Radiology: Artificial Intelligence

ORIGINAL RESEARCH

Accuracy of Radiomics in Predicting *IDH* Mutation Status in Diffuse Gliomas: A Bivariate Meta-Analysis

Gianfranco Di Salle, MD • Lorenzo Tumminello, MSc • Maria Elena Laino, MD • Sherif Shalaby, MD • Gayane Aghakhanyan, MD • Salvatore Claudio Fanni, MD • Maria Febi, MD • Jorge Eduardo Shortrede, PhD • Mario Miccoli, PhD • Lorenzo Faggioni, MD • Mirco Cosottini, MD • Emanuele Neri, MD

From Academic Radiology, Department of Translational Research on New Technologies in Medicine and Surgery (G.D.S., L.T., G.A., S.C.F., M.F., J.E.S., L.F., E.N.), Department of Clinical and Experimental Medicine (M.M.), and Neuroradiology Unit, Department of Translational Research on New Technologies in Medicine and Surgery (M.C.), University of Pisa, Via Roma 67, 56126 Pisa, Italy; Artificial Intelligence Center, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy (M.E.L.); The Shrewsbury and Telford Hospital NHS Trust, Shrewsbury, England (S.S.); and Italian Society of Medical and Interventional Radiology, SIRM Foundation, Milan, Italy (E.N.). Received November 22, 2022; revision requested January 9, 2023; revision received September 12; accepted October 24. **Address correspondence to** L.T. (email: *lorenzo.tumminello@unipi.it*).

Partial funding from the Horizon 2020 project PRIMAGE, grant agreement no. 826494.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2024; 6(1):e220257 • https://doi.org/10.1148/ryai.220257 • Content codes: AI NR

Purpose: To perform a systematic review and meta-analysis assessing the predictive accuracy of radiomics in the noninvasive determination of isocitrate dehydrogenase *(IDH)* status in grade 4 and lower-grade diffuse gliomas.

Materials and Methods: A systematic search was performed in the PubMed, Scopus, Embase, Web of Science, and Cochrane Library databases for relevant articles published between January 1, 2010, and July 7, 2021. Pooled sensitivity and specificity across studies were estimated. Risk of bias was evaluated using Quality Assessment of Diagnostic Accuracy Studies-2, and methods were evaluated using the radiomics quality score (RQS). Additional subgroup analyses were performed according to tumor grade, RQS, and number of sequences used (PROSPERO ID: CRD42021268958).

Results: Twenty-six studies that included 3280 patients were included for analysis. The pooled sensitivity and specificity of radiomics for the detection of *IDH* mutation were 79% (95% CI: 76, 83) and 80% (95% CI: 76, 83), respectively. Low RQS scores were found overall for the included works. Subgroup analyses showed lower false-positive rates in very low RQS studies (RQS < 6) (meta-regression, $z = -1.9$; $P = .02$) compared with adequate RQS studies. No substantial differences were found in pooled sensitivity and specificity for the pure grade 4 gliomas group compared with the all-grade gliomas group (81% and 86% vs 79% and 79%, respectively) and for studies using single versus multiple sequences (80% and 77% vs 79% and 82%, respectively).

Conclusion: The pooled data showed that radiomics achieved good accuracy performance in distinguishing *IDH* mutation status in patients with grade 4 and lower-grade diffuse gliomas. The overall methodologic quality (RQS) was low and introduced potential bias.

Supplemental material is available for this article.

Glioblastoma (GBM) is a highly lethal brain tumor and the most common and aggressive among diffuse gliomas. It belongs to class 4 in the World Health Organization classification of brain tumors, which includes the most biologically aggressive types that have inherent heterogeneity in histopathology, microscopic anatomy, and genetic features (1,2). GBM exhibits substantial genetic heterogeneity, with various genes associated with the disease (3) and others dividing patients into different prognostic subgroups according to their methylation status (4,5).

The mutation of isocitrate dehydrogenase (*IDH*), present in approximately 12% of grade 4 glioma cases, was initially considered a prognostic factor for GBM (6). It has been found to be associated with longer overall survival and better response to chemotherapy with temozolomide compared with patients with wild-type *IDH1* or *IDH2* (7). Thus, *IDH*-mutated grade 4 gliomas were classified as a separate category of GBM in the World Health Organization 2016 Classification of Tumors of the Central Nervous System and as astrocytomas in the 2021 update. Genotyping for *IDH* is therefore essential for diagnostic workup and prognostic evaluation of patients with high-grade gliomas and may play a role in selecting patients for targeted therapies. Indeed, preliminary evidence is emerging about *IDH*-specific therapeutic interventions in patients with gliomas (8–10), increasing the need for in vivo mutational assessment. However, noninvasive techniques for *IDH* genotyping are currently lacking, limiting their application in preoperative settings.

In recent years, radiomics has emerged as a quantitative approach of artificial intelligence for the analysis of imaging data, enabling the extraction of numerous features and their correlation with clinical information, potentially revolutionizing clinical decision-making (11). Recently, mathematical analysis instruments in radiomics have been used in multiple endeavors to noninvasively determine *IDH* mutation status in diffuse gliomas, primarily relying on baseline MRI and occasionally other imaging techniques. An initial comprehensive analysis of these studies was conducted in early 2020 (12), predating the publication of the revised glioma classification. Subsequently, the volume of evidence on this topic has more than doubled,

Abbreviations

AUC = area under the ROC curve, GBM = glioblastoma, *IDH* = isocitrate dehydrogenase, ROC = receiver operating characteristic, RQS = radiomics quality score

Summary

In this meta-analysis of 26 studies with 3280 patients, radiomics techniques achieved good accuracy performance in distinguishing isocitrate dehydrogenase mutation status in patients with grade 2–4 gliomas.

Key Points

- According to a meta-analysis of 26 studies, pooled sensitivity and specificity achieved in distinguishing isocitrate dehydrogenase mutation status in patients with diffuse gliomas were 81% and 79%, respectively.
- Low radiomics quality scores (RQS) were found overall (mean = 10.6 ± 3.3 [SD]), and it was observed that very low RQS scores influenced diagnostic accuracy metrics (false-positive rates meta-regression, z = -1.9; P = .02), leading to substantial overestimation.
- Subgroup analyses showed no evidence of differences in pooled sensitivity and specificity for the grade 4 gliomas group compared with the all-grade gliomas group (z = 0.125 and -0.013; P = .90 and .99, respectively).

Keywords

Neuro-Oncology, Radiomics, Integration, Application Domain, Glioblastoma, IDH Mutation, Radiomics Quality Scoring

indicating the necessity for an updated overview. In this study, we conducted a state-of-the-art systematic review and meta-analysis to assess the predictive accuracy of radiomics in the noninvasive determination of *IDH* mutation status in grade 4 and lower-grade diffuse gliomas.

