iRADIOLOGY

REVIEW OPEN ACCESS

Utilizing Radiomics as Predictive Factor in Brain Metastasis Treated With Stereotactic Radiosurgery: Systematic Review and Radiomic Quality Assessment

Abdulrahman Umaru[1,2] | Hanani Abdul Manan[3] | Ramesh Kumar Athi Kumar[4] | Siti Khadijah Hamsan[4] | Noorazrul Yahya[1]

[1]Diagnostic Imaging and Radiotherapy (CODTIS), Faculty of Health Sciences, National University of Malaysia, Kuala Lumpur, Malaysia | [2]Department of Medical Radiography, Ahmadu Bello University, Zaria, Nigeria | [3]Department of Radiology, Faculty of Medicine, National University of Malaysia, Kuala Lumpur, Malaysia | [4]Gamma Knife Centre, Faculty of Medicine National University of Malaysia, Kuala Lumpur, Malaysia

Correspondence: Noorazrul Yahya (azrulyahya@ukm.edu.my)

Received: 30 November 2024 | **Revised:** 24 January 2025 | **Accepted:** 7 February 2025

Funding: The authors received no specific funding for this work.

Keywords: brain metastasis | deep learning | machine learning | MRI | predictive modeling | radiomics | radiomics quality score | stereotactic radiosurgery

ABSTRACT

Radiomics and machine learning (ML) are increasingly utilized to predict treatment response by uncovering latent information in medical images. This study systematically reviews radiomics studies on brain metastasis treated with stereotactic radiosurgery (SRS) and quantifies their radiomic quality score (RQS). A systematic search on Scopus, Web of Science, and PubMed was conducted to identify original studies on radiomics for predicting treatment response, adhering to predefined patient, intervention, comparator, and outcome (PICO) criteria. No restrictions were placed on language or publication date. Two independent reviewers assessed eligible studies, and the RQS was calculated based on Lambin's guidelines. The Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines were followed. Seventeen studies involving 2744 patients met the inclusion criteria out of 200 identified. All studies were retrospective and utilizing various MRI scanners models with different field strength. The average RQS across studies was low (39.2%), with a maximum score of 19 points (52.7%). Radiomic-based models demonstrated superior predictive accuracy compared to clinical or visual assessment, with AUC values ranging from 0.74 to 0.92. Integration of clinical features such as Karnofsky performance status, dose, and isodose line further improved model performance. Deep learning models achieved the highest predictive accuracy, with AUC of 0.92. Radiomics demonstrate significant potential in predicting treatment outcomes with high accuracy, offering opportunities to advance personalized management for BM. To facilitate clinical adoption, future studies must prioritize adherence to standardized guidelines and robust model validation to ensure reproducibility.

1 | Introduction

Brain metastases (BM) are devastating complication of systemic malignancies, affecting approximately 30% of cancer patients [1, 2]. Radiation-based treatments are the primary therapeutic option for most patients, as only a small fraction qualify for alternative interventions [3, 4]. Stereotactic radiosurgery (SRS) and hypo-fractionated stereotactic radiation therapy (SRT) are highly conformal noninvasive treatments delivered in one to five fractions. This technique effectively spares healthy tissue and preserves cognitive function, offering significant advantage over whole-brain radiation therapy (WBRT) [4]. Both the

Abbreviations: BM, brain metastases; IBSI, image biomarker standardization initiative; MRI, magnetic resonance imaging; RQS, radiomic quality score; SRS, stereotactic radiosurgery; WBRT, whole-brain radiation therapy.

American Society of Radiation Oncology and the International Stereotactic Radiosurgery Society recommend SRS and SRT as preferred treatment modalities for BM [5–8]. However, despite these advancements, treatment failure still occurs in 10%–30% of cases [9].

Magnetic resonance imaging (MRI) plays a critical role in the diagnosis and post-treatment monitoring of BM due to its superior soft tissue contrast [2, 3]. Traditionally, treatment response is assessed often via MRI by evaluating changes in the tumor size and anatomy [10, 11]. However, these changes often manifest only after a considerable delay, highlighting the need for reliable biomarkers that can predict the treatment response earlier. Early prediction can guide timely decision-making and improve outcomes [10].

Radiomics, a quantitative image analysis technique, facilitates the high-throughput extraction of data from medical images. This approach provides valuable insights into tumor phenotypes, classification, and treatment response prediction [12, 13]. By uncovering relationships in tumor characteristics that are not visually apparent, radiomics offers a deeper understanding of tumor characteristics [14, 15]. The process involve several stages, including image acquisition and preprocessing, tumor delineation, features extraction and selection, and model development and validation [12, 16].

Over the past decade, radiomics has demonstrated significant potential in predicting treatment outcomes across various tumor types [17–24]. Its noninversive nature, reliance on routine diagnostic images, and ability to analyze the entire tumor rather than isolated regions make it particularly advantageous [25, 26]. By enabling early therapy adjustments or salvage treatments, radiomics hold promise for improving survival rate and quality of life in patients [9]. Additionally, radiomics can enhance personalized treatment strategies by analyzing unique radiomic features of brain metastases from different primary cancer types. This allows for tailoring treatment plans to individual tumors, improving therapeutic efficacy, minimizing side effects, and revolutionize BM management.

Despite its potential, radiomics research faces challenges related to standardization, reproducibility, and study quality. To address these issues, the radiomic quality score (RQS) was introduced in 2017 to harmonize reporting practices, standardize methodologies, and enhance research rigor [27, 28].

