



Systematic review of the radiomics quality score applications: an EuSoMII Radiomics Auditing Group Initiative

Gaia Spadarella¹ · Arnaldo Stanzione¹ · Tugba Akinci D'Antonoli² · Anna Andreychenko³ · Salvatore Claudio Fanni⁴ · Lorenzo Ugga¹ · Elmar Kotter⁵ · Renato Cuocolo^{6,7}

Received: 27 May 2022 / Revised: 31 August 2022 / Accepted: 19 September 2022
© The Author(s) 2022

Abstract

Objective The main aim of the present systematic review was a comprehensive overview of the Radiomics Quality Score (RQS)–based systematic reviews to highlight common issues and challenges of radiomics research application and evaluate the relationship between RQS and review features.

Methods The literature search was performed on multiple medical literature archives according to PRISMA guidelines for systematic reviews that reported radiomic quality assessment through the RQS. Reported scores were converted to a 0–100% scale. The Mann-Whitney and Kruskal-Wallis tests were used to compare RQS scores and review features.

Results The literature research yielded 345 articles, from which 44 systematic reviews were finally included in the analysis. Overall, the median of RQS was 21.00% (IQR = 11.50). No significant differences of RQS were observed in subgroup analyses according to targets (oncological/not oncological target, neuroradiology/body imaging focus and one imaging technique/more than one imaging technique, characterization/prognosis/detection/other).

Conclusions Our review did not reveal a significant difference of quality of radiomic articles reported in systematic reviews, divided in different subgroups. Furthermore, low overall methodological quality of radiomics research was found independent of specific application domains. While the RQS can serve as a reference tool to improve future study designs, future research should also be aimed at improving its reliability and developing new tools to meet an ever-evolving research space.

Key Points

- Radiomics is a promising high-throughput method that may generate novel imaging biomarkers to improve clinical decision-making process, but it is an inherently complex analysis and often lacks reproducibility and generalizability.
- The Radiomics Quality Score serves a necessary role as the de facto reference tool for assessing radiomics studies.
- External auditing of radiomics studies, in addition to the standard peer-review process, is valuable to highlight common limitations and provide insights to improve future study designs and practical applicability of the radiomics models.

Keywords Systematic review · Diagnostic imaging · Big data · Radiomics · Radiomics quality score

✉ Arnaldo Stanzione
arnaldo.stanzione@unina.it

¹ Department of Advanced Biomedical Sciences, University of Naples “Federico II”, Naples, Italy

² Institute of Radiology and Nuclear Medicine, Cantonal Hospital Baselland, Liestal, Switzerland

³ Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow, Russia

⁴ Department of Translational Research, Academic Radiology, University of Pisa, Pisa, Italy

⁵ Department of Radiology, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁶ Department of Medicine, Surgery, and Dentistry, University of Salerno, Baronissi, Italy

⁷ Augmented Reality for Health Monitoring Laboratory (ARHeMLab), Department of Electrical Engineering and Information Technology, University of Naples “Federico II”, Naples, Italy

Abbreviations

IBSI	Image biomarker standardization initiative
ICC	Intraclass correlation coefficient
IQR	Interquartile range
RQS	Radiomics quality score

Introduction

The overwhelming enthusiasm toward radiomics is emphasized by the ever-growing number of publications in the field [1, 2]. This high-throughput strategy to mine quantitative data from medical images searching for novel biomarkers and to generate decision-support models is deemed a feasible approach to overcome the limitations of conventional image interpretation, particularly in oncology [3–5]. The potential applications of radiomics are seemingly endless across all imaging modalities, and according to a survey study, the future physicians are confident that advanced computer-aided image analyses will revolutionize radiology for the best [6–9].

Nevertheless, after nearly a decade of research, translation of radiomics into clinical practice remains a distant prospect, and there are many unanswered questions about the potential availability of commercial radiomics tools [10]. Additionally, reasonable concerns have also been raised that we might be overlooking negative, unpublished, but potentially valuable results, i.e., publication bias [11].

Radiomics is a complex multi-step process, and within each step there are methodological challenges to overcome in order to ensure the robustness of model's findings, while reproducibility and generalizability are often compromised [12–14]. Aiming to untangle this methodological complexity and streamline the structure of radiomics pipelines, a set of recommendations was released in 2017 along with a proposal of a “quality seal” for published results named Radiomics Quality Score (RQS) [15]. Although there is still room for improvement, the RQS has been embraced by the scientific community and has been mainly used to assess the methodological quality of previously published radiomics studies in the setting of systematic reviews [16].

The RQS consists of 16 items, with a total score ranging from – 8 to + 36 points. The percentage score is derived from the absolute score and obtained by dividing the total score by 36 [17]. The RQS items may also be grouped into six domains [18]. Domain 1 covers protocol quality and reproducibility in image and segmentation (items 1–4), domain 2 reporting of feature reduction and validation (items 5 and 12), domain 3 biological/clinical validation and utility (items 6, 7, 13, and 14), domain 4 performance index (items 8, 9, and 10), domain 5 demonstration of a higher level of evidence (items 11–15), and domain 6 open science (item 16).

In the present work, we aim to provide a comprehensive overview of RQS-based systematic reviews to highlight

common issues and unique challenges in the vast array of radiomics applications.

Methods

The study was registered on the International Prospective Register of Systematic Reviews database with the registration number CRD42021292310.