Materials and Methods

Database Search Strategy

We performed a systematic search of PubMed, Scopus, Embase, Web of Science, and Cochrane Library databases for articles relevant to the application of radiomics in the noninvasive determination of *IDH* status in grade 4 gliomas. Details on the string used in the search can be found in Appendix S1. All reviews were performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, or PRISMA, 2020 guidelines (13). Since we expected heterogeneity of studies and low level of available evidence, we did not formulate PICO (population, intervention, comparison, and outcome) questions. This review was registered in the PROSPERO database (ID: CRD42021268958). All original research articles reporting on humans, written and published (including those distributed online first) from 2016 to July 2021 that respected the following criteria, were included: English language; studies including MRI, PET, or CT as imaging techniques; patients with a diagnosis of GBM; specified number of patients; and pathologic results proven with either surgery or biopsy. Exclusion criteria included review articles, conference papers and editorials or commentaries; patients younger than 18 years old; patients previously treated with surgery or radiation therapy; language other than English; no GBM diagnosis; no radiomics applications; no *IDH* mutation detection; and no statistical models for *IDH* mutational status assessment.

Importantly, the time lapse considered for inclusion of articles exactly matches the validity period of the World Health Organization 2016 classification. As the nomenclature of gliomas in publications was adapted to the World Health Organization 2016 classification, the search string was also worded accordingly.

To avoid neglecting literature data about GBM, we chose to include studies that had mixed cohorts of grade 2–4 gliomas in our meta-analysis when data about GBM were not dissociable from the rest of the cohort. A more conservative, grade 4–only subgroup meta-analysis was further performed (see below and Appendix S1).

First, two reviewers (radiologists G.D.S. with 2 years of experience and S.S. with 7 years of experience) independently screened the titles and abstracts following the inclusion and exclusion criteria. Discrepancies were resolved by a third reviewer (radiologist M.E.L. with 10 years of experience). In a second step, the reviewers retrieved the full-text articles of the selected abstracts and performed an independent second-step selection.

Quality Assessment

The quality of the included studies was evaluated by two readers (radiologists S.C.F. with 3 years of experience and M.F. with 1 year of experience) by using the Quality Assessment of Diagnostic Accuracy Studies-2 criteria, as proposed by Whiting et al (14), and the radiomics quality score (RQS), as proposed by Lambin et al (15).

Data Collection and Preparation

We collected the relevant data from the included full-text articles into an evidence table (Table). From each publication, we specified the following information: first author's name, year of the publication, number of patients, tumor grade, type of *IDH* mutation, and whether single or multiple sequences and/or modalities were used for radiomics feature extraction. Two reviewers independently conducted data extraction, and all discrepancies between them were resolved at a consensus meeting.

The standard 2 × 2 contingency table was constructed and included the true-positive, false-positive, false-negative, and true-negative values. When the data were insufficient to complete the 2 × 2 table, corresponding authors were contacted through email to request the masked data.

Each study was labeled as "low" or "good" quality based on its overall RQS, using the median value as cutoff. In addition, studies with RQS of less than 6 were categorized as "very low quality" studies, as compared with "adequate quality" studies. This data-driven categorization was determined by a subset of the RQS distribution that exhibited notably low scores (see Results section). We assessed tumor grade homogeneity as an additional factor by categorizing studies into those that incorporated only grade 4 gliomas versus those

Essential Histologic and Imaging Characteristics of Included Studies

Study	Patient Population			Type of *IDH* Mutation	Tumor Grade	Single or Multiple Sequences and/or Modality for Features Extraction	Imaging Modality*
	Total	*IDH* Mutation	*IDH* Wild Type				
Alis et al, 2020 (22)	142	48	94	IDH-1	3–4	Multiple	T2WI FLAIR, C+ T1WI, DWI
Chen et al, 2017 (23)	46	13	33	IDH-1	3–4	Multiple	RS-fMRI, DTI
Chen et al, 2018 (24)	46	13	33	IDH-1	3–4	Multiple	T1, RS-fMRI, DTI
Choi et al, 2020 (25)	136	15	121	IDH-1	4	Single	T2WI
Chougule et al, 2020 (26)	151	70	81	IDH-1	All	Multiple	C+ T1WI, T2WI, FLAIR
Han et al, 2019 (27)	42	21	21	IDH-1	2–4	Single	(C±T1WI), T2WI
Haubold et al, 2019 (28)	42	16	26	IDH-1	All	Multiple	C+ T1, FLAIR, MRF
Hsieh et al, 2017 (29)	39	7	32	IDH-1	4	Single	C+T1WI
Huang et al, 2021 (30)	38	8	30	IDH-1	All	Single	(C ± T1WI), T2WI, FLAIR
Kihira et al, 2021 (31)	111	23	88	IDH-1	2–4	Multiple	C+T1WI, FLAIR, DWI
Lewis et al, 2019 (32)	97	78	19	Not specified	2–4	Multiple	C ± T1WI, T2WI, ADC
Li et al, 2018 (33)	225	20	205	IDH-1	4	Multiple	C ± T1WI, T2WI, FLAIR
Li et al, 2019 (34)	127	51	76	IDH-1/2	2–4	Single	^{18}F-FDG PET
Lo et al, 2020 (35)	39	7	32	IDH-1	4	Single	C+T1WI
Lohmann et al, 2018 (36)	84	58	26	Not specified	2–4	Single	FET-PET/MRI
Lu et al, 2018 (37)	244	103	141	Not specified	2–4	Multiple	C+T1WI, T2, FLAIR, DWI
Niu et al, 2020 (38)	182	79	103	Not specified	3–4	Single	C+T1WI
Peng et al, 2021 (39)	105	50	55	Not specified	2–4	Multiple	C+T1WI, T2, ASL
Rathore et al, 2020 (40)	105	100	210	Not specified	2–4	Multiple	C ± T1WI, T2, FLAIR
Sakai et al, 2020 (41)	100	22	78	IDH-1	2–4	Single	C+T1WI, FLAIR, (DWI)
Santinha et al, 2021 (42)	77	51	26	IDH-1/2	2–4	Multiple	C ± T1WI, T2, FLAIR
Su et al, 2020 (43)	122	33	89	IDH-1	4	Single	(C ± T1WI), FLAIR
Sudre et al, 2020 (44)	333	151	182	IDH-1/2	2–4	Single	C-T1WI, T2WI, FLAIR, (DSC)
Verduin et al, 2021 (45)	185	17	168	IDH-1	4	Multiple	C+T1WI, T2WI
Wu et al, 2017 (46)	105	72	33	IDH-1	Not specified	Multiple	C+T1WI, FLAIR
Zhang et al, 2017 (47)	152	60	92	Not specified	Not specified	Multiple	C ± T1WI, T2WI, FLAIR

Note.—ADC = apparent diffusion coefficient, ASL = arterial spin labeling, DSC = dynamic susceptibility contrast, C+ = contrast-enhanced, C± = before and after contrast medium administration, DTI = diffusion tensor imaging, DWI = diffusion-weighted imaging, FET = O-(2-[18F]fuoroethyl)-L-tyrosine, FLAIR = fluid-attenuated inversion recovery, ^{18}F-FDG = fluorine 18–labeled fluorodeoxyglucose, *IDH* = isocitrate dehydrogenase, MRF = MR fingerprinting, RS-fMRI = resting-state functional MRI, T1WI = T1-weighted imaging, T2WI = T2-weighted imaging.