This systematic review aims to evaluate radiomics studies focused on BM treated with SRS and specifically assessing their predictive performance and adherence to the RQS. By analyzing both methodological quality and clinical utility, this review seeks to advance the application of radiomics in neuro-oncology.

2 | Material and Methods

2.1 | Search Strategy

This systematic review adhered to Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines. A comprehensive literature search was performed using the patient, intervention, comparator, and outcome (PICO) framework across three databases: Scopus, Web of Science, and PubMed. Boolean operators and MeSH terms were employed to maximize search sensitivity. Details of the keywords and search strings are provided in Supporting Information S1: Files A and B, respectively.

2.2 | Inclusion and Exclusion Criteria

The literature search was performed without restriction on data or language. Studies were included if they met the following criteria: (1) use of gamma knife or CyberKnife radiosurgery, (2) involvement of patients with brain metastases, (3) utilization of MRI for treatment planning, and (4) application of radiomics for features extraction. Exclusion criteria included studies that (1) employed conventional radiotherapy, (2) did not utilize radiomics features (e.g., studies utilizing images directly without radiomic features extraction), (3) were nonhuman studies (e.g., phantom study), and (4) were review articles.

Two independent reviewers (A.U. and N.Y.) screened the articles for eligibility based on their title, abstract, and full texts, adhering to predefined inclusion and exclusion criteria. Discrepancies were resolved through consensus following the PRISMA guidelines to ensure transparency. The search and selection process were completed in February 2024.

2.3 | Data Extraction

Data extraction was performed using an excel spreadsheet, after duplicate studies were removed, titles and abstract were screened independently by two reviewers. Studies that passed this initial screening were assessed through full text review. Data were extracted and categorized into the following: (1) patient characteristics: authors, publication year, study design, number of patients, lesions, and primary cancer type, (2) imaging details: MRI system used, field strength, segmentation software, radiomics software, number of features extracted, and feature selection methods; and (3) outcome: accuracy, sensitivity, and AUC and any unique metrics. A summary of extracted data is presented in Tables 1–3.

2.4 | Assessment of Radiomics Quality Score

The RQS, proposed by Lambin et al. [40] was used to assess the quality of the radiomic workflow and reporting standards in the included studies. The RQS comprises of 16 components grouped into 6 key domains: domain 1: protocol quality and segmentation stability (items 1, 2, 3, and 4), domain 2: features selection and validation (items 5 and 12), domain 3: biological, clinical validation, and utility (items 6, 7, 13, and 14), domain 4: model performance index (items 8, 9, and 10), domain 5: high level of evidence (items 11 and 15), and domain 6: open science and data sharing (item 16). Each study was assessed using the original RQS checklist. The total RQS score was calculated and expressed as a percentage, with a maximum score of 36 points

TABLE 1 | Summary of image segmentation and radiomics features extraction.

Authors	Segmentation software	Segmentation method	Radiomics software	Features reduction/ selection	Numbers of extracted and selected features	Training data (%)	Validation data (%)	Testing data (%)	Clinical features	Type of algorism used	Software used for ML
Kawahara et al. [17]	GK planning system	Manual	IBEX	LASSO	7/740	55	15	30	NO	ML	MATLAB
Liao et al. [11]	NR	NR	NR	Forward selection	1763	70	30	NR	YES	ML	Not mentioned
Jiang et al. [22]	ITK-SNAP	Manual	Pyradiomics	Correlation analysis	1733	75	25	NR	Not mentioned	ML	Python
Mouraviev et al. [23]	GK planning system	Manual	Pyradiomics	Random forest features important matrix	440/12	NR	NR	NR	YES	ML	NR
Zheng et al. [29]	ITK-SNAP	Manual	Pyradiomics	Cox proportional hazard regression	236	NR	NR	NR	YES	ML	R
DeVries et al. [4]	Planning system	Manual	Pyradiomics	Wilcoxon rank, Kruskal–Wallis and Pearson correlation	107	NR	NR	NR	YES	ML	MATLAB
Chang et al. [30]	Planning system	Manual	Not mentioned	LASSO/mRMR	2851	NR	NR	NR	YES	ML	NR
Park et al. [31]	IBEX	Manual	IBEX	ICC/LASSO	362	NR	NR	NR	YES	ML	R
Karama et al. [32]	ITK/3DSLICER	Manual	Local binary pattern	Pearson correlation	15/220	NR	NR	NR	Not mentioned	ML	R
Jaberipour et al. [9]	3D slicer	Manual	Pyradiomics	Pearson correlation/ mRMR	100/800	83	NR	17	YES	ML	NR
H. Wang et al. [33]	Planning system/3D slicer	Manual	PyRadiomics	LASSO	10/125	82	NR	18	YES	ML	Python
Karami et al. [10]	3D slicer	Semi-automatic	Matlab	Pearson correlation/ Mann–Whitney *U* test/ANOVA	9/3072	NR	NR	NR	Not mentioned	ML	Matlab
Kawahara et al. [34]	Planning system	Manual	IBEX	LASSO	16/1962	60	15	25	NO	ML	MATLAB
Du et al. [35]	MITK	Manual	Pyradiomics	LASSO/mRMR	200/19,377	82	9	9	YES	ML	Python

(Continues)

TABLE 1 | (Continued)

