).

In the initial stage, the raters assessed the GBCA-free sequences of the development set (*n* = 31) consisting of precontrast 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 postcontrast 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? (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 postcontrast images: eloquent brain (feature 3), presence of necrosis (modified feature 7), multifocality (feature 9), T1/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 precontrast and additional postcontrast 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 decision-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 postcontrast images. Ultimately, they identified necrosis, nonenhancing tumor margins, diffusion, and T2 inhomogeneity as the most influential features in predicting enhancement quality (EQ), which correlated well with postcontrast 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.

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 precontrast 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 postcontrast 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

Enhancement Prediction Decision Tree (final version)

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

only modification made was the incorporation of the pTEM into the decision tree.

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

Test set

The performance of human raters predicting 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).

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 postcontrast 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.

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-wildtype 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.

Interrater 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 interrater 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
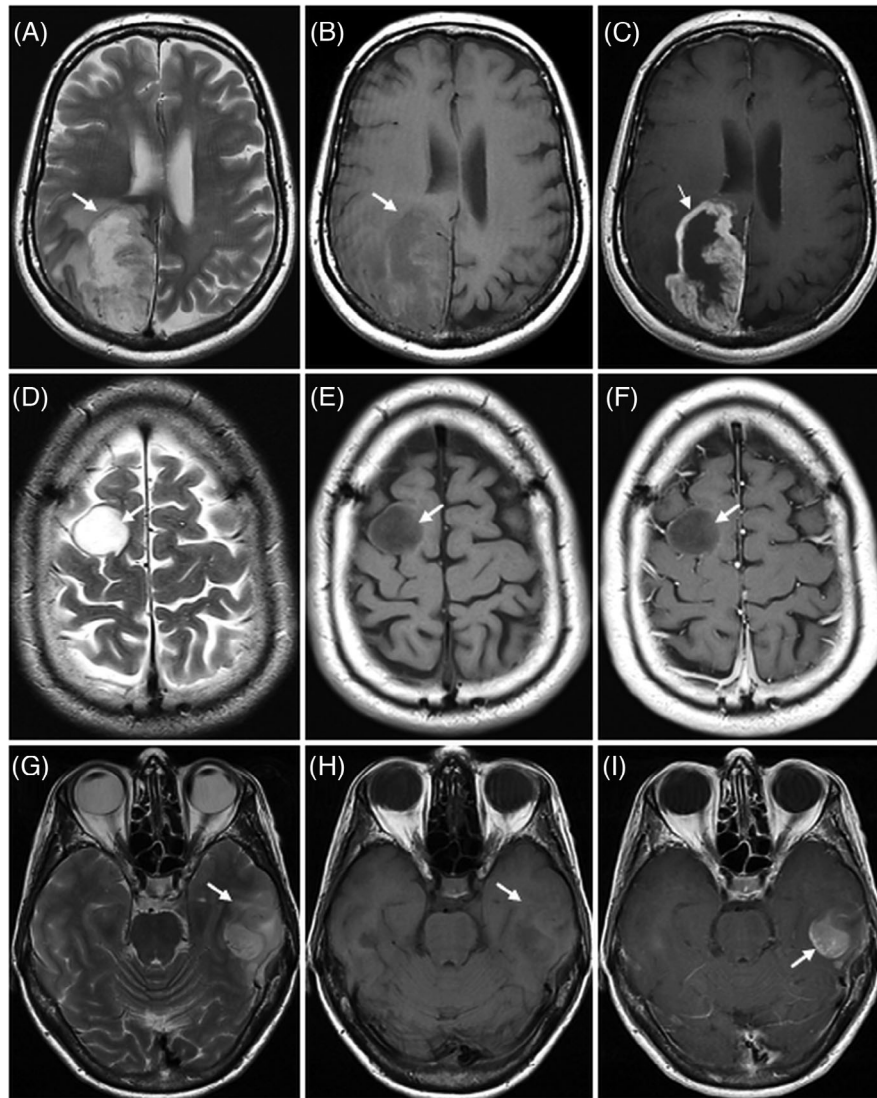
FIGURE 4 Demonstrative cases for the evaluation of necrosis, T2 inhomogeneity, and nonenhancing tumor margins. Necrosis (a-c; glioblastoma, IDH-wildtype): right parietal lesion (white arrows) with central necrosis characterized by irregular and thick margins, and internal characteristics of T2 hyperintensity (a) and T1 hypointensity (b). Contrast-enhanced image (c) shows marked rim enhancement. T2 homogeneity and well-defined margins (d-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 precontrast T1-weighted image (e) compatible with nonenhancing glioma. T2 inhomogeneity and ill-defined margins (g-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 precontrast T1-weighted image (h), compatible with marked solid enhancement.

features (diffusion restriction, EQ) was assessed with linearly weighted Cohen's kappa. This methodology was also used for the intrarater intergroup agreement analysis, which compares the predicted and true enhancement features per rater using GBCA-free and GBCA-enhanced datasets, respectively.

Group interrater 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: .01-.20 slight, .21-.40 fair, .41-.60 moderate, .61-.80 substantial, and .81-.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 $p < .05$.