* In this column, parentheses indicate the modality used for the development of the radiomics model from which the data for this research was extracted. If there are no modalities within parentheses, it means that a combined model was developed and used.

including all-grade gliomas. Furthermore, we arranged studies according to the acquisition parameters, such as single versus multiple sequences and/or modalities, used for radiomics feature extraction.

Moreover, as we found potential dataset overlap among articles, we conducted a reduced-sample meta-analysis (as a sensitivity analysis [16]) using only the articles with an independent dataset—that is, without any potential overlap among training sets.

Statistical Analysis

The κ statistics were calculated to assess the agreement between the two raters during the full manuscript review process. We categorized the strength of agreement measured by the κ statistic as no (< 0.0), none to slight (0.0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.0) agreement.

Diagnostic forest plots of sensitivity and specificity were obtained for the included studies. As the forest plot is a univariate analysis, we additionally explored the bivariate relationship and variation between sensitivity and specificity estimates using crosshairs plot (17) and ROCellipse plot (18). Bivariate meta-analysis (19) with a random-effects model was used to estimate pooled sensitivities and specificities across studies. This model jointly analyzes the pairs of logit-transformed sensitivity and specificity from studies, incorporating the inherent correlation between them and reducing the potential bias in the estimate of the CIs and heterogeneity. The summary receiver operating characteristic (ROC) curve was derived, and the area under the ROC curve (AUC) was estimated. The pooled sensitivity and specificity, the positive and negative likelihood ratios, and the diagnostic odds ratios (20) were obtained with the 95% CI estimates.

The χ^2 tests were performed on the original data to separately assess heterogeneity of sensitivities and specificities. To address the impact of heterogeneity on the meta-analysis (21), we also assessed I^2 and Cochran Q of positive and negative likelihood ratios and the associated diagnostic odds ratios (20). Subgroup analyses were performed to qualitatively highlight differences in the pooled outcomes in settings with different RQS values (low or good and very low or adequate, as previously defined above), tumor grades (grade 4 vs mixed grade) and acquisition parameters (single vs multiple sequences and/or modalities used for radiomics feature extraction).

Additionally, meta-regression analyses were performed to quantify the effect of these categorical covariates and of RQS considered as a continuous variable using separate regression models. A meta-regression was also used to investigate the relationship between study sample size and accuracy outcomes. The comparative summary ROC curves were assessed.

$P < .05$ was considered statistically significant. Statistical analyses were performed with the open-source package *mada* (18) written in R software (R Project for Statistical Computing) using RStudio, version 1.4.1106 (Posit) (18).

Results

Study Selection

We initially identified 919 relevant articles. After removing duplicates, 577 titles and abstracts were reviewed, and 531 studies were excluded according to the eligibility criteria. The remaining 46 studies were potentially appropriate and were assessed for eligibility according to the inclusion criteria. After the full-text review, 33 articles were considered eligible for the meta-analysis. During data extraction, we contacted the authors of nine articles to obtain missing data. As we received complete answers in only two cases, we excluded the remaining seven due to insufficient data. Finally, 26 studies were selected for this meta-analysis (Fig 1).

Cohen κ was calculated to assess the agreement between two raters during the full manuscript review process and showed an almost perfect agreement (κ = 0.93).

Study Characteristics

A total of 3280 patients were described in the 26 included studies, including training, validation, and test sets. The different models were trained on 2527 patients, with a median 82 patients (IQR: 45.5–127). All studies included patients with GBM as defined in the World Health Organization 2016 classification; 18 (69.6%) also included grade 3 gliomas, 14 (53.9%) also included grade 2 gliomas, and three (11.5%) included all four grades of gliomas. In two studies, the grade of lower-grade gliomas was not specified. The isoform of *IDH* tested for mutation was specified in 19 studies (73%): both *IDH1* and *IDH2* in three studies (11.5% [three of 26]) and only *IDH1* in 16 studies (61.5% [16 of 26]).

Quality Assessment

The total RQS was calculated for each article and each component. The RQS (mean ± SD) was 10.6 ± 3.3 (29.4% of the possible maximum value of 36); the maximum score was 14 (38.9%), and the minimum score was 4 (11%). Importantly, five articles had scores of 4–5, while all the others scored 9 or higher. This observation led to categorization of this subgroup as a very low quality RQS subgroup. The RQS basic adherence rates of the studies were calculated according to the six key domains proposed by Park et al (48). Details about the adherence to the single domains are reported in Appendix S1 (see also Table S1).

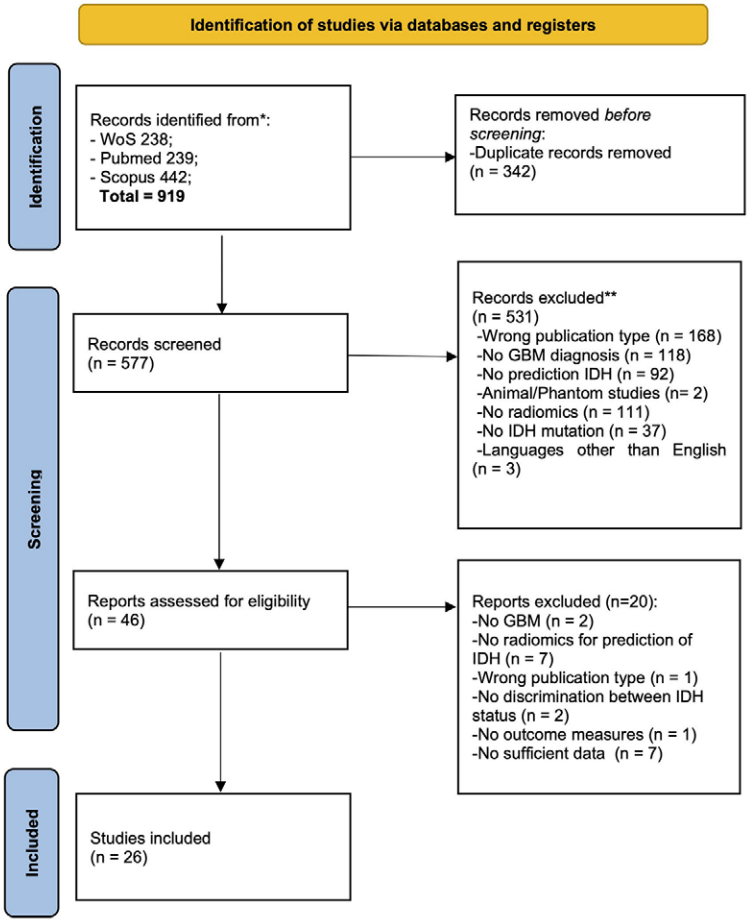
Data about the risk of bias and applicability concerns, calculated according to the Quality Assessment of Diagnostic Accuracy Studies-2, for the 26 studies are reported in Appendix S1 (see also Fig S1 and Table S2).

