IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement

Ji Eun Park[1] · Donghyun Kim[2] · Ho Sung Kim[1] · Seo Young Park[3] · Jung Youn Kim[4] · Se Jin Cho[1] · Jae Ho Shin[5] · Jeong Hoon Kim[6]

Abstract

Objectives To evaluate radiomics studies according to radiomics quality score (RQS) and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) to provide objective measurement of radiomics research.

Materials and methods PubMed and Embase were searched for studies published in high clinical imaging journals until December 2018 using the terms "radiomics" and "radiogenomics." Studies were scored against the items in the RQS and TRIPOD guidelines. Subgroup analyses were performed for journal type (clinical vs. imaging), intended use (diagnostic vs. prognostic), and imaging modality (CT vs. MRI), and articles were compared using Fisher's exact test and Mann-Whitney analysis.

Results Seventy-seven articles were included. The mean RQS score was 26.1% of the maximum (9.4 out of 36). The RQS was low in demonstration of clinical utility (19.5%), test-retest analysis (6.5%), prospective study (3.9%), and open science (3.9%). None of the studies conducted a phantom or cost-effectiveness analysis. The adherence rate for TRIPOD was 57.8% (mean) and was particularly low in reporting title (2.6%), stating study objective in abstract and introduction (7.8% and 16.9%), blind assessment of outcome (14.3%), sample size (6.5%), and missing data (11.7%) categories. Studies in clinical journals scored higher and more frequently adopted external validation than imaging journals.

Conclusions The overall scientific quality and reporting of radiomics studies is insufficient. Scientific improvements need to be made to feature reproducibility, analysis of clinical utility, and open science categories. Reporting of study objectives, blind assessment, sample size, and missing data is deemed to be necessary.

Key Points

- *The overall scientific quality and reporting of radiomics studies is insufficient.*
- *The RQS was low in demonstration of clinical utility, test-retest analysis, prospective study, and open science.*
- *Room for improvement was shown in TRIPOD in stating study objective in abstract and introduction, blind assessment of outcome, sample size, and missing data categories.*

Ji Eun Park and Donghyun Kim contributed equally to this work.

✉ Ho Sung Kim
   radhskim@gmail.com

1 Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 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, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea

4 Department of Radiology, Kangbuk Samsung Medical Center, Seoul, South Korea

5 St. Vincent Hospital, College of Medicine, The Catholic University of Korea, Suwon, South Korea

6 Department of Neurosurgery, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea

Abbreviations

RQS Radiomics quality score,
TRIPOD Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

Introduction

Radiomics research has been rapidly expanding ever since Gilles et al declared "images are data" [1]. Sophisticated bioinformatics tools are applied to reduce data dimensionality and select features from high-dimensional data, and models with potential diagnostic or prognostic utility are typically developed [1–3]. Although radiomics research shows great potential, its current use is confined to the academic literature, without real-world clinical applications. High quality in science and reporting may present strategies for radiomics to become an effective imaging biomarker able to cross the "translational gap" [4, 5] for use in guiding clinical decisions.

The quality of scientific research articles consists of two elements: the quality of the science and the quality of the report [6], and deficiencies in either may hamper translation of biomarkers to patient care [7]. With regard to the quality of the science, a system of metrics in the form of the radiomics quality score (RQS) was developed by the expert opinions of Lambin et al [2], to determine the validity and completeness of radiomics studies. The RQS consists of 16 components that consider radiomics-specific high-dimensional data and modeling and accounts for image protocol and feature reproducibility, biologic/clinical validation and utility, performance index, high level of evidence, and open science. With regard to the quality of reporting, radiomics research is a model-based approach and reporting according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) initiative [8] is desirable.

To our knowledge, the quality of the science and reporting in radiomics research studies is largely unknown. A RQS study from the score developer [3] reported an average score of less than 50% over 41 radiomics studies, but the RQS score is underutilized because many investigators and peer reviewers are unfamiliar with it. Prediction model studies in the clinical domain showed suboptimal quality of reporting according to TRIPOD [9], but whether the radiomics domain is good or bad at reporting has not been studied. In this study, we evaluated radiomics studies using RQS and TRIPOD items to evaluate their scientific quality and assessed whether the score and degree of adherence depend on the study design or journal type. The purpose of the study was therefore to evaluate the quality of the science and reporting of radiomics studies according to RQS and TRIPOD.

Materials and methods

### Article search strategy and study selection

A search was conducted for all potentially relevant original research papers using radiomics analysis published up until December 3, 2018. The search terms used to find radiomics studies were "radiomic" OR "radiogenomic" in the MEDLINE (National Center for Biotechnology Information, NCBI) and EMBASE databases. The eligible articles were high-impact factor medical journals ranked higher than 7.0 according to the 2018 edition of the Journal Citation reports, as well as those published in radiology journal of *Radiology* and *European Radiology*. The impact factor of 7.0 was chosen as it was considered that articles published in journals above 7.0 would be representative of the reporting of high-quality clinical studies on radiomics analysis. Imaging journals were chosen because they are the highest-ranked US and non-US general radiology, given the impact and status of the two journals. The inclusion process is shown in Fig. 1. Study selection and data extraction are shown in Supplementary Materials 1.

Analysis of method quality based on RQS

The RQS score with 16 components is defined in Supplementary Table 1 [2]. The reviewers extracted the data using a predetermined RQS evaluation according to six domains. Domain 1 covers the 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 imaging at multiple time points (1 point). Domain 2 covers 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 covers the 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 4 covers the reporting of biological/clinical validation and utility: multivariate analysis with non-radiomics features (1 point), biological correlates (1 point), comparison with the gold standard (2 points), and potential clinical utility (2 points). Domain 5 covers the 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) covers open science, with open availability of source code and data (4 points).

The six domains and topics which were subject to further discussions until a consensus was reached were in Supplementary Materials 1.

Fig. 1 Flow diagram of the study selection process. *Note*: Non-relevant to radiomics indicate the analytic methods are volumetric measurement and locations and not categorized into radiomics analysis
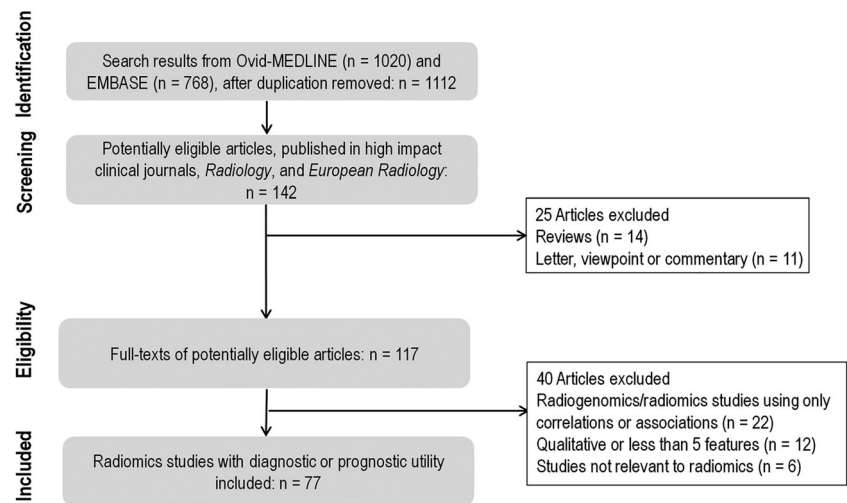
Analysis of reporting completeness based on TRIPOD statement