Authors	Segmentation software	Segmentation method	Radiomics software	Features reduction/selection	Numbers of extracted and selected features	Training data (%)	Validation data (%)	Testing data (%)	Clinical features	Type of algorism used	Software used for ML
Jalalifar et al. [36]	2D UNet, 3D UNet and MSGA	Manual and automatic	Pyradiomic	mRMR	7/3436	74	7	19	NO	ML	NR
Carloni et al. [37]	Planning system	Manual	Pyradiomic and SOPHiA	LASSO	1763	66	NR	34	YES	ML	Python
Liao et al. [38]	Treatment planning system	Manual	Neuro-image biomarker analysis	Univariate Cox propotional analysis and chi square & SFS	1793	70	NR	30	YES	DL	DeepSurv

Note: NR is used when the value is not given (not reported).
Abbreviations: DL, deep learning; ICC, interclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; ML, machine learning; mRMR, maximum relevance and minimum redundance; SFS, sequential forward selection.

(100%) [40–42]. Details of RQS evaluation and radiomic workflow diagram are presented in Supporting Information S1: File C and Figure 1, respectively.

3 | Results

3.1 | Study Selection and Characteristics of Included Studies

The literature search yielded 200 studies across 3 consulted databases: Scopus (174), PubMed (24), and Web of Science (2). After removing 16 duplicates records and excluding 130 studies based on title review, 54 studies were assessed further. Abstract screening excluded 29 studies for reasons detailed in Figure 2. An additional eight studies were excluded for primarily focusing on tumor classification rather than prediction. Ultimately, 17 studies met the inclusion criteria based on the PICO framework and were included in this review. The PRISMA flow diagram in Figure 2 summarizes the selection process.

3.2 | General Characteristics of the Studies

The 17 included studies, all retrospective, were published between 2019 and 2023, encompassing 2744 patients (mean: 161 patients per study, range 28–831 patients), and 7377 lesions. Six studies (35%) included patients with multiple primary cancer type, whereas others focused on specific cancers: nonsmall cell lung cancer (18%), lung cancer (12%), melanoma (12%), and breast cancer (6%). Three studies (18%) did not specify the primary cancer type. Study endpoints included treatment response, survival, clinical outcome, and BRAF mutation status prediction in one study.

Regarding treatment modalities, nine studies (52.9%) utilized gamma knife, three (17.6%) employed LINAC, and five (29.4%) did not specify the SRS modality. MRI system varied: 11.8% used GE, 23.5% Siemens, and 29.4% Philips, with 11.8% combining Siemens and GE systems. Four studies (23.5%) did not specify the MRI system. Most studies used 1.5T field strength (94.1%), and all utilized contrast-enhanced T1-weighted images (CE-T1W) for feature extraction (100%). Some studies also incorporated T2-weighted (35.3%) and T2-FLAIR sequences (35.3%), reporting improved model performance with combined sequences. However, no significant performance gains were noted with 3T MRI system (Table 3).

3.3 | RQS Assessment

The average RQS for the 17 studies was 14.1 points (39.2% of the total score), with the highest being 19 points (52.7%) [39].

- Domain 1: All the studies reported a well document image protocol and image segmentation methods (100%) but nonconducted phantom assessments.
- Domain 2: All studies demonstrated good compliance with features selection (100%) and validation using same data

TABLE 2 | Characteristics of studies and MRI machine used included in the review.

Authors	Year	Country and region	Study design	No. of patients	No of metastases	Primary cancer	Treatment	MRI machine	Field strength	Pulse sequences
Kawahara et al. [17]	2021	UK	R	45	115	Melanoma	GK model 4C	Siemens syngo MR	1.5	FLASH T1WCE
Liao et al. [11]	2021	Taiwan, China	R	256	976	NCSLC	GK	Not specified	Not mentioned	T1WCE, T2 WI.
Jiang et al. [22]	2022	China	R	137	213	LCBM	GK	Siemens magneton Skyra	3	CE-T1W, T1W T2WI T2 FLAIR
Mouraviev et al. [23]	2020	Canada	R	87	408	Not specified	GK and GK icon	Philips ingenia	1.5	CE-T1W T2W FLAIR
Zheng et al. [29]	2021	China	R	44	81	BCBM	GK Perfexion	GE-signa	1.5	CE-T1W, T2W DWI
DeVries et al. [4]	2022	Netherlands	R	99	123	Multiple	Novalis LINAC	Siemens and 4 GE	1.5 and 1	CE-T1WI
Chang et al. [30]	2021	USA	R	831	3596	Multiple	Not specified	Not specified	Not mentioned	CE-T1W T2 FLAIR
Park et al. [31]	2021	Republic of Korea	R	83	118	LCBM	GK perfexion	Philips gyroscan intera	1.5	CE-T1W T2W
Karami et al. [32]	2019	Canada	R	38	38	Multiple	Not specified (hf-SRT)	Philips ingenia	1.5	T1-WCE T2 FLAIR
Jaberipour et al. [9]	2021	Canada	R	120	171	Not specified	LINAC	Philips ingenia	1.5	T1-WCE T2 FLAIR
H. Wang et al. [33]	2021	USA	R	28	179	Multiple	GK perfexion	Siemens	1.5	CE-TWI
Karami et al. [10]	2019	Canada	R	100	133	Multiple	Not specified (hf_SRT)	Philips ingenia	1.5	CE-T1W T2FLAIR
Kawahara et al. [34]	2023	Japan	R	30	220	Melanoma	GK	Siemens syngo	1.5	CE-T1W FLASH
Du et al. [35]	2023	China	R	337	337	Multiple	Gamma Knife	GE signa	1.5	CE-T1W
Jalalifar et al. [36]	2022	Canada	R	124	156	Not specified	Not specified (hf_SRT)	No details	Not mentioned	CE-T1W T2 FLAIR
Carloni et al. [37]	2023	Italy	R	148	276	NSCLC	CyberKnife	Siemens and GE	1.5	CE-TW1
Liao et al. [38]	2023	Taiwan, China	R	237	237	NSCLC	GK	Not specified	Not mentioned	CE-T1W, T1WI, T2WI