Diagnostic Accuracy of Radiomics in Predicting *IDH* Mutation

Paired diagnostic forest plots of sensitivity and specificity, the weighted crosschair plot, and the ROCellipse plot of the 26 included studies are shown in Figure 2, Figure S2A, and Figure S2B, respectively. The pooled sensitivity of radiomics for the detection of *IDH* mutation was 79% (95% CI: 76, 83) with heterogeneity ($\chi^2 = 49$; $P < .001$), and the pooled specificity was 80% (95% CI: 76, 83) with heterogeneity ($\chi^2 = 59.4$; $P < .001$). The overall positive likelihood ratio was 3.9 (95% CI: 3.3, 4.6), and the negative likelihood ratio was 0.28 (95% CI: 0.23, 0.34). The pooled diagnostic odds ratio was 17.8 (95% CI: 12, 26). Figure 3 shows the summary ROC curve of pooled sensitivity and specificity with an AUC of 0.85, indicating good performance for the prediction of *IDH* mutation.

Heterogeneity

The χ^2 test provided evidence of significant heterogeneity for sensitivity ($\chi^2 = 49$; $df = 24$; $P < .001$) and specificity ($\chi^2 =$



*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Figure 1: Flow diagram for selection pipeline according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 statement. GBM = glioblastoma, *IDH* = isocitrate dehydrogenase, WoS = Web of Science. (Adapted, under a CC BY 4.0 license, from reference 13.)

59.4; *df* = 24; *P* < .001). For positive and negative likelihood ratio and diagnostic odds ratios, Cochran *Q* was 28.3 (*df* = 25; *P* = .30), 21.7 (*df* = 25; *P* = .65), and 22.5 (*df* = 25; *P* = .61), respectively; Higgins I^2 was 11.5%, 0%, and 0%, respectively.

Subgroup Analyses

Radiomics quality score.— In low RQS settings (lower half of the distribution, 13 primary studies), the pooled sensitivity for detecting *IDH* mutations was 80% (95% CI: 72, 86) with heterogeneity (χ^2 = 21; *P* = .048), and the pooled specificity was 83% (95% CI: 78, 87) without heterogeneity (χ^2 = 17; *P* = .14). In good RQS settings (13 primary studies), the pooled sensitivity for detecting *IDH* mutation was 78% (95% CI: 70, 84) with heterogeneity (χ^2 = 25; *P* = .02), and the pooled specificity was 78% (95% CI: 74, 82) with heterogeneity (χ^2 = 39; *P* < .001). Figure 4A depicts the comparison of low versus good RQS, with summary ROC curves showing that the summary estimates for both groups are separated, but the confidence regions overlap. The calculated AUC for low and good RQS was 85% and 84%, respectively.

We conducted additional analyses to compare very low versus adequate RQS studies. In studies with very low RQS (five primary studies), the pooled sensitivity for detecting *IDH* mutation was 87% (95% CI: 69, 95) with heterogeneity (χ^2 = 11; *P* = .03), and the pooled specificity was 87% (95% CI: 78, 92) without heterogeneity (χ^2 = 4.4; *P* = .35). In studies with adequate RQS (21 primary studies), the pooled sensitivity for detecting *IDH* mutation was 77% (95% CI: 72, 82) with heterogeneity (χ^2 = 31.5; *P* = .048), and the pooled specificity was

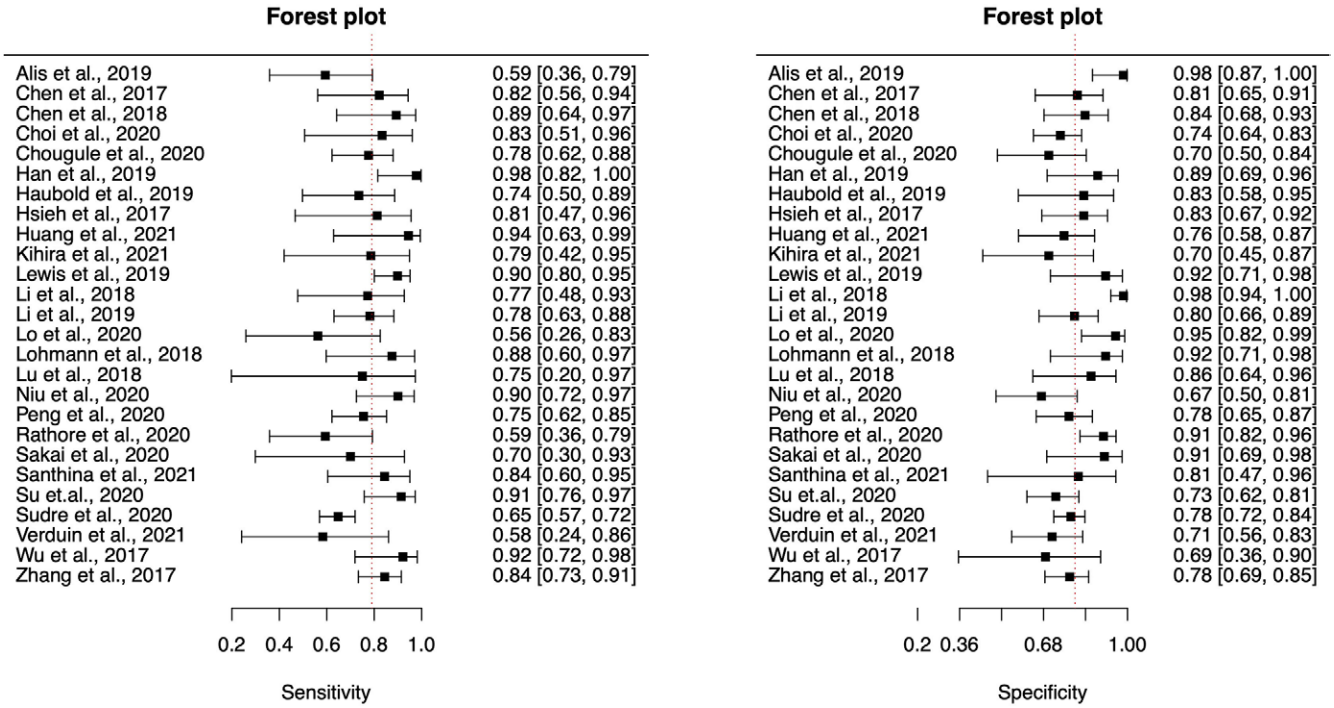


Figure 2: Forest plots of sensitivity and specificity with 95% CIs per study. Vertical red dashed lines denote summary estimates of sensitivity and specificity.