The TRIPOD checklist was applied to each article to determine the completeness of reporting. The details of the checklist are described elsewhere [8], but it consists of 22 main criteria with 37 items. First, the type of prediction model was decided, whether the radiomics model was development only (type 1a), development and validation using resampling (type 1b), random split-sample validation (type 2a), nonrandom split-sample validation (type 2b), validation using separate data (type 3), or validation only (type 4). The details for TRIPOD checklist and data extraction are shown in Supplementary Materials 1.

Analysis of the role of radiologists

To demonstrate the role of radiologists in the radiomics studies, the analysis was undertaken to calculate the position and number of radiologist among the author lists. The radiologists include general radiologists and nuclear medicine radiologists. First, the main authors, either first or corresponding author, were checked. When the radiologists are not main authors, the position and number of radiologists among the author's lists were checked. The position is checked for the first appearance (i.e., 3rd and 5th author are radiologists among 8 authors, the position was checked as 3/8, 0.37).

Statistical analysis

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 to RQS (for 0–16 criteria)

and each item scored in TRIPOD were counted (range, 0–35 items) and calculated in a descriptive manner using proportions (%). The TRIPOD item 5c ("if done" item) and the validation items 10c, 10e, 12, 13c, 17, and 19a were excluded from both the numerator and denominator when the overall adherence rate was calculated. For all included articles, the total RQS score was calculated (score range, − 8 to 36) and expressed as mean ± standard deviation. A graphical display for the proportion of studies was adopted from the suggested graphical display for Quality Assessment of Diagnostic Accuracy Studies-2 results [10].

Subgroup analyses were performed to determine whether the reporting quality differed according to intended use (diagnostic or prognostic), journal type (clinical or imaging journal), and imaging modality (CT or MRI). Additionally, we compared RQS between radiogenomics studies and non-radiogenomics studies. Before subgroup analysis, the RQS was plotted for each journal to observe whether there was a systematic difference between journals (Supplementary Figure 2). As no systematic difference was observed between journals, the journal was not adjusted for in the analysis. The nonparametric Mann-Whitney U test was used to compare the RQS score in each group. Fisher's exact test was used to compare proportions in RQS and TRIPOD for small sample sizes in each group. All statistical analyses were performed using SPSS (SPSS version 22; SPSS) and R (R version 3.3.3; R Foundation for Statistical Computing), and a *p* value < .05 was considered statistically significant.

Results

Characteristics of the included studies

Seventy-seven articles [11–87] were finally analyzed. The journal, impact factor, study topic, intended use, imaging

modality, number of patients, and model type are summarized in Supplementary Table 2. The number and characteristics of the included radiomics studies are provided in Table 1 and Fig. 2. The mean patient number was 232 (standard deviation, 248.7; range, 38–2029). The studies were published in 2014 (1 article), 2016 (8 articles), 2017 (16 articles), and 2018 (52 articles). There were 25 articles published in high IF clinical journals and 52 articles in imaging journals (14 in *Radiology* and 38 in *European Radiology*). Most articles were oncologic studies (90.9%). Radiomics analysis was most frequently studied as a diagnostic biomarker (80.5%), then as a prognostic (19.5%) biomarker. MRI was the most studied modality (66.0%), followed by CT (26.0%), and PET or US (each 4.0%). Analysis of the validation methods revealed that external validation was missing in 63 out of 77 studies (81.8%). In the oncologic studies, the study purposes most frequently included histopathologic grade and differential diagnosis

(51.9%), followed by molecular or genomic classification (21.4%), survival prediction (12.8%), and assessment of treatment response (11.4%).

RQS according to the six key domains

Table 2 summarizes the results. The averaged RQS of the 77 studies expressed as a percentage of the ideal score according to the six key domains is shown in Fig. 3. The mean RQS score of the 36 studies was 9.40 (standard deviation, 5.60), which was 26.1% of the ideal score of 36. The lowest score was − 5, and the highest score was 21 (58.3% of the ideal quality score).

In domain 1, all studies except one reported well-documented image acquisition protocols or the use of publicly available image databases. Multiple segmentations by two readers were performed in 35 of the 77 studies (45.4%),

Table 1 Characteristics of the 77 included radiomics studies with diagnostic or prognostic utility

Article characteristics		Number of articles
Patient number	232 (standard deviation 248.7; range, 38–2029)	
Journal type		
	Clinical journal	25 (32.5)
	Imaging journal	52 (67.5)
Study topic		
	Oncology	70 (90.9)
	Not oncology	7 (9.1)
Intended use		
	Diagnostic	62 (80.5)
	Prognostic	15 (19.5)
Imaging type		
	CT	20 (26.0)
	MRI	51 (66.0)
	PET	3 (4.0)
	US	3 (4.0)
Validation type		
	No validation (type 1a)	8 (10.3)
	Validation using resampling (type 1b)	15 (19.5)
	Random split-sample validation (type 2a)	15 (19.5)
	Nonrandom split-sample validation (type 2b)	20 (26)
	External validation (type 3)	14 (18.2)
	Validation only (type 4)	0
	Uncertain whether random or nonrandom split sample (2a or 2b)	5 (6.5)
Topic in oncology (n = 70)		
	Histopathologic grade/differential diagnosis	40[†] (51.9)
	Molecular/genomic classification	15 (21.4)
	Survival prediction	9[†] (12.8)
	Response to treatment	8 (11.4)

Note: numbers in parentheses are percentages. [†] Two studies overlap in both histopathological grade and survival

