Korean Journal of Radiology

KJR

Check for updates

Quality of Radiomics Research on Brain Metastasis: A Roadmap to Promote Clinical Translation

Chae Jung Park[1]*, Yae Won Park[2]*, Sung Soo Ahn[2], Dain Kim[3], Eui Hyun Kim[4], Seok-Gu Kang[4], Jong Hee Chang[4], Se Hoon Kim[5], Seung-Koo Lee[2]

[1]Department of Radiology, Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yongin Severance Hospital, Yonsei University College of Medicine, Seoul, Korea; [2]Department of Radiology and Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Yonsei University College of Medicine, Seoul, Korea; [3]Department of Psychology, Yonsei University, Seoul, Korea; Departments of [4]Neurosurgery and [5]Pathology, Yonsei University College of Medicine, Seoul, Korea

Objective: Our study aimed to evaluate the quality of radiomics studies on brain metastases based on the radiomics quality score (RQS), Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) checklist, and the Image Biomarker Standardization Initiative (IBSI) guidelines.
Materials and Methods: PubMed MEDLINE, and EMBASE were searched for articles on radiomics for evaluating brain metastases, published until February 2021. Of the 572 articles, 29 relevant original research articles were included and evaluated according to the RQS, TRIPOD checklist, and IBSI guidelines.
Results: External validation was performed in only three studies (10.3%). The median RQS was 3.0 (range, -6 to 12), with a low basic adherence rate of 50.0%. The adherence rate was low in comparison to the "gold standard" (10.3%), stating the potential clinical utility (10.3%), performing the cut-off analysis (3.4%), reporting calibration statistics (6.9%), and providing open science and data (3.4%). None of the studies involved test-retest or phantom studies, prospective studies, or cost-effectiveness analyses. The overall rate of adherence to the TRIPOD checklist was 60.3% and low for reporting title (3.4%), blind assessment of outcome (0%), description of the handling of missing data (0%), and presentation of the full prediction model (0%). The majority of studies lacked pre-processing steps, with bias-field correction, isovoxel resampling, skull stripping, and gray-level discretization performed in only six (20.7%), nine (31.0%), four (3.8%), and four (13.8%) studies, respectively.
Conclusion: The overall scientific and reporting quality of radiomics studies on brain metastases published during the study period was insufficient. Radiomics studies should adhere to the RQS, TRIPOD, and IBSI guidelines to facilitate the translation of radiomics into the clinical field.
Keywords: *Machine learning; Brain metastasis; Quality improvement; Radiomics; Radiomics quality score*

INTRODUCTION

Brain metastases are the most common type of intracranial tumors in adults [1], and these substantially affect the overall prognosis of patients with underlying cancers [2]. Although the clinical outcomes of patients with brain metastases have largely improved due to

advances in neuro-oncology treatment, there is still room for improvement, because the management of brain metastases is often complex and controversial [3]. As MRI is the imaging modality of choice for brain metastasis [4], many studies have attempted to use MRI to address many questions in the field of brain metastasis. Radiomics is a recently introduced method that enables data mining from

MRI; therefore, many studies have used MRI with radiomics in patients with brain metastases. As radiomics is more likely to be used with MRI in the field of brain metastases in the near future, it is highly desirable to assess the quality of published studies to evaluate the current status.

Radiomics involves the exploitation of MRI data to extract high-dimensional quantitative imaging features, which can be used to support clinical decision-making [5,6]. Although previous radiomics studies in the field of neuro-oncology have mainly focused on gliomas [7-12], there has been an increase in the number of studies on brain metastases. Indeed, radiomics studies have demonstrated promising results in the discrimination of brain metastasis from other tumors [13-20], identification of primary tumor types in patients with brain metastases [21-24], prediction of specific genetic mutations [25-29], prediction of survival [30,31], differentiation between radiation necrosis and brain metastasis [32-35], and prediction of outcome after radiosurgery [36-41]. However, such studies on brain metastases are confined within the limits of experimental settings, without translation into real-world clinical settings [42]. To lessen this translational gap, scientific and reporting quality must be high, which will enable a standardized evaluation of performance and increase the reproducibility and clinical utility of radiomics.

The radiomics quality score (RQS) is determined using a scoring system that incorporates the crucial aspects of radiomics studies and assesses their quality [43,44]. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) checklist is also a well-known tool for assessing the quality of prediction models [45]. Additionally, the Image Biomarkers Standardization Initiative (IBSI) provides a comprehensive and detailed review of each mandatory step for radiomics analyses, including nomenclature of radiomics features, general schemes, and datasets for calibration [46]. To the best of our knowledge, no study has evaluated radiomics studies on brain metastases according to the criteria of these three quality assessment tools. Assessing the quality of current radiomics studies on brain metastases may further promote the use of radiomics as a clinical tool. Therefore, in light of the RQS, TRIPOD checklist, and IBSI guidelines, this study aimed to evaluate the quality of radiomics studies that focused on brain metastases.

MATERIALS AND METHODS

Systematic Search Strategy and Study Selection

PubMed MEDLINE (n = 516) and EMBASE (n = 56) databases were searched to collect all original research papers on radiomics analysis published until February 2021, and the following terms were used for the search: ("brain metastasis" OR "brain metastases") AND ("radiomic" OR "radiogenomic" OR "texture"). From a total of 572 papers retrieved using the specified search terms, 39 duplicate articles were removed. Of the remaining articles, 504 were excluded because they were non-radiomics studies (n = 100), conference abstracts (n = 25), not in the field of interest (n = 255), non-human (n = 6), non-brain (n = 20), review articles (n = 60), technical notes (n = 18), editorial articles (n = 1), erratum (n = 1), comments (n = 5), or case reports (n = 11), or included < 50 brain metastases (n = 2). Finally, 29 articles were included in the analysis (Fig. 1).

Analysis of Method Quality Based on RQS

The RQS was determined using 16 items classified into six domains, as reported in previous studies (Supplementary Table 1) [47-49]. Two reviewers (with 8 and 10 years of experiences in neuroradiology) independently evaluated the domain scores of all included articles (Supplement, Radiomics Quality Scoring) after first achieving consensus on the evaluation criteria through discussion. If disagreements occurred between the two reviewers, a final decision was made after a consensus was reached.