Article search strategy

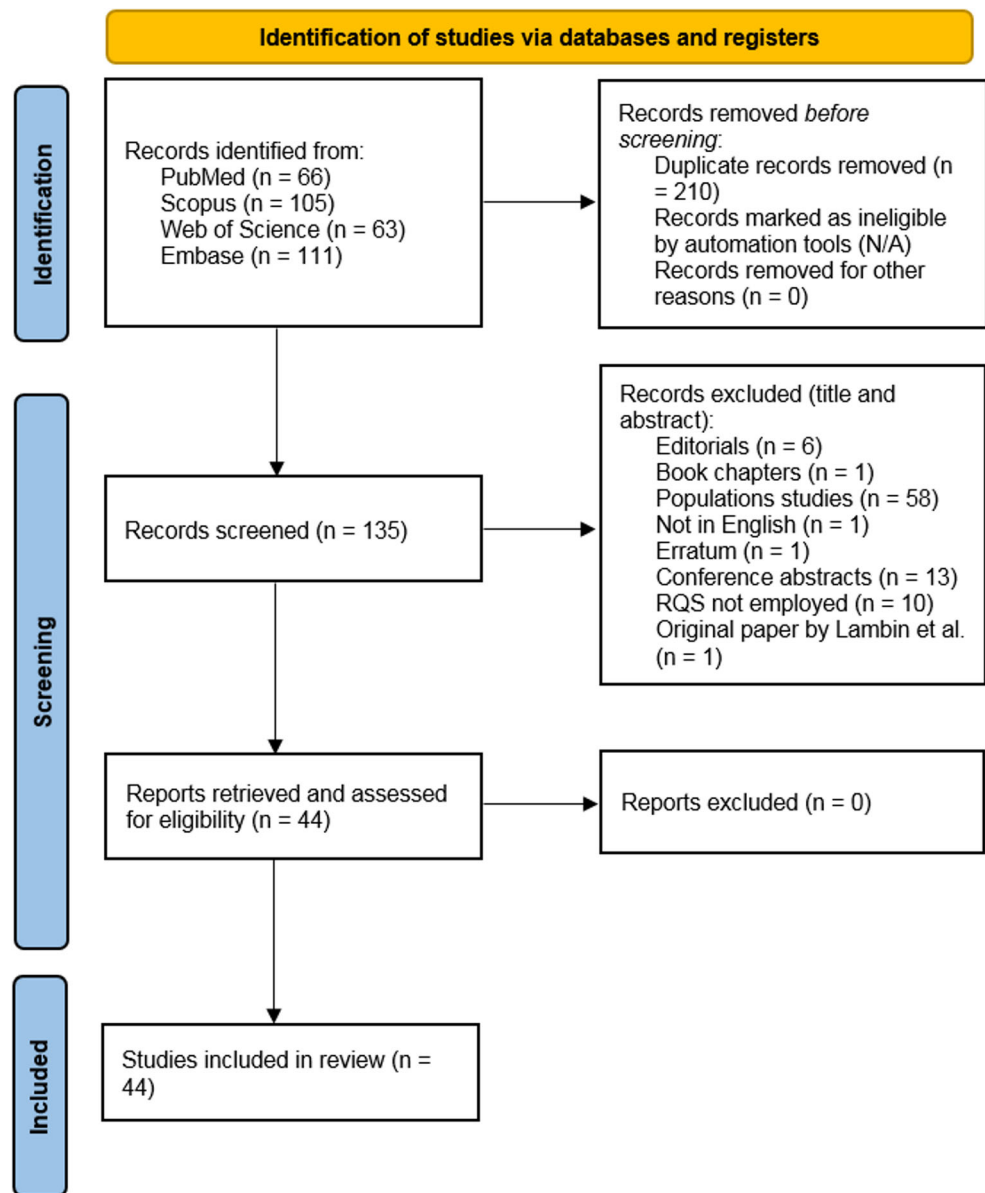
The literature search was performed according to PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines in the electronic databases (PubMed, Web of Science, Embase, and Scopus) using the following search query: (“radiomics” OR “radiomic”) AND “quality” AND “score”. The systematic reviews that reported radiomic quality assessment performed according to the RQS and published until December 31, 2021, were included. Letters, editorials, duplicates, original articles, literature reviews, and RQS systematic reviews published in languages other than English were excluded from the analysis. The included articles were selected by consensus of four radiologists experienced in radiomics/texture analysis, systematic literature review, and RQS assessment. In Fig. 1, the results of the article selection are shown.

Data extraction and analysis

The RQS comprises six domains (image protocol, radiomics features extraction, data analysis and statistics, model validation, clinical validity, and open science) and 16 items. By assessing each item, a final score will be determined, which is presented on a scale of – 8 to 36 and can be converted to a percentage (where scores below 0 accepted as 0 and 36 equals 100%), as reported by Lambin et al [15]. Details of the RQS domains and items along with the scores can be found in the [supplementary materials](#). The same group of radiologists (A.S. and L.U.: 5 years of experience, C.F. and T.A.D.: 2 years each) also extracted the data from all included studies and collected the median or mean of RQS from included studies.

Moreover, the included studies were classified based on the following characteristics: (1) oncological versus non-oncological target; (2) neuroradiology versus body imaging focus; (3) single versus multiple imaging modalities; (4) aim of the studies that included to systematic reviews (characterization, detection, prognosis prediction, or other).

Fig. 1 The literature research flow diagram. Adapted from: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews [19]



Statistical analysis

All the analyses were performed using the mean RQS percentage scores reported in each systematic review, after conversion of the median values to corresponding means [20]. When necessary, raw data from included studies were retrieved to calculate mean RQS percentage scores. The relation between the study quality and article subgroups was tested. The normality of the data distribution was assessed with the Kolmogorov-Smirnov test. To compare variables with a non-normal distribution, a Mann-Whitney test was performed. The Kruskal-Wallis test was used to compare multiple continuous variables. Continuous variables are presented as median and interquartile range (IQR), categorical ones as count and

percentage. All statistical analyses were performed using SPSS (SPSS version 27; SPSS). Alpha level was set to 0.05.

Results

Literature review

The initial literature research resulted in 345 articles, of which 210 were duplicates. Finally, 44 studies were selected from the remaining 135 because 91 articles did not meet the inclusion criteria. The study flowchart is shown in Fig. 1 and all systematic reviews included in this study are listed in Table 1.