Fig. 2 Summary charts of the 77 included radiomics studies are displayed according to disease, biomarker design, imaging type, and topic in oncological studies
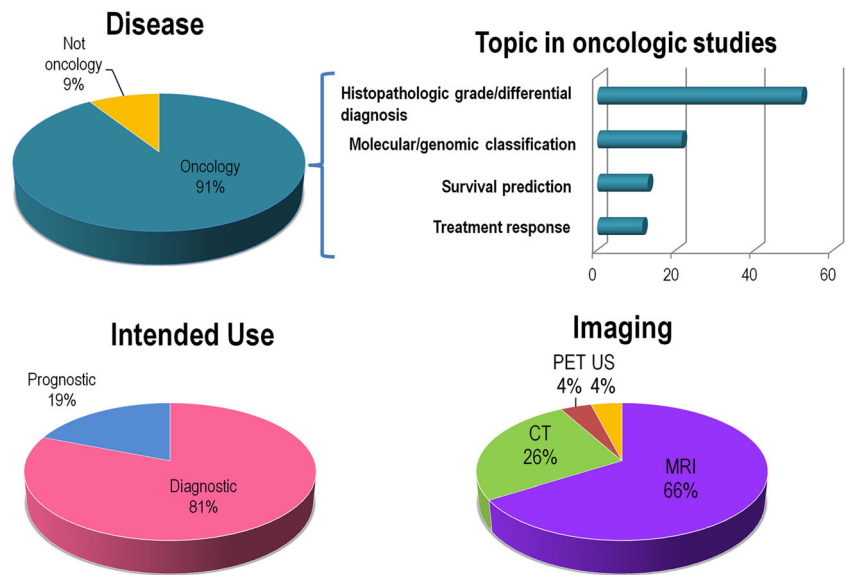
including six studies with automatic segmentation. Notably, only five studies [11, 12, 20, 26, 47] conducted imaging at multiple time points and tested feature robustness. No articles conducted a phantom study.

In domain 2, most studies adopted appropriate feature reduction or adjustment for multiple testing (74/77, 96.1%). 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 lasso, elastic net, random forest, recursive feature elimination, and support vector machine. Validation was performed without retraining from

Table 2 Radiomics quality score according to the six key domains

	Basic adherence rate	Mean score	Percentage of the ideal score (%)
Total (ideal score 36)	38.7%	9.40 ± 5.60	26.1
Domain 1: Protocol quality and stability in image and segmentation (0 to 5 points)		0.40 ± 0.54	8
Protocol quality (2)	76 (98.7%)	1.09 ± 0.33	54.5
Test-retest (1)	5 (6.5%)	0.05 ± 0.25	5
Phantom study (1)	0 (0%)	0	0
Multiple segmentation (1)	35 (45.4%)	0.45 ± 0.50	45
Domain 2: Feature selection and validation (− 8 to 8 points)		1.61 ± 2.91	20.1
Feature reduction or adjustment of multiple testing (− 3 or 3)	74 (96.1%)	2.76 ± 1.16	92
Validation (− 5, 2, 3, 4, or 5)	54 (70.1%)	0.46 ± 3.61	9.2
Domain 3: Model performance index (0 to 5 points)		0.70 ± 0.78	14
Discrimination statistics (2)	76 (98.7%)	1.54 ± 0.53	77
Calibration statistics (2)	23 (29.9%)	0.34 ± 0.55	17
Cut-off analysis (1)	16 (20.8%)	0.21 ± 0.41	21
Domain 4: Biologic/clinical validation and utility (0 to 6 points)		0.53 ± 0.76	8.8
Non-radiomics features (1)	39 (50.6%)	0.50 ± 0.50	50
Biologic correlates (1)	22 (28.6%)	0.28 ± 0.45	28
Comparison to "gold standard" (2)	36 (46.7%)	0.93 ± 1.00	46.5
Potential clinical utility (2)	15 (19.5%)	0.39 ± 0.79	19.5
Domain 5: High level of evidence (0 to 8 points)		0.14 ± 0.97	1.7
Prospective study (7)	3 (3.9%)	0.27 ± 1.36	3.8
Cost-effective analysis (1)	0 (0%)	0	0
Domain 6: Open science and data (0 to 4 points)	3 (3.9%)	0.08 ± 0.35	2
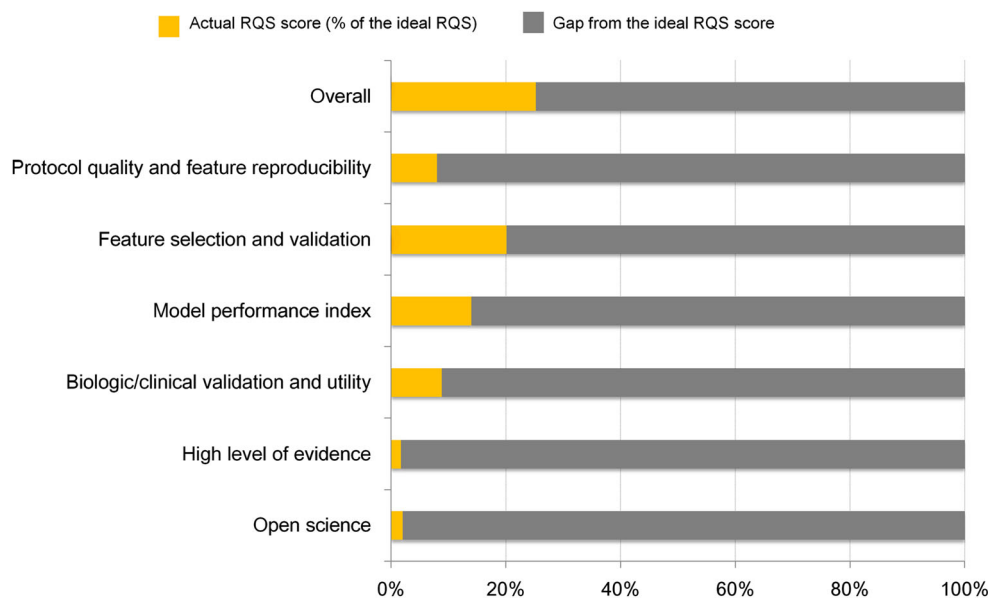
the same or a different institute in 70.1% of studies (54 out of 77).

In domain 3, all studies used discriminative statistics, but one study [23] provided hazard ratios and p values from a log-rank test for survival analysis instead of the C-index.

In domain 4, half of the studies evaluated relationships between the radiomics features and non-radiomics features (50.6%), but only 28.6% of studies found biological correlates of radiomics to provide a more holistic model and imply biological relevance. Less than half of the studies (46.7%) compared results with an existing gold standard. By contrast, in terms of clinical utility, only 15 studies (19.5%) analyzed a net improvement in health outcomes using decision curve analysis or other statistical tools.

Surprisingly, studies were deficient in demonstrating a high level of evidence such as a prospective design or cost-effectiveness analysis. Only three studies [21, 48, 55] (3.9%) included prospective validation, and no studies conducted cost-effective analysis. For domain 6, only three studies [33, 34, 46] (3.9%) made their code and/or data publicly available.

Both feature reduction and validation were missing from the study [25] with the lowest score. Meanwhile, seven studies with the highest scores [12, 14, 21, 26, 48, 54, 55] (three articles with a RQS score of 16, 1 article with 18, 1 article with 19, and 1 article with 21) earned additional points by using publicly available images [12], multiple segmentation [12, 14, 26], test-retest analysis [12, 26], and validation using three or more datasets [12, 21, 26, 55], demonstrating potential clinical utility using decision curve analysis [14, 54] and conducting prospective validation [21, 48, 55], with all studies fulfilling requirements for image protocol quality, feature reduction, and use of a discrimination index.