Additional discussions were required to reach consensus on the following topics: issues of "validation" (domain 2), "comparison with the gold standard" (domain 3), and "detection and discussion of biologic correlation" (domain 3) (Supplement, Consensus Reached for RQS Scoring) due to the distinct nature of studies on brain metastasis.

Analysis of Reporting Completeness Based on the TRIPOD Checklist

The TRIPOD checklist, consisting of 37 items in 22 criteria, was applied to each article to determine the completeness of the report [45]. The analysis type of the prediction model was determined as follows: development only, type 1a; development and validation using resampling, type 1b; random split-sample development and validation, type 2a; nonrandom split-sample development and validation, type 2b; development and validation using separate data, type 3; or validation only, type 4. Details of the TRIPOD checklist
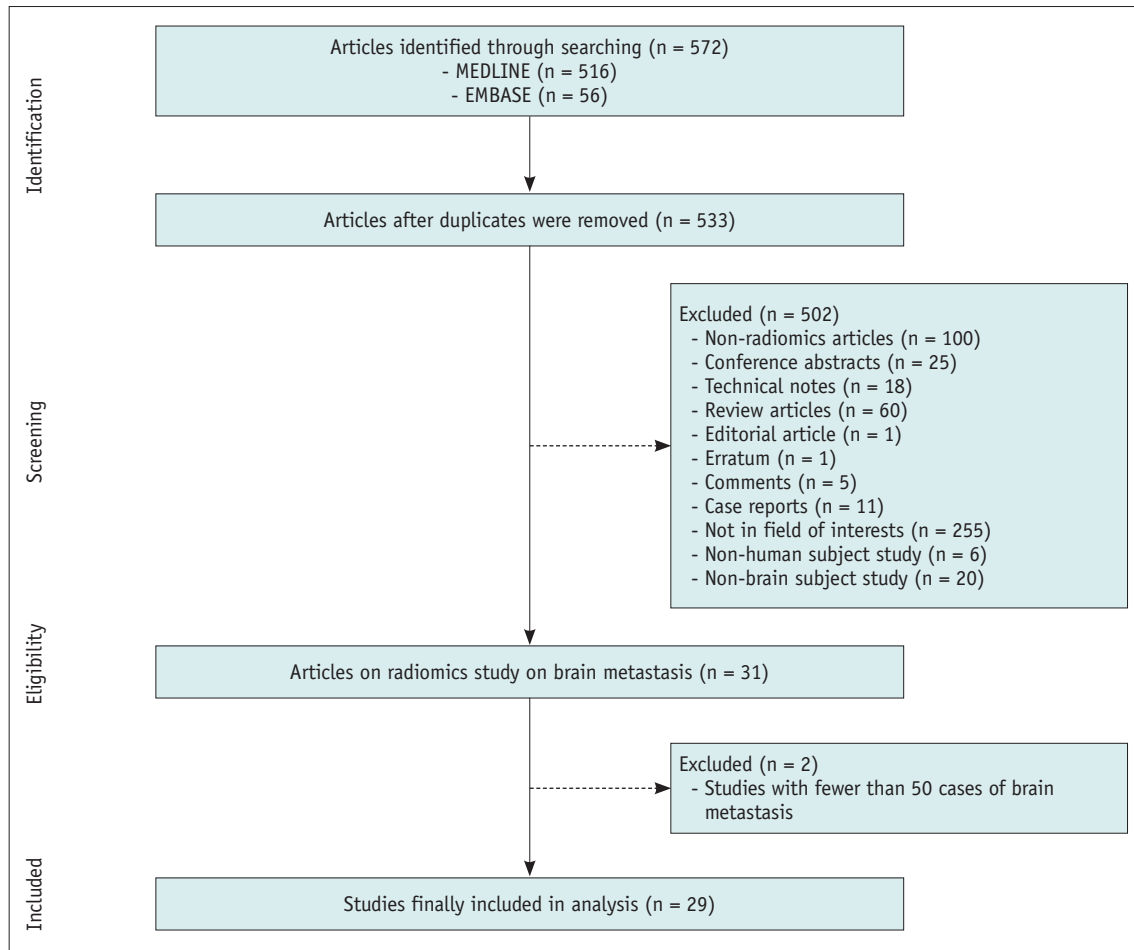
Fig. 1. Flow chart of the study selection process.

and data extraction are shown in Supplement (Reporting Completeness Based on TRIPOD Statement).

Reporting of Image Processing and Radiomics Feature Extraction according to the IBSI Guidelines

The IBSI guidelines provide a total of 76 items for complete reports on image processing and image biomarker extraction (https://ibsi.readthedocs.io/en/latest/04_Radiomics_reporting_guidelines_and_nomenclature.html#reporting-guidelines) [50]. Among these items, the two reviewers evaluated all the journals according to the six items related to the pre-processing: 1) bias-field correction algorithm, 2) isovoxel resampling, 3) skull stripping, 4) gray-level discretization, 5) signal intensity normalization, and 6) use of IBSI guideline-compliant radiomics feature extraction software. These six items were chosen because they are the most commonly performed pre-processing methods in the field of radiomics. In addition, the segmentation method (i.e., manual, semi-automatic, or

automatic segmentation) was retrieved from each journal. Evaluations were discussed to reach a consensus. The detailed assessment criteria for each item are described in Supplement (Assessment of pre-processing steps according to the IBSI guidelines).

Statistical Analysis

All statistical analyses were performed using R software (version 4.0.2; R Foundation for Statistical Computing).

In total, 29 articles were reviewed. In the cases where a score of one point per item was obtained, the study was considered to have basic adherence to each item of the RQS, TRIPOD checklist, or IBSI guideline. The basic adherence rate (%) was calculated as the proportion of the number of articles with basic adherence to the total number of articles. Sixteen items of the RQS were scored. The RQS for each item was reported as the median and range values. The percentage of the ideal score (%) was calculated as the ratio of the mean score to the ideal score for each item, and

the total RQS (-8 to 36) was calculated for all articles. A total of 35 items on the TRIPOD checklist was scored. When calculating the overall adherence rate, validation items (10c, 10e, 12, 13, 17, and 19a) and "if done" items (5c and 11) were excluded from both the denominator and numerator. Seven items from the IBSI guidelines are scored.