Table 1 Characteristics of the included systematic reviews

First author	Year	Journal	Organ system	<i>n</i> of studies	Mean %
Abdurixiti [21]	2021	British Journal of Radiology	Lung	6	35
Abunahel [22]	2020	European Radiology	Pancreas	72	29
Bhandari [23]	2020	Abdominal Radiology	Kidney	13	31
Bhandari [24]	2020	American Journal of Neuroradiology	Brain	14	29
Calabrese [25]	2021	Journal of Cancer Research and Clinical Oncology	Breast	10	30
Carbonara [26]	2021	Journal of Oncology	Head and Neck	8	21
Castillo [27]	2020	Cancers	Prostate	13	51
Chen [28]	2021	European Journal of Nuclear Medicine and Molecular Imaging	Lung	10	30
Chetan [29]	2020	European Radiology	Lung	14	21
Crombe [30]	2020	European Journal of Radiology	Soft tissue	52	18
Davey [31]	2021	European Journal of Radiology	Breast	41	18
Fornacon-wood [32]	2020	Lung Cancer	Lung	43	21
Granzier [33]	2019	European Journal of Radiology	Breast	16	12
Harding-theobald [34]	2021	Alimentary Pharmacology and Therapeutics	Liver	54	25
Janssen [35]	2021	Annals of Surgery	Pancreas	23	21
Kao [36]	2021	In Vivo	Esophagus	7	26
Kao [37]	2021	Diagnostics	Lung	7	39
Kendrick [38]	2021	Frontiers in Oncology	Prostate	17	23
Kim [39]	2021	Neuro-Oncology Advances	Brain	7	3
Kozikowskim [40]	2021	European Urology Focus	Bladder	8	41
Lecointre [41]	2021	European Journal of Surgical Oncology	Uterus	17	15
Muhlbauer [42]	2021	Cancers	Kidney	113	14
Nardone [43]	2021	Radiologia Medica	Multiorgan	48	21
Park [44]	2020	European Radiology	Multiorgan	77	26
Park [18]	2020	BMC Cancer	Brain	51	22
Ponsiglione [45]	2021	European Radiology	Cardiovascular	53	12
Sanduleanu [17]	2018	Radiotherapy and Oncology	Multiorgan	41	22
Shi [46]	2021	European Journal of Radiology	Lung	28	19
Spadarella [47]	2021	European Journal of Radiology	Pharynx	24	21
Staal [48]	2021	Clinical Colorectal Cancer	Large bowel	76	13
Stanzione [49]	2020	European Journal of Radiology	Prostate	73	23
Tabatabaei [50]	2021	Oncology	Brain	18	76
Ugga [51]	2021	Neuroradiology	Brain	23	19
Ursprung [52]	2020	European Radiology	Kidney	57	9
Valdora [53]	2018	Breast Cancer Research and Treatment	Breast	17	33
Wakabayashi [54]	2019	Hepatology International	Liver	23	23
Walls [55]	2021	Clinical Oncology	Lung	44	17
Wang [56]	2020	European Radiology	Hematology	45	14
Wang [57]	2021	Cancers	Liver	22	28
Wesdorp [58]	2020	European Journal of Nuclear Medicine and Molecular Imaging	Gastrointestinal	60	23
Wesdorp [59]	2021	Surgical Oncology	Gastrointestinal	14	19
Won [60]	2021	European Journal of Radiology	Brain	25	15
Won [61]	2020	Korean Journal of Radiology	Brain	26	10
Zhong [62]	2021	European Radiology	Bone	12	20

Study features and subgroup analysis

Study features are summarized in Table 2. Additional details are reported in the [supplementary materials](#). The median of RQS was 21.00% (IQR = 11.50). In 36 systematic reviews, quality assessment was performed by 2 or more readers (36/44, 81%). Discrepancies were evaluated in different ways: 11/44 studies assessed agreement intraclass correlation coefficient (ICC) or Cohen's kappa, and 2 authors reported the mean of RQS score, while 23 authors chose consensus for reproducibility evaluation. The remaining studies (8/44, 18%) did not specify the reproducibility test. As shown in Fig. 2, the highest mean RQS score of 27.50% reported in systematic reviews published in the year 2018 while the lowest RQS was reported in 2019. Most of the review articles focused on oncological radiomics studies (40/44, 90%); ten out of forty-four (22.7%) reviews were focused on neuroradiology radiomics articles. Twenty-five percent of systematic reviews included 50 or more studies in the main analysis (11/44), with a range between 6 and 113 articles included. Furthermore, the systematic reviews with a body imaging topic included 33 articles on average, while neuro-imaging reviews covered a mean of 20 studies. Notably, 38% (17/44) of articles were focused on one imaging technique, in which most of them selected MRI (16/44, 36%). In Fig. 2, mean RQS% of selected systematic reviews in each year were reported, while in Fig. 3, the mean RQS% of each review included are described. The mean RQS% separated according to the systematic review characteristics is shown in Figs. 4 and 5.

The results of the subgroup analysis according to the systematic review features did not demonstrate any significant difference between subgroups (Figs. 4 and 5).

Discussion

In recent years, the number of published radiomics studies has been increasing exponentially, notably in the field of oncological imaging [63]. This is mainly due to the promising results in this area, made possible thanks to the use of artificial intelligence/machine learning approaches instead of classical

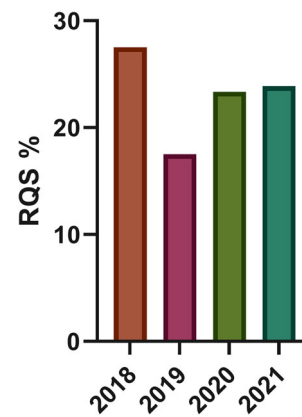


Fig. 2 Bar plot reporting the mean RQS percentage score by year of publication

statistical tests and expert systems, capable of analyzing such a large amount of quantitative data and producing classification or prediction models. As a result of the overwhelming number of studies in this field, the need for providing research guidelines has arisen to ensure better standardization and homogenization. In this context, the image biomarker standardization initiative (IBSI), an independent international collaboration, has been working toward standardizing the extraction of image biomarkers from acquired imaging. IBSI provides an image biomarker nomenclature and specific feature definitions, as well as a general image processing workflow, tools for verifying radiomics software implementations, and reporting guidelines for radiomics studies [64]. Together with the need for standardization, the need for a tool for qualitative assessment and comparison of extremely heterogeneous radiomics methodologies has also arisen. In relation to this question, Lambin et al introduced the radiomics quality score (RQS) in 2017 [15]. The RQS followed previous efforts that did not focus on radiomics, such as the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement published in 2015 [65]. The aim of the RQS is to evaluate the methodological quality of radiomics-based investigations, identifying high-quality results as well as issues limiting their value and applicability. However, as stated by the creators of the RQS themselves,