Completeness in reporting a radiomics-based multivariable prediction model using TRIPOD

The mean number of TRIPOD items reported was 18.51 ± 3.96 (standard deviation; range, 11–26) when all 35 items were considered. The adherence rate for TRIPOD was $57.8\% \pm 10.9\%$ (standard deviation; range, 33–78%) when "if relevant" and "if done" items were excluded from both the numerator and denominator. The completeness of reporting individual TRIPOD items is shown in Table 3. The detailed results are shown in Supplementary Materials 2.

Subgroup analysis

The results of the subgroup analysis are shown in Table 4. Prognostic studies showed a trend for a higher RQS score than diagnostic studies (11.83 ± 5.03 vs. 8.93 ± 5.52), but this was not statistically significant. Prognostic studies received a higher score than diagnostic studies in comparison with a "gold standard" ($p < .001$) and using cut-off analysis ($p < .001$). This was reflected in the TRIPOD items, with the prognostic studies showing higher adherence rates in "describing risk group" ($p = .007$) and "report unadjusted association between predictors and outcome" (if done, $p = .017$).

Studies in clinical journals also showed significantly higher RQS scores than those in imaging journals (12.2 ± 5.23 vs. 8.03 ± 5.17, $p = .001$). They achieved a higher score in protocol quality ($p = .018$), test-retest analysis ($p < .001$), validation ($p = .012$), multivariable analysis with non-radiomics features ($p = .036$), finding biologic correlates ($p = .009$), and conducted prospective study ($p = .011$). In the reporting quality, studies in clinical journals well reported the study design or source of data ($p = .047$) and reported unadjusted association.

Table 3 Adherence to individual TRIPOD items in radiomics studies

	All articles ($n = 77$)
Total (35 items)	18.5 (57.8%)
Title and abstract	
1. Title: identify developing/validating a model, target population, and the outcome	2 (2.6)
2. Abstract: provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions	6 (7.8)
Introduction	
3a. Explain the medical context and rationale for developing/validating the model	72 (93.5)
3b. Specify the objectives, including whether the study describes the development/validation of the model or both	13 (16.9)
Methods	
4a. Source of data: describe the study design or source of data (randomized trial, cohort, or registry data)	45 (58.4)
4b. Source of data: specify the key dates	64 (83.1)
5a. Participants: specify key elements of the study setting including number and location of centers	49 (63.6)
5b. Participants: describe eligibility criteria for participants (inclusion and exclusion criteria)	67 (87.0)
5c. Participants: give details of treatment received, *if relevant* ($n = 12$)	8 out of 12
6a. Outcome: clearly define the outcome, including how and when assessed	70 (90.9)
6b. Outcome: report any actions to blind assessment of the outcome	11 (14.3)
7a. Predictors: clearly define all predictors, including how and when assessed	77 (100)
7b. Predictors: report any actions to blind assessment of predictors for the outcome and other predictors	24 (32.1)
8. Sample size: explain how the study size was arrived at	5 (6.5)
9. Missing data: describe how missing data were handled with details of any imputation method	9 (11.7)
10a. Statistical analysis methods: describe how predictors were handled	77 (100)
10b. Statistical analysis methods: specify type of model, all model-building procedures (any predictor selection), and method for internal validation	65 (84.4)
10d. Statistical analysis methods: specify all measures used to assess model performance and if relevant, to compare multiple models (discrimination and calibration)	21 (27.3)
11. Risk groups: provide details on how risk groups were created, *if done* (yes or no, $n = 77$)	16 (20.8)
Results	
13a. Participants: describe the flow of participants, including the number of participants with and without the outcome. A diagram may be helpful	71 (92.2)
13b. Participants: describe the characteristics of the participants, including the number of participants with missing data for predictors and outcome	68 (88.3)
14a. Model development: specify the number of participants and outcome events in each analysis	69 (89.6)
14b. Model development: report the unadjusted association between each candidate predictor and outcome, *if done* (yes or no, $n = 77$)	12 (15.6)
15a. Model specification: present the full prediction model to allow predictions for individuals (regression coefficients, intercept)	29 (37.7)
15b. Model specification: explain how to the use the prediction model (nomogram, calculator, etc)	24 (31.2)
16. Model performance: report performance measures (with confidence intervals) for the prediction model	52 (67.5)
Discussion	
18. Limitations: Discuss any limitations of the study	76 (98.7)
19b. Interpretation: Give an overall interpretation of the results	77 (100)
20. Implications: Discuss the potential clinical use of the model and implications for future research	76 (98.7)
For validation (types 2a, 2b, 3, and 4)	$n = 54$
10c. Methods-Statistical analysis methods: describe how the predictions were calculated	33 (66.1)
10e. Methods-Statistical analysis methods: describe any model updating (recalibration), *if done*	0
12. Methods-Identify any differences from the development data in setting, eligibility criteria, outcome, and predictors	49 (90.7)
13c. Results-show a comparison with the development data of the distribution of important variables	41 (75.9)
17. Results-Model updating: report the results from any model updating, if done	0
19a. Discussion-Interpretation: discuss the results with reference to performance in the development data and any other validation data	46 (85.2)

Table 4 Subgroup analysis of RQS and TRIPOD items in radiomics studies according to the intended use, impact factor, and imaging modality