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.
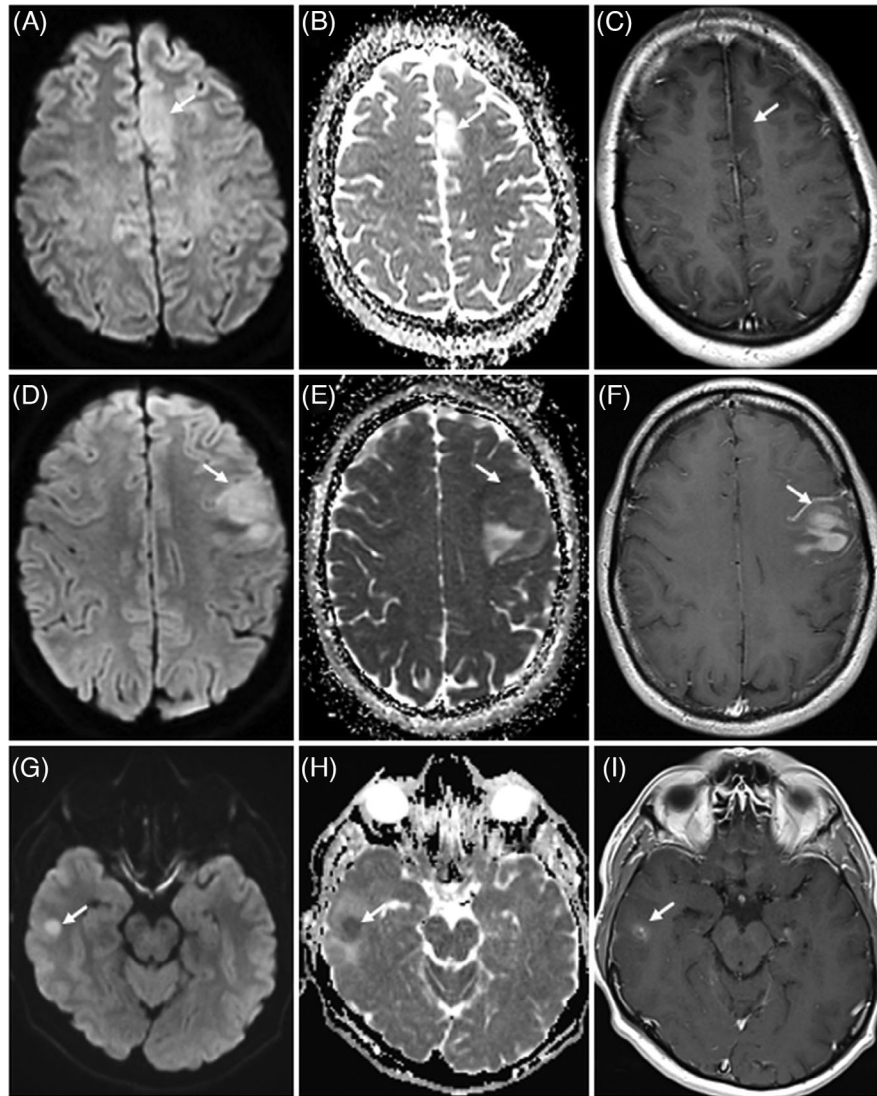
FIGURE 5 Demonstrative cases for the evaluation of diffusion. Facilitated diffusion (a-c; low-grade oligodendroglioma, IDH-mutant and 1p/19q-codeleted): Left frontal lesion (white arrows) with hyperintense signal on b-1000 map of diffusion-weighted imaging (DWI) (a) and higher apparent diffusion coefficient (ADC) signal (b) than the normal cortex. There is no enhancement on the contrast-enhanced image (c). Dubious diffusion (d-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-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).

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 .68/.64 to .95/.97, respectively, for raters 1 and 2. The results for binary pEQ were .84/.74 and .97/.97, respectively.

Test set: Prediction performance of human raters

Table 3 describes the cohort characteristics of the test set.

The overall accuracy of the pEQ (marked, mild, or no enhancement) was .86 (95% CI: .81-.90), .89 (95% CI: .85-.92), and .92 (95% CI: .89-.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") was assessed: .89 (95% CI: .85-.92), .92 (95% CI: .89-.95), and .93 (95% CI: .89-.95). The overall accuracy of pTEM (solid, rim, or no enhancement) was .84 (95% CI: .79-.88), .88 (95% CI: .84-.92), and .89 (95% CI: .85-.92) for raters 1, 2, and 3, respectively. Table 4 lists rater-based confusion
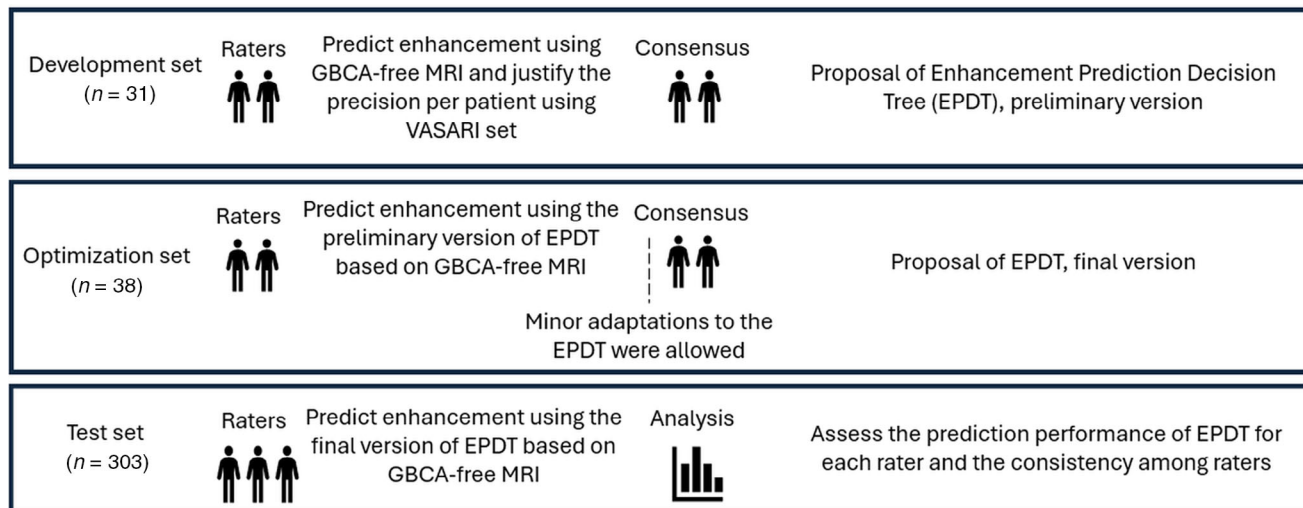
FIGURE 6   Flowchart demonstrates all steps of the study design. GBCA, gadolinium-based contrast agent; *n*, number; VASARI, Visually AcceSAble Rembrandt Images.