Note: R mean retrospective study.
Abbreviations: CE, contrast enhanced; GE, general electric; GK, gamma knife; LCBM, lung cancer brain metastasis; LINAC, linear accelerator; NSCLC, non-small cell lung cancer; T1WI, T1 weighted images.

(88.2%), but only 17.6%, and 5.9% validated models using internal and external datasets, respectively.

- Domain 3: Multivariate analysis using nonradiomics data was performed in 82.4% of studies. However, there was poor adherence to biological correlates (29.4%) and clinical utility (5.9%), with no comparison to gold standards.
- Domain 4: Discrimination and calibration statistics were well reported.

TABLE 3 | Summary of findings.

Authors	AUC				Accuracy and sensitivity (%)				Univariate analysis (C-index at 95% CI)	Multivariate analysis (C-index at 95% CI)
	Radiomic	External	Clinical	Combined	Radiomics	Clinical	Combined	Visual		
Kawahara et al. [17]	0.87	—	—	—	80, 83	—	—	54, 44	—	—
Liao et al. [11]	0.86	—	0.80	0.95	85, 85	76, 69	90, 91	—	—	—
Jiang et al. [22]	0.852	—	—	0.922	76.2	—	—	—	—	—
Mouraviev et al. [23]	0.793	—	0.669	0.718	—	—	—	—	—	—
DeVries et al. [4]	0.74	—	0.71	0.77	—	—	—	—	—	—
Karami et al. [32]	0.80	—	—	—	82, 84	—	—	—	—	—
Jabaripour et al. [9]	0.76	—	0.62	0.87	77, 69	63, 62	87, 85	—	—	—
H. Wang et al. [33]	0.82	—	—	—	83%	—	—	—	—	—
Karami et al. [10]	0.82	—	—	—	82, 80	—	—	—	—	—
Du et al. [39]	0.93	0.90	0.67	0.94	93, 90	—	87, 85	—	—	—
Jalalifar et al. [36]	0.81	80, 82.5	—	—	—	—	—	—	—	—
Liao et al. [38]	0.92	86	—	—	—	—	—	—	—	—
Kawahara et al. [34]	0.82	—	—	—	83.1, 78.8	—	—	—	—	—
Zheng et al. [29]	—	—	—	—	—	—	—	—	0.64	0.79
Chang et al. [30]	—	—	—	—	—	—	—	—	0.652	—
Park et al. [31]	—	—	—	—	—	—	—	—	—	0.79 (testing) 0.78 (validation)
Carloni et al. [37]	—	—	—	—	—	—	—	—	SR = 0.70 PyR = 0.63	—

Note: Not all studies reported the value for external validation, clinical features, C-index sensitivity and specificity and therefore represented by "—".
Abbreviations: IMC1, informational measurement of correlation; PyR; PyRadiomics; RT, tumor radiomic value; RTE, tumor Edema radiomic value; SR, SOPHiA Radiomics.