RESULTS

Characteristics of Radiomics Studies on Brain Metastasis

The characteristics of the 29 included radiomics studies [13,15-19,22,24-26,28-33,35,37-40,51] are documented in Table 1. In the included studies, the median number of patients included was 77 (range, 24–439). The types of journals were clinical journals (7, 24.1%), imaging journals (18, 62.1%), and computer/neuroscience journals (4, 13.8%). Radiomics studies were either diagnostic (21, 72.4%) or prognostic (8, 27.6%). Except for eight prognostic studies covering prediction of survival and prediction of outcome after radiosurgery, all other studies were diagnostic. Detailed characteristics of the included studies are presented in Supplementary Table 2.

Except for one study using positron emission tomography, 28 studies used MRI. Twenty-three (79.3%) studies used conventional images for feature extraction, and five (17.2%) used additional advanced images. Most studies performed manual segmentation (62.1%). Only three studies performed an external validation. Multiple brain metastases (> 1 lesion per patient) were included in the dataset of 20 studies (69.0%). Among them, seven studies demonstrated the statistical analysis that was relevant to the handling of cluster-correlated data. In three studies, patients with multiple metastases were assigned to the validation set or the training set to prevent bias from cluster correlation [18,22,31]. Correction for false discovery rate [30], correlation tests between the features [24,27], and marginal proportional-hazards models were used for multiple lesions within the same patient [40]. In the remaining nine studies, only one lesion per patient constituted the dataset.

Rate of Basic Adherence to the Reporting Quality according to the Six Key Domains

The rate of basic adherence to reporting quality for the 16 items (six domains) of the RQS is documented in Table 2. In terms of domain 1, 20 studies (69.0%) included well-documented image protocols. Nine studies (31.0%) involved multiple segmentations performed by different readers to

Table 1. Characteristics of the 29 Included Radiomics Studies

Article Characteristics	Data
No. of patients, median (range)	77 (24–439)
Journal type	
Clinical journal	7 (24.1)
Imaging journal	18 (62.1)
Computer science/neuroscience journal	4 (13.8)
Biomarker	
Diagnostic	21 (72.4)
Predictive	0 (0.0)
Prognostic	8 (27.6)
Topics in brain metastasis	
Differentiation of brain metastasis and other brain tumors	8 (27.6)
Differentiation of brain metastasis and radiation necrosis after radiosurgery	4 (13.8)
Prediction of primary tumor type in patients with brain metastasis	4 (13.8)
Prediction of specific gene mutations in patients with brain metastasis	5 (17.2)
Prediction of outcome of radiosurgery in patients with brain metastasis	6 (20.7)
Prediction of survival in patients with brain metastasis	2 (6.9)
Sequence used for feature extraction*	
Conventional images	23 (82.1)
Advanced images (diffusion-weighted image or diffusion tensor image)	5 (17.9)
Segmentation method	
Automatic	1 (3.4)
Semi-automatic	9 (31.0)
Manual	18 (62.1)
Both automatic and manual	1 (3.4)
External validation	
Performed	3 (10.3)
Not performed	26 (89.7)
Magnetic field strength (Tesla)*	
1.5T	10 (35.7)
3T	11 (39.3)
1.5T & 3T	6 (21.4)
Not illustrated	1 (3.6)

Data are the number of studies with a percentage in parentheses unless specified otherwise. *Relevant for 28 radiomics studies that used MRI except for one study that used PET.

evaluate segmentation reliability [14-17,23,26,29,35,40]. No study has used a test-retest approach or a phantom study.

In terms of domain 2, 24 studies (82.8%) performed feature reduction or adjustment for multiple tests. Ten studies (34.5%) performed validation using a dataset

Table 2. Radiomics Quality Score according to the Six Key Domains

	Basic Adherence Rate (%)*	Median (Range)	Percentage of the Ideal Score (%)†
Total 16 items (ideal score 36)	50.0	3 (-6–12)	9.6 (3.4)
Domain 1: protocol quality and stability in image and segmentation (0 to 5 points)	69.0 (20)	1 (0–2)	20.0 (1.0)
Protocol quality (2 points)	76.0 (19)	1 (0–1)	34.5 (0.7)
Multiple segmentations (1 point)	16.0 (4)	0 (0–1)	31.0 (0.3)
Test-retest (1 point)	0 (0)	0 (0–0)	0 (0)
Phantom study (1 point)	0 (0)	0 (0–0)	0 (0)
Domain 2: feature selection and validation (-8 to 8 points)	82.8 (24)	-2 (-8–6)	-4.3 (-0.3)
Feature reduction or adjustment of multiple testing (-3 or 3 points)	82.8 (24)	3 (-3–3)	65.5 (2.0)
Validation (-5, 2, 3, 4, or 5 points)	37.9 (11)	-5 (-5–3)	-46.2 (-2.3)
Domain 3: biologic/clinical validation and utility (0 to 6 points)	55.2 (16)	1 (0–4)	16.7 (1.0)
Non-radiomics features (1 point)	44.8 (13)	0 (0–1)	44.8 (0.4)
Biologic correlations (1 point)	13.8 (4)	0 (0–1)	13.8 (0.1)
Comparison to "gold standard" (2 points)	10.3 (3)	0 (0–2)	8.6 (0.2)
Potential clinical utility (2 points)	10.3 (3)	0 (0–2)	10.3 (0.2)
Domain 4: model performance index (0 to 5 points)	89.7 (26)	2 (0–4)	35.2 (1.8)
Cut-off analysis (1 point)	3.4 (1)	0 (0–1)	3.4 (0.0)
Discrimination statistics (2 points)	89.7 (26)	2 (0–2)	81.0 (1.6)
Calibration statistics (2 points)	6.9 (2)	0 (0–2)	5.2 (0.1)
Domain 5: high level of evidence (0 to 8 points)	0 (0)	0 (0–0)	0 (0)
Prospective study (7 points)	0 (0)	0 (0–0)	0 (0)
Cost-effectiveness analysis (1 point)	0 (0)	0 (0–0)	0 (0)
Domain 6: open science and data (0 to 4 points)	3.4 (1)	0 (0–1)	0.9 (0.0)

*Numbers in parentheses are the number of studies, †Numbers in parentheses are the mean values.

obtained from the same institution, gaining two points [13,15,17,22,27,29-32,38]. One study (3.4%) performed validation using a dataset from another institute, earning three points [14]. In the remaining 18 studies (36%), validation was missing.