78% (95% CI: 75, 81) with heterogeneity (χ^2 = 51; P < .001). Figure 4B depicts the comparison of very low versus adequate RQS, with summary ROC curves showing that the summary estimates and confidence regions for both groups are well separated. The AUC was 0.92 for very low RQS and 0.83 for adequate RQS.

Meta-regression using RQS as a continuous covariate showed no evidence of association with the pooled outcomes. Subsequent analyses were carried out based on categorical RQS subgroups. Meta-regression using very low versus adequate RQS as a categorical covariate showed a significant regression coefficient for the false-positive rates (low RQS, z = -1.9; P = .02), indicating that the false-positive rates are lower for the very low RQS studies and higher for the studies with adequate RQS (Table S3). Meta-regression using low versus good RQS as a covariate showed borderline regression coefficient for the false-positive rates (low RQS, z = -1.7; P = .085) (Table S3).

Tumor grade, number of sequences and/or modalities, and patient overlap.— The additional subgroup analyses showed no significant association between tumor grade, number of sequences or modalities used, and the pooled sensitivity and specificity. A sensitivity analysis with only the articles without any potential dataset overlap was substantially similar to the comprehensive analysis. Detailed description of these results is available in Appendix S1.

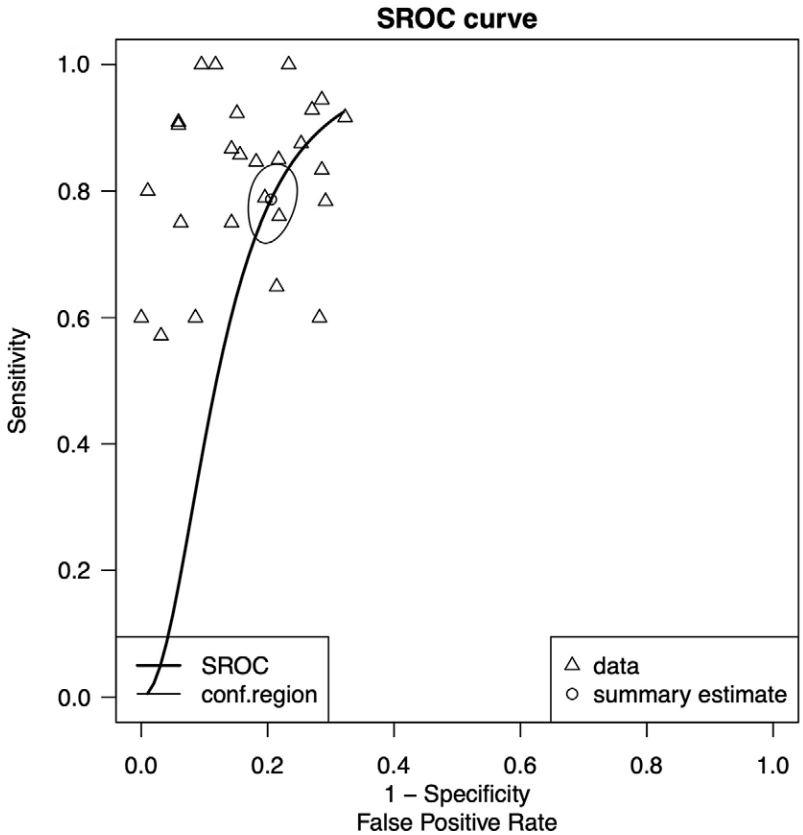


Figure 3: Graph of summary receiver operating characteristic (SROC) curve of pooled sensitivity and specificity of all studies included in the meta-analysis (26 studies), with area under the ROC curve of 0.85, indicating good performance of radiomics analysis for predicting isocitrate dehydrogenase mutation. Conf = confidence.