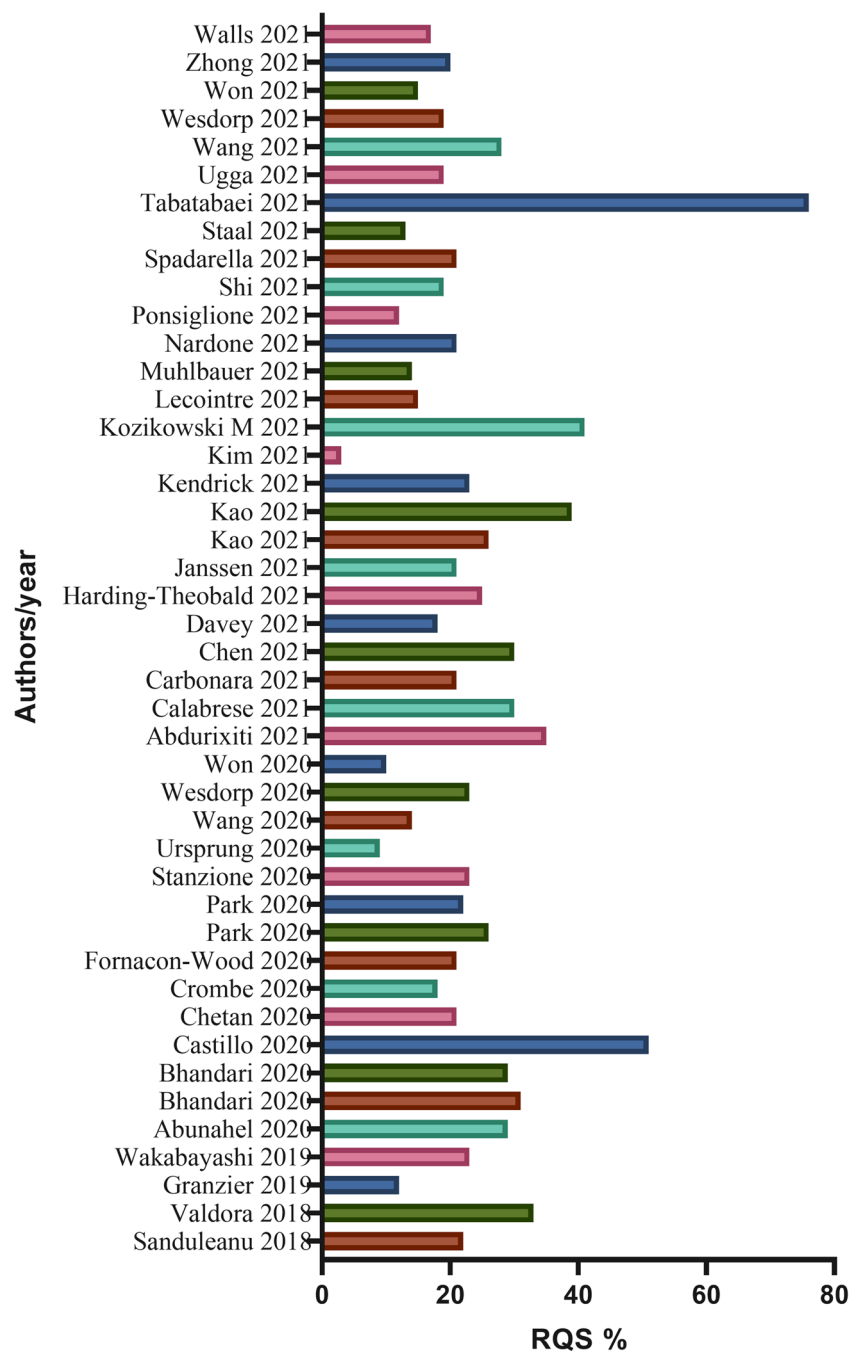
Table 2 Median RQS percentage scores for review subgroups

Characteristics	Statistical analysis	Median RQS% (<i>n</i> of studies)	<i>p</i> value
Body/neuroradiology/other*	Anova Kruskal-Wallis	23.6 (31)/23.6 (10)/23.0 (3)	0.586
Oncology/not oncology	Mann-Whitney	27.3 (40)/20.3 (4)	0.396
Single modality/2 or more modalities	Mann-Whitney	27.3 (17)/20.3 (27)	0.277
Characterization/detection/prognosis/other**	Anova Kruskal-Wallis	27.2 (14)/27.0 (4)/21.3 (10)/19.6 (13)	0.413

*Systematic reviews not covered by neuroradiology or body imaging categories

**Systematic reviews not covered by characterization/detection/prognosis target

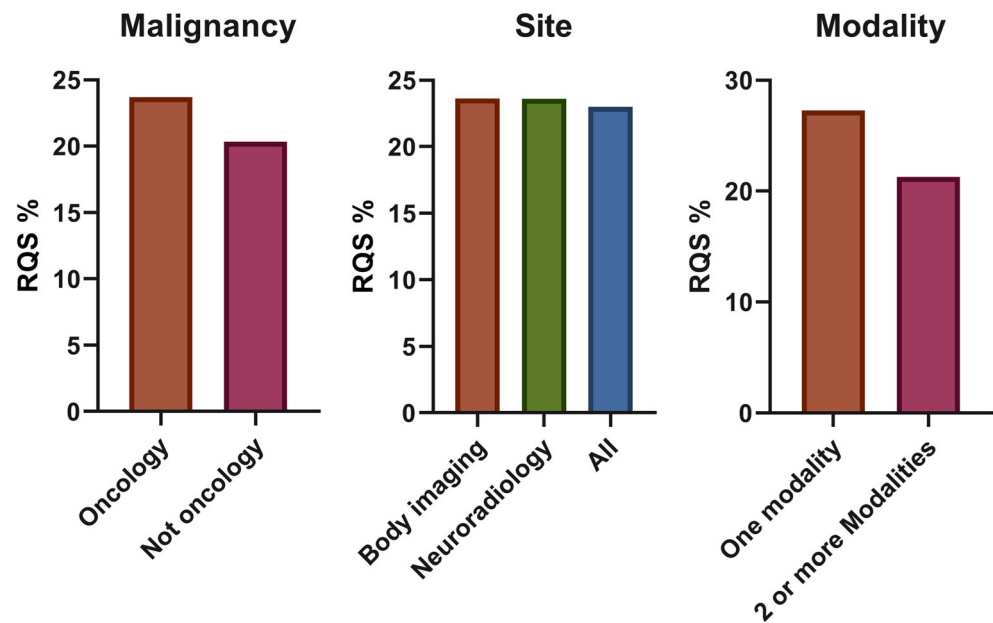
Fig. 3 Mean of RQS percentage score of each review included in this study