For domain 3, 13 studies (44.8%) performed multivariate analyses with non-radiomics features. Four studies (13.8%) earned 1 point each for using biological correlation components [23,29,33,41]. Only three studies (10.3%) compared radiomics-based methods with the "gold standard" method (e.g., neuroradiologists' decision to distinguish brain metastasis from other tumors) [14,17,20]. Three studies (10.3%) provided potential applications of radiomics for the prediction of mutation status or progression-free survival through decision curve analysis or nomograms [26,31,41].

In terms of domain 4, only one study (3.4%) performed a cut-off analysis to stratify patients with significantly different progression-free survival [31]. Among the 26 studies (90.0%) that used discrimination statistics or a resampling method, 21 studies (80.8%) used both

discrimination statistics and a resampling method [13-17,22-29,31-33,35,37-40]. Only two studies (6.9%) used calibration statistics with statistical significance [31,41], with only one study also applying a resampling method [41].

In terms of domain 5, none of the studies were prospective or executed a cost-effectiveness analysis. For domain 6, only one study (3.4%) made the code publicly available, which defined image pre-processing and feature extraction.

Assessment of the RQS

The median RQS for the 25 radiomics studies was 3.0 (range, -6 to 12). The mean score was 3.4, which was 9.6% of the ideal score (Table 2, Fig. 2). When considering each domain, the mean score and percentage of the ideal score were the lowest in domain 2 (feature selection and validation) and highest in domain 4 (model performance). Among the six domains, domains 5 (high level of evidence) and 6 (open science and data) had mean scores of 0 and 0.03, respectively. In five studies (17.2%), neither feature selection nor validation was performed [19-21,34,41].
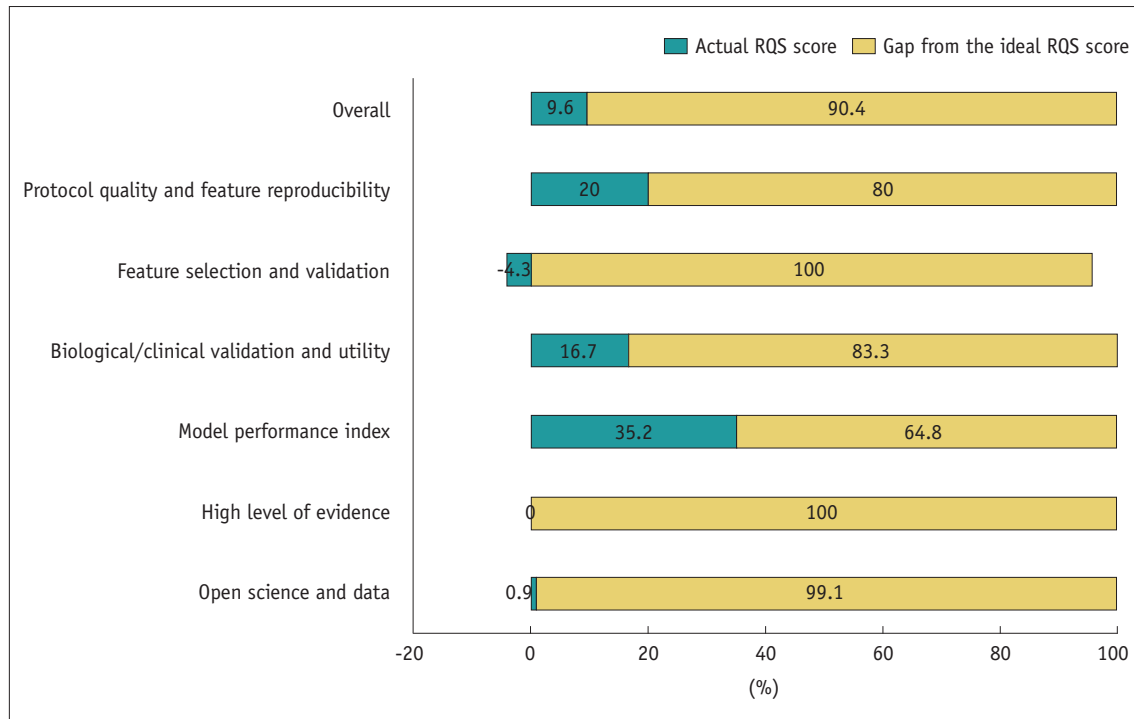
Fig. 2. RQS assessment results according to the six key domains. RQS = radiomics quality score

Completeness of Reporting a Radiomics-Based Multivariable Prediction Model Using the TRIPOD Checklist

Of the 35 items in the TRIPOD checklist, the mean number of items reported ± standard deviation was 17.2 ± 3.8 (range, 6–26). When we excluded "if relevant" and "if done" (item 5c) items from both the numerator and denominator, the adherence rate to this checklist was 60.3%. The adherence rates to each TRIPOD checklist are shown in Table 3.

Reporting of Image Pre-Processing and Radiomics Feature Extraction according to the IBSI Guidelines

Among the 29 studies, 20 (69.0%) performed signal intensity normalization. Only six studies (20.7%) performed bias-field correction (five performed N4 bias correction while one performed N3 bias correction), and nine (31.0%) performed isovoxel resampling, four (13.8%) performed skull stripping, and four (13.8%) performed gray-level discretization (Table 4, Fig. 3). The software used for the extraction of radiomics features included Pyradiomics (27.6%), Matlab (27.6%), LIFEx (13.8%), IBEX (6.9%), and others (13.8%). Two studies (6.9%) did not mention the software used. Among these, Pyradiomics and LIFEx adhered to the IBSI guidelines (Table 4, Fig. 3).

