RESEARCH ARTICLE

Open Access

A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features

Ji Eun Park[1], Ho Sung Kim[1*] , Donghyun Kim[2], Seo Young Park[3], Jung Youn Kim[4], Se Jin Cho[1] and Jeong Hoon Kim[5]

Abstract

Background: To evaluate radiomics analysis in neuro-oncologic studies according to a radiomics quality score (RQS) system to find room for improvement in clinical use.

Methods: Pubmed and Embase were searched up the terms radiomics or radiogenomics and gliomas or glioblastomas until February 2019. From 189 articles, 51 original research articles reporting the diagnostic, prognostic, or predictive utility were selected. The quality of the methodology was evaluated according to the RQS. The adherence rates for the six key domains were evaluated: image protocol and reproducibility, feature reduction and validation, biologic/clinical utility, performance index, a high level of evidence, and open science. Subgroup analyses for journal type (imaging vs. clinical) and biomarker (diagnostic vs. prognostic/predictive) were performed.

Results: The median RQS was 11 out of 36 and adherence rate was 37.1%. Only 29.4% performed external validation. The adherence rate was high for reporting imaging protocol (100%), feature reduction (94.1%), and discrimination statistics (96.1%), but low for conducting test-retest analysis (2%), prospective study (3.9%), demonstrating potential clinical utility (2%), and open science (5.9%). None of the studies conducted a phantom study or cost-effectiveness analysis. Prognostic/predictive studies received higher score than diagnostic studies in comparison to gold standard ($P < .001$), use of calibration ($P = .02$), and cut-off analysis ($P = .001$).

Conclusions: The quality of reporting of radiomics studies in neuro-oncology is currently insufficient. Validation is necessary using external dataset, and improvements need to be made to feature reproducibility, demonstrating clinical utility, pursuits of a higher level of evidence, and open science.

Keywords: Radiomics, Neuro-oncology, Quality, Score

---

* Correspondence: radhskim@gmail.com
[1]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 43 Olympic-ro 88, Songpa-Gu, Seoul 05505, South Korea
Full list of author information is available at the end of the article

Background

Radiomics is a powerful tool for developing and testing medical hypotheses, involving the use of high-dimensional quantitative imaging features for predictive purposes. The extraction of high-throughput quantitative features and the use of sophisticated bioinformatics tools enables the development of models with potential diagnostic, prognostic, or predictive utility in cancer studies [1–3]. In the field of neuro-oncology, a large number of radiomics studies have demonstrated their diagnostic, prognostic, and predictive use in differential diagnosis [4, 5], molecular classification [6–8], survival analysis [9, 10], and treatment response to antiangiogenic treatment [11].

Although radiomics research shows great potential, its current use is rather confined to the academic literature, without real-world clinical applications. This is in part due to a lack of efficient and effective strategies for biomarker translation [12], which hampers the effective development of radiomics as an imaging biomarker to cross the 'translational gap' for use in guiding clinical decisions [13, 14]. A standardized evaluation of the performance, reproducibility, and/or clinical utility of radiomics biomarkers is needed; with regard to the great need for qualified reporting, a system of metrics to determine the validity and completeness of radiomics studies was developed by Lambin et al. [2] in the form of the radiomics quality score (RQS). The RQS is comprised of 16 components, chosen to emulate the Transparent Reporting of a multivariable prediction model for Individual Prognosis OR Diagnosis (TRIPOD) initiative [15]. These are applied to a radiomics-specific design that considers high-dimensional data and modeling, and emphasizes clinical adoption of modeling research as in the TRIPOD guidelines.

A subsequent RQS study from the developer [3] reported an average score less than 50% over 41 radiomics studies using various modalities and conditions, including ultrasound, computed tomography, positron emission tomography, and magnetic resonance imaging (MRI). However, the results do not represent the field of neuro-oncology as both disease and imaging modality varies. Most of neuro-oncologic imaging studies are based on MRI, which is particularly challenging in generalizability and robustness in radiomics analysis as it has non-standardized pixel-values and large variations in signal intensities. To our knowledge, the quality of the science and reporting in radiomics research studies in the neuro-oncologic imaging is largely unknown.

In this study, we evaluated the radiomics analysis conducted in previous publications, summarizing six domains from the RQS: image protocol and feature reproducibility, feature reduction and validation, biologic/clinical validation and utility, performance index, high level of evidence, and open science. Our intention was to promote the quality of radiomics research studies as diagnostic, prognostic, and/or predictive biomarkers, to allow radiomics to become a viable tool for medical decision-making by facilitating the combined analysis of clinical data and high-throughput imaging features. The purpose of our study was to evaluate the quality of reporting radiomics-based analysis in neuro-oncologic studies using RQS.
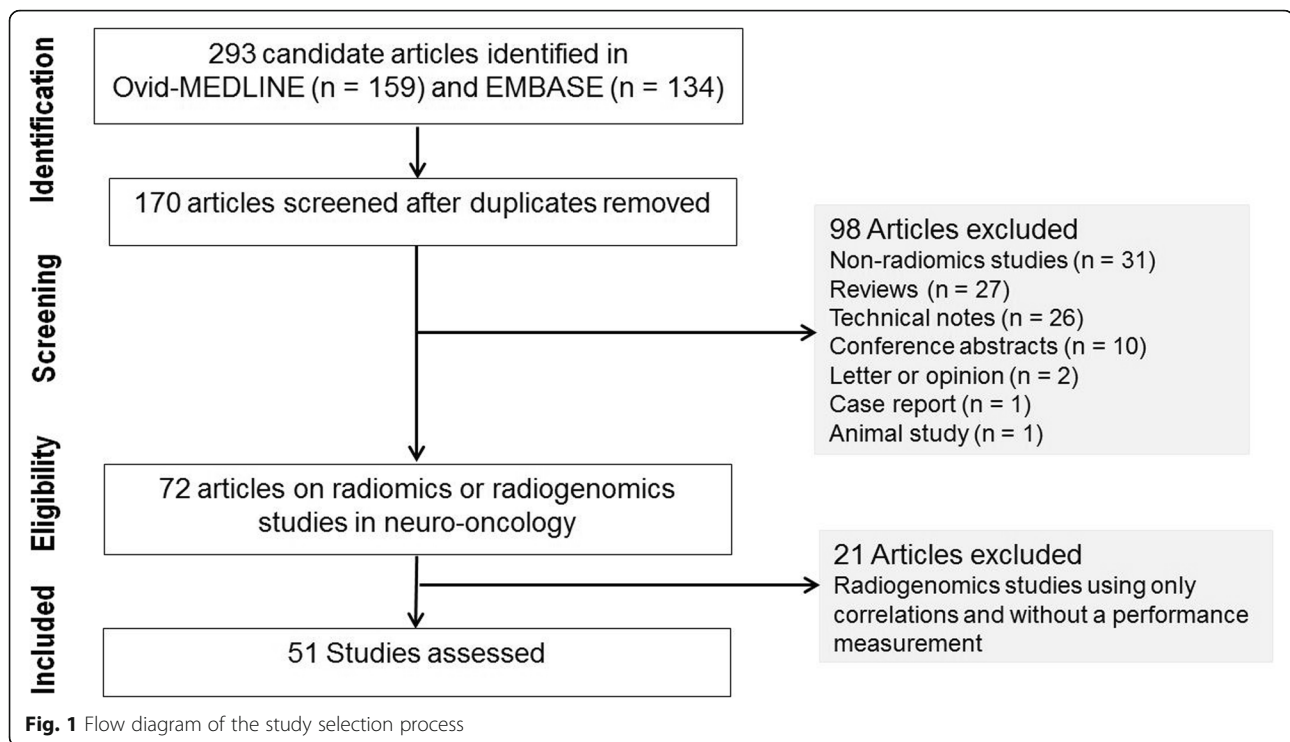
Materials and methods