this score was not conceived as an external auditing tool to express a qualitative appraisal in absolute terms or to conduct systematic reviews, but rather as a practical checklist to guide researchers in study designing and to give them the possibility to justify any methodology noncompliance [16]. However, in practice, this tool has become the de facto standard for systematic reviews of the literature focused on radiomics quality assessment as confirmed by our findings. In any case, it should be acknowledged that some alternatives have been proposed, even though their use is usually sporadic [66, 67]. Additionally, several checklists

have been presented in the recent literature, including the Checklist for Artificial Intelligence in Medical Imaging, Minimum Information for Medical AI Reporting checklist, and currently under development artificial intelligence extensions of the Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis statement and the Prediction model Risk Of Bias Assessment Tool [68–70]. However, these are not tailored for use in radiomics specifically, but are focused on machine learning modeling and correct management of the data in relation to model bias. Also, their nature as

Fig. 4 RQS percentage score for different review subgroups



checklists does not allow a formal methodological quality score, but rather an unweighted assessment of overall adherence to the included items.

Regarding the use of RQS to perform an external assessment of methodological quality in radiomics studies, its potential lack of reproducibility may represent an issue. Only in a minority of the studies included in our systematic review, the authors performed an assessment of the RQS's inter-reader reproducibility, either through the intraclass correlation coefficient or Cohen's K. In several cases, a consensus approach was employed with multiple raters, which may represent a

valid solution to ensure the score's reliability. The assessment of the RQS's reproducibility, also accounting for differences in raters' experience levels, may represent an avenue of future research of itself if its use for systematic study quality auditing will continue. It would also be ideal to identify a standard practice on this topic, either requiring inclusion of an inter-reader reproducibility analysis or a consensus approach for all future RQS-based reviews. This would mitigate concerns regarding possible biases in the final scores.

Another limitation of the RQS pertains to its use for deep learning-based studies. Several authors have used the RQS to assess the quality of this type of research, but the RQS items are not perfectly suited for this task. On one hand, it can be argued that computer vision neural networks, especially when based on convolutions, essentially extract quantitative features that can be assimilated to typical radiomics parameters. However, the processing of this data diverges from the classical feature processing, selection, and model tuning pipeline of radiomics. Probably, the appropriateness of the RQS should be evaluated on a case-by-case basis for deep learning research. In the future, it could be appropriate to develop dedicated tools tailored to address both classical machine learning and deep learning radiomics analyses, sharing part of the items but diverging as necessary to avoid biases [41]. In this setting, the information contained in the previously mentioned healthcare artificial intelligence modeling checklists could prove valuable to complement the original RQS.

It should also be noted that the average RQS across all included reviews was low (median = 21.00%; IQR = 11.50). This, not only, supports the conclusions drawn by each individual review that the methodological quality and/or thoroughness of its presentation within scientific studies is still far from ideal but also raises questions about the

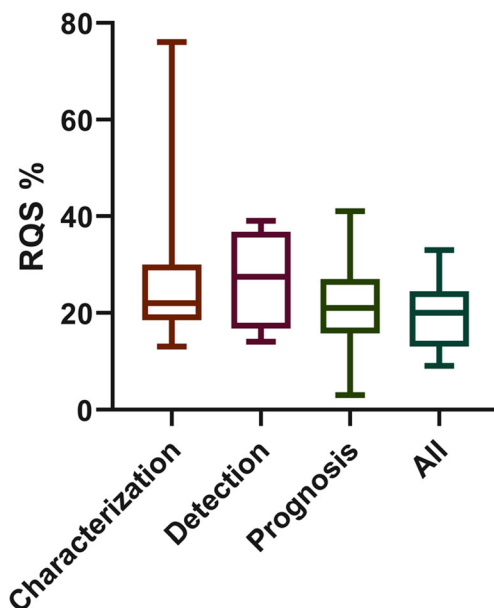


Fig. 5 Box plot depicting the distribution of RQS percentage score by aim of the studies included in the systematic reviews

appropriateness of RQS as a qualitative quality measure of radiomics research. The former is also supported by a recent investigation of methodological issues in machine learning research across different domains, including medicine and radiology, which also supports that inappropriate use of data analysis techniques indeed constitutes a critical issue [71], while the latter stems from the low variance of reported RQSs. This limitation of the current landscape of radiomics research represents undoubtedly one of the main factors preventing the translation of these tools to clinical decision support systems. Furthermore, as awareness of this problem grows throughout the medical imaging community, skepticism in the general public will only increase. This negative perception will probably persist for some time even if the quality and reliability of radiomics research improve in the near future. Researchers active in this field should therefore be particularly incentivized in improving the presentation and clarity of their methods and ease the reproduction of their experiments to foster a more positive environment and facilitate rather than hinder the adoption of radiomics-based software in clinical practice. As shown in our review, unfortunately, this does not seem to be currently the case. In this setting, journals and reviewers will probably need to take a more active role in raising the bar for minimal quality of radiomics research to be published. Guiding researchers toward a greater focus on investigations aiming at improved clinical outcomes rather than technical feasibility alone would also be a positive development. Some editors and journals have already begun to move in this direction, and it is desirable for this trend to spread at least to the more visible publications in our field [68, 72, 73].