DISCUSSION

In this study, radiomics studies on brain metastasis were evaluated with respect to scientific and reporting quality. Overall, the scientific and reporting quality of radiomics studies on brain metastases was suboptimal, with a low overall adherence rate for the RQS (50.0%) and TRIPOD checklist (60.3%) items. Only three studies performed external validation, indicating that most of the studies lacked generalizability. According to the RQS, none of the studies involved test-retest or phantom studies, prospective studies, or cost-effectiveness analyses. The majority of the studies did not address biologic correlations, compare radiomics-based methods to the gold standard method, state the potential clinical utility, perform cut-off analysis or calibration statistics, and provide open science and data. According to the TRIPOD checklist, the majority of studies did not report the title adequately. No study has described the blind assessment of outcomes and handling of missing data or presented a full prediction model. The detailed pre-processing steps were missing in most of the studies.

The basic adherence rate to the RQS items was low (50.0%). Specifically, the evaluation of feature robustness was insufficient, with only a few studies performing multiple segmentations, while no studies performed phantom studies

Table 3. Adherence of Radiomics Studies to Individual Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis Checklist Items

	Total (35 Items)	Data (%)
Title and abstract (n = 29)		
Title	1. Identify developing/validating a model, target population, and the outcome	1 (3.4)
Abstract	2. Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions	23 (79.3)
Introduction (n = 29)		
Background and objectives	3a. Explain the medical context and rationale for developing/validating the model	26 (89.7)
	3b. Specify the objectives, including whether the study describes the development/validation of the model or both	24 (82.8)
Methods (n = 29 except for item 5c, for which n = 12)		
Source of data	4a. Describe the study design or source of data (randomized trial, cohort, or registry data)	26 (89.7)
	4b. Specify the key dates	25 (86.2)
Participants	5a. Specify key elements of the study setting including number and location of centers	4 (13.8)
	5b. Describe eligibility criteria for participants (inclusion and exclusion criteria)	22 (75.9)
	5c. Give details of treatment received, if relevant (n = 12)	9 out of 12
Outcome	6a. Clearly define the outcome, including how and when assessed	20 (69.0)
	6b. Report any actions to blind assessment of the outcome	0 (0)
Predictors	7a. Clearly define all predictors, including how and when assessed	28 (96.6)
	7b. Report any actions to blind assessment of predictors for the outcome and other predictors	7 (24.1)
Sample size	8. Explain how the study size was arrived at	11 (37.9)
Missing data	9. Describe how missing data were handled with details of any imputation method	0 (0)
Statistical analysis methods	10a. Describe how predictors were handled	29 (100)
	10b. Specify type of model, all model-building procedures (any predictor selection), and method for internal validation	27 (93.1)
	10d. Specify all measures used to assess model performance and if relevant, to compare multiple models (discrimination and calibration)	26 (89.7)
Risk groups	11. Provide details on how risk groups were created, if done	3 (10.3)
Results (n = 29 except for item 14b, for which n = 20)		
Participants	13a. Describe the flow of participants, including the number of participants with and without the outcome. A diagram may be helpful	10 (34.5)
	13b. Describe the characteristics of the participants, including the number of participants with missing data for predictors and outcome	21 (72.4)
Model development	14a. Specify the number of participants and outcome events in each analysis	26 (89.7)
	14b. Report the unadjusted association between each candidate predictor and outcome, if done (yes or no, n = 20)	13 (44.8)
Model specification	15a. Present the full prediction model to allow predictions for individuals (regression coefficients, intercept)	0 (0)
	15b. Explain how to the use the prediction model (nomogram, calculator, etc.)	3 (10.3)
Model performance	16. Report performance measures (with confidence intervals) for the prediction model	22 (75.9)

Table 3. Adherence of Radiomics Studies to Individual Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis Checklist Items (Continued)

	Total (35 Items)	Data (%)
Discussion		
Limitations	18. Discuss any limitations of the study	26 (89.7)
Interpretation	19b. Give an overall interpretation of the results	28 (96.6)
Implications	20. Discuss the potential clinical use of the model and implications for future research	24 (82.8)
For validation (types 2a, 2b, 3, and 4) (n = 7 except for items 10e and17, for which n = 0)		
Methods-statistical analysis methods	10c. Methods-Statistical analysis methods: describe how the predictions were calculated	6 (85.7)
	10e. Describe any model updating (recalibration), if done (n = 0)	0 (0)
Methods	12. Identify any differences from the development data in setting, eligibility criteria, outcome, and predictors	4 (57.1)
Results	13c. Show a comparison with the development data of the distribution of important variables	4 (57.1)
Results-model updating	17. Report the results from any model updating, if done (n = 0)	0 (0)
Discussion-interpretation	19a. Discuss the results with reference to performance in the development data and any other validation data	2 (28.6)

Data are the number of studies.

Table 4. Quality of Image Processing and Radiomics Feature Extraction according to the Image Biomarker Standardization Initiative Guidelines

Pre-Processing Performed	Number of Studies (%)
Bias-field correction	6 (20.7)*
Isovoxel resampling	9 (31.0)
Skull stripping	4 (13.8)
Gray-level discretization	4 (13.8)
Signal intensity normalization	20 (69.0)
Software for feature extraction	
Pyradiomics	9 (31.0)
Matlab	8 (27.6)
LIFEx	4 (13.8)
IBEX	2 (6.9)
Others	4 (13.8)
N/A	2 (6.9)

*One study performed N3 bias correction, and the remaining five studies performed N4 bias correction. N/A = not available

or collected images at multiple time points. A considerable proportion of the studies executed feature extraction, which is essential to avoid overfitting [52]. However, only three studies performed external validation, whereas 18 studies did not perform validation at all. As validation is the key process to allow radiomics to be integrated as a reliable tool for clinical practice, future radiomics studies should include a validation step. Most studies did not describe biological correlations or comparisons with the gold standard. However, there are no established gold standards for the prediction of outcomes following radiosurgery or prediction of primary tumor types, except for pathologic confirmation, which inevitably decreases the adherence rate. Few studies have provided the potential clinical utility of radiomics and made the code publicly available. None of the studies were prospective and performed cut-off analysis, calibration statistics, and cost-effectiveness analysis, resulting in low adherence rates. Thus, the quality of radiomics studies on brain metastases has the potential to improve several categories of RQS.