Radiomics quality score	Mean score (n = 77)	Diagnostic (n = 65)	Prognostic (n = 12)	p	Clinical (n = 25)	Imaging (n = 52)	p	CT+ (n = 20)	MRI+ (n = 51)	p
	9.40±5.60	8.93±5.52	11.83±5.03	.106	12.2±5.23	8.03±5.17	.001	11.8±3.71	9±5.2	.055
Domain 1: Protocol quality and stability in image and segmentation (0 to 5 points)										
Protocol quality (2)	1.09±0.33	1.08±0.31	1.08±0.27	.968	1.2±0.4	1.02±0.24	.018	1.10±0.3	1.08±0.33	.508
Test-retest (1)	0.05±0.25	0.06±0.24	0.08±0.27	.792	0.2±0.4	0	<.001	0.15±0.36	0.04±0.19	.028
Phantom study (1)	0	0	0	NA	0	0	NA	0	0	NA
Multiple segmentation (1)	0.45±0.50	0.49±0.49	0.25±0.43	.125	0.35±0.48	0.5±0.5	.253	0.65±0.48	0.39±0.48	.301
Domain 2: Feature selection and validation (−8 to 8 points)										
Feature reduction or adjustment of multiple testing (−3 or 3)	2.76±1.16	2.72±1.25	3±0	.463	2.76±1.17	2.77±1.19	.987	3±0	2.88±0.93	.352
Validation (−5, 2, 3, 4, or 5)	0.46±3.61	0.34±3.68	1.08±2.84	.872	1.56±3.48	−0.07±3.5	.012	2±2.61	0.15±3.58	.491
Domain 3: Model performance index (0 to 5 points)										
Discrimination statistics (2)	1.54±0.53	1.55±0.49	1.5±0.64	.967	1.64±0.56	1.5±0.5	.192	1.24±0.48	1.64±0.51	.011
Calibration statistics (2)	0.34±0.55	0.29±0.51	0.58±0.64	.095	0.4±0.49	0.31±0.57	.264	0.6±0.67	0.23±0.47	.667
Cut-off analysis (1)	0.21±0.41	0.12±0.33	0.67±0.47	<.001	0.24±0.43	0.19±0.39	.637	0.35±0.48	0.18±0.38	.039
Domain 4: Biologic/clinical validation and utility (0 to 6 points)										
Non-radiomics features (1)	0.50±0.50	0.46±0.49	0.75±0.43	.069	0.68±0.47	0.42±0.49	.036	0.65±0.48	0.47±0.49	.08
Biologic correlates (1)	0.28±0.45	0.32±0.47	0.08±0.27	.095	0.48±0.49	0.19±0.39	.009	0.2±0.4	0.33±0.47	.115
Comparison to 'gold standard' (2)	0.93±1.00	0.77±0.97	1.83±0.55	<.001	1.2±0.98	0.81±0.98	.109	1±1	0.94±0.99	.122
Potential clinical utility (2)	0.39±0.79	0.4±0.8	0.33±0.74	.798	0.4±0.8	0.38±0.78	.943	0.7±0.95	0.27±0.69	.086
Domain 5: High level of evidence (0 to 8 points)										
Prospective study (7)	0.27±1.36	0.76±1.20	0.58±1.93	.402	0.84±2.27	0	.011	0	0.27±1.35	.178
Cost-effective analysis (1)	0	0	0	NA	0	0	NA	0	0	NA
Domain 6: Open science and data (0 to 4 points)	0.08±0.35	0.09±0.38	0	.391	0.2±0.56	0.01±0.13	.06	0	0.09±0.40	.095
TRIPOD items	All articles (n = 77)	Diagnostic (n = 65)	Prognostic (n = 12)	p	Clinical (n = 25)	Imaging (n = 52)	p	CT+ (n = 20)	MRI+ (n = 51)	p
Title and abstract										
1. Title	2 (2.6)	2 (3)	0	1	2 (8)	0	.102	2 (10)	0	.07
2. Abstract	6 (7.8)	4 (6.1)	2 (16.7)	.233	3 (12)	3 (5.7)	.383	5 (25)	1 (1.9)	.006
Introduction										
3a. The medical context and rationale	72 (93.5)	60 (92.3)	12 (100)	1	25 (100)	47 (90.3)	.167	20 (100)	48 (94)	.055
3b. The objectives	13 (16.9)	9 (13.8)	4 (33.3)	.111	8 (32)	5 (9.6)	.022	5 (25)	7 (13.7)	.298
Methods										
4a. The study design or source of data	45 (58.4)	37 (56.9)	8 (66.7)	.751	19 (76)	26 (50)	.047	8 (40)	33 (64.7)	.067
4b. The key dates	64 (83.1)	55 (84.6)	9 (75)	.415	18 (72)	46 (88.4)	.103	17 (85)	42 (82.3)	1
5a. The study setting	49 (63.6)	43 (66.1)	6 (50)	.335	19 (76)	30 (57.6)	.136	11 (55)	33 (64.7)	.587
5b. Eligibility criteria	67 (87.0)	55 (84.6)	12 (100)	.346	23 (92)	44 (84.6)	.485	18 (90)	45 (88.2)	1
5c. Details of treatment, if relevant (n = 12)	8 out of 12	1 (1.5)	7 (58.3)	1	5 (20)	3 (5)	1	0	8 (15.8)	.09
6a. Define the outcome	70 (90.9)	61 (93.8)	9 (75)	.071	21 (84)	49 (94.2)	.205	18 (90)	47 (92.1)	1
6b. Blind assessment of the outcome	11 (14.3)	9 (13.8)	2 (16.7)	.678	3 (12)	8 (15.3)	1	2 (10)	6 (11.8)	1
7a. Define all predictors	77 (100)	65 (100)	12 (100)	NA	25 (100)	52 (100)	NA	20 (100)	51 (100)	NA
7b. Blind assessment of predictors	24 (31.2)	20 (30.8)	4 (33.3)	1	4 (16)	22 (42.3)	.038	9 (45)	12 (23.5)	.17
8. Sample size	5 (6.5)	4 (6.1)	1 (8.3)	.582	2 (8)	3 (5.7)	.657	1 (5)	4 (7.8)	1
9. Missing data	9 (11.7)	8 (12.3)	1 (8.3)	1	6 (24)	3 (5.7)	.051	1 (5)	6 (11.8)	.664
10a. How predictors were handled	77 (100)	65 (100)	12 (100)	NA	25 (100)	52 (100)	NA	20 (100)	51 (100)	NA
10b. Modeling	21 (27.3)	15 (23.1)	6 (50)	.09	10 (40)	12 (23.1)	.177	10 (50)	9 (17.6)	.017

Table 4 (continued)

10d. Model performance: both discrimination and calibration										
11. Risk groups, *if done* (yes or no, n = 77)	16 (20.8)	8 (12.3)	8 (66.7)	*<.001*	6 (24)	10 (19.2)	.765	7 (35)	9 (17.6)	.128
Results										
13a. The flow of participants	71 (92.2)	62 (95.4)	9 (75)	*.045*	20 (80)	51 (98.1)	*.012*	19 (95)	46 (90.2)	.668
13b. The characteristics of the participants	68 (88.3)	57 (87.7)	11 (91.7)	1	21 (84)	47 (90.4)	.46	20 (100)	42 (82.3)	.053
14a. Number of predictors and outcome in each analysis	69 (89.6)	62 (95.3)	7 (58.3)	*.002*	18 (72)	51 (98.1)	*.001*	18 (90)	45 (88.2)	1
14b. Unadjusted association, *if done* (yes or no, n = 77)	12 (15.6)	7 (10.7)	5 (41.7)	*.017*	7 (28)	5 (9.6)	*.048*	3 (15)	8 (15.7)	1
15a. The full prediction model	29 (37.7)	23 (35.4)	6 (50)	.351	10 (40)	19 (36.5)	.805	12 (60)	14 (27.4)	*.014*
15b. How to the use the model	24 (31.2)	19 (29.2)	5 (41.7)	.499	6 (24)	18 (34.6)	.435	10 (50)	13 (25.4)	.055
16. Performance measures	52 (67.5)	42 (64.6)	10 (83.3)	.317	17 (68)	35 (67.3)	1	14 (70)	34 (66.7)	1
Discussion										
18. Limitations	76 (98.7)	64 (83.1)	12 (100)	1	24 (96)	52 (100)	.325	19 (95)	51 (100)	.28
19b. Interpretation	77 (100)	65 (100)	12 (100)	NA	25 (100)	52 (100)	NA	20 (100)	51 (100)	NA
20. Implications	76 (98.7)	64 (98.4)	12 (100)	1	25 (100)	51 (98.1)	1	19 (95)	51 (100)	.281
For Validation (types 2a, 2b, 3, and 4)	n = 54	n = 44	n = 10		n = 20	n = 34		n = 18	n = 34	
10c. How the predictions were calculated	33 (66.1)	26 (59.1)	7 (70)	0.723	11 (55)	24 (70.5)	.809	14 (77.8)	18 (52.9)	.13
10e. Recalibration, *if done*	0	0	0	NA	0	0	NA	0	0	NA
12. Methods: Differences from the development data	49 (90.7)	39 (88.6)	10 (100)	0.571	19 (95)	19 (88.2)	.640	16 (88.9)	31 (91.1)	1
13c. Results: Comparison with the development data	41 (75.9)	31 (70.4)	10 (100)	1	15 (75)	26 (76.4)	1	14 (77.8)	25 (73.5)	1
17. Results: Model updating, *if done*	0	0	0	NA	0	0	NA	0	0	NA
19a. Discussion: Interpretation of validation results	46 (85.2)	38 (86.4)	8 (80)	0.631	17 (85)	29 (85.2)	1	16 (88.9)	28 (82.3)	.698