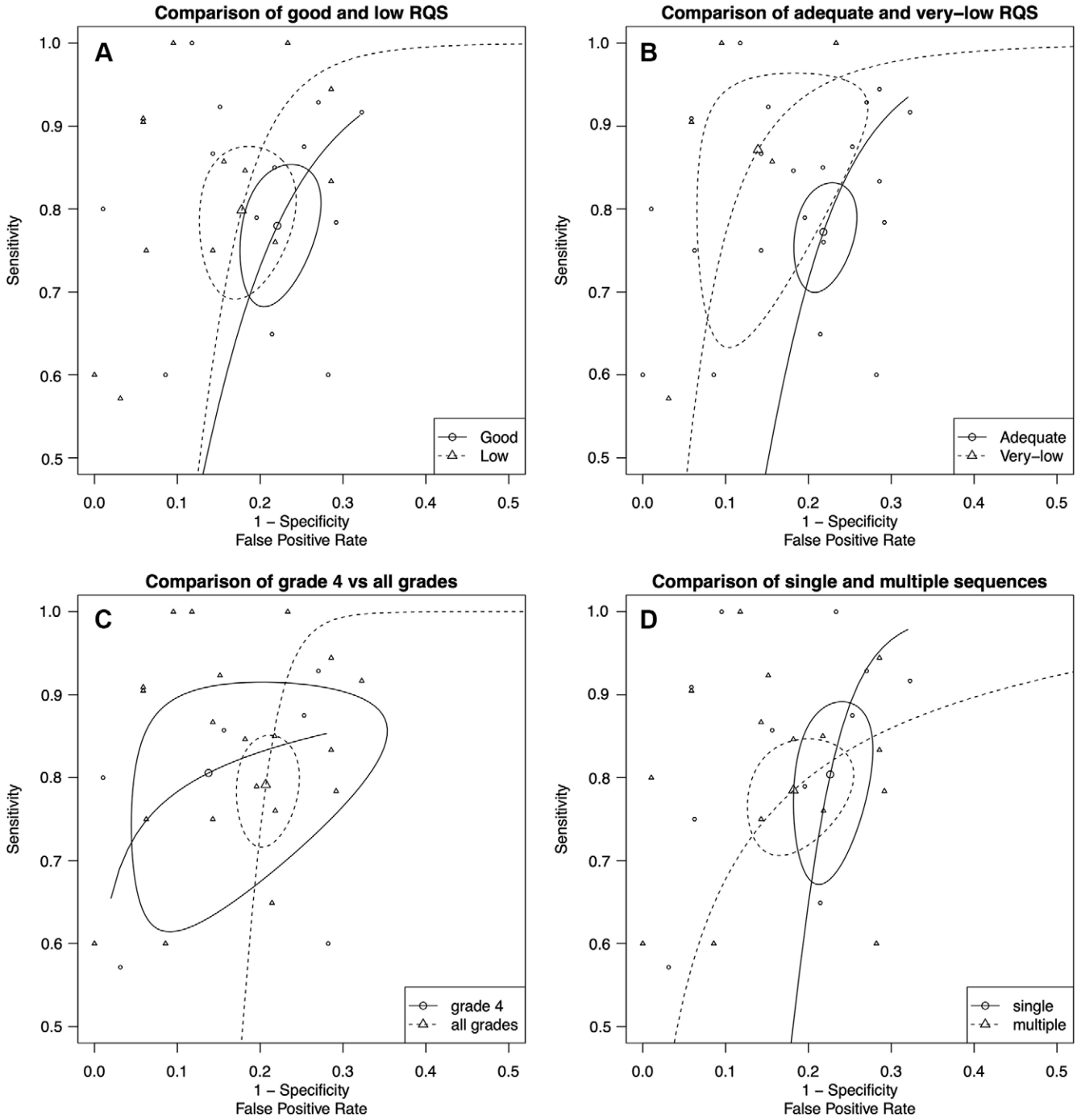


Figure 4: **(A)** Comparison of low and good radiomics quality score (RQS) studies with summary receiver operating characteristic (ROC) curves showing that the summary estimates for both groups are separated, but the confidence regions are overlapped. The area under the ROC curve (AUC) for predicting isocitrate dehydrogenase mutation was 0.85 for low RQS and 0.84 for good RQS. **(B)** Comparison of very low and adequate RQS studies with summary ROC curves showing that the summary estimates for both groups are well separated, but the confidence regions are slightly overlapped. The AUC was 0.92 for low RQS and 0.83 for high RQS. **(C)** Comparison of grade 4 and mixed-grade (2–4) glioma studies with summary ROC curves. Note the overlapping summary estimates and confidence regions for both groups. The AUC was 0.87 for grade 4 gliomas and 0.82 for mixed-grade gliomas. **(D)** Comparison of summary ROC curves for single versus multiple sequences and/or modalities. Note the distinct overlap of the summary estimates and confidence regions. Nonetheless, slightly higher false-positive rates were detected in studies using modalities with multiple sequences. The AUC was 0.88 for single sequence modalities and 0.79 for multiple sequence modalities.

Discussion

We conducted a bivariate meta-analysis to leverage the pooled sensitivity and specificity of diagnostic accuracy studies that applied radiomics to predict *IDH* mutation status in diffuse gliomas. Our results indicate that radiomics techniques achieve good accuracy performance in distinguishing *IDH* mutation status in patients with diffuse gliomas, with pooled sensitivity and specificity of 81% and 79%, respectively. Tumor grade, number of sequences or modalities, and methodologic quality according to the RQS were identified as potentially prominent

sources of variability and were used as criteria for conceptualizing the subgroup analysis. Among these, only the RQS was found to be a potential source of heterogeneity.

Pooled outcomes measured in our meta-analysis were lower than reported in a previous work. In the span of 2 years from the only other meta-analysis about radiomics-based *IDH* prediction in gliomas (12), the number of articles eligible for such analysis has almost tripled, with sensitivity decreasing from 88% to 79% and specificity decreasing from 87% to 80%. A plausible explanation of this decrease might be that our meta-analysis was grade 4 oriented, with systematic exclusion of studies including only grade 1–3 gliomas. Grade 4 gliomas, with a dramatically lower prevalence of *IDH* mutations compared with lower grades (49), might have introduced class imbalance-related issues, with variations in the sampling strategies affecting final accuracy measurements (50). Additionally, the number of articles lacking model validation was four of nine (44%) in Zhao et al (12) and four of 26 (15%) in the present work. A separate validation set is essential to avoid overfitting and outcome metric overestimation. In this respect, the negative association we found between sample size and sensitivity suggests that results are more influenced by overfitting than by the benefit of training set expansion. Last, we performed subgroup analyses using data from training sets, although the sample size was admittedly too low, leading to possibly unreliable conclusions. Contrary to the previous study, our subgroup analysis was conducted using outcomes from validation sets. Comparing the outcomes of pure grade 4 versus all-grade studies suggested that the inclusion of articles with mixed cohorts did not substantially alter the main outcomes. Similarly to Spadarella et al (51), we found globally low RQS scores (15) and hypothesized that methodologic quality might introduce a bias in the pooled outcomes. In the absence of established cutoffs for the RQS, we categorized our cohort based on data-driven hypotheses, and we found that an RQS of less than 6 (corresponding with the very low quality subgroup) predicts higher specificity, suggesting possible overestimation. Despite continuous improvements of the data analysis algorithms, witnessed by the overall increase in adherence to domain 4 of RQS, organizational milestones, such as the availability of multicenter data and prospective study designs, are still insufficient and therefore hamper research quality in radiomics. A last subgroup analysis, based on training data for algorithms, found no macroscopic differences between single and multiple sequence radiomic pipelines, possibly due to the absence of widespread standards for radiomics pipelines and inappropriate data dimensionality reduction techniques (52).

Our results showed that radiomics techniques do not outperform advanced imaging in accurately predicting *IDH* mutational status in gliomas. A meta-analysis by Suh et al (53) showed that a combination of diffusion- and perfusion-weighted MRI allows for the noninvasive prediction of *IDH* mutational status in all-grade gliomas with a summary sensitivity of 86% and specificity of 87%. Sensitivity was further enhanced in a subgroup using 2-hydroxyglutarate MR spectroscopy, which detects a metabolite selectively accumulated in *IDH*-mutant gliomas. Compared with MR spectroscopy, radiomics does not require 2-hydroxyglutarate-specific sequences optimization and complex postprocessing implementation; furthermore, effective algorithms could be widely available in centers without MR spectroscopy expertise. The lack of incremental benefit from sporadic use of perfusion- and diffusion-weighted imaging in the studies included in the present work, compared with the qualitative evaluation of these sequences made in Suh et al (53), should be reevaluated in future studies in light of more extensive evidence. Similarly, good *IDH* discrimination performances have also been reported for nonradiomics artificial intelligence applications, such as deep learning (54) and liquid biopsy (55). Deep learning techniques can achieve high diagnostic accuracy, but current state-of-the-art applications lack explainability and work as "black boxes," thus raising doubts about biologic significance and generalizability. Unlike radiomics, no distinct features can be isolated, compared across studies, or assessed for potential clinical significance in future research. Likewise, though promising, liquid biopsy must face challenges such as reagents availability in clinical routine and difficulty in obtaining cerebrospinal fluid samples in a preoperative setting (55).