Article search strategy and study selection

To identify all potentially relevant original research papers published in the neuro-oncology field, database search was conducted in the MEDLINE (National Center for Biotechnology Information, NCBI) and EMBASE databases up from any time until February 28, 2019. The search terms used to find radiomics studies were "glioma" OR "glioblastoma" AND "radiomic" OR "radiogenomic". The search identified 293 candidate articles. Retrieved articles were screened for eligibility. After removal of 122 duplicates, screening of the abstracts of the remaining 171 articles was performed. Abstract review further excludes 98 articles for the following reasons: 31 non-radiomics studies, 27 reviews, 26 technical notes, 10 conference abstracts, 2 letter or opinion, 1 case report, and 1 animal study. Full-text reviews of the 73 potentially eligible articles were performed by two experienced reviewer (H.S.K., with 20 years of experience in neuro-oncologic imaging and J.H.K., with 25 years of experience in neurosurgery) selected articles in consensus that tested performance in respect to a diagnostic, prognostic, or predictive biomarker study aspect. Radiogenomics studies searching only for correlations and not containing a performance measurement for a diagnostic, prognostic, or predictive biomarker, were excluded. This process further removed 21 articles as they did not comprise performance tests for diagnostic, prognostic, or predictive utility. Finally, 52 articles were included in the main analysis (Fig. 1).

Data extraction and analysis

Three reviewers (J.E.P., D.K., who had 5 and 2 years of experience in neuro-oncologic imaging and H.S.K.) evaluated the eligible radiomics performance studies. Before performing their analysis, a seminar was convened in which the reviewers participated, to review and discuss the items listed in the RQS and to ensure they all had a clear knowledge of RQS.

The detailed RQS score with 16 components is defined elsewhere [2] (Additional file 1: Table S1). The reviewers extracted the data using a predetermined RQS evaluation according to six domains. Domain 1 covered protocol quality and reproducibility in image and segmentation: well-documented image protocols (1 point) and/or usage of public image protocols (1 point), multiple segmentations (1 point), phantom study (1 point), and test-retest analysis with

Fig. 1 Flow diagram of the study selection process

imaging at multiple time points (1 point). Domain 2 covered the reporting of feature reduction and validation: feature reduction or adjustment for multiple testing (3 or – 3 points) and validation (– 5 to 5 points). Domain 3 covered the reporting of biological/clinical validation and utility: multivariate analysis with non-radiomics features (1 point), biologic correlates (1 point), comparison to the gold standard (2 points), and potential clinical utility (2 points). Domain 4 covered reporting of the performance index: reporting of discrimination statistics (1 point) with resampling (1 point), calibration statistics (1 point) with resampling (1 point), and application of cut-off analyses (1 point). Domain 5 covered demonstration of a higher level of evidence: by conducting a prospective study (7 points) or cost-effectiveness analysis (2 points). The final domain (domain 6) covered open science, with open availability of source code and data (4 points).

Each article was evaluated by two of the three independent reviewers. Disagreements between any two reviewers were discussed at a research meeting attended by all three reviewers and an additional statistical reviewer. The following topics were, subject to some initial disagreements and were discussed until a consensus was reached.

1) Multiple segmentation (domain 1): when there were two or more readers, the article earned an additional point as segmentation variability was considered. Automatic segmentation using a convolutional neural network or other automatic software earned a point as the method pursued better segmentation reproducibility.

2) Validation (domain 2): a definition of missing validation (– 5 points) was applied when the article performed cross-validation or nested cross-validation using only the training data, as the validation needs to be performed without retraining and without adaptation of the previous cut-off value from the training data.

3) Multivariate analysis with non-radiomics features (domain 3): When the main study endpoint was survival and the selected radiomics features were further correlated with non-radiomic features, i.e. MGMT (O-6-methylguanine-DNA methyltransferase) promoter methylation status, the article earned an additional point. However, if the main outcome was prediction of IDH (iso-dehydroxygenase) mutation and the radiomics features were selected for IDH mutation only, the article would not gain any additional score, as this would not provide a more holistic radiomics model.

4) Comparison with the gold standard (domain 3): As there is no TNM staging in neuro-oncology, the well-known parameters of age, Karnofsky performance score, extent of resection [16], IDH, or MGMT status were considered as gold standards for survival analysis.