Note: + 6 studies with US (*n* = 3) or PET (*n* = 3) were excluded. *P* value is italicized when the *p* value is less than 0.05

Meanwhile, studies in imaging journal more frequently reported blind assessment of predictors ($p = .038$), the flow of participants ($p = .012$), and number of predictors and outcomes ($p = .001$).

Studies utilizing CT tended to have higher RQS than those using MRI (11.8 ± 3.71 vs. 9 ± 5.2), but this trend was not statistically significant. Studies using CT received a higher score in test-retest analysis ($p = .028$), discrimination statistics with resampling or cross-validation ($p = .011$), and cut-off analysis ($p = .039$) than those using MRI. In the TRIPOD items, studies using CT clearly stated study objective and setting in the abstract ($p = .006$) and described both discrimination and calibration ($p = .017$) and more provided the full prediction model ($p = .014$) than those using MRI.

There were 15 articles studied radiogenomics (19.6% among total, 21.4% among oncologic studies). There was no significant difference in radiomics quality score (Mann-Whitney U test, $p = .862$) between that of radiomics studies (median 10.5, interquartile range 5.0–13.0) and radiogenomics studies (median 10.0, interquartile range 4.25–14.7) and according to each domain.

Role of radiologists in radiomics studies

The results are shown in the Supplementary Table 3. There were 18 articles (23.4%) that radiologists were not the main authors. Three articles (3.9%) did not have radiologists in the author list. When radiologists were not the main authors, the relative position of radiologist in the author list was 0.5, which indicates middle position in the entire author lists.

Discussion

In this study, radiomics studies were evaluated in respect to the quality of both the science and the reporting, using RQS and TRIPOD guidelines. Radiomics studies were insufficient in regard to both the quality of the science and the reporting, with an average score of 26.1% of the ideal RQS and 57.8% of the maximum adherence rate to the TRIPOD reporting guidelines. No study conducted a phantom study or cost-effective analysis and a high level of evidence for radiomics studies, with further limitations being demonstrated in the openness to data and code. Half of the items that the TRIPOD statement deems necessary to report in multivariable prediction model publications were not completely recorded in the radiomics studies. Our results imply that radiomics studies require significant improvement in both scientific and reporting quality.

The six key RQS domains used in this study were designed to support the integration of the RQS in radiomics approaches. Adopted from the consensus statement of the FDA-NIH Biomarker Working Group [4], the three aspects of technical validation, biological/clinical validation, and assessment of cost-effectiveness for imaging biomarker standardization were included in domains 1 (image protocol and feature reproducibility), 4 (biologic/clinical validation), and 5 (high level of evidence), respectively. With regard 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 testing was conducted in one study [47], 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 [88]. For clinical validation, prospective testing of an imaging biomarker in clinical populations is required [89], and only three reports covered a prospective study in the field of neuro-oncology. After biological/clinical validation, the cost-effectiveness of radiomics needs to be studied to ensure that it provides good value for money compared with the other currently available biomarkers. From the current standpoint, the time when radiomics will achieve this end seems far away, and technical and clinical validation is still required.

Validation in TRIPOD pursues external validation, which was performed in only 18.2% of the reports covered in the present study, while independent validation including internal validation is acceptable in RQS, which accounts for 70.1% of the studies. In the reporting of radiomics studies, the highly problematic TRIPOD items existed. In the title, only two radiomics studies explicitly wrote the term "development" or "validation" with the target population and outcome. There are several elements that should be present in the abstract, and one of these was missing in 92.2% of studies, as they did not explicitly describe "development" or "validation" in the study objective, or did not describe whether the study design was a randomized controlled trial, a cohort study, or a case-control design. Furthermore, reporting of the sample size calculation and the handling of missing data were often poorly conducted. These results are similar to the findings of a previous systematic review of TRIPOD adherence that examined publications using clinical multivariable prediction models [9]. Differing from the findings on the clinical multivariable prediction models, the radiomics studies were excellent (100%) in the criterion involving the definition of all predictors and how quantitative features are handled. However, the reporting of blindness to the outcome was insufficient (32.1%). The results for blindness were similar to the adherence to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) [90, 91], which stated that blinding of both predictors and outcome is still insufficient in both clinical and imaging studies.

Subgroup analysis may provide more specific guidance in radiomics research. Studies in clinical journals showed significantly higher RQS scores, especially in test-retest analysis, multivariable analysis with non-radiomics features, finding biologic correlates, and pursuit of prospective study design. In the TRIPOD items, they more clearly defined the data source and study design, i.e., a consecutive retrospective design or case-control design, and the study setting, i.e., tertiary hospital or general population. In terms of validation, all three prospective studies were in the clinical journal, and both external and independent validation was performed more frequently in the clinical journal group. These findings imply that high-impact clinical journals pursue precision in research, clarification of epidemiological background, and independent validation, which demonstrates the room for improvement in radiomics studies.

Of note, radiologists played the main role as either first or corresponding authors in the radiomics studies. There were 23.4% articles that radiologists were not the main authors, but most studies work collaboratively among radiologists, other clinicians, and physicists.

This study has some potential limitations. The first is the relatively small sample size, especially with an impact factor below 7. This was placed to permit in-depth analysis of the radiomics applications. Second, radiomics is still a developing imaging biomarker, and the suggested RQS may be too "ideal." The criteria of phantom study and multiple imaging acquisitions may be unrealistic for clinical situations. Third, the adoption of TRIPOD items to radiomics studies can be rather strict. For example, most studies are case-control and retrospective study designs, and a clear description of "case-control" is not commonly given in imaging journals. Nonetheless, clear stating of the participants and study setting is important for study transportability, and studies in imaging journals need to pursue this. Fourth, we considered internal validation with a random sample or split sample as independent validation, while the TRIPOD statement only considers external validation for validation of a pre-existing model. When the rates of open science and open data increase in the field of radiomics, the true validation of a model should become easier to perform.