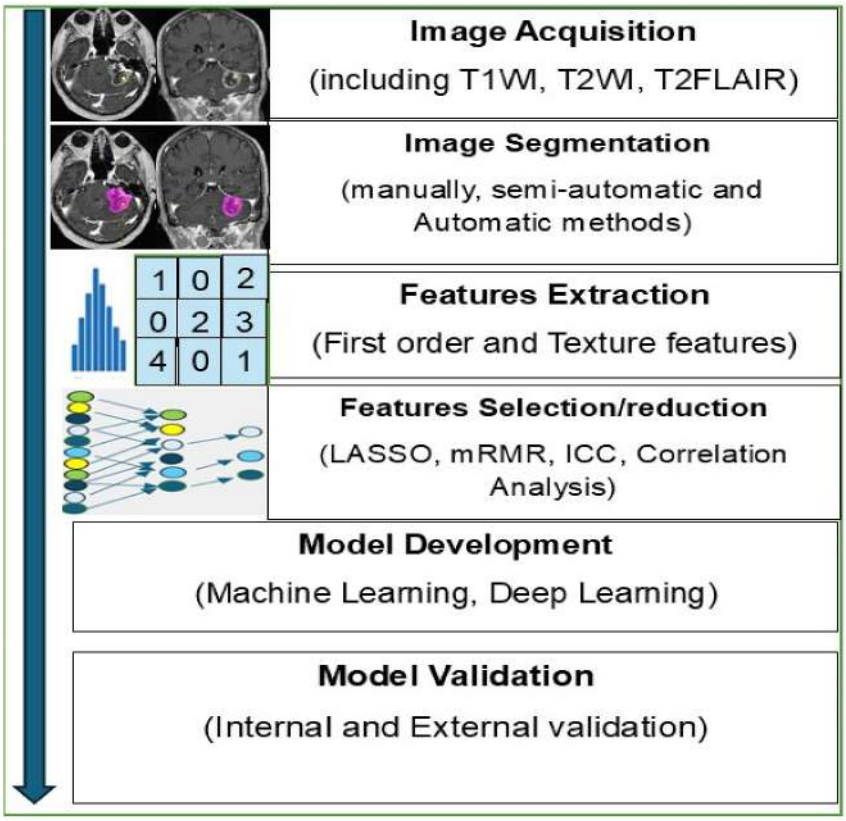

FIGURE 1 | Radiomics flow diagram for implementation of radiomics study. All studies utilized different approach from the image acquisition, segmentation process, radiomics features extraction and selection technique, model development to model validation. The arrow demonstrates the direction of flow. ICC, interclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; mRMR, minimum redundancy maximum relevance (*Images Sourced:* HCTM KL, 2023).

- Domain 5: Cut-off analysis and cost-effectiveness were not addressed, and all studies were retrospective studies.
- Domain 6: Data sharing was noted in 64.7% of studies.

Studies published in 2022 and 2023 demonstrated improved RQS scores [22, 34, 36, 37]. The basic adherence rate to the reporting of RQS is presented in Table 4 [40].

3.4 | Predicting Treatment Response

Radiomics models demonstrated strong predictive capabilities for treatment response and local failure (LF), achieving AUCs ranging from 0.74 to 0.87. These models significantly outperformed visual evaluations, with accuracies and sensitivities of 80% and 83%, respectively, compared to 54% and 44% for visual evaluations [17]. Radiomics models also exhibited 7%–25% higher AUCs than clinical models.

Integrating radiomic and clinical features further enhanced model performance. Key clinical features included Karnofsky performance status (KPS), presence of extracranial metastasis, number and volume of brain metastasis (BM), tumor volume, and treatment using WBRT [23, 38, 39]. Additional contributors included primary tumor type, edema index, radiation dose, isodose line, tumor site, histology, systemic therapy, and neurological symptoms [4, 9, 23, 33]; detailed radiomics features utilized by each study is available in Supporting Information S1: File D.

Deep learning (DL) models improved predictive power, achieving an AUC of 0.92, comparable to combined radiomic-clinical models [38]. Tumor size and scanner variability influence model performance, with smaller tumors (< 7.5 cc) yielding better result [4], combining data from similar scanners (e.g., Magnetom/Expert) improve performance (AUC of 0.84), whereas pairing different scanners (e.g., Avanto scanner), reduces performance (AUC to 0.77) [4]. Segmentation accuracy had minimal impact, with AUCs of 0.78 for less accurate and 0.81 for more accurate manual ROIs segmentation [36]. The radiomics platform also affected outcomes, with SOPHiA Radiomics outperforming PyRadiomics (median C-Index of 0.70 vs. 0.63) [37].

4 | Discussion

This systematic review synthesized the current evidence on the utilization of radiomics in predicting treatment outcomes in brain metastasis treated with SRS, alongside an evaluation of the methodological rigor of the included studies using the RQS. Radiomics demonstrated a growing role of predicting treatment response and other clinical endpoints while uncovering unique imaging phenotypes unique associated with brain metastases