5) Potential clinical utility (domain 3): According to the consensus statement of the FDA-NIH Biomarker Working Group [13], 'clinical utility' is thought to be achieved when a biomarker leads to net improvement of health outcomes or provides information useful for prevention, diagnosis, treatment, and management of a

Park *et al. BMC Cancer* (2020) 20:29

Page 4 of 11

disease [13, 17, 18]. For example, a study earned additional points if decision curve analysis was performed and demonstrated net improvement. Discussion of the potential utility of radiomics without proper analysis did not earn additional points.

The multicentricity of data source was further investigated by one reviewer (J.E.P.). The study was evaluated whether the model validation was conducted using external dataset. The multicentricity was evaluated whether the data source was from single center, multi-center, and public data, when constructing the training set. In addition, whether data acquisition was on 1.5 Tesla or 3.0 Tesla magnet was evaluated.

Statistical analysis

The total RQS score was calculated for each article and for each component. For all included articles, the total RQS score was calculated (score range, − 8–36) and expressed as median and interquartile range.

For the six domains in the RQS (protocol quality and segmentation, feature selection and validation, biologic/clinical validation and utility, model performance index, high level of evidence, and open science and data), basic adherence was assigned when a score of at least 1 point was obtained without minus points. The basic adherence rate was then calculated in a descriptive manner using proportions, and the proportion (%) of articles that fulfilled each reporting domain was determined.

A graphical display for the proportion of studies with a basic adherence rate was adopted from the suggested graphical display for Quality Assessment of Diagnostic Accuracy Studies–2 results [19].

Subgroup analyses were performed to determine whether the reporting quality differed according to intended use (diagnostic or prognostic) and published journal (imaging journal or clinical journal). Before subgroup analysis, the RQS was plotted for each journal to observe whether there was a systematic difference between each journals (Additional file 2: Figure S2), and with no systematic difference being observed between journals, this effect was not considered. The nonparametric Mann-Whitney U test was used to compare the RQS score in each group. All statistical analyses were performed using SPSS (SPSS version 22; SPSS, Chicago, IL) and *R* (R version 3.3.3; R Foundation for Statistical Computing, Vienna, Austria), and a *P* value < .05 was considered statistically significant.

Results

Characteristics of Radiomics studies in Neuro-oncology

Fifty-one articles [4–11, 20–61] were finally analyzed (Additional file 3: Table S2). The characteristics of the articles are summarized in Table 1. Most studies included gliomas (90.1%), with either glioblastomas only (47.1%) or lower grade gliomas only (LGG; 29.4%).

Other non-glial tumors were studied in 9.8%. The study purposes included molecular or genomic classification (49.0%), survival prediction (25.5%), differential diagnosis of gliomas from non-glial tumors (11.8%), histopathologic grading (9.8%), and assessment of response to treatment (5.9%). One study predicted the occurrence of epilepsy in LGG patients using radiomics analysis [38]. Radiomics analysis was most frequently studied as a diagnostic biomarker (70.6%), followed by use as a prognostic biomarker (25.5%), and as a predictive (7.8%) biomarker. Analysis of the validation methods revealed that external validation was missing in 36 out of 51 studies (70.6%). Among 51 studies, 66.7% ([34/51]) of studies were from single center, 29.4% ([15/51]) studies were from public data (TCIA), and 3.9% ([2/51]) of studies were from multicenter data source. In regard of magnetic strength, 58.8% ([30/51]) of studies utilized 3 Tesla magnet, while 41.2% ([21/51]) of studies utilized both 1.5 Tesla and 3.0 Tesla magnet (also in the Additional file 3: Table S2).

Basic adherence rate of the reporting quality according to the six key domains

Table 2 summarizes the basic adherence rate of the radiomics studies according to the six key domains. In domain 1, all studies reported well-documented image acquisition protocols or use of publicly available image databases. Multiple segmentations by two readers or automatic segmentation were performed in 14 of the 51 studies (27.4%). Notably, only one study [10] conducted imaging at multiple time points and tested feature robustness. After test-retest analysis, 37.0% of the radiomic features (386 out of the 1043 extracted features in the study) were stable and reproducible [10] over three different sessions on the same machine.

In domain 2, most studies adopted appropriate feature reduction or adjustment for multiple testing (48 out of 51, 94.1%). The studies used either false discovery rate with univariate logistic regression or two-sample t-tests (for binary outcomes), and a variety of statistical and machine learning methods such as Least Absolute Shrinkage Selector Operator (LASSO), random forest, recursive feature elimination, and support vector machines. Many of studies performed validation using datasets from the same or a different institute (35 out of 51, 68.6%). Six studies earned the full 5 points for validation [21, 24, 31, 36, 54, 55], using data from three datasets from distinct institutes or public dataset.

In domain 3, many of studies performed multivariate analysis of the radiomics features with non-radiomic features (62.7%), and most of the studies found biological correlates (74.5%) to provide a more holistic model and imply biological relevance. Less than half of the studies (37.2%) compared results with an existing gold standard.

Park *et al. BMC Cancer* (2020) 20:29

Page 5 of 11

Table 1 Characteristics of the 51 neuro-oncologic radiomics studies with study design of diagnostic, prognostic, or predictive biomarker

Article characteristics	Number of articles*
Patient number	153 (standard deviation 82.2; range 32–439)
Journal type	
Clinical journal	20 (39.2)
Imaging journal	31 (60.8)
Study inclusion	
Gliomas	46 (90.1)
Glioblastoma only	24 (47.1)
Lower grade gliomas only	15 (29.4)
Gliomas (WHO grade I to IV)	7 (13.7)
Other tumors	5 (9.8)
Study intent	
Differential diagnosis	6 (11.8)
Histopathological grade	5 (9.8)
Molecular/genomic classification	25† (49.0)
Survival	13† + (25.5)
Response to treatment	3† + (5.9)
Others	1 (2)
Biomarker	
Diagnostic	36† (70.6)
Prognostic	13† + (25.5)
Predictive	4+ (7.8)
External validation	
Yes	15 (29.4)
No	36 (70.6)

*numbers in parentheses are percentages
†Two studies overlap in both molecular classification and survival (diagnostic and prognostic marker)
+ one study overlaps in both prognostic and predictive biomarker

By contrast, in terms of clinical utility, only one study [50] analyzed a net improvement in health outcomes using decision curve analysis or other statistical tools.