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[a] (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[a] (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[a] (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
Enhancement quality, VASARI feature 4[a] (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 precontrast images No enhancement: no difference between precontrast and postcontrast images
Thickness of enhancing margin, modified VASARI feature 11[a] (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

Note: The table 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.
[a]https://wiki.cancerimagingarchive.net/display/Public/VASARI±Research±Project18

WILEY | 681

matrices and sensitivity, specificity, and positive and negative predictive values.

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 .89 (95% CI: .85-.93), .91 (95% CI: .87-.95), and .94 (95% CI: .91-.97) for raters 1, 2, and 3, respectively. Similarly, the results for binary pEQ (.92 [95% CI: .88-.95], .95 [95% CI: .91-.97], and .94 [95% CI: .91-.97]) and pTEM (.87 [95% CI: .82-.91], .91 [95% CI: .87-.94], and .91 [95% CI: .87-.94]) also increased (for details, see Table 6).

Inter- and intrarater agreement for EQ and TEM

Group interrater agreement analysis (Figure 7) revealed moderate agreement for both pEQ and tEQ (Kendall's W .42 [95% CI: .36-.48] and .55 [95% CI: .48-.62]) and substantial and almost perfect agreement for the pTEM and tTEM, respectively (Fleiss' kappa .66 [95% CI: .60-.71] and .83 [95% CI: .79-.88]). The results were further improved for both pEQ and tEQ (Fleiss' kappa .65 [95% CI: .57-.73] and .87 [95% CI: .82-.93]) when binary analysis for EQ (enhancing/nonenhancing) was applied. Pairwise interrater agreements were substantial (≥.61 [95% CI: .50-.72]) and almost perfect (≥.82 [95% CI: .75-.89]) for the predicted and true features, respectively (see Table 7). Comparison analysis of agreements showed significant differences between the evaluation agreements of predicted and true features (compare Figure 7 and Table 7 for details). Intrarater intergroup agreement analysis comparing predicted and true enhancement features per rater demonstrated substantial to almost perfect agreements (≥.68 [95% CI: .61-.75]) for each rater without significant differences in agreement among all raters (p-values: EQ .10, binary EQ .22, TEM .07) (see Figure 8 and Table 8).

Agreement analysis for single imaging features involved in EPDT

Group interrater agreement analysis revealed almost perfect agreement for necrosis identification in both GBCA-enhanced and GBCA-free datasets (Fleiss' kappa .85 [95% CI: .80-.90] and .83 [95% CI: .78-.88]). The agreement for other imaging features was moderate (≥.43 [95% CI: .28-.57]). Notably, the availability of postcontrast images did not significantly influence imaging feature agreement (p-values: necrosis .49, diffusion restriction .81, T2 inhomogeneity .63, nonenhancing

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 $n = 221$)

Note: The table 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.

tumor margins .65) (see Figure 9 and Table 9). A pairwise interrater agreement was almost perfect for necrosis (≥.80 [95% CI: .73-.88]) and fair to moderate for other features (≥.33 [95% CI: .25-.41]) in both GBCA-enhanced and GBCA-free datasets. However, the results for T2 inhomogeneity and nonenhancing tumor margins increased to substantial-almost perfect (≥.71) level after applying PABAK analysis (Table 10). Intrarater agreement in the assessment of necrosis was almost perfect (≥.82 [95% CI: .75-.89]) for all three raters. This analysis in the evaluation of other features showed substantial-almost perfect agreements (≥.66 [95% CI: .55-.77]) for raters 2 and 3 while being fair to moderate (≥.35 [95% CI: .18-.53]) for rater 1. Applying PABAK analysis revealed better agreement for T2 inhomogeneity and nonenhancing tumor margins (≥.78; Figure 10 and Table 11).

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

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. Intrarater intergroup 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.

TABLE 4 Human prediction performance results for enhancement features.