This study had limitations. First, due to the limited number of studies with a pure grade 4 cohort, we included studies with both grade 4 and lower-grade gliomas. Second, substantial variability existed in the imaging protocols and MRI sequences used. Our subgroup analysis oversimplified this variability into a categorical distinction (single vs multiple sequences and/or modality). Additionally, the validation pipeline and machine learning technique to fit the final models were highly variable among the studies. Based on the numerosity of each study sample, different strategies of model validation were used, ranging from leave-one-out cross-validation to independent test sets. Therefore, differential propensity to overfitting is likely to have influenced the estimated effect magnitude. The reliance on the RQS as the only available tool for the quality assessment and the data-driven approach in choosing the cutoffs for the subgroup analysis posed further limitations to our study. Last, there was potential dataset overlap among the included primary studies. Based on the anonymized, geographical information provided by the single studies, we inferred that patients' data were not used in multiple studies. Yet, eight articles included anonymized data from The Cancer Imaging Archive dataset in their training set. A sensitivity analysis conducted without this subset was not substantially different from the results of our comprehensive database. The literature provides limited evidence regarding a standardized and rigorous method to handle overlapping datasets or patients in meta-analyses, especially in the radiomics field. While recommending caution in the interpretation of the results, we nonetheless chose not to exclude articles with potentially overlapping datasets, both to avoid arbitrary decisions and to preserve the information provided by different radiomic pipelines.

In conclusion, radiomics is a promising tool for the determination of *IDH* mutational status in grade 4 and lower-grade diffuse gliomas. However, in recent years, the performance of radiomics-based algorithms has not improved to the point of overcoming conventional approaches, limiting their widespread use in clinical routine. Missed compliance to several quality criteria is a further remarkable caveat for the translation of the predictive algorithms into clinical practice.

Author contributions: Guarantors of integrity of entire study, **G.D.S., L.T., M.M., L.F., E.N.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **G.D.S., L.T., S.S., S.C.F., J.E.S., L.F., E.N.**; clinical studies, **S.S., J.E.S.**; statistical analysis**, L.T., G.A., S.C.F., M.M.**; and manuscript editing, **G.D.S., L.T., G.A., S.C.F., M.F., J.E.S., L.F., M.C., E.N.**

Disclosures of conflicts of interest: G.D.S. No relevant relationships. **L.T.** No relevant relationships. **M.E.L.** No relevant relationships. **S.S.** No relevant relationships. **G.A.** No relevant relationships. **S.C.F.** No relevant relationships. **M.F.** No relevant relationships. **J.E.S.** No relevant relationships. **M.M.** Statistical associate editor of the journal *Heroin Addiction and Related Clinical Problems* and of the journal *Updates Surgery*. **L.F.** No relevant relationships. **M.C.** No relevant relationships. **E.N.** Consultant to the editor for *Radiology: Artificial Intelligence*.