In domain 4, all studies used discriminative statistics, but two studies [40, 41] provided hazard ratios and *P* values from a log-rank test for survival analysis instead of the C-index.

Surprisingly, studies were deficient in demonstrating a high level of evidence such as a prospective design or cost-effectiveness analysis. One two studies partly performed validation using a prospective registry [24, 29], but the study per se was conducted in a retrospective manner. For domain 6, only three studies [4, 20, 47] made their code and/or data publicly available. The adherence rate according to the six key domains is shown in Fig. 2.

Assessment of the Radiomics quality score

The median RQS score of the 51 studies was 11.0, (interquartile range [IQR], 3–12.75), which was 30.5% of the ideal score of 36 (Table 3). The lowest score was – 6

and the highest score was 16 (44.4% of the ideal quality score). Compared with the ideal score, the RQS of the selected studies was lowest in the high level of evidence domain and open science and data domain (0%), followed by biological/clinical validation, and feature reproducibility in image and segmentation.
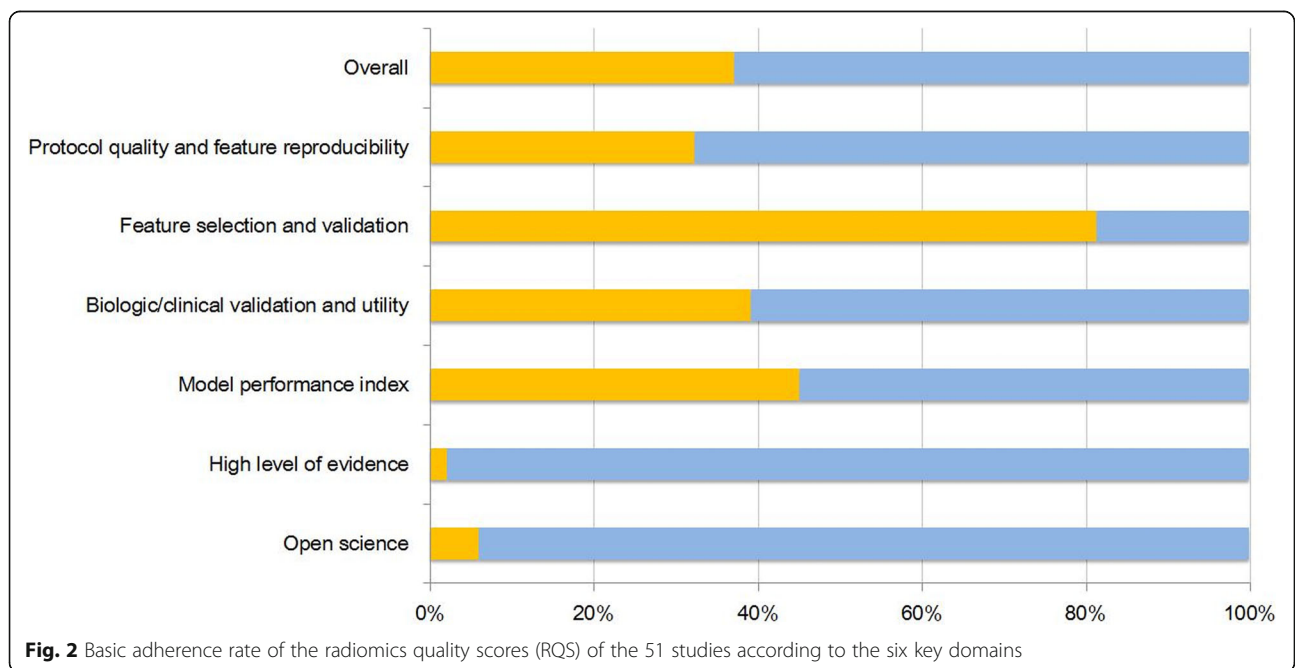
Both feature reduction and validation were missing from the study [51] with the lowest score. Meanwhile, studies with the highest scores [10, 24, 29, 31, 36, 55] earned additional points by using publicly available images from the TCIA (The Cancer Imaging Archive) [24, 31, 36, 55], registry or trial data [24, 29], multiple segmentation [31], test-retest analysis [10], and calculation of calibration index and bootstrapping [10], with all studies fulfilling requirements for image protocol quality, feature reduction, validation, and use of a discrimination index.

Subgroup analysis

The results of the subgroup analysis according to the journal type and biomarker study design are shown in

Table 2 Basic adherence rate according to the six key domains

	Basic adherence rate
Total 16 items	37.1%
Domain 1: Protocol quality and stability in image and segmentation	32.3%
Protocol quality	51 (100%)
Test-retest	1 (2%)
Phantom study	0 (0%)
Multiple segmentation	14 (27.4%)
Domain 2: Feature selection and validation	81.4%
Feature reduction or adjustment of multiple testing	48 (94.1%)
Validation	35 (68.6%)
Domain 3: Biologic/clinical validation and utility	39.2%
Multivariate analysis with non-radiomics features	32 (62.7%)
Biologic correlates	28 (74.5%)
Comparison to 'gold standard'	19 (37.2%)
Potential clinical utility	1 (2%)
Domain 4: Model performance index	45.1%
Discrimination statistics	49 (96.1%)
Calibration statistics	7 (13.7%)
Cut-off analysis	13 (25.5%)
Domain 5: High level of evidence	2%
Prospective study	2 (3.9%)
Cost-effective analysis	0 (0%)
Domain 6: Open science and data	3 (5.9%)

Fig. 2 Basic adherence rate of the radiomics quality scores (RQS) of the 51 studies according to the six key domains

Park *et al. BMC Cancer*   (2020) 20:29

Page 7 of 11

Table 3 Radiomics quality score and subgroup analysis according to journal type and biomarker study design