In the reporting of radiomics studies according to the TRIPOD checklist, there were several highly problematic items. Only one study included the terms "development" and "validation" with the target population and outcome mentioned in the title. Similarly, most of the studies lacked the description of "development" and "validation" in the abstract and did not describe the specific study design. None of the studies reported handling of missing data. Furthermore, only a few studies reported their actions for blind assessment of predictors. These results of suboptimal
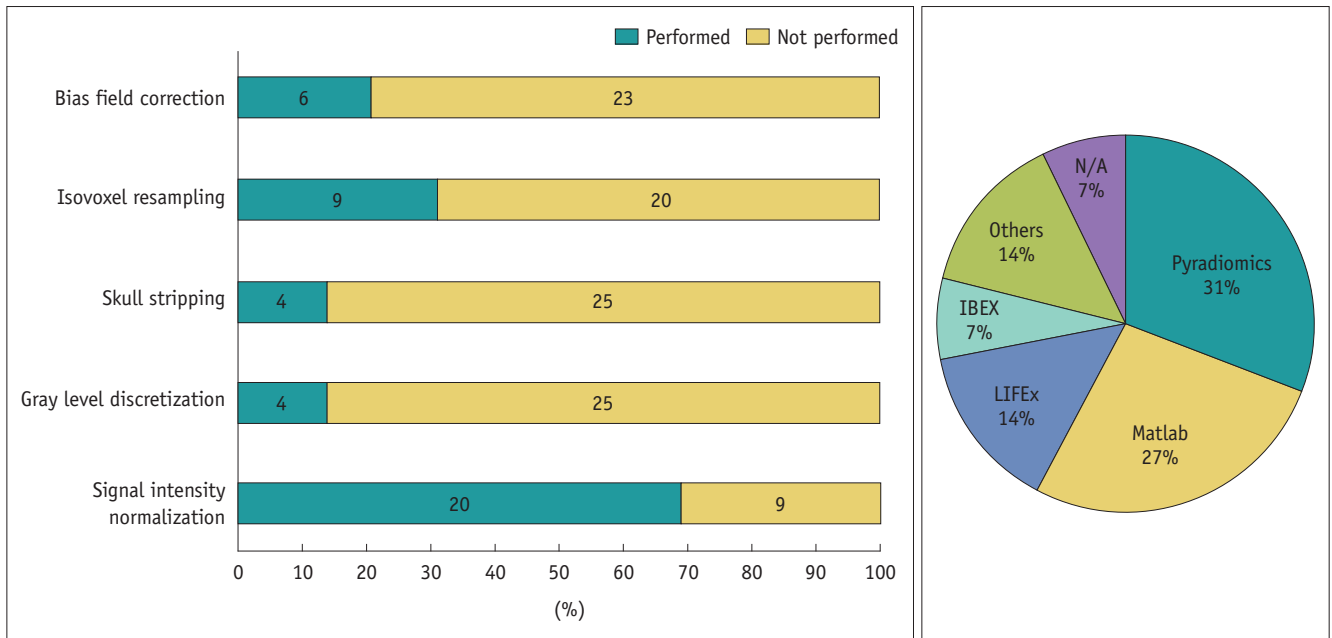
Fig. 3. Reporting of image pre-processing and radiomics feature extraction according to Image Biomarker Standardization Initiative. N/A = not available

reporting are in line with those of previous studies that investigated multivariable prediction model studies and oncologic studies according to TRIPOD [47,53]. Meanwhile, since the current TRIPOD checklists focus on regression-based prediction model approaches, several studies that were irrelevant to the regression analyses were not applicable to the assessment of TRIPOD. Specifically, the checklists associated with model development and specification are based on regression analyses. The use of TRIPOD in artificial intelligence and machine learning (ML) studies is limited, as these studies frequently did not entail regression analyses [54]. Therefore, the development of a new version of the TRIPOD statement specific to ML (TRIPOD-ML) is currently in progress, focusing on ML prediction algorithms and building an established methodology for prediction research [54].

Adherence to the nomenclature of standard radiomics features and calculations described by the IBSI are important for improving the reproducibility of studies [52,55]. However, only a few studies have reported the quality of radiomics studies that assessed adherence to the IBSI of each journal [44,56]. In our study, we focused on the pre-processing steps in IBSI and reported the number of published radiomics studies on brain metastases following the guidelines. Bias-field correction aims to reduce low-frequency intensity nonuniformity [57], while isotropic resampling reduces directional biases [58] and gray-level

discretization cluster pixels according to intensity values to enhance feature reproducibility [59]. Image intensity normalization is also a necessary step in increasing repeatability [60]. However, the majority of the studies did not execute all pre-processing steps and only performed a few. Radiomics studies on brain metastases should focus on pre-processing, which will increase the reproducibility of radiomics features.

Among the 20 studies that enrolled multiple brain metastases of more than one lesion per patient to constitute the dataset, only seven studies demonstrated the statistical analysis relevant to the handling of cluster-correlated data. Assignment of the respective metastases either to the validation set or to the training set for patients with multiple metastases was most frequently used to avoid bias from cluster correlation. As brain metastases tend to be multiple and multiple lesions from one patient may correlate with each other, specific considerations in the statistical analysis should be considered in future studies.

Our study had several limitations. First, the sample size was relatively small. Second, radiomics is still a developing imaging biomarker, and the suggested RQS and TRIPOD criteria may be too idealistic and strict. Phantom studies and multiple imaging acquisitions have rarely been applied to real-world practice. Additionally, it may be too extensive to provide the whole unadjusted association between each candidate predictor and outcome and the full prediction

model. However, considering these details is necessary to improve the scientific and reporting quality of radiomics to enable the clinical translation of radiomics in the future.