EQ (enhancement categories: 0 = no, 1 = mild, 2 = marked)

Results	Rater 1				Rater 2				Rater 3			
Confusion matrix		**Predicted**				**Predicted**				**Predicted**		
		0	1	2		0	1	2		0	1	2
	True				True				True			
	0	53	1	7	0	46	0	18	0	60	2	8
	1	8	0	10	1	2	0	9	1	0	5	0
	2	17	0	207	2	3	2	223	2	12	1	215
Sensitivity												
0	.68				.90				.83			
1	.00				.00				.63			
2	.92				.89				.96			
Specificity												
0	.96				.93				.96			
1	.94				.96				1.00			
2	.78				.91				.84			
PPV												
0	.87				.72				.86			
1	.00				.00				1.00			
2	.92				.98				.94			
NPV												
0	.90				.98				.95			
1	1.00				.99				.99			
2	.78				.64				.89			

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

	Rater 1			Rater 2			Rater 3		
Confusion matrix		**Predicted**			**Predicted**			**Predicted**	
		0	1		0	1		0	1
	True			True			True		
	0	53	8	0	46	18	0	60	10
	1	25	217	1	5	234	1	12	221
Sensitivity	.68			.90			.83		
Specificity	.96			.93			.96		
PPV	.87			.72			.86		
NPV	.90			.98			.95		

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

TABLE 4 (Continued)

Rater 1

Confusion matrix	True	Predicted 0	1	2
	0	53	3	5
	1	6	186	10
	2	19	6	15
Sensitivity	True			
0	.68			
1	.95			
2	.50			
Specificity				
0	.96			
1	.85			
2	.91			
PPV				
0	.87			
1	.92			
2	.38			
NPV				
0	.90			
1	.91			
2	.94			

Rater 2

Confusion matrix	True	Predicted 0	1	2
	0	46	1	17
	1	1	190	5
	2	4	8	31
Sensitivity	True			
0	.90			
1	.95			
2	.58			
Specificity				
0	.93			
1	.94			
2	.95			
PPV				
0	.72			
1	.97			
2	.72			
NPV				
0	.98			
1	.92			
2	.92			

Rater 3

Confusion matrix	True	Predicted 0	1	2
	0	60	6	4
	1	7	190	7
	2	5	4	20
Sensitivity	True			
0	.83			
1	.95			
2	.65			
Specificity				
0	.96			
1	.86			
2	.97			
PPV				
0	.86			
1	.93			
2	.69			
NPV				
0	.95			
1	.90			
2	.96			

Note: The table describes the prediction performance for each rater in evaluating glioma enhancement features. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using the gadolinium-based contrast agent (GBCA)-free dataset (precontrast 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 postcontrast T1-weighted sequences).

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

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

Confusion matrix

EQ (enhancement categories: 0 = no, 1 = mild, 2 = marked)

	Predicted 0	Predicted 1	Predicted 2	Possible failed EPDT steps
True 0	53	8 [LGG/HGG n = 4 (7%)/4 (2%), ODG/Astro/GB n = 3 (9%)/2 (4%)/3 (1%)]	17 [LGG/HGG n = 5 (9%)/12 (5%), ODG/Astro/GB n = 4 (12%)/6 (13%)/7 (3%)]	0-1, 0-2, 1-0: dubious diffusion or T2 inhomogeneity
True 1	1 [LGG/HGG n = 1 (2%), ODG n = (3%)]	0	10 [LGG/HGG n = 4 (7%), HGG n = 6 (2%), ODG n = 2 (6%), Astro n = 4 (8%), GB n = 4 (2%)]	1-2: necrosis/diffusion restriction or well-defined margins
True 2	7 [LGG/HGG n = 4 (7%)/3 (1%), ODG/Astro n = 4 (12%)/3 (6%)]	2	207	2-0: necrosis/diffusion restriction or well-defined margins

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

	Predicted 0	Predicted 1	Possible failed EPDT steps
True 0	53	8 [LGG/HGG n = 5 (9%)/3 (1%), ODG/Astro n = 5 (15%)/3 (6%)]	0-1, 1-0: necrosis/diffusion restriction or dubious diffusion or T2 inhomogeneity
True 1	25 [LGG/HGG n = 9 (17%)/16 (6%), ODG/Astro/GB n = 7 (21%)/8 (17%)/10 (5%)]	217	

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

	Predicted 0	Predicted 1	Predicted 2	Possible failed EPDT steps
True 0	53	6 [LGG n = 6 (11%), ODG/Astro/GB n = 1 (3%)/3 (6%)/2 (1%)]	19 [LGG/HGG n = 9 (17%)/10 (4%), ODG/Astro/GB n = 6 (18%)/5 (10%)/8 (4%)]	0-1, 1-0, 1-2, 2-1: necrosis
True 1	3 [LGG/HGG n = 2 (4%)/1 (<1%), ODG/Astro n = 2 (6%)/1 (2%)]	15	10 [LGG n = 10 (19%), ODG/Astro/GB n = 1 (3%)/1 (2%)/8 (4%)]	0-2, 2-0: diffusion restriction or dubious ADC or T2 inhomogeneity
True 2	5 [LGG/HGG n = 3 (6%)/2 (<1%), ODG/Astro n = 3 (9%)/2 (4%)]	2 [LGG/HGG n = 2 (4%)/4 (2%), ODG/Astro/GB n = 1 (3%)/1 (2%)/4 (2%)]	186	

Note: The table describes the subgroup analysis results exploring failed enhancement predictions based on histopathological diagnosis for Rater 1. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using the gadolinium-based contrast agent (GBCA)-free dataset (precontrast 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 postcontrast T1-weighted sequences). Total sample size: LGG/HGG n = 54 (18%)/249 (82%), ODG/Astro/GB n = 34 (11%)/48 (16%)/221 (73%).

Abbreviations: ADC, apparent diffusion coefficient; Astro, astrocytoma; EPDT, enhancement prediction decision tree; EQ, enhancement quality; GB, glioblastoma; HGG, high-grade (grade 3/4) glioma; IDH-mutant and 1p/19q-codeleted; IDH-wildtype; LGG, low-grade (grade 2) glioma; n, number; ODG, oligodendroglioma; TEM, thickness of enhancing margin.