Radiomics quality score	Median score (Interquartile range)	Clinical (n = 20)	Imaging (n = 31)	P	Diagnostic (n = 34)	Prognostic /Predictive† (n = 17)	P
Total 36 points	11.00 (3–12.5)	11.5 (3–13)	10 (3.5–11.5)	.57	10 (3–11)	12 (6–14)	.07
Domain 1: Protocol quality and stability in image and segmentation (0 to 5 points)	1 (1–2)	1.5 (1–2)	1 (1–2)		1 (1–2)	1 (1–2)	
Protocol quality (2)	1 (1–1.5)	1 (1–1.25)	1 (1–1)	.95	1 (1–1)	1 (1–1.75)	.662
Test-retest (1)	0 (0–0)	0 (0–0)	0 (0–0)	.228	0 (0–0)	0 (0–0)	.169
Phantom study (1)	0 (0–0)	0 (0–0)	0 (0–0)	NA	0 (0–0)	0 (0–0)	NA
Multiple segmentation (1)	0 (0–1)	0 (0–0.25)	0 (0–1)	.76	0 (0–1)	0 (0–1)	.277
Domain 2: Feature selection and validation (−8 to 8 points)	5 (−2–6)	5 (−2–6)	5 (−2–5)		5 (−2–6)	5 (−2–6)	
Feature reduction or adjustment of multiple testing (−3 or 3)	3 (3–3)	3 (3–3)	3 (3–3)	.849	3 (3–3)	3 (3–3)	1
Validation (−5, 2, 3, 4, or 5)	2 (−5–3)	2 (−5–3)	2 (−5–2)	.975	2 (−5–3)	2 (−5–3)	.833
Domain 3: Biologic/clinical validation and utility (0 to 6 points)	2 (1–3)	2 (1–3)	2 (1–2)		2 (1–2)	3 (2–4)	
Non-radiomics features (1)	1 (0–1)	1 (0–1)	1 (0–1)	.799	1 (0–1)	1 (0–1)	.159
Biologic correlates (1)	1 (0–1)	1 (0–1)	1 (0–1)	1	1 (1–1)	1 (0–1)	**.001**
Comparison to 'gold standard' (2)	0 (0–2)	0 (0–2)	0 (0–2)	.97	0 (0–0)	2 (2–2)	**<.001**
Potential clinical utility (2)	0 (0–0)	0 (0–0)	0 (0–0)	.44	0 (0–0)	0 (0–0)	.50
Domain 4: Model performance index (0 to 5 points)	2 (1–2)	2 (1–3)	2 (1–3)		2 (1–2)	2 (1–3)	
Discrimination statistics (2)	2 (1–2)	1.5 (1–2)	2 (1–2)	.75	2 (1–2)	2 (1–2)	.99
Calibration statistics (2)	0 (0–0)	0 (0–0)	0 (0–0)	.84	0 (0–0)	0 (0–1)	**.02**
Cut-off analysis (1)	0 (0–0.5)	0 (0–0)	0 (0–1)	.48	0 (0–0)	1 (0–1)	**.001**
Domain 5: High level of evidence (0 to 8 points)	0 (0–0)	0 (0–0)	0 (0–0)		0 (0–0)	0 (0–0)	
Prospective study (7)	0 (0–0)	0 (0–0)	0 (0–0)	.08	0 (0–0)	0 (0–0)	.634
Cost-effective analysis (1)	0 (0–0)	0 (0–0)	0 (0–0)	NA	0 (0–0)	0 (0–0)	NA
Domain 6: Open science and data (0 to 4 points)	0 (0–0)	0 (0–0)	0 (0–0)	**.03**	0 (0–0)	0 (0–0)	1

† Two studies overlap in both diagnostic and prognostic marker were classified as prognostic use

Table 3. Studies in clinical journal showed a trend for a higher RQS score than those in imaging journals (median 11.5 vs. 10), but this was not statistically significant. Studies in clinical journals (n = 20) were similar with those in imaging journals (n = 31) in most of the RQS score except for open science and data (P = .03). All three studies [4, 20, 47] made their code and/or data publicly available were published in the same clinical journal (*Neuro-Oncology*).

Prognostic/predictive studies showed a trend for a higher RQS score than diagnostic studies (median 12 vs. 10), but this was not statistically significant. Prognostic studies received a higher score than diagnostic studies in comparison with biologic correlates (P = .001) and comparison to a 'gold standard' (P < .001). Also, prognostic/predictive studies used calibration statistics (P = .02) and

cut-off analysis (P = .001) more frequently than diagnostic studies, which is potentially useful for future modeling studies.

Discussion

In this study, radiomics studies in neuro-oncologic imaging were evaluated in respect to the quality of both the science and the reporting, using radiomics quality score. Overall, radiomics studies still have room for improvement, with basic adherence rate of 37.1% out of total 16 items. In terms of protocol quality, radiomics studies were particularly deficient for testing image stability using test-retest analysis and a phantom study. Selected radiomics features were often correlated to non-radiomics features or biological phenotype, but linking them to clinical validation and achievement of clinical

utility was insufficient. The radiomics models were often measured with discriminative statistics, while calibration statistics and cut-off analyses were underutilized. A high level of evidence for radiomics studies is critically lacking, with further limitations being demonstrated in their openness to data and code. Our results imply that a low quality of reporting may hamper the use of radiomics utilities as a tool for clinical decision-making, and several key domains in radiomics studies require significant improvement.

The six key domains used in this study were designed to support the integration of the RQS to facilitate the use in radiomics approaches. Adopted from the consensus statement of the FDA-NIH Biomarker Working Group [13], the three aspects of technical validation, biological/clinical validation, and assessment of cost-effectiveness for imaging biomarker standardization were included in domains 1, 3, and 5, respectively. With regards to technical validation, radiomics approaches are yet to become a reliable measure for the testing of hypotheses in clinical cancer research, with insufficient data supporting their precision or technical bias. Precision analysis using repeatability and reproducibility test was conducted in one study [10], but reproducibility needs to be tested using different geographical sites and different equipment. Furthermore, none of the evaluated studies reported analysis of technical bias using a phantom study, which describes the systemic difference between the measurements of a parameter and its real values [62]. According to our results, further technical validation needs to be achieved before radiomics analysis can be related to clinical outcomes.