In conclusion, the overall scientific and reporting quality of radiomics studies on brain metastases published during the study period was insufficient, and this low quality may hamper the use of radiomics in the clinical field. The RQS, TRIPOD checklist, and IBSI guidelines should be adhered to make radiomics a more robust decision-making tool in clinical practice.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2021.0421.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Yae Won Park, Seung-Koo Lee. Data curation: Chae Jung Park, Sung Soo Ahn. Formal analysis: Chae Jung Park, Dain Kim. Funding acquisition: Yae Won Park, Eui Hyun Kim, Seok-Gu Kang. Investigation: Yae Won Park, Jong Hee Chang, Se Hoon Kim. Methodology: Jong Hee Chang, Seung-Koo Lee. Project administration: Se Hoon Kim, Seung-Koo Lee. Resources: Jong Hee Chang, Se Hoon Kim. Software: Chae Jung Park, Dain Kim. Supervision: Yae Won Park, Sung Soo Ahn, Seung-Koo Lee. Validation: Dain Kim, Eui Hyun Kim, Seok-Gu Kang. Visualization: Sung Soo Ahn, Seung-Koo Lee. Writing—original draft: Chae Jung Park, Yae Won Park. Writing—review & editing: Chae Jung Park, Yae Won Park, Sung Soo Ahn.

ORCID iDs

Chae Jung Park
  https://orcid.org/0000-0002-5567-8658
Yae Won Park
  https://orcid.org/0000-0001-8907-5401
Sung Soo Ahn
  https://orcid.org/0000-0002-0503-5558
Dain Kim
  https://orcid.org/0000-0003-3425-0182
Eui Hyun Kim
  https://orcid.org/0000-0002-2523-7122
Seok-Gu Kang
  https://orcid.org/0000-0001-5676-2037
Jong Hee Chang
  https://orcid.org/0000-0003-1509-9800
Se Hoon Kim
  https://orcid.org/0000-0001-7516-7372
Seung-Koo Lee
  https://orcid.org/0000-0001-5646-4072

REFERENCES

1. Cagney DN, Martin AM, Catalano PJ, Redig AJ, Lin NU, Lee EQ, et al. Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: a population-based study. *Neuro Oncol* 2017;19:1511-1521
2. Nayak L, Lee EQ, Wen PY. Epidemiology of brain metastases. *Curr Oncol Rep* 2012;14:48-54
3. Suh JH, Kotecha R, Chao ST, Ahluwalia MS, Sahgal A, Chang EL. Current approaches to the management of brain metastases. *Nat Rev Clin Oncol* 2020;17:279-299
4. Kaufmann TJ, Smits M, Boxerman J, Huang R, Barboriak DP, Weller M, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. *Neuro Oncol* 2020;22:757-772
5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278:563-577
6. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006
7. Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, Kim SH, et al. Machine learning and radiomic phenotyping of lower grade gliomas: improving survival prediction. *Eur Radiol* 2020;30:3834-3842
8. Park CJ, Han K, Kim H, Ahn SS, Choi YS, Park YW, et al. Radiomics risk score may be a potential imaging biomarker for predicting survival in isocitrate dehydrogenase wild-type lower-grade gliomas. *Eur Radiol* 2020;30:6464-6474
9. Li ZZ, Liu PF, An TT, Yang HC, Zhang W, Wang JX. Construction of a prognostic immune signature for lower grade glioma

that can be recognized by MRI radiomics features to predict survival in LGG patients. *Transl Oncol* 2021;14:101065

10. Park CJ, Han K, Kim H, Ahn SS, Choi D, Park YW, et al. MRI features may predict molecular features of glioblastoma in isocitrate dehydrogenase wild-type lower-grade gliomas. *AJNR Am J Neuroradiol* 2021;42:448-456

11. Wang J, Zheng X, Zhang J, Xue H, Wang L, Jing R, et al. An MRI-based radiomics signature as a pretreatment noninvasive predictor of overall survival and chemotherapeutic benefits in lower-grade gliomas. *Eur Radiol* 2021;31:1785-1794

12. Park YW, Choi D, Park JE, Ahn SS, Kim H, Chang JH, et al. Differentiation of recurrent glioblastoma from radiation necrosis using diffusion radiomics with machine learning model development and external validation. *Sci Rep* 2021;11:2913

13. Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J Magn Reson Imaging* 2019;50:519-528

14. Bae S, An C, Ahn SS, Kim H, Han K, Kim SW, et al. Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. *Sci Rep* 2020;10:12110

15. Chen C, Ou X, Wang J, Guo W, Ma X. Radiomics-based machine learning in differentiation between glioblastoma and metastatic brain tumors. *Front Oncol* 2019;9:806

16. Ortiz-Ramón R, Ruiz-España S, Mollá-Olmos E, Moratal D. Glioblastomas and brain metastases differentiation following an MRI texture analysis-based radiomics approach. *Phys Med* 2020;76:44-54

17. Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, et al. Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. *Cancer Lett* 2019;451:128-135

18. Sachdeva J, Kumar V, Gupta I, Khandelwal N, Ahuja CK. Segmentation, feature extraction, and multiclass brain tumor classification. *J Digit Imaging* 2013;26:1141-1150

19. Tateishi M, Nakaura T, Kitajima M, Uetani H, Nakagawa M, Inoue T, et al. An initial experience of machine learning based on multi-sequence texture parameters in magnetic resonance imaging to differentiate glioblastoma from brain metastases. *J Neurol Sci* 2020;410:116514

20. Zhang Y, Chen C, Tian Z, Feng R, Cheng Y, Xu J. The diagnostic value of MRI-based texture analysis in discrimination of tumors located in posterior fossa: a preliminary study. *Front Neurosci* 2019;13:1113

21. Béresová M, Larroza A, Arana E, Varga J, Balkay L, Moratal D. 2D and 3D texture analysis to differentiate brain metastases on MR images: proceed with caution. *MAGMA* 2018;31:285-294

22. Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. *Radiology* 2019;290:479-487

23. Li Z, Mao Y, Li H, Yu G, Wan H, Li B. Differentiating brain metastases from different pathological types of lung cancers using texture analysis of T1 postcontrast MR. *Magn Reson Med* 2016;76:1410-1419

24. Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D. Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol* 2018;28:4514-4523

25. Ahn SJ, Kwon H, Yang JJ, Park M, Cha YJ, Suh SH, et al. Contrast-enhanced T1-weighted image radiomics of brain metastases may predict EGFR mutation status in primary lung cancer. *Sci Rep* 2020;10:8905

26. Chen BT, Jin T, Ye N, Mambetsariev I, Daniel E, Wang T, et al. Radiomic prediction of mutation status based on MR imaging of lung cancer brain metastases. *Magn Reson Imaging* 2020;69:49-56