References

1. 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.
2. Holland EC. Glioblastoma multiforme: the terminator. Proc Natl Acad Sci USA 2000;97(12):6242–6244.
3. Urbańska K, Sokołowska J, Szmidt M, Sysa P. Glioblastoma multiforme - an overview. Contemp Oncol (Pozn) 2014;18(5):307–312.
4. Shah N, Lin B, Sibenaller Z, et al. Comprehensive Analysis of MGMT Promoter Methylation: Correlation with MGMT Expression and Clinical Response in GBM. PLoS One 2011;6(1):e16146.
5. Mosrati MA, Malmström A, Lysiak M, et al. TERT promoter mutations and polymorphisms as prognostic factors in primary glioblastoma. Oncotarget 2015;6(18):16663–16673.
6. Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. Science 2008;321(5897):1807–1812.
7. Qi ST, Yu L, Gui S, et al. IDH mutations predict longer survival and response to temozolomide in secondary glioblastoma. Cancer Sci 2012;103(2):269–273.7
8. Pirozzi CJ, Yan H. The implications of IDH mutations for cancer development and therapy. Nat Rev Clin Oncol 2021;18(10):645–661.
9. Platten M, Bunse L, Wick A, et al. A vaccine targeting mutant IDH1 in newly diagnosed glioma. Nature 2021;592(7854):463–468.
10. Li S, Wang C, Chen J, et al. Signaling pathways in brain tumors and therapeutic interventions. Signal Transduct Target Ther 2023;8(1):8.
11. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. Insights Imaging 2020;11(1):91.
12. Zhao J, Huang Y, Song Y, et al. Diagnostic accuracy and potential covariates for machine learning to identify IDH mutations in glioma patients: evidence from a meta-analysis. Eur Radiol 2020;30(8):4664–4674.
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
14. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–536.
15. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–762.
16. Lunny C, Pieper D, Thabet P, Kanji S. Managing overlap of primary study results across systematic reviews: practical considerations for authors of overviews of reviews. BMC Med Res Methodol 2021;21(1):140.
17. Phillips B, Stewart LA, Sutton AJ. 'Cross hairs' plots for diagnostic meta-analysis. Res Synth Methods 2010;1(3-4):308–315.
18. Doebler P, Holling H, Sousa-Pinto B. Meta-Analysis of Diagnostic Accuracy with mada. https://cran.r-project.org/web/packages/mada/vignettes/mada.pdf. Published 2012. Accessed September 2021.
19. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58(10):982–990.
20. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56(11):1129–1135.
21. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309(6965):1351–1355.
22. Alis D, Bagcilar O, Senli YD, et al. Machine learning-based quantitative texture analysis of conventional MRI combined with ADC maps for assessment of IDH1 mutation in high-grade gliomas. Jpn J Radiol 2020;38(2):135–143.
23. Chen L, Zhang H, Thung KH, et al. Multi-label Inductive Matrix Completion for Joint MGMT and IDH1 Status Prediction for Glioma Patients. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, eds. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2017. Vol 10434. Springer; 2017:450–458.
24. Chen L, Zhang H, Lu J, et al. Multi-Label Nonlinear Matrix Completion With Transductive Multi-Task Feature Selection for Joint MGMT and IDH1 Status Prediction of Patient With High-Grade Gliomas. IEEE Trans Med Imaging 2018;37(8):1775–1787.
25. Choi Y, Nam Y, Lee YS, et al. IDH1 mutation prediction using MR-based radiomics in glioblastoma: comparison between manual and fully automated deep learning-based approach of tumor segmentation. Eur J Radiol 2020;128:109031.
26. Chougule T, Shinde S, Santosh V, Saini J, Ingalhalikar M. On Validating Multimodal MRI Based Stratification of IDH Genotype in High Grade Gliomas Using CNNs and Its Comparison to Radiomics. In: Mohy-ud-Din H, Rathore S, eds. Radiomics and Radiogenomics in Neuro-oncology. Vol 11991. Springer;2020:53–60.
27. Han L, Wang S, Miao Y, et al. MRI texture analysis based on 3D tumor measurement reflects the IDH1 mutations in gliomas - A preliminary study. Eur J Radiol 2019;112:169–179.
28. Haubold J, Demircioglu A, Gratz M, et al. Non-invasive tumor decoding and phenotyping of cerebral gliomas utilizing multiparametric 18F-FET PET-MRI and MR Fingerprinting. Eur J Nucl Med Mol Imaging 2020;47(6):1435–1445.
29. Hsieh KLC, Chen CY, Lo CM. Radiomic model for predicting mutations in the isocitrate dehydrogenase gene in glioblastomas. Oncotarget 2017;8(28):45888–45897.
30. Huang WY, Wen LH, Wu G, et al. Comparison of Radiomics Analyses Based on Different Magnetic Resonance Imaging Sequences in Grading and Molecular Genomic Typing of Glioma. J Comput Assist Tomogr 2021;45(1):110–120.
31. Kihira S, Tsankova NM, Bauer A, et al. Multiparametric MRI texture analysis in prediction of glioma biomarker status: added value of MR diffusion. Neurooncol Adv 2021;3(1):vdab051.
32. Lewis MA, Ganeshan B, Barnes A, et al. Filtration-histogram based magnetic resonance texture analysis (MRTA) for glioma IDH and 1p19q genotyping. Eur J Radiol 2019;113:116–123.
33. Li ZC, Bai H, Sun Q, et al. Multiregional radiomics profiling from multiparametric MRI: Identifying an imaging predictor of IDH1 mutation status in glioblastoma. Cancer Med 2018;7(12):5999–6009.
34. Li L, Mu W, Wang Y, et al. A Non-invasive Radiomic Method Using 18F-FDG PET Predicts Isocitrate Dehydrogenase Genotype and Prognosis in Patients With Glioma. Front Oncol 2019;9:1183.
35. Lo CM, Weng RC, Cheng SJ, Wang HJ, Hsieh KLC. Computer-aided diagnosis of isocitrate dehydrogenase genotypes in glioblastomas from radiomic patterns. Medicine (Baltimore) 2020;99(8):e19123.
36. Lohmann P, Lerche C, Bauer EK, et al. Predicting IDH genotype in gliomas using FET PET radiomics. Sci Rep 2018;8(1):13328.
37. Lu CF, Hsu FT, Hsieh KLC, et al. Machine Learning-Based Radiomics for Molecular Subtyping of Gliomas. Clin Cancer Res 2018;24(18):4429–4436.
38. Niu L, Feng WH, Duan CF, Liu YC, Liu JH, Liu XJ. The Value of Enhanced MR Radiomics in Estimating the IDH1 Genotype in High-Grade Gliomas. Biomed Res Int 2020;2020:4630218.
39. Peng H, Huo J, Li B, et al. Predicting Isocitrate Dehydrogenase (IDH) Mutation Status in Gliomas Using Multiparameter MRI Radiomics Features. J Magn Reson Imaging 2021;53(5):1399–1407.
40. Rathore S, Mohan S, Bakas S, et al. Multi-institutional noninvasive in vivo characterization of IDH, 1p/19q, and EGFRvIII in glioma using neuro-Cancer Imaging Phenomics Toolkit (neuro-CaPTk). Neurooncol Adv 2021;2(Suppl 4):iv22–iv34.
41. Sakai Y, Yang C, Kihira S, et al. MRI Radiomic Features to Predict IDH1 Mutation Status in Gliomas: A Machine Learning Approach using Gradient Tree Boosting. Int J Mol Sci 2020;21(21):8004.
42. Santinha J, Matos C, Figueiredo M, Papanikolaou N. Improving performance and generalizability in radiogenomics: a pilot study for prediction of IDH1/2 mutation status in gliomas with multicentric data. J Med Imaging (Bellingham) 2021;8(3):031905.
43. Su X, Sun H, Chen N, et al. A radiomics-clinical nomogram for preoperative prediction of IDH1 mutation in primary glioblastoma multiforme. Clin Radiol 2020;75(12):963.e7–963.e15.
44. Sudre CH, Panovska-Griffiths J, Sanverdi E, et al. Machine learning assisted DSC-MRI radiomics as a tool for glioma classification by grade and mutation status. BMC Med Inform Decis Mak 2020;20(1):149.
45. Verduin M, Primakov S, Compter I, et al. Prognostic and Predictive Value of Integrated Qualitative and Quantitative Magnetic Resonance Imaging Analysis in Glioblastoma. Cancers (Basel) 2021;13(4):722.

46. Wu G, Chen Y, Wang Y, et al. Sparse Representation-Based Radiomics for the Diagnosis of Brain Tumors. IEEE Trans Med Imaging 2018;37(4):893–905.
47. Zhang X, Tian Q, Wu YX, et al. IDH mutation assessment of glioma using texture features of multimodal MR images. In: Armato SG, Petrick NA, eds. Proceedings. Vol 10134. SPIE; 2017.
48. Park JE, Kim HS, Kim D, et al. A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features. BMC Cancer 2020;20(1):29.
49. Eckel-Passow JE, Lachance DH, Molinaro AM, et al. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. N Engl J Med 2015;372(26):2499–2508.
50. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intell Data Anal 2002;6(5):429–449.
51. Spadarella G, Stanzione A, Akinci D'Antonoli T, et al. Systematic review of the radiomics quality score applications: an EuSoMII Radiomics Auditing Group Initiative. Eur Radiol 2023;33(3):1884–1894.
52. Lecler A, Duron L, Balvay D, et al. Combining Multiple Magnetic Resonance Imaging Sequences Provides Independent Reproducible Radiomics Features. Sci Rep 2019;9(1):2068.
53. Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: a systemic review and meta-analysis. Eur Radiol 2019;29(2):745–758.
54. Zlochower A, Chow DS, Chang P, Khatri D, Boockvar JA, Filippi CG. Deep Learning AI Applications in the Imaging of Glioma. Top Magn Reson Imaging 2020;29(2):115–121.
55. Fujioka Y, Hata N, Akagi Y, et al. Molecular diagnosis of diffuse glioma using a chip-based digital PCR system to analyze IDH, TERT, and H3 mutations in the cerebrospinal fluid. J Neurooncol 2021;152(1):47–54.