Based on the results reported in the RQS systematic reviews included in this investigation, some common trends emerge. Some points were lacking in all or almost all instances, such as cost-effectiveness and decision curve analyses. Prospectively designed studies are also very rare, which is a common situation across radiology research compared to other clinical specialties. More worrisome, there is still a relevant number of studies that do not perform a validation of a final model, without retraining (e.g., as done in cross-validation). While cross-validation is a valuable tool to extract more information from smaller datasets and provide a better estimate of general performance of a pipeline, it is also true that it does not provide a univocal assessment of a model's deployment in a real-world setting. The pairing of cross-validation for model development and pipeline tuning and external validation of a definite model on a diverse dataset is probably the best solution. However, understandably, dataset size has to be adequate to allow both the training and external validation data to appropriately represent the model's general population target. It should also be noted that the RQS also addresses some items only superficially, such as feature reduction. It does not include an assessment of the appropriateness of the

techniques applied or the resulting dataset's size in comparison to the number of instances available for training. This could lead to an overestimation of the study's RQS score, as feature reduction accounts for either a -3 or $+3$ score out of the maximum of 36. Finally, we wish to highlight the lack of openness in many of the radiomics studies. Sharing the models and, ideally, the data used to train them is essential to allow correct assessment of their validity and validation on data from institutions different from those where they were developed. These steps are essential to grow trust in radiomics research and allow development of clinical decision support tools integrating these types of models.

Our systematic review presents some limitations that should be acknowledged. We did not aggregate the singular item data from each RQS-based review included in our study. This was partly due to the significant divergence in methods used to perform the rating (consensus, single reader). Also, it was not our intention to substitute the original studies in their topic-specific assessment, but rather to provide a wider overview of the current radiomics research state of the art. Therefore, we chose to aggregate the overall RQS percentage scores to obtain this result.

In conclusion, our review confirms the common sentiment that radiomics research quality must be increased in the near future as it is currently unsatisfactory independently of the study topic. External auditing of these investigations, in addition to the standard peer-review process, is valuable to highlight common limitations and provide insights to improve future study designs. The RQS serves a necessary role as the de facto reference tool for this task, but future research should be aimed at improving its reliability and developing new tools to meet an ever-evolving research space.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-022-09187-3>.

Acknowledgements The European Society of Medical Imaging Informatics supports the Radiomics Auditing Group Initiative with the aim to evaluate and improve the quality of radiomics studies.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Declarations

Guarantor The scientific guarantor of this publication is Prof. Renato Cuocolo, MD, PhD.

Conflict of interest Renato Cuocolo serves as an editorial board member of European Radiology and European Radiology Experimental.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because it is a systematic literature review.

Ethical approval Institutional Review Board approval was not required because it is a systematic literature review.

Methodology

- systematic review
- performed at multiple institutions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Pinto dos Santos D, Dietzel M, Baessler B (2021) A decade of radiomics research: are images really data or just patterns in the noise? *Eur Radiol* 31:1–4. <https://doi.org/10.1007/s00330-020-07108-w>
- Song J, Yin Y, Wang H et al (2020) A review of original articles published in the emerging field of radiomics. *Eur J Radiol* 127:108991. <https://doi.org/10.1016/j.ejrad.2020.108991>
- Bera K, Braman N, Gupta A et al (2022) Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol* 19:132–146. <https://doi.org/10.1038/s41571-021-00560-7>
- La Greca S-EA, Vuong D, Tschanz F et al (2021) Systematic review on the association of radiomics with tumor biological endpoints. *Cancers (Basel)* 13:3015. <https://doi.org/10.3390/cancers13123015>
- Qi Y, Zhao T, Han M (2022) The application of radiomics in predicting gene mutations in cancer. *Eur Radiol*. <https://doi.org/10.1007/s00330-021-08520-6>
- Jia Y, Yang J, Zhu Y et al (2021) Ultrasound-based radiomics: current status, challenges and future opportunities. *Med Ultrason*. <https://doi.org/10.11152/mu-3248>
- Li W, Liu H, Cheng F et al (2021) Artificial intelligence applications for oncological positron emission tomography imaging. *Eur J Radiol* 134:109448. <https://doi.org/10.1016/j.ejrad.2020.109448>
- Corrias G, Micheletti G, Barberini L et al (2022) Texture analysis imaging “what a clinical radiologist needs to know”. *Eur J Radiol* 146:110055. <https://doi.org/10.1016/j.ejrad.2021.110055>
- Pinto dos Santos D, Giese D, Brodehl S et al (2019) Medical students’ attitude towards artificial intelligence: a multicentre survey. *Eur Radiol* 29:1640–1646. <https://doi.org/10.1007/s00330-018-5601-1>
- Cuocolo R, Imbriaco M (2021) Machine learning solutions in radiology: does the emperor have no clothes? *Eur Radiol* 31:3783–3785. <https://doi.org/10.1007/s00330-021-07895-w>
- Buvat I, Orhac F (2019) The dark side of radiomics: on the paramount importance of publishing negative results. *J Nucl Med* 60:1543–1544. <https://doi.org/10.2967/jnumed.119.235325>
- Rizzo S, Botta F, Raimondi S et al (2018) Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2:36. <https://doi.org/10.1186/s41747-018-0068-z>
- Lafata KJ, Wang Y, Konkel B et al (2021) Radiomics: a primer on high-throughput image phenotyping. *Abdom Radiol (NY)*. <https://doi.org/10.1007/s00261-021-03254-x>
- van Timmeren JE, Cester D, Tanadini-Lang S et al (2020) Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging* 11:91. <https://doi.org/10.1186/s13244-020-00887-2>
- 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. <https://doi.org/10.1038/nrclinonc.2017.141>
- Guiot J, Vaidyanathan A, Deprez L et al (2022) A review in radiomics: making personalized medicine a reality via routine imaging. *Med Res Rev* 42:426–440. <https://doi.org/10.1002/med.21846>
- 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. <https://doi.org/10.1016/j.radonc.2018.03.033>
- Park JE, Kim HS, Kim D et al (2020) A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features. *BMC Cancer* 20:29. <https://doi.org/10.1186/s12885-019-6504-5>
- Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71. <https://doi.org/10.1136/bmj.n71>
- Wan X, Wang W, Liu J, Tong T (2014) Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol* 14:135. <https://doi.org/10.1186/1471-2288-14-135>
- Abdurixiti M, Nijjati M, Shen R et al (2021) Current progress and quality of radiomic studies for predicting EGFR mutation in patients with non-small cell lung cancer using PET/CT images: a systematic review. *Br J Radiol* 94:20201272. <https://doi.org/10.1259/bjr.20201272>
- Abunahel BM, Pontre B, Kumar H, Petrov MS (2021) Pancreas image mining: a systematic review of radiomics. *Eur Radiol* 31:3447–3467. <https://doi.org/10.1007/s00330-020-07376-6>
- Bhandari A, Ibrahim M, Sharma C et al (2021) CT-based radiomics for differentiating renal tumours: a systematic review. *Abdom Radiol (NY)* 46:2052–2063. <https://doi.org/10.1007/s00261-020-02832-9>
- Bhandari AP, Liang R, Koppen J et al (2021) Noninvasive determination of IDH and 1p19q status of lower-grade gliomas using MRI radiomics: a systematic review. *AJNR Am J Neuroradiol* 42:94–101. <https://doi.org/10.3174/ajnr.A6875>
- Calabrese A, Santucci D, Landi R et al (2021) Radiomics MRI for lymph node status prediction in breast cancer patients: the state of art. *J Cancer Res Clin Oncol* 147:1587–1597. <https://doi.org/10.1007/s00432-021-03606-6>
- Carbonara R, Bonomo P, Di Rito A et al (2021) Investigation of radiation-induced toxicity in head and neck cancer patients through radiomics and machine learning: a systematic review. *J Oncol* 2021:1–9. <https://doi.org/10.1155/2021/5566508>
- Castillo TJM, Arif M, Niessen WJ et al (2020) Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. *Cancers (Basel)* 12:1606. <https://doi.org/10.3390/cancers12061606>