Along with technical performance assessment, imaging biomarkers need to pass clinical performance assessment in multicenter studies before they can be considered for clinical use [13, 63]. For clinical validation, prospective testing of an imaging biomarker in clinical populations is required [64], but until now little studies have conducted a prospective study in the field of neuro-oncology. After biological/clinical validation, the cost-effectiveness of radiomics needs to be studied to ensure it provides good value for money compared with the other currently available biomarkers. However, no study has conducted cost-effectiveness analysis and only one study demonstrated net benefit improvement. Also, external validation is conducted in only 29.4% of the total studies. From the current standpoint, the clinical use of radiomics may seem far away, and technical and clinical validation is still required.

Subgroup analysis demonstrated that prognostic and predictive radiomics studies showed better quality than diagnostic studies in regard to comparison to 'gold standard', use of calibration statistics, and use of cut-off analysis. These are important for adoption of radiomics

modeling, by demonstrating point estimates of prediction and actual data using calibration and by applying cut-off in the future studies. These measures further emphasize utility of radiomics modeling in clinic.

Biological validation of imaging biomarkers occurs relatively late in the development process [13]. In terms of biological validation, 74.5% of studies related radiomics features to biological correlates, such as the molecular subtype of IDH mutation or MGMT methylation status, while 62.7% of studies performed multivariate analysis using both radiomics and non-radiomic features. Nonetheless, biological validation of imaging-pathologic correlations [13] is not currently available for radiomics, as extensive spatial heterogeneity exists [1] and co-localized pathology data are currently not achievable. The unclear relationship with tumor biology is probably the reason why radiomics approaches have not influenced clinical decision-making or exhibited potential clinical utility.

It is important to adhere to the standardization of the radiomics features nomenclature and calculation according to the IBSI (International Biomarker Standardization Initiative) to improve reproducibility of scientific researches. Only 3 of studies made their code open, and many of studies did not provide detailed descriptions of the calculation of radiomics features and did not permit to clarify the details of radiomics calculation. Future studies are needed in terms of adherence to the standardization of radiomics features. Also, studies have utilized public data sources such as the TCIA did not made their own data publicly available. Further determination of the reproducibility of radiomics techniques requires collaborative multicenter studies, which would benefit greatly from the open availability of data and models.

Our study had several limitations. First, the publication of radiomics studies is not limited to the neuro-oncologic field or to MRI. Limits were placed on the search to permit in-depth analysis of applications for a particular disease where radiomics research seems to be most actively performed. Given the impact and number of populations of the selected studies, our results may actually represent a higher than average quality. Second, radiomics is still a developing imaging biomarker and the suggested RQS may be too 'ideal' to be qualified. The criteria of phantom study and multiple imaging acquisitions may become unrealistic in clinical situation. Also, segmentation stability earned score when it was performed by two readers or more, but recent development of deep learning segmentation may provide more robust result than manual segmentation. However, aiming for a higher level of evidence is necessary for the future use of radiomics approaches in future clinical trials and in the clinic.

Park *et al. BMC Cancer*   (2020) 20:29

Page 9 of 11

Conclusion

The quality of reporting of radiomics studies in neuro-oncology is currently not sufficient. Validation is necessary using external dataset, and improvements need to be made to feature reproducibility, analysis of the clinical utility, pursuits of a higher level of evidence in study design, and open science.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12885-019-6504-5.

Additional file 1: Table S1. The six key domains of the radiomics quality score.

Additional file 2: Figure S1. RQS score according to the journal shows no definite systematic differences between the journals.

Additional file 3: Table S2. Characteristics of the included studies.

Abbreviations

IDH: Isodehydroxygenase; LGG: Lower grade glioma; MGMT: O-6-methylguanine-DNA methyltransferase; MRI: Magnetic resonance imaging; RQS: Radiomics quality score; TRIPOD: Multivariable prediction model for Individual Prognosis OR Diagnosis

Authors' contributions

All authors commented and approved the final manuscript. JEP: Data curation, data analysis, manuscript writing, editing, statistical analysis. DK: Data analysis, statistical analysis. HSK: Manuscript editing, conceptual feedback and project integrity. SYP: Data provision, statistical analysis. JYK: Data analysis. SJC: Data analysis. JHK: clinical oversight, conceptual feedback and project integrity.

Availability of data and materials

All data are used in this study is provided in the electronic supplemental materials.

Ethics approval and consent to participate

Not applicable. Institutional Review Board approval was not required because of the nature of our study, which was a study based on research articles.

Consent for publication

Not applicable. There are no details on individuals reported within the manuscript, so we don't have the consent for publication.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 43 Olympic-ro 88, Songpa-Gu, Seoul 05505, South Korea. [2]Department of Radiology, Inje University Busan Paik Hospital, Busan, South Korea. [3]Department of Clinical Epidemiology and Biostatistics, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea. [4]Department of Radiology, Kangbuk Samsung Medical Center, Seoul, South Korea. [5]Department of Neurosurgery, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea.