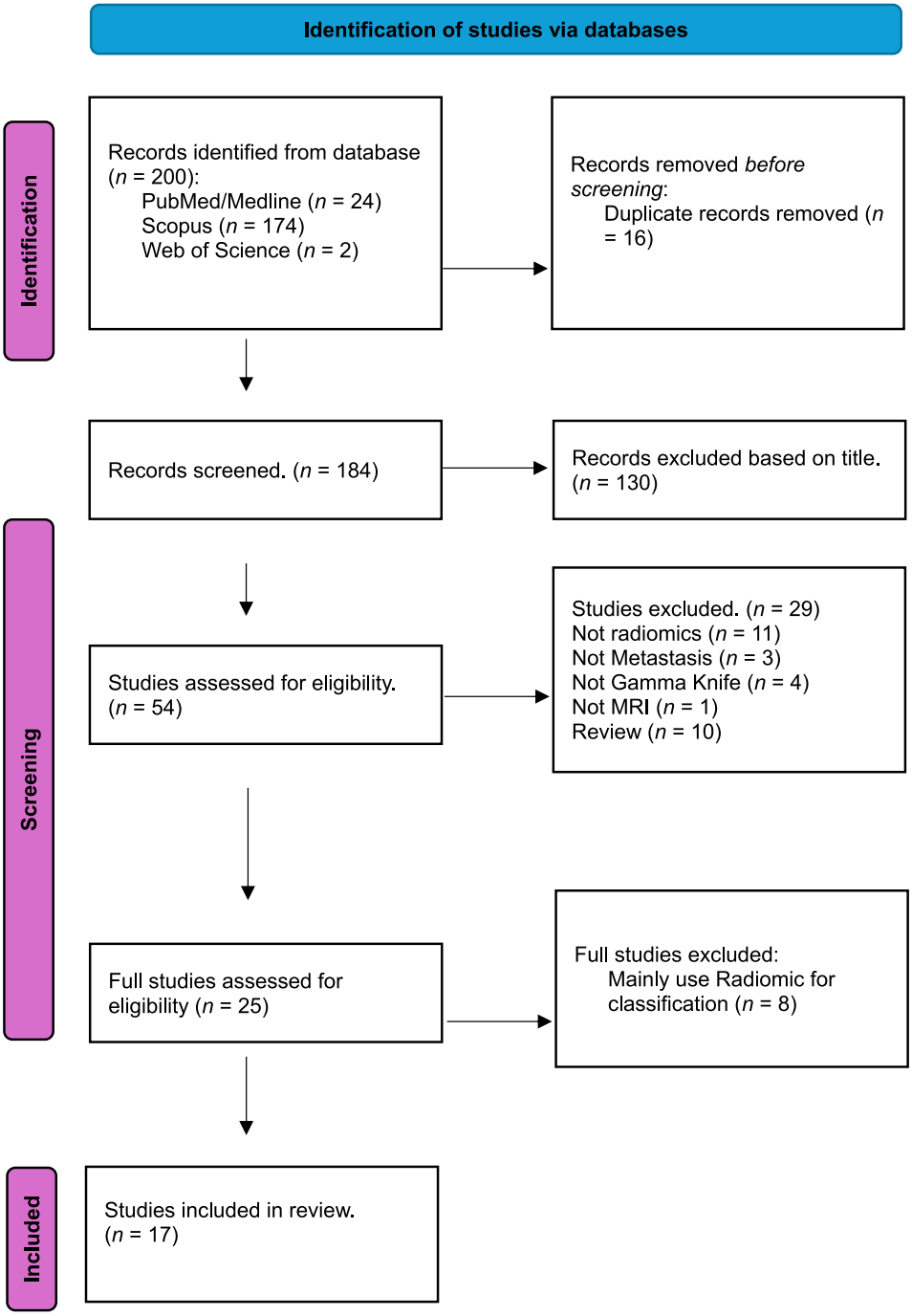

FIGURE 2 | Flow diagram of preferred reporting items for systematic review and meta-analysis.

originating from different primary cancers [9, 23, 31]. This is critical, as primary tumor origin may significantly influence the biology, progression, and treatment response of brain metastases. These findings underscore the potential of radiomics to provide deeper insights into tumor heterogeneity, enabling more personalized treatment strategies tailored to the molecular and histopathological characteristics of individual metastases.

The RQS assessment revealed significant methodological and reporting gaps, with an average adherence rate of 14.1 (39.2% of the maximum score). Although recent studies demonstrated improve compliance with RQS guidelines, notable strengths included well-documented imaging protocols, features reduction/selection methods, and consistent image segmentation practices. However, critical deficiencies contributed to the overall low RQS. A major limitation common to most radiomics studies was the universal absence of phantom studies, which are crucial for assessing intra-scanner variability and ensuring the reproducibility of radiomics features across different imaging platforms [43, 44]. The lack of phantom study may stem from the logistics and financial challenges associated with conducting these experiments, as well as retrospective nature of the studies. Another significant gap was the limited use of external dataset for model validation, with only 17.6% of studies validating models internally and just 5.9% using external datasets. Despite this, the only study with external validation demonstrated

TABLE 4 | Basic adherence rate according to six key domains.

Six key domains of RQS	Adherence rate (%)
Domain 1: Protocol quality and stability in images segmentation	
Image protocol quality	100
Multiple segmentation	100
Phantom study	0
Multiple time point imaging (test retest)	47.1
Domain 2: Feature selection and validation	
Features reduction or adjustment	100
Validation	5.9
Domain 3: Biologic/clinical validation and utility	
Multivariate analysis with non-radiomic features	82.4
Biologic correlates	29.4
Comparison to "gold standard"	0
Potential clinical utility	5.9
Domain 4: Model performance index	
Discrimination statistics	100
Calibration statistics	100
Cut-off analysis	0
Domain 5: High level of evidence	
Prospective study	0
Cost-effective analysis	0
Domain 6: Open science and data	
Open science and data	64.7

promising performance. Jalalifar et al. [36] reported an AUC of 80%, on internal validation and 82.5% on external validation. These findings demonstrate the strong predictive performance of radiomic models beyond their training dataset.

The reliance on single-center retrospective datasets hinders the generalizability of radiomics models likely due to data sharing challenges. External validation must be prioritized in features studies to ensure models perform robustly in diverse clinical setting, thereby enhancing clinical applicability [35, 37].

Additionally, limited reporting on clinical utility and absence of cost effectiveness analyses restricts radiomics translation into clinical practice. Addressing these gaps requires multidisciplinary collaboration among researchers, clinicians, radiologist, oncologist, data scientist, bioinformatics, and health economists, to assess the feasibility and economic implication of integrating radiomics into clinical workflow, particularly in resource limited setting. Future studies should prioritize assessing the feasibility and economic implications of integrating radiomics into clinical workflows. This includes detailed cost-effectiveness analyses to determine whether radiomics models can optimize resources utilization and reduce unnecessary interventions. Moreover, interdisciplinary teams can jointly design and implement comprehensive studies to develop robust models. Such collaboration ensures that radiomics studies address real-world challenges and provide actionable insights that benefits both clinician and patients.