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

EQ (enhancement categories: 0 = no, 1 = mild, 2 = marked)

Results		Rater 1				Rater 2				Rater 3			
			Predicted				**Predicted**				**Predicted**		
		True	0	1	2	True	0	1	2	True	0	1	2
Confusion matrix	0		39	0	3		34	0	10		40	2	6
	1		5	0	8		1	0	8		0	4	0
	2		13	0	201		3	1	212		7	1	209
Sensitivity	0	.68				.89				.85			
	1	.00				.00				.57			
	2	.95				.92				.97			
Specificity	0	.99				.96				.96			
	1	.95				.97				1.00			
	2	.77				.90				.85			
PPV	0	.93				.77				.83			
	1	.00				.00				1.00			
	2	.94				.98				.96			
NPV	0	.92				.98				.97			
	1	.00				1.00				.99			
	2	.80				.66				.88			

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

		Rater 1			Rater 2			Rater 3		
			Predicted			**Predicted**			**Predicted**	
		True	0	1	True	0	1	True	0	1
Confusion matrix	0		39	3		34	10		40	8
	1		18	209		4	221		7	214
Sensitivity		.68			.89			.85		
Specificity		.99			.96			.96		
PPV		.93			.77			.83		
NPV		.92			.98			.97		

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

(Continues)

TABLE 6 (Continued)

Rater 1

Confusion matrix	True	Predicted 0	Predicted 1	Predicted 2
	0	39	1	2
	1	5	183	9
	2	13	5	12

	True
Sensitivity	
0	.68
1	.97
2	.52
Specificity	
0	.99
1	.83
2	.93
PPV	
0	.93
1	.93
2	.40
NPV	
0	.92
1	.92
2	.95

Rater 2

Confusion matrix	True	Predicted 0	Predicted 1	Predicted 2
	0	34	1	10
	1	1	188	5
	2	3	5	23

	True
Sensitivity	
0	.89
1	.97
2	.60
Specificity	
0	.96
1	.92
2	.97
PPV	
0	.77
1	.97
2	.74
NPV	
0	.98
1	.93
2	.94

Rater 3

Confusion matrix	True	Predicted 0	Predicted 1	Predicted 2
	0	40	4	4
	1	6	188	7
	2	1	3	16

	True
Sensitivity	
0	.85
1	.96
2	.59
Specificity	
0	.96
1	.82
2	.98
PPV	
0	.83
1	.94
2	.80
NPV	
0	.97
1	.90
2	.96

Note: The table describes the prediction performance for each rater in evaluating glioma enhancement features after excluding oligodendroglioma cases. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using the gadolinium-based contrast agent (GBCA)-free dataset (precontrast 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 postcontrast T1-weighted sequences).

Abbreviations: EQ, enhancement quality; NPV, negative predictive value; PPV, positive predictive value; TEM, thickness of enhancing margin.
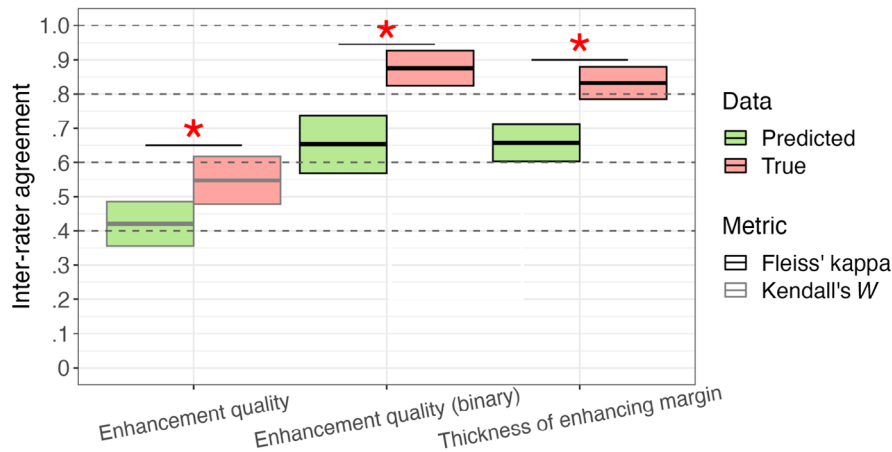
FIGURE 7 Group interrater agreement analysis in true (red color bars) and predicted (green color bars) enhancement quality and thickness of enhancing margin among all three raters. Red stars show significant differences between evaluation agreements of true and predicted datasets. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using gadolinium-based contrast agent (GBCA)-free dataset (precontrast 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 postcontrast T1-weighted sequences).
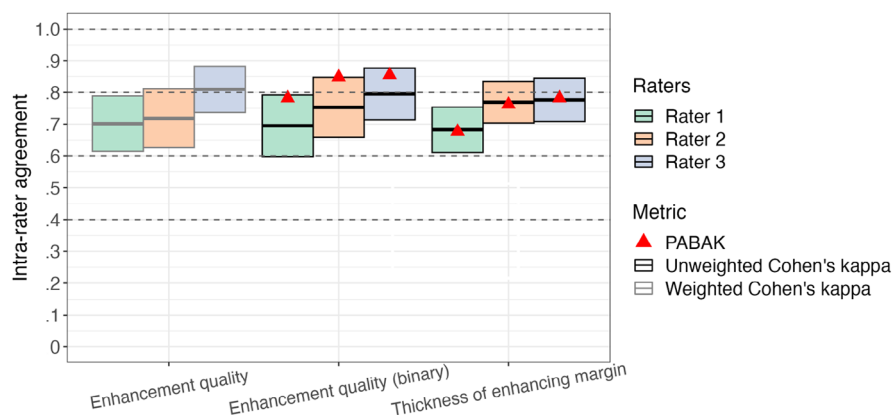
FIGURE 8 Intrarater intergroup agreement analysis shows the comparison between true and predicted enhancement features per rater. Prevalence-adjusted and bias-adjusted kappa (PABAK) values (red triangles) are comparable with unweighted Cohen's kappa for thickness of enhancing margin, indicating a negligible 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. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using gadolinium-based contrast agent (GBCA)-free dataset (precontrast 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 postcontrast T1-weighted sequences).