References

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, They Are Data. Radiology. 2016;278(2):563–77.
2. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue R, Even AJG, Jochems A, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–62.
3. Sanduleanu S, Woodruff HC, de Jong EEC, van Timmeren JE, Jochems A, Dubois L, Lambin P. Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. Radiother Oncol. 2018;127(3):349–60.
4. Kang D, Park JE, Kim YH, Kim JH, Oh JY, Kim J, Kim Y, Kim ST, Kim HS. Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation. Neuro-oncology. 2018;20(9):1251–61.
5. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, Kim EH, Kim SH, Lee SK. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. Eur Radiol. 2018;28(9):3832–9.
6. Han Y, Xie Z, Zang Y, Zhang S, Gu D, Zhou M, Gevaert O, Wei J, Li C, Chen H, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. J Neurooncol. 2018;140(2):297–306.
7. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep. 2017;7(1):5467.
8. Zhang X, Tian Q, Wang L, Liu Y, Li B, Liang Z, Gao P, Zheng K, Zhao B, Lu H. Radiomics strategy for molecular subtype stratification of lower-grade Glioma: detecting IDH and TP53 mutations based on multimodal MRI. J Magn Reson Imaging. 2018;48(4):916–26.
9. Kickingereder P, Burth S, Wick A, Gotz M, Eidel O, Schlemmer HP, Maier-Hein KH, Wick W, Bendszus M, Radbruch A, et al. Radiomic profiling of Glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. Radiology. 2016;280(3):880–9.
10. Kickingereder P, Neuberger U, Bonekamp D, Piechotta PL, Gotz M, Wick A, Sill M, Kratz A, Shinohara RT, Jones DTW, et al. Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma. Neuro-oncology. 2018;20(6):848–57.
11. Kickingereder P, Gotz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, Sill M, Nowosielski M, Schlemmer HP, Radbruch A, et al. Large-scale Radiomic profiling of recurrent Glioblastoma identifies an imaging predictor for stratifying anti-Angiogenic treatment response. Clin Cancer Res. 2016;22(23):5765–71.
12. Waterton JC, Pylkkanen L. Qualification of imaging biomarkers for oncology drug development. Eur J Cancer (Oxford, England : 1990). 2012;48(4):409–15.
13. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017;14(3):169–86.
14. Sung NS, Crowley WF Jr, Genel M, Salber P, Sandy L, Sherwood LM, Johnson SB, Catanese V, Tilson H, Getz K, et al. Central challenges facing the national clinical research enterprise. Jama. 2003;289(10):1278–87.
15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med. 2015;162(10):735–6.
16. Marko NF, Weil RJ, Schroeder JL, Lang FF, Suki D, Sawaya RE. Extent of resection of glioblastoma revisited: personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. J Clin Oncol. 2014;32(8):774–82.
17. Hayes DF, Allen J, Compton C, Gustavsen G, Leonard DG, McCormack R, Newcomer L, Pothier K, Ransohoff D, Schilsky RL, et al. Breaking a vicious cycle. Sci Transl Med. 2013;5(196):196cm196.
18. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete and transparent reporting. J Clin Oncol. 2012;30(34):4223–32.

19. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. Group Q-: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.

20. Akbari H, Bakas S, Pisapia JM, Nasrallah MP, Rozycki M, Martinez-Lage M, Morrissette JJD, Dahmane N, O'Rourke DM, Davatzikos C. In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. Neuro-oncology. 2018;20(8):1068–79.

21. Arita H, Kinoshita M, Kawaguchi A, Takahashi M, Narita Y, Terakawa Y, Tsuyuguchi N, Okita Y, Nonaka M, Moriuchi S, et al. Lesion location implemented magnetic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. Sci Rep. 2018;8(1): 11773.

22. Artzi M, Liberman G, Blumenthal DT, Aizenstein O, Bokstein F, Ben Bashat D. Differentiation between vasogenic edema and infiltrative tumor in patients with high-grade gliomas using texture patch-based analysis. J Magn Reson Imaging. 2018. https://doi.org/10.1002/jmri.25939. [Epub ahead of print].

23. Eichinger P, Alberts E, Delbridge C, Trebeschi S, Valentinitsch A, Bette S, Huber T, Gempt J, Meyer B, Schlegel J, et al. Diffusion tensor image features predict IDH genotype in newly diagnosed WHO grade II/III gliomas. Sci Rep. 2017;7(1):13396.

24. Grossmann P, Narayan V, Chang K, Rahman R, Abrey L, Reardon DA, Schwartz LH, Wen PY, Alexander BM, Huang R, et al. Quantitative imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab. Neuro-oncology. 2017;19(12):1688–97.

25. Hsieh KL, Chen CY, Lo CM. Radiomic model for predicting mutations in the isocitrate dehydrogenase gene in glioblastomas. Oncotarget. 2017;8(28): 45888–97.

26. Hu LS, Ning S, Eschbacher JM, Baxter LC, Gaw N, Ranjbar S, Plasencia J, Dueck AC, Peng S, Smith KA, et al. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. Neuro-oncology. 2017;19(1):128–37.

27. Ingrisch M, Schneider MJ, Norenberg D, Negrao de Figueiredo G, Maier-Hein K, Suchorska B, Schuller U, Albert N, Bruckmann H, Reiser M, et al. Radiomic analysis reveals prognostic information in T1-weighted baseline magnetic resonance imaging in patients with Glioblastoma. Investig Radiol. 2017;52(6):360–6.

28. Kickingereder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, Wick A, Eidel O, Schlemmer HP, Radbruch A, et al. Radiogenomics of Glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. Radiology. 2016;281(3):907–18.

29. Kim JY, Park JE, Jo Y, Shim WH, Nam SJ, Kim JH, Yoo RE, Choi SH, Kim HS. Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. Neuro Oncol. 2019;21(3):404–14.

30. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, Zhai G. A deep learning-based Radiomics model for prediction of survival in Glioblastoma Multiforme. Sci Rep. 2017;7(1):10353.

31. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, Wang G, Liang C, Li ZC. A fully-automatic multiparametric Radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in Glioblastoma Multiforme. Sci Rep. 2017;7(1):14331.

32. Li Y, Liu X, Qian Z, Sun Z, Xu K, Wang K, Fan X, Zhang Z, Li S, Wang Y, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. Eur Radiol. 2018;28(7):2960–8.

33. Li Y, Liu X, Xu K, Qian Z, Wang K, Fan X, Li S, Wang Y, Jiang T. MRI features can predict EGFR expression in lower grade gliomas: a voxel-based radiomic analysis. Eur Radiol. 2018;28(1):356–62.

34. Li Y, Qian Z, Xu K, Wang K, Fan X, Li S, Jiang T, Liu X, Wang Y. MRI features predict p53 status in lower-grade gliomas via a machine-learning approach. NeuroImage Clinical. 2018;17:306–11.

35. Li Y, Qian Z, Xu K, Wang K, Fan X, Li S, Liu X, Wang Y, Jiang T. Radiomic features predict Ki-67 expression level and survival in lower grade gliomas. J Neuro-Oncol. 2017;135(2):317–24.

36. Li ZC, Bai H, Sun Q, Li Q, Liu L, Zou Y, Chen Y, Liang C, Zheng H. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. Eur Radiol. 2018;28(9):3640–50.

37. Liu TT, Achrol AS, Mitchell LA, Rodriguez SA, Feroze A, Iv M, Kim C, Chaudhary N, Gevaert O, Stuart JM, et al. Magnetic resonance perfusion image features uncover an angiogenic subgroup of glioblastoma patients

with poor survival and better response to antiangiogenic treatment. Neuro-oncology. 2017;19(7):997–1007.