In conclusion, the overall scientific quality and reporting of radiomics studies is insufficient, with the scientific quality showing the greatest deficiencies. Scientific improvements need to be made to feature reproducibility, analysis of clinical utility, and open science. Reporting of study objectives, blind assessment, sample size, and missing data is deemed to be necessary. Our intention is to promote the quality of radiomics research studies as diagnostic and prognostic prediction models, and the above criteria and items need to be pursued for radiomics to become a viable tool for medical decision-making.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Jeong Hoon Kim.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise (Seo Young Park, 8 years of experience).

Informed consent Written informed consent was not required because of the nature of our study, which was a study based on research articles.

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

Methodology
• retrospective
• cross-sectional study
• performed at one institution

References

1. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
2. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762
3. Sanduleanu S, Woodruff HC, de Jong EEC et al (2018) Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. Radiother Oncol 127:349–360
4. O'Connor JP, Aboagye EO, Adams JE et al (2017) Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 14:169–186
5. Sung NS, Crowley WF Jr, Genel M et al (2003) Central challenges facing the national clinical research enterprise. JAMA 289:1278–1287
6. Choi YJ, Chung MS, Koo HJ, Park JE, Yoon HM, Park SH (2016) Does the reporting quality of diagnostic test accuracy studies, as defined by STARD 2015, affect citation? Korean J Radiol 17:706–714
7. Waterton JC, Pylkkanen L (2012) Qualification of imaging biomarkers for oncology drug development. Eur J Cancer 48:409–415
8. Moons KG, Altman DG, Reitsma JB et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 162:W1–W73
9. Heus P, Damen JAAG, Pajouheshnia R et al (2018) Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. BMC Med 16:120
10. Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536

11. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006

12. Hawkins S, Wang H, Liu Y et al (2016) Predicting malignant nodules from screening CT scans. J Thorac Oncol 11:2120–2128

13. Huang Y, Liu Z, He L et al (2016) Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. Radiology 281:947–957

14. Huang YQ, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 34:2157–2164

15. Kickingereder P, Bonekamp D, Nowosielski M et al (2016) Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. Radiology 281:907–918

16. Kickingereder P, Burth S, Wick A et al (2016) Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. Radiology 280:880–889

17. Kickingereder P, Gotz M, Muschelli J et al (2016) Large-scale Radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. Clin Cancer Res 22:5765–5771

18. Li H, Zhu Y, Burnside ES et al (2016) MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, oncotype DX, and PAM50 gene assays. Radiology 281:382–391

19. Nie K, Shi L, Chen Q et al (2016) Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. Clin Cancer Res 22:5256–5264

20. Coroller TP, Agrawal V, Huynh E et al (2017) Radiomic-based pathological response prediction from primary tumors and lymph nodes in NSCLC. J Thorac Oncol 12:467–476

21. Grossmann P, Narayan V, Chang K et al (2017) Quantitative imaging biomarkers for risk stratification of patients with recurrent glioblastoma treated with bevacizumab. Neuro Oncol 19:1688–1697

22. Hu LS, Ning S, Eschbacher JM et al (2017) Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. Neuro Oncol 19:128–137

23. Liu TT, Achrol AS, Mitchell LA et al (2017) Magnetic resonance perfusion image features uncover an angiogenic subgroup of glioblastoma patients with poor survival and better response to antiangiogenic treatment. Neuro Oncol 19:997–1007

24. Liu Z, Zhang XY, Shi YJ et al (2017) Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Clin Cancer Res 23:7253–7262

25. Lohmann P, Stoffels G, Ceccon G et al (2017) Radiation injury vs. recurrent brain metastasis: combining textural feature radiomics analysis and standard parameters may increase (18)F-FET PET accuracy without dynamic scans. Eur Radiol 27:2916–2927

26. Rios Velazquez E, Parmar C, Liu Y et al (2017) Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer Res 77:3922–3930

27. Song SH, Park H, Lee G et al (2017) Imaging phenotyping using Radiomics to predict micropapillary pattern within lung adenocarcinoma. J Thorac Oncol 12:624–632

28. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD (2017) Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. Eur Radiol 27:4082–4090

29. Wu S, Zheng J, Li Y et al (2017) A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. Clin Cancer Res 23:6904–6911

30. Yu J, Shi Z, Lian Y et al (2017) Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. Eur Radiol 27:3509–3522

31. Yuan M, Zhang YD, Pu XH et al (2017) Comparison of a radiomic biomarker with volumetric analysis for decoding tumour phenotypes of lung adenocarcinoma with different disease-specific survival. Eur Radiol 27:4857–4865

32. Zhang B, Tian J, Dong D et al (2017) Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 23:4259–4269

33. Zhou H, Vallieres M, Bai HX et al (2017) MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro Oncol 19:862–870

34. Akbari H, Bakas S, Pisapia JM et al (2018) In vivo evaluation of EGFRvIII mutation in primary glioblastoma patients via complex multiparametric MRI signature. Neuro Oncol 20:1068–1079

35. Bae S, Choi YS, Ahn SS et al (2018) Radiomic MRI phenotyping of glioblastoma: improving survival prediction. Radiology 289:797–806

36. Beukinga RJ, Hulshoff JB, Mul VEM et al (2018) Prediction of response to neoadjuvant chemotherapy and radiation therapy with baseline and restaging (18)F-FDG PET imaging biomarkers in patients with esophageal cancer. Radiology 287:983–992

37. Bickelhaupt S, Jaeger PF, Laun FB et al (2018) Radiomics based on adapted diffusion kurtosis imaging helps to clarify most mammographic findings suspicious for cancer. Radiology 287:761–770

38. Chen T, Ning Z, Xu L et al (2018) Radiomics nomogram for predicting the malignant potential of gastrointestinal stromal tumours preoperatively. Eur Radiol 29:1074–1082

39. Chen Y, Chen TW, Wu CQ et al (2018) Radiomics model of contrast-enhanced computed tomography for predicting the recurrence of acute pancreatitis. Eur Radiol 29:4408–4417

40. Cui Y, Yang X, Shi Z et al (2018) Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Eur Radiol 29:1211–1220

41. Dong F, Li Q, Xu D et al (2018) Differentiation between pilocytic astrocytoma and glioblastoma: a decision tree model using contrast-enhanced magnetic resonance imaging-derived quantitative radiomic features. Eur Radiol 29:3968–3975

42. Dong Y, Feng Q, Yang W et al (2018) Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. Eur Radiol 28:582–591

43. Guo J, Liu Z, Shen C et al (2018) MR-based radiomics signature in differentiating ocular adnexal lymphoma from idiopathic orbital inflammation. Eur Radiol 28:3872–3881