Lack of comparison to the gold standards and inadequate reporting on biological correlates further impede the integration of radiomics with other biomarkers, which is critical for advancing precision medicine. The retrospective nature of the studies, while common in radiomics research, limits the robustness of the findings. Prospective studies designed with standardized imaging protocols and adherence to IBSI guidelines are essential for ensuring reliability in future research. Self-reported RQS could improve transparency and enhance the study quality. Although RQS provides a good framework, emerging tools such as METRICS (methods for evaluating the reliability, integrity, and completeness of statistics) and TRIPOT (transparent reporting of an interpretation of prediction outcomes for diagnosis) may complement the limitations of RQS and are recommend for consideration in future studies. These finding are consistent with previous review by Ismail et al. [45] who identified similar limitations in radiomics studies in head and neck cancers.

Despite the limitations in the reporting system highlighted above, radiomics models consistently demonstrated superior performance in predicting treatment responses compared to clinical and visual evaluation. The accuracy of radiomics studies ranged from 76% to 85% with AUCs values between 0.7 and 0.92, significantly outperforming clinical models (AUC: 0.62–0.80) [4, 9, 11, 17, 23, 39]. Kawahara et al. [17] reported an accuracy of 80% using radiomics, compared to 54% with visual evaluation. These findings highlight radiomics potential to advance personalized treatment strategies.

Radiomics can identify features associated with poor treatment response, guiding adaptive radiotherapy plans such as dose escalation or integrating additional therapies for high-risk patients. Studies have shown that integrating radiomic and clinical features further improve model performance, with AUC values increasing up to 25% compared to clinical model alone [4, 9, 11, 23, 39]. However, not all clinical features were significant, features such as Karnofsky performance status, dose and isodose line improved performance, whereas others such as gender and primary tumor location did not [39]. These findings underscore the need for careful selection of clinical features to avoid redundancy. By providing data-driven insights into tumor behavior, radiomics can empower clinicians to make inform decisions regarding treatment adjustments, ultimately improving patient outcomes. Future studies incorporating long-term follow-up could further enhance the ability to predict sustained treatment outcomes, offering more comprehensive insights into patient outcome.

Notably, DL models have demonstrated significant potential in radiomics research, often surpassing traditional machine learning models in predictive performance due to their ability to learn complex, high dimensional representations of data. For instance, a study reported an AUC of 0.92 using DL, a

performance comparable to combined radiomics and clinical models, highlighting deep learning potential to enhancing prediction accuracy [38]. The strength of DL lies in its capacity to automatically learn complex high-dimensional features representation, making it particularly well suited for tasks involving large scale datasets. However, its effectiveness is contingent on the availability of large datasets. As insufficient data increases the risk of overfitting. In contrast ML tends to perform robustly with smaller datasets provided meaningful features are extracted, reducing susceptibility to over fitting.

These limitations can be mitigated through strategies such as data augmentation, which artificially expands training dataset by introducing minor variations, and transfer learning, where pre-trained models on large dataset are fine-tuned for specific tasks. Transfer learning enables models to leverage previously learned features and adapt to smaller datasets. Future research should focus on balancing the strengths of DL with its challenges by optimizing these techniques. Furthermore, available online data, collaboration, data sharing, and synthetic data generation should be explored to further enhance dataset availability. Additionally, DL requires considerable computational resources and careful hyperparameter tuning. Overall, although DL holds immense promise for future radiomics applications, addressing challenges associated with small datasets is critical to unlock the full potential of DL models.

Tumor size also influence radiomic model performance, with smaller tumors (< 7.5 cc) being predicted more accurately than the larger ones [4]. Scanner variability is another critical factor impacting model performance. Model trained on data from closely matched scanner such as Magnetom/Expert, achieved better performance (AUC of 0.84) than those combining data from different MRI models, such as Avanto with other GE scanners (AUC of 0.77) [4]. This underscores the importance of harmonizing imaging protocols and adherence to IBSI guidelines and RQS, including phantom study and external validation, to address inter-scanner variability. Interestingly, variation in field strength did not significantly affect model performance, further emphasizing the importance of phantom study to strengthen the inter scanner variability.

Pulse sequences choice also affected model accuracy. Jiang et al. [22] reported that T2-weighed imaging (T2WI) outperformed T2-FLAIR, CE-T1WI, and TWI in terms of AUC (0.725, 0.704, 0.657, 0.557, respectively), likely due to T2WI's ability to highlight edema in certain pathologic conditions. However, Zhang et al. [46] reported neither T2WI nor T2 FLAIR could reliably distinguish radiation necrosis from tumor progression. Similarly Park et al. [31] suggested that CE-T1W texture features were more valuable than T2WI, T2-FLAIR, or T1WI. Combining features from multiple pulse sequences has shown to enhance model performance, supporting the idea that different pulse sequence contributes unique complementary information [47]. These findings highlight the importance of a multi-modal approach, combining diverse imaging modalities and pulse sequences to improve radiomics model performance and provide a more comprehensive understanding of tumor characteristics.

Surprisingly, segmentation accuracy had minimal impact model performance (AUC 0.78 vs. 0.81) [36] suggesting that radiomics model are robust enough to tolerate some minor segmentation inconsistencies. This finding supports the use of simpler, automated segmentation methods, which could be more practical in clinical setting [48]. However, variability in features extraction across radiomics platforms affect model outcomes. For instance, SOPHiA Radiomics outperformed Pyradiomics, achieving a median C-index of 0.70 compared to 0.63. This is a critical aspect of radiomics studies, and the finding underscores the importance of utilizing platforms compliant with the IBIS guidelines to ensure consistency and reliability in results. This view is supported by Fornacon-Wood et al. [49] and Foy et al. [50], which emphasize the need for standardized features extraction practices to enhance reproducibility and clinical applicability.