38. Liu Z, Wang Y, Liu X, Du Y, Tang Z, Wang K, Wei J, Dong D, Zang Y, Dai J, et al. Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas. NeuroImage Clinical. 2018;19:271–8.

39. Lu CF, Liu FT, Hsieh KL, Kao YJ, Cheng SJ, Hsu JB, Tsai PH, Chen RJ, Huang CC, Yen Y, et al. Machine learning-based Radiomics for molecular subtyping of Gliomas. Clin Cancer Res. 2018;24(18):4429–36.

40. McGarry SD, Hurrell SL, Kaczmarowski AL, Cochran EJ, Connelly J, Rand SD, Schmainda KM, LaViolette PS. Magnetic resonance imaging-based Radiomic profiles predict patient prognosis in newly diagnosed Glioblastoma before therapy. Tomography. 2016;2(3):223–8.

41. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. Eur Radiol. 2017;27(10):4188–97.

42. J-b Q, Liu Z, Zhang H, Shen C, Wang X-c, Tan Y, Wang S, Wu X-f, Tian J. Grading of Gliomas by using Radiomic features on multiple magnetic resonance imaging (MRI) sequences. Med Sci Monit. 2017;23:2168–78.

43. Rathore S, Akbari H, Rozycki M, Abdullah KG, Nasrallah MP, Binder ZA, Davuluri RV, Lustig RA, Dahmane N, Bilello M, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. Sci Rep. 2018;8(1):5087.

44. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, Liu ZC, Nan HY, Sun Q, Sun YZ, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. J Magn Reson Imaging. 2018;48(6):1518–28.

45. Xi YB, Guo F, Xu ZL, Li C, Wei W, Tian P, Liu TT, Liu L, Chen G, Ye J, et al. Radiomics signature: a potential biomarker for the prediction of MGMT promoter methylation in glioblastoma. J Magn Reson Imaging. 2018;47(5):1380–7.

46. Yu J, Shi Z, Lian Y, Li Z, Liu T, Gao Y, Wang Y, Chen L, Mao Y. Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. Eur Radiol. 2017;27(8):3509–22.

47. Zhou H, Vallieres M, Bai HX, Su C, Tang H, Oldridge D, Zhang Z, Xiao B, Liao W, Tao Y, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro-oncology. 2017;19(6):862–70.

48. Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. J Magn Reson Imaging. 2019;50(2):519–28.

49. Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, Kim SH, Lee SK. Radiomic MRI Phenotyping of Glioblastoma: improving survival prediction. Radiology. 2018;289(3):797–806.

50. Wang Q, Li Q, Mi R, Ye H, Zhang H, Chen B, Li Y, Huang G, Xia J. Radiomics Nomogram building from multiparametric MRI to predict grade in patients with Glioma: a cohort study. J Magn Reson Imaging. 2019;49(3):825–33.

51. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, Vagal A. Diagnostic accuracy of MRI texture analysis for grading gliomas. J Neuro-Oncol. 2018;140(3):583–9.

52. Dong F, Li Q, Xu D, Xiu W, Zeng Q, Zhu X, Xu F, Jiang B, Zhang M. Differentiation between pilocytic astrocytoma and glioblastoma: a decision tree model using contrast-enhanced magnetic resonance imaging-derived quantitative radiomic features. Eur Radiol. 2019;29(8):3968–75.

53. Lee MH, Kim J, Kim ST, Shin HM, You HJ, Choi JW, Seol HJ, Nam DH, Lee JI, Kong DS. Prediction of IDH1 mutation status in the Glioblastoma using the machine learning technique based on the quantitative Radiomic data. World Neurosurg. 2019;125:e688–96.

54. Li ZC, Bai H, Sun Q, Zhao Y, Lv Y, Zhou J, Liang C, Chen Y, Liang D, Zheng H. Multiregional radiomics profiling from multiparametric MRI: identifying an imaging predictor of IDH1 mutation status in glioblastoma. Cancer medicine. 2018;7(12):5999–6009.

55. Liu X, Li Y, Qian Z, Sun Z, Xu K, Wang K, Liu S, Fan X, Li S, Zhang Z, et al. A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. NeuroImage Clinical. 2018;20:1070–7.

56. Lohmann P, Lerche C, Bauer EK, Steger J, Stoffels G, Blau T, Dunkl V, Kocher M, Viswanathan S, Filss CP, et al. Predicting IDH genotype in gliomas using FET PET radiomics. Sci Rep. 2018;8(1):13328.

57. Ren Y, Zhang X, Rui W, Pang H, Qiu T, Wang J, Xie Q, Jin T, Zhang H, Chen H, et al. Noninvasive prediction of IDH1 mutation and ATRX expression loss in low-grade Gliomas using multiparametric MR Radiomic features. J Magn Reson Imaging. 2019;49(3):808–17.

Park *et al. BMC Cancer*   (2020) 20:29

Page 11 of 11

58. Sanghani P, Ang BT, King NKK, Ren H. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. Surg Oncol. 2018;27(4):709–14.

59. Su C, Jiang J, Zhang S, Shi J, Xu K, Shen N, Zhang J, Li L, Zhao L, Zhang J, et al. Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. Eur Radiol. 2019;29(4):1986–96.

60. Wu S, Meng J, Yu Q, Li P, Fu S. Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas. J Cancer Res Clin Oncol. 2019;145(3):543–50.

61. Zinn PO, Singh SK, Kotrotsou A, Hassan I, Thomas G, Luedi MM, Elakkad A, Elshafeey N, Idris T, Mosley J, et al. A Coclinical Radiogenomic validation study: conserved magnetic resonance Radiomic appearance of Periostin-expressing Glioblastoma in patients and Xenograft models. Clin Cancer Res. 2018;24(24):6288–99.

62. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, Guimaraes AR, Filice R, Zhang Z, Sullivan DC, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res. 2015;24(1):9–26.

63. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Aerts HJ, Bendriem B, Bendtsen C, Boellaard R, Boone JM, Cole PE, et al. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology. 2011;259(3):875–84.

64. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, Statistics Subcommittee of the NCIEWGoCD. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst. 2005; 97(16):1180–4.

Publisher's Note