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 EQ using T2-FLAIR images,[23] which demonstrated high accuracy levels (area under the curve: .81 [95% CI: .71-.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

TABLE 7   Pairwise interrater agreement results for the assessment of the true/predicted enhancement.

EQ/TEM	Raters 1 and 2	Raters 1 and 3	Raters 2 and 3
tEQ[a]	.90; 95% CI: .86-.95	.89; 95% CI: .85-.94	.92; 95% CI: .87-.96
pEQ[a]	.66; 95% CI: .57-.78	.70; 95% CI: .61-.79	.61; 95% CI: .51-.71
p-value[b]	<.001	<.001	<.001
tEQ (binary)			
_[c]	.87; 95% CI: .79-.94	.87; 95% CI: .80-.94	.88; 95% CI: .82-.95
_[d]	.91	.91	.92
pEQ (binary)			
_[c]	.68; 95% CI: .58-.78	.68; 95% CI: .58-.77	.61; 95% CI: .50-.72
_[d]	.78	.76	.74
p-value[b]	<.001	<.001	<.001
tTEM			
_[c]	.84; 95% CI: .78-.91	.82; 95% CI: .75-.89	.84; 95% CI: .78-.89
_[d]	.84	.82	.83
pTEM			
_[c]	.69; 95% CI: .62-.76	.64; 95% CI: .57-.71	.64; 95% CI: .57-.72
_[d]	.68	.64	.64
p-value[b]	<.001	<.001	<.001

Note: The table describes the pairwise interrater agreement results for enhancement features based on assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using GBCA-free dataset (precontrast 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 postcontrast T1-weighted sequences).
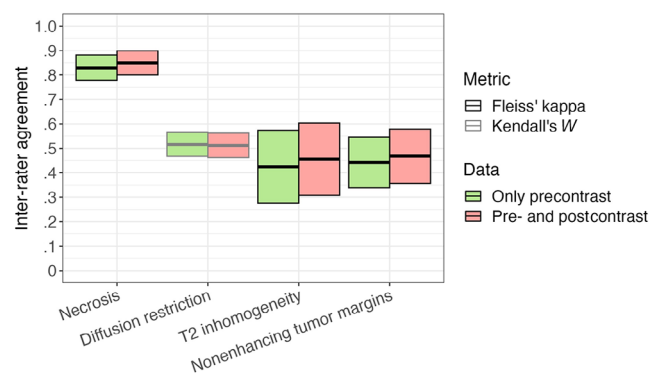
Abbreviations: CI, confidence interval; pEQ/tEQ, predicted/true enhancement quality; pTEM/tTEM, predicted/true thickness of enhancing margin.

[a]Weighted Cohen's kappa.

[b]Comparison of the agreements in true- and predicted EQ/TEM.

[c]Unweighted Cohen's kappa.

[d]Prevalence-adjusted and bias-adjusted kappa.

FIGURE 9   Group interrater 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. GBCA-free dataset includes precontrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and postcontrast T1-weighted sequences.

Dice scores in low-grade gliomas (.58 [95% CI: .49-.68] vs. .65 [95% CI: .63-.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 noninferior 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

TABLE 8 Intrarater intergroup agreement results for assessment of the true/predicted enhancement.

tEQ/tTEM vs. pEQ/pTEM	Rater 1	Rater 2	Rater 3
EQ[a]	.70; 95% CI: .61-.79	.72; 95% CI: .63-.81	.81; 95% CI: .74-.88
_p_-value[b]	.10		
EQ (binary)			
_[c]	.69; 95% CI: .60-.79	.75; 95% CI: .66-.85	.80; 95% CI: .71-.88
_[d]	.78	.85	.85
_p_-value[b]	.22		
TEM			
_[c]	.68; 95% CI: .61-.75	.77; 95% CI: .70-.83	.78; 95% CI: .71-.85
_[d]	.68	.76	.78
_p_-value[b]	.07		

Note: The table describes the intrarater intergroup agreement results comparing predicted and true enhancement features per rater. Predicted enhancement features, representing the index test, are based on assessing enhancement prediction decision tree using gadolinium-based contrast agent (GBCA)-free dataset (precontrast 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 postcontrast T1-weighted sequences).
Abbreviations: CI, confidence interval; pEQ/tEQ, predicted/true enhancement quality; pTEM/tTEM, predicted/true thickness of enhancing margin.
[a]Weighted Cohen's kappa.
[b]Comparison of the agreements among all raters.
[c]Unweighted Cohen's kappa.
[d]Prevalence-adjusted and bias-adjusted kappa.

FIGURE 10 Intrarater 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 shows significant differences in agreement among the raters. GBCA-free dataset includes precontrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and postcontrast T1-weighted sequences.