28. Chen Q, Zhang L, Mo X et al (2021) Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 49:345–360. <https://doi.org/10.1007/s00259-021-05509-7>
29. Chetan MR, Gleeson FV (2021) Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur Radiol* 31:1049–1058. <https://doi.org/10.1007/s00330-020-07141-9>
30. Crombè A, Fadli D, Italiano A et al (2020) Systematic review of sarcomas radiomics studies: bridging the gap between concepts and clinical applications? *Eur J Radiol* 132:109283. <https://doi.org/10.1016/j.ejrad.2020.109283>
31. Davey MG, Davey MS, Boland MR et al (2021) Radiomic differentiation of breast cancer molecular subtypes using pre-operative breast imaging – a systematic review and meta-analysis. *Eur J Radiol* 144:109996. <https://doi.org/10.1016/j.ejrad.2021.109996>
32. Fornaçon-Wood I, Faivre-Finn C, O'Connor JPB, Price GJ (2020) Radiomics as a personalized medicine tool in lung cancer: separating the hope from the hype. *Lung Cancer* 146:197–208. <https://doi.org/10.1016/j.lungcan.2020.05.028>
33. Granzier RWY, van Nijnatten TJA, Woodruff HC et al (2019) Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: a systematic review. *Eur J Radiol* 121:108736. <https://doi.org/10.1016/j.ejrad.2019.108736>
34. Harding-Theobald E, Louissaint J, Maraj B et al (2021) Systematic review: radiomics for the diagnosis and prognosis of hepatocellular carcinoma. *Aliment Pharmacol Ther* 54:890–901. <https://doi.org/10.1111/apt.16563>
35. Janssen BV, Verhoef S, Westorp NJ et al (2022) Imaging-based machine-learning models to predict clinical outcomes and identify biomarkers in pancreatic cancer. *Ann Surg* 275:560–567. <https://doi.org/10.1097/SLA.0000000000005349>
36. Kao Y-S, Hsu Y (2021) A meta-analysis for using radiomics to predict complete pathological response in esophageal cancer patients receiving neoadjuvant chemoradiation. *In Vivo* 35:1857–1863. <https://doi.org/10.21873/invivo.12448>
37. Kao Y-S, Lin K-T (2021) A meta-analysis of computerized tomography-based radiomics for the diagnosis of COVID-19 and viral pneumonia. *Diagnostics* 11:991. <https://doi.org/10.3390/diagnostics11060991>
38. Kendrick J, Francis R, Hassan GM et al (2021) Radiomics for identification and prediction in metastatic prostate cancer: a review of studies. *Front Oncol* 11. <https://doi.org/10.3389/fonc.2021.771787>
39. Kim HY, Cho SJ, Sunwoo L et al (2021) Classification of true progression after radiotherapy of brain metastasis on MRI using artificial intelligence: a systematic review and meta-analysis. *Neurooncol Adv* 3. <https://doi.org/10.1093/nojnl/vdab080>
40. Kozikowski M, Suarez-Ibarrola R, Osiecki R et al (2021) Role of radiomics in the prediction of muscle-invasive bladder cancer: a systematic review and meta-analysis. *Eur Urol Focus*. <https://doi.org/10.1016/j.euf.2021.05.005>
41. Lecointre L, Dana J, Lodi M et al (2021) Artificial intelligence-based radiomics models in endometrial cancer: a systematic review. *Eur J Surg Oncol*. <https://doi.org/10.1016/j.ejso.2021.06.023>
42. Mühlbauer J, Egen L, Kowalewski K-F et al (2021) Radiomics in renal cell carcinoma—a systematic review and meta-analysis. *Cancers (Basel)* 13:1348. <https://doi.org/10.3390/cancers13061348>
43. Nardone V, Reginelli A, Grassi R et al (2021) Delta radiomics: a systematic review. *Radiol Med* 126:1571–1583. <https://doi.org/10.1007/s11547-021-01436-7>
44. Park JE, Kim D, Kim HS et al (2020) Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol* 30:523–536. <https://doi.org/10.1007/s00330-019-06360-z>
45. Ponsiglione A, Stanzione A, Cuocolo R et al (2022) Cardiac CT and MRI radiomics: systematic review of the literature and radiomics quality score assessment. *Eur Radiol* 32:2629–2638. <https://doi.org/10.1007/s00330-021-08375-x>
46. Shi L, Zhao J, Peng X et al (2021) CT-based radiomics for differentiating invasive adenocarcinomas from indolent lung adenocarcinomas appearing as ground-glass nodules: A systematic review. *Eur J Radiol* 144:109956. <https://doi.org/10.1016/j.ejrad.2021.109956>
47. Spadarella G, Calareso G, Garanzini E et al (2021) MRI based radiomics in nasopharyngeal cancer: systematic review and perspectives using radiomic quality score (RQS) assessment. *Eur J Radiol* 140:109744. <https://doi.org/10.1016/j.ejrad.2021.109744>
48. Staal FCR, van der Reijdt DJ, Taghavi M et al (2021) Radiomics for the prediction of treatment outcome and survival in patients with colorectal cancer: a systematic review. *Clin Colorectal Cancer* 20: 52–71. <https://doi.org/10.1016/j.clcc.2020.11.001>
49. Stanzione A, Gambardella M, Cuocolo R et al (2020) Prostate MRI radiomics: a systematic review and radiomic quality score assessment. *Eur J Radiol* 129:109095. <https://doi.org/10.1016/j.ejrad.2020.109095>
50. Tabatabaei M, Razaeei A, Sarrami AH et al (2021) Current status and quality of machine learning-based radiomics studies for glioma grading: a systematic review. *Oncology* 99:433–443. <https://doi.org/10.1159/000515597>
51. Ugga L, Perillo T, Cuocolo R et al (2021) Meningioma MRI radiomics and machine learning: systematic review, quality score assessment, and meta-analysis. *Neuroradiology* 63:1293–1304. <https://doi.org/10.1007/s00234-021-02668-0>
52. Ursprung S, Beer L, Bruining A et al (2020) Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma—a systematic review and meta-analysis. *Eur Radiol* 30:3558–3566. <https://doi.org/10.1007/s00330-020-06666-3>
53. Valdora F, Houssami N, Rossi F et al (2018) Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat* 169:217–229. <https://doi.org/10.1007/s10549-018-4675-4>
54. Wakabayashi T, Ouhmich F, Gonzalez-Cabrera C et al (2019) Radiomics in hepatocellular carcinoma: a quantitative review. *Hepatol Int* 13:546–559. <https://doi.org/10.1007/s12072-019-09973-0>
55. Walls GM, Osman SOS, Brown KH et al (2022) Radiomics for predicting lung cancer outcomes following radiotherapy: a systematic review. *Clin Oncol* 34:e107–e122. <https://doi.org/10.1016/j.clon.2021.10.006>
56. Wang H, Zhou Y, Li L et al (2020) Current status and quality of radiomics studies in lymphoma: a systematic review. *Eur Radiol* 30:6228–6240. <https://doi.org/10.1007/s00330-020-06927-1>
57. Wang Q, Li C, Zhang J et al (2021) Radiomics models for predicting microvascular invasion in hepatocellular carcinoma: a systematic review and radiomics quality score assessment. *Cancers (Basel)* 13:5864. <https://doi.org/10.3390/cancers13225864>
58. Westorp NJ, Hellingman T, Jansma EP et al (2021) Advanced analytics and artificial intelligence in gastrointestinal cancer: a systematic review of radiomics predicting response to treatment. *Eur J Nucl Med Mol Imaging* 48:1785–1794. <https://doi.org/10.1007/s00259-020-05142-w>
59. Westorp NJ, van Goor VJ, Kemna R et al (2021) Advanced image analytics predicting clinical outcomes in patients with colorectal liver metastases: a systematic review of the literature. *Surg Oncol* 38:101578. <https://doi.org/10.1016/j.suronc.2021.101578>
60. Won SY, Park YW, Ahn SS et al (2021) Quality assessment of meningioma radiomics studies: bridging the gap between exploratory research and clinical applications. *Eur J Radiol* 138:109673. <https://doi.org/10.1016/j.ejrad.2021.109673>
61. Won SY, Park YW, Park M et al (2020) Quality reporting of radiomics analysis in mild cognitive impairment and Alzheimer's