44. Horvat N, Veeraraghavan H, Khan M et al (2018) MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. Radiology 287:833–843

45. Hu HT, Wang Z, Huang XW et al (2018) Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. Eur Radiol 29:2890–2901

46. Kang D, Park JE, Kim YH et al (2018) Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation. Neuro Oncol 20:1251–1261

47. Kickingereder P, Neuberger U, Bonekamp D et al (2018) Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma. Neuro Oncol 20:848–857

48. Kim JY, Park JE, Jo Y et al (2018) Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. Neuro Oncol 21:404–414

49. Kniep HC, Madesta F, Schneider T et al (2018) Radiomics of brain MRI: utility in prediction of metastatic tumor type. Radiology 180946

50. Multiparametric ultrasomics of significant liver fibrosis: a machine learning-based analysis. Eur Radiol 29:1496–1506

51. Li Y, Liu X, Qian Z et al (2018) Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. Eur Radiol 28:2960–2968

52. Li Y, Liu X, Xu K et al (2018) MRI features can predict EGFR expression in lower grade gliomas: a voxel-based radiomic analysis. Eur Radiol 28:356–362

53. Li ZC, Bai H, Sun Q et al (2018) Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. Eur Radiol 28:3640–3650

54. Liang W, Yang P, Huang R et al (2018) A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. Clin Cancer Res 25:584–594

55. Liu H, Zhang C, Wang L et al (2018) MRI radiomics analysis for predicting preoperative synchronous distant metastasis in patients with rectal cancer. Eur Radiol 29:4418–4426

56. Lu CF, Hsu FT, Hsieh KL et al (2018) Machine learning-based radiomics for molecular subtyping of gliomas. Clin Cancer Res 24:4429–4436

57. Lv W, Yuan Q, Wang Q et al (2018) Robustness versus disease differentiation when varying parameter settings in radiomics features: application to nasopharyngeal PET/CT. Eur Radiol 28:3245–3254

58. Meng X, Xia W, Xie P et al (2018) Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. Eur Radiol 29:3200–3209

59. Naganawa S, Enooku K, Tateishi R et al (2018) Imaging prediction of nonalcoholic steatohepatitis using computed tomography texture analysis. Eur Radiol 28:3050–3058

60. Niu J, Zhang S, Ma S et al (2018) Preoperative prediction of cavernous sinus invasion by pituitary adenomas using a radiomics method based on magnetic resonance images. Eur Radiol 29:1625–1634

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

62. Park YW, Oh J, You SC et al (2018) Radiomics and machine learning may accurately predict the grade and histological subtype in meningiomas using conventional and diffusion tensor imaging. Eur Radiol 29:4068–4076

63. She Y, Zhang L, Zhu H et al (2018) The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules. Eur Radiol 28:5121–5128

64. Shi Z, Zhu C, Degnan AJ et al (2018) Identification of high-risk plaque features in intracranial atherosclerosis: initial experience using a radiomic approach. Eur Radiol 28:3912–3921

65. Su C, Jiang J, Zhang S et al (2018) Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. Eur Radiol 29:1986–1996

66. Suh HB, Choi YS, Bae S et al (2018) Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. Eur Radiol 28:3832–3839

67. Sun H, Chen Y, Huang Q et al (2018) Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: a radiomics analysis. Radiology 287:620–630

68. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C (2018) Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. Radiology 181352

69. Wu M, Tan H, Gao F et al (2018) Predicting the grade of hepatocellular carcinoma based on non-contrastenhanced MRI radiomics signature. Eur Radiol 29:2802–2811

70. Yang L, Dong D, Fang M et al (2018) Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? Eur Radiol 28:2058–2067

71. Yin P, Mao N, Zhao C et al (2018) Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features. Eur Radiol 29:1841–1847

72. Zhang S, Song G, Zang Y et al (2018) Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery. Eur Radiol 28:3692–3701

73. Zhang Y, Zhang B, Liang F et al (2018) Radiomics features on noncontrast-enhanced CT scan can precisely classify AVM-related hematomas from other spontaneous intraparenchymal hematoma types. Eur Radiol 29:2157–2165

74. Zhang Z, Yang J, Ho A et al (2018) A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. Eur Radiol 28:2255–2263

75. Zhu X, Dong D, Chen Z et al (2018) Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. Eur Radiol 28:2772–2778

76. Zinn PO, Singh SK, Kotrotsou A et al (2018) A coclinical radiogenomic validation study: conserved magnetic resonance radiomic appearance of periostin-expressing glioblastoma in patients and xenograft models. Clin Cancer Res 24:6288–6299

77. Choe J, Lee SM, Do KH et al (2019) Prognostic value of radiomic analysis of iodine overlay maps from dual-energy computed tomography in patients with resectable lung cancer. Eur Radiol 29:915–923

78. Hu T, Wang S, Huang L et al (2019) A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules. Eur Radiol 29:439–449

79. Ji GW, Zhang YD, Zhang H et al (2019) Biliary tract cancer at CT: a radiomics-based model to predict lymph node metastasis and survival outcomes. Radiology 290:90–98

80. Kontos D, Winham SJ, Oustimov A et al (2019) Radiomic phenotypes of mammographic parenchymal complexity: toward augmenting breast density in breast cancer risk assessment. Radiology 290:41–49

81. Qu J, Shen C, Qin J et al (2019) The MR radiomic signature can predict preoperative lymph node metastasis in patients with esophageal cancer. Eur Radiol 29:906–914

82. Tan X, Ma Z, Yan L, Ye W, Liu Z, Liang C (2019) Radiomics nomogram outperforms size criteria in discriminating lymph node metastasis in resectable esophageal squamous cell carcinoma. Eur Radiol 29:392–400

83. Wei J, Yang G, Hao X et al (2019) A multi-sequence and habitat-based MRI radiomics signature for preoperative prediction of MGMT promoter methylation in astrocytomas with prognostic implication. Eur Radiol 29:877–888

84. Bonekamp D, Kohl S, Wiesenfarth M et al (2018) Radiomic machine learning for characterization of prostate lesions with MRI: comparison to ADC values. Radiology 289:128–137

85. Park H, Lim Y, Ko ES et al (2018) Radiomics signature on magnetic resonance imaging: association with disease-free survival in patients with invasive breast cancer. Clin Cancer Res 24:4705–4714

86. Sun R, Limkin EJ, Vakalopoulou M et al (2018) A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol 19:1180–1191

87. Wang K, Lu X, Zhou H et al (2018) Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 68:729–741

88. Kessler LG, Barnhart HX, Buckler AJ et al (2015) The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 24:9–26

89. McShane LM, Altman DG, Sauerbrei W et al (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 97:1180–1184

90. Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L (2014) Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. Evid Based Med 19:47–54

91. Korevaar DA, Wang J, van Enst WA et al (2015) Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. Radiology 274:781–789