TABLE 9 Group interrater agreement results for assessment of single imaging features.

Features involved in EPDT	GBCA-enhanced dataset	GBCA-free dataset
Necrosis[a]	.85; 95% CI: .80-.90	.83; 95% CI: .78-.88
_p_-value[b]	.49	
Diffusion restriction[c]	.52; 95% CI: .46-.56	.52; 95% CI: .47-.56
_p_-value[b]	.81	
T2 inhomogeneity[a]	.46; 95% CI: .31-.60	.43; 95% CI: .28-.57
_p_-value[b]	.63	
Nonenhancing tumor margins[a]	.47; 95% CI: .36-.58	.44; 95% CI: .34-.56
_p_-value[b]	.65	

Note: The table describes the interrater agreement results for single imaging features based on assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets among all raters. GBCA-free dataset includes precontrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and postcontrast T1-weighted sequences.
Abbreviations: CI, confidence interval; EPDT, enhancement prediction decision tree.
[a]Fleiss' kappa.
[b]Comparison of the agreements between the assessments of GBCA-enhanced and GBCA-free datasets.
[c]Kendall's _W_.

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 interrater 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
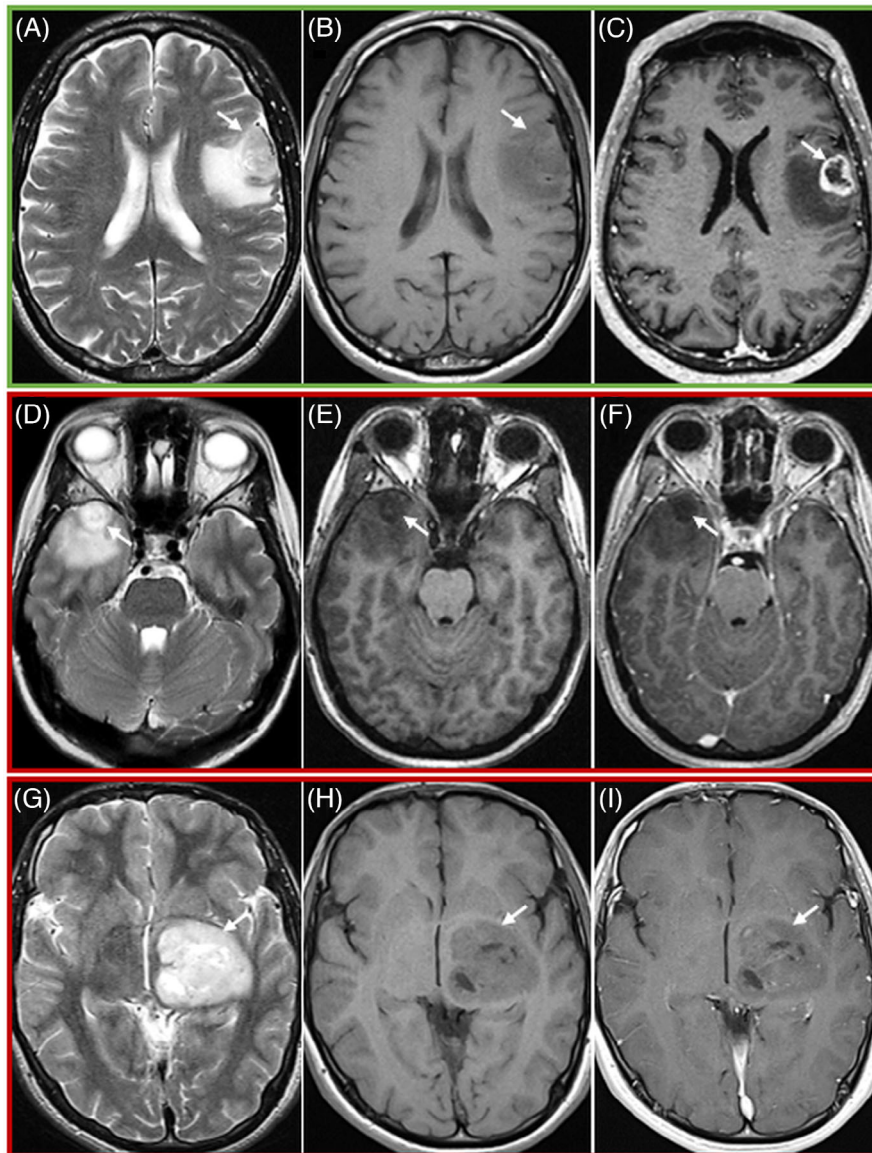
FIGURE 11 Demonstrative cases illustrating successful (green) versus failed (red) prediction of glioma enhancement based on the evaluation of necrosis using the gadolinium-based contrast-free dataset. Successful prediction of both enhancement quality (pEQ) and thickness of enhancing margin (pTEM; a-c; glioblastoma, IDH-wildtype): T2-weighted (a) and precontrast T1-weighted (b) images 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-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 precontrast T1-weighted image (e) compatible with nonenhancing glioma. Failed prediction of enhancement patterns (g-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).

administration is not available, for example, in radiology units of low- and middle-income countries.

When introducing a new diagnostic tool, such as the EPDT, achieving high interrater agreement is crucial to guarantee consistency and ensure that decision criteria are easily reproducible. Various studies[30–33] found high interrater agreements when assessing the EPDT's most consistent feature, necrosis, with kappa values ranging from .71 to .96, thus supporting our findings. The remaining three features of the EPDT (diffusion restriction, T2 inhomogeneity, and nonenhancing tumor margin) yielded fair to moderate interrater 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

WILEY | 691

TABLE 10 Pairwise interrater agreement results for assessment of single imaging features.