- disease: a roadmap for moving forward. *Korean J Radiol* 21:1345. <https://doi.org/10.3348/kjr.2020.0715>
62. Zhong J, Hu Y, Si L et al (2021) A systematic review of radiomics in osteosarcoma: utilizing radiomics quality score as a tool promoting clinical translation. *Eur Radiol* 31:1526–1535. <https://doi.org/10.1007/s00330-020-07221-w>
 63. Shur JD, Doran SJ, Kumar S et al (2021) Radiomics in oncology: a practical guide. *Radiographics* 41:1717–1732. <https://doi.org/10.1148/rg.2021210037>
 64. Zwanenburg A, Vallières M, Abdalah MA et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295:328–338. <https://doi.org/10.1148/radiol.2020191145>
 65. Collins GS, Reitsma JB, Altman DG, Moons K (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 13:1. <https://doi.org/10.1186/s12916-014-0241-z>
 66. Morland D, Triumbari EKA, Boldrini L et al (2022) Radiomics in oncological PET imaging: a systematic review—Part 1, Supradiaphragmatic cancers. *Diagnostics* 12:1329. <https://doi.org/10.3390/diagnostics12061329>
 67. Morland D, Triumbari EKA, Boldrini L et al (2022) Radiomics in oncological PET imaging: a systematic review—Part 2, Infradiaphragmatic cancers, Blood Malignancies, Melanoma and Musculoskeletal Cancers. *Diagnostics* 12:1330. <https://doi.org/10.3390/diagnostics12061330>
 68. Mongan J, Moy L, Kahn CE (2020) Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2:e200029. <https://doi.org/10.1148/ryai.2020200029>
 69. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH (2020) MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Informatics Assoc* 27:2011–2015. <https://doi.org/10.1093/jamia/ocaa088>
 70. Collins GS, Dhiman P, Andaur Navarro CL et al (2021) Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 11:e048008. <https://doi.org/10.1136/bmjopen-2020-048008>
 71. Kapoor S, Narayanan A (2022) Leakage and the reproducibility crisis in ML-based science. Available via <https://arxiv.org/abs/2207.07048>
 72. Halligan S, Menu Y, Mallett S (2021) Why did European Radiology reject my radiomic biomarker paper? How to correctly evaluate imaging biomarkers in a clinical setting. *Eur Radiol* 31:9361–9368. <https://doi.org/10.1007/s00330-021-07971-1>
 73. Pinto dos Santos D (2022) Radiomics in endometrial cancer and beyond - a perspective from the editors of the *EJR*. *Eur J Radiol* 150:110266. <https://doi.org/10.1016/j.ejrad.2022.110266>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.