27. Park YW, An C, Lee J, Han K, Choi D, Ahn SS, et al. Diffusion tensor and postcontrast T1-weighted imaging radiomics to differentiate the epidermal growth factor receptor mutation status of brain metastases from non-small cell lung cancer. *Neuroradiology* 2021;63:343-352

28. Shofty B, Artzi M, Shtrozberg S, Fanizzi C, DiMeco F, Haim O, et al. Virtual biopsy using MRI radiomics for prediction of BRAF status in melanoma brain metastasis. *Sci Rep* 2020;10:6623

29. Wang G, Wang B, Wang Z, Li W, Xiu J, Liu Z, et al. Radiomics signature of brain metastasis: prediction of EGFR mutation status. *Eur Radiol* 2021;31:4538-4547

30. Bhatia A, Birger M, Veeraraghavan H, Um H, Tixier F, McKenney AS, et al. MRI radiomic features are associated with survival in melanoma brain metastases treated with immune checkpoint inhibitors. *Neuro Oncol* 2019;21:1578-1586

31. Zhao S, Hou D, Zheng X, Song W, Liu X, Wang S, et al. MRI radiomic signature predicts intracranial progression-free survival in patients with brain metastases of ALK-positive non-small cell lung cancer. *Transl Lung Cancer Res* 2021;10:368-380

32. Larroza A, Moratal D, Paredes-Sánchez A, Soria-Olivas E, Chust ML, Arribas LA, et al. Support vector machine classification of brain metastasis and radiation necrosis based on texture analysis in MRI. *J Magn Reson Imaging* 2015;42:1362-1368

33. Lohmann P, Kocher M, Ceccon G, Bauer EK, Stoffels G, Viswanathan S, et al. Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis. *Neuroimage Clin* 2018;20:537-542

34. Lohmann P, Stoffels G, Ceccon G, Rapp M, Sabel M, Filss CP, et al. Radiation injury vs. recurrent brain metastasis: combining textural feature radiomics analysis and standard parameters may increase 18F-FET PET accuracy without dynamic scans. *Eur Radiol* 2017;27:2916-2927

35. Peng L, Parekh V, Huang P, Lin DD, Sheikh K, Baker B, et al. Distinguishing true progression from radionecrosis after stereotactic radiation therapy for brain metastases with machine learning and radiomics. *Int J Radiat Oncol Biol Phys* 2018;102:1236-1243

36. Huang CY, Lee CC, Yang HC, Lin CJ, Wu HM, Chung WY, et al. Radiomics as prognostic factor in brain metastases treated

with Gamma Knife radiosurgery. *J Neurooncol* 2020;146:439-449

37. Karami E, Soliman H, Ruschin M, Sahgal A, Myrehaug S, Tseng CL, et al. Quantitative MRI biomarkers of stereotactic radiotherapy outcome in brain metastasis. *Sci Rep* 2019;9:19830

38. Kawahara D, Tang X, Lee CK, Nagata Y, Watanabe Y. Predicting the local response of metastatic brain tumor to gamma knife radiosurgery by radiomics with a machine learning method. *Front Oncol* 2021 Jan [Epub]. https://doi.org/10.3389/fonc.2020.569461

39. Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee YK, Karam I, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. *Neuro Oncol* 2020;22:797-805

40. Park JH, Choi BS, Han JH, Kim CY, Cho J, Bae YJ, et al. MRI texture analysis for the prediction of stereotactic radiosurgery outcomes in brain metastases from lung cancer. *J Clin Med* 2021;10:237

41. Zheng Y, Geng D, Yu T, Xia W, She D, Liu L, et al. Prognostic value of pretreatment MRI texture features in breast cancer brain metastasis treated with Gamma Knife radiosurgery. *Acta Radiol* 2021;62:1208-1216

42. Tong E, McCullagh KL, Iv M. Advanced imaging of brain metastases: from augmenting visualization and improving diagnosis to evaluating treatment response. *Front Neurol* 2020;11:270

43. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749-762

44. Won SY, Park YW, Ahn SS, Moon JH, Kim EH, Kang SG, et al. Quality assessment of meningioma radiomics studies: bridging the gap between exploratory research and clinical applications. *Eur J Radiol* 2021;138:109673

45. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-W73

46. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020;295:328-338

47. Park JE, Kim D, Kim HS, Park SY, Kim JY, Cho SJ, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol* 2020;30:523-536

48. Park JE, Kim HS, Kim D, Park SY, Kim JY, Cho SJ, 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;20:29

49. Won SY, Park YW, Park M, Ahn SS, Kim J, Lee SK. Quality reporting of radiomics analysis in mild cognitive impairment and Alzheimer's disease: a roadmap for moving forward. *Korean J Radiol* 2020;21:1345-1354

50. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. arXiv [Preprint]. 2016 [cited 2021 May 1]. Available at: https://doi.org/10.1148/radiol.2020191145

51. Stefano A, Comelli A, Bravatà V, Barone S, Daskalovski I, Savoca G, et al. A preliminary PET radiomics study of brain metastases using a fully automatic segmentation method. *BMC Bioinformatics* 2020;21:325

52. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives. *Korean J Radiol* 2019;20:1124-1137

53. Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Med* 2018;16:120

54. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019;393:1577-1579

55. Moradmand H, Aghamiri SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *J Appl Clin Med Phys* 2020;21:179-190

56. Crombé A, Fadli D, Italiano A, Saut O, Buy X, Kind M. Systematic review of sarcomas radiomics studies: bridging the gap between concepts and clinical applications? *Eur J Radiol* 2020;132:109283

57. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 2010;29:1310-1320

58. Whybra P, Parkinson C, Foley K, Staffurth J, Spezi E. Assessing radiomic feature robustness to interpolation in 18F-FDG PET imaging. *Sci Rep* 2019;9:9649

59. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik JC, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS One* 2019;14:e0213459

60. Scalco E, Belfatto A, Mastropietro A, Rancati T, Avuzzi B, Messina A, et al. T2w-MRI signal normalization affects radiomics features reproducibility. *Med Phys* 2020;47:1680-1691