EPDT imaging features	Raters 1 and 2	Raters 1 and 3	Raters 2 and 3
Necrosis_GBCA-enhanced			
_[a]	.85; 95% CI: .78-.91	.84; 95% CI: .77-.90	.86; 95% CI: .80-.92
_[b]	.87	.86	.87
Necrosis_GBCA-free			
_[a]	.87; 95% CI: .81-.93	.80; 95% CI: .73-.88	.82; 95% CI: .75-.88
_[b]	.88	.82	.83
p-value[c]	.63	.42	.26
Diffusion restriction_GBCA-enhanced[d]	.42; 95% CI: .34-.50	.33; 95% CI: .25-.41	.42; 95% CI: .33-.52
Diffusion restriction_GBCA-free[d]	.42; 95% CI: .34-.49	.39; 95% CI: .31-.47	.37; 95% CI: .28-.47
p-value[c]	.94	.17	.21
T2 inhomogeneity_GBCA-enhanced			
_[a]	.41; 95% CI: .24-.59	.39; 95% CI: .22-.57	.58; 95% CI: .40-.76
_[b]	.82	.81	.88
T2 inhomogeneity_GBCA-free			
_[a]	.43; 95% CI: .27-.60	.38; 95% CI: .22-.55	.49; 95% CI: .30-.67
_[b]	.80	.77	.87
p-value[c]	.84	.90	.36
Nonenhancing tumor margins_GBCA-enhanced			
_[a]	.43; 95% CI: .28-.57	.47; 95% CI: .34-.61	.49; 95% CI: .37-.62
_[b]	.74	.73	.71
Nonenhancing tumor margins_GBCA-free			
_[a]	.36; 95% CI: .20-.52	.36; 95% CI: .21-.51	.58; 95% CI: .45-.70
_[b]	.74	.71	.78
p-value[c]	.46	.15	.24

Note: The table describes the pairwise interrater agreement results for single imaging features based on assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets. GBCA-free dataset includes precontrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and postcontrast T1-weighted sequences.

Abbreviations: CI, confidence interval; EPDT, enhancement prediction decision tree.

[a]Unweighted Cohen's kappa.

[b]Prevalence-adjusted and bias-adjusted kappa.

[c]Comparison of the agreements between GBCA-enhanced and GBCA-free assessments.

[d]Weighted Cohen's kappa.

studies reported agreement values for these imaging features, ranging from .36 to .85 for diffusion[31–34] and .77 to .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, for example, when applied to other types of brain lesions than adult-type diffuse gliomas or when predicting enhancement characteristics in posttreatment scenarios, such as distinguishing treatment-related

changes from tumor progression. The posttreatment situation can lead to confounders, especially, for example, 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 posttreatment settings or for other tumor types, such as grade 1 and pediatric gliomas or

TABLE 11 Intrarater agreement results for assessment of single imaging features.

EPDT imaging features	Rater 1	Rater 2	Rater 3
Necrosis			
_[a]	.82; 95% CI: .75-89	.91; 95% CI: .87-96	.83; 95% CI: .77-.90
_[b]	.84	.92	.85
p-value[c]	.10		
Diffusion restriction[d]	.56; 95% CI: .49-.64	.75; 95% CI: .67-.82	.76; 95% CI: .70-.82
p-value[c]	.10		
T2 inhomogeneity			
_[a]	.45; 95% CI: .30-.60	.90; 95% CI: .79-.99	.67; 95% CI: .51-.82
_[b]	.78	.97	.91
p-value[c]	.22		
Nonenhancing tumor margin			
_[a]	.35; 95% CI: .18-.53	.86; 95% CI: .77-.93	.66; 95% CI: .55-.77
_[b]	.78	.93	.80
p-value[c]	.07		

Note: The table describes the intrarater agreement results for single imaging features between assessments of gadolinium-based contrast agent (GBCA)-enhanced and GBCA-free datasets per rater. GBCA-free dataset includes precontrast T1-weighted, T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted imaging sequences. GBCA-enhanced dataset includes GBCA-free and postcontrast T1-weighted sequences.
Abbreviations: CI, confidence interval; EPDT, enhancement prediction decision tree.
[a]Unweighted Cohen's kappa.
[b]Prevalence-adjusted and bias-adjusted kappa.
[c]Comparison of the agreements among all raters.
[d]Weighted Cohen's kappa.

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 technique 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, as well as assessing 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.

ORCID

Vera C. Keil https://orcid.org/0000-0001-8699-3506

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: a prospective, single institution, case series. Oper Neurosurg. 2021;20:151–63.

8. Haddad AF, Young JS, Morshed RA, et al. FLAIRectomy: resecting beyond the contrast margin for glioblastoma. Brain Sci. 2022;12:544.

9. Fraum TJ, Ludwig DR, Bashir MR, et al. Gadolinium-based contrast agents: a comprehensive risk assessment. J Magn Reson Imaging. 2017;46:338–53.

10. Blomqvist L, Nordberg GF, Nurchi VM, et al. 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, et al. Neuroimaging safety during pregnancy and lactation: a review. Neuroradiology. 2021;63:837–45.

13. 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.

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, et al. 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 from: https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project. Accessed 19 Dec 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, 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.

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 inter-observer 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, 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.

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.