Despite these promising findings, significant limitations remain. Key gaps include the absence of phantom studies to assess intra-scanner variability, external validation, cost effectiveness analyses, and consideration for clinical utility, alongside the retrospective nature of the studies. Variability in imaging protocols, and scanner types across studies poses a major challenge, as it impacts the reproducibility and comparability of results. Future studies should prioritize standardized imaging protocol to enhance reproducibility and ensure consistent outcome across centers. Multicenter collaborations are also essential for accessing diverse datasets, thereby strengthening the external validity and generalizability of radiomics models. Given the growing importance of these findings, we recommend that features reviews or meta-analyses include detailed imaging acquisition protocol as reference points. Prospective studies, and strict adherence to IBSI guidelines are critical for advancing the clinical utility of radiomics models.

In conclusion, although radiomics demonstrates significant potential for predicting treatment outcomes in brain metastasis, addressing methodological and reporting gaps is vital to improving the robustness, reproducibility, and clinical translation of these models. Future studies should strictly adhere to RQS and IBIS guidelines.

Author Contributions

Abdulrahman Umaru: conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), writing – original draft (equal), writing – review and editing (equal). **Hanani Abdul Manan:** conceptualization (equal), methodology (equal), supervision (equal), writing – review and editing (equal). **Ramesh Kumar Athi Kumar:** conceptualization (equal), methodology (equal), supervision (equal), writing – review and editing (equal). **Siti Khadijah Hamsan:** conceptualization (equal), methodology (equal), supervision (equal), writing – review and editing (equal). **Noorazrul Yahya:** conceptualization (equal), methodology (equal), project administration (equal), resources (equal), supervision (equal), writing – original draft (equal), writing – review and editing (equal).

Acknowledgments

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

1. P. Sacks and M. Rahman, "Epidemiology of Brain Metastases," *Neurosurgery Clinics of North America* 31, no. 4 (2020): 481–488, https://doi.org/10.1016/j.nec.2020.06.001.

2. V. Salvestrini, C. Greco, A. E. Guerini, et al., "The Role of Feature-Based Radiomics for Predicting Response and Radiation Injury After Stereotactic Radiation Therapy for Brain Metastases: A Critical Review by the Young Group of the Italian Association of Radiotherapy and Clinical Oncology (yAIRO)," *Translational Oncology* 15, no. 1 (2022): 101275, https://doi.org/10.1016/j.tranon.2021.101275.

3. M. Kocher, M. I. Ruge, N. Galldiks, and P. Lohmann, "Applications of Radiomics and Machine Learning for Radiotherapy of Malignant Brain Tumors," *Strahlentherapie und Onkologie* 196, no. 10 (2020): 856–867, https://doi.org/10.1007/s00066-020-01626-8.

4. D. A. DeVries, F. Lagerwaard, J. Zindler, et al., "Performance Sensitivity Analysis of Brain Metastasis Stereotactic Radiosurgery Outcome Prediction Using MRI Radiomics," *Scientific Reports* 12, no. 1 (2022): 20975, https://doi.org/10.1038/s41598-022-25389-7.

5. Y. J. Cha, W. I. Jang, M. S. Kim, et al., "Prediction of Response to Stereotactic Radiosurgery for Brain Metastases Using Convolutional Neural Networks," *Anticancer Research* 38, no. 9 (2018): 5437–5445, https://doi.org/10.21873/anticanres.12875.

6. R. Hu, X. Li, X. Zhou, and S. Ding, "Development and Validation of a Competitive Risk Model in Patients With Rectal Cancer: Based on SEER Database," *European Journal of Medical Research* 28, no. 1 (2023): 362, https://doi.org/10.1186/s40001-023-01357-3.

7. E. Rosenblatt and E. Zubizarreta, *Radiotherapy in Cancer Care: Facing the Global Challenge* (International Atomic Energy Agency [IAEA], 2017), https://www.iaea.org/publications/10627/radiotherapy-in-cancer-care-facing-the-global-challenge.

8. C. Borek, "Antioxidants and Radiation Therapy," *Journal of Nutrition* 134, no. 11 (2004): 3207S–3209S, https://doi.org/10.1093/jn/134.11.3207S.

9. M. Jaberipour, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, "A Priori Prediction of Local Failure in Brain Metastasis After Hypo-Fractionated Stereotactic Radiotherapy Using Quantitative MRI and Machine Learning," *Scientific Reports* 11, no. 1 (2021): 21620, https://doi.org/10.1038/s41598-021-01024-9.

10. E. Karami, H. Soliman, M. Ruschin, et al., "Quantitative MRI Biomarkers of Stereotactic Radiotherapy Outcome in Brain Metastasis," *Scientific Reports* 9, no. 1 (2019): 19830, https://doi.org/10.1038/s41598-019-56185-5.

11. C.-Y. Liao, C.-C. Lee, H.-C. Yang, et al., "Enhancement of Radiosurgical Treatment Outcome Prediction Using MRI Radiomics in Patients With Non-Small Cell Lung Cancer Brain Metastases," *Cancers* 13, no. 16 (2021): 4030, https://doi.org/10.3390/cancers13164030.

12. F. C. R. Staal, D. J. van der Reijd, M. Taghavi, D. M. J. Lambregts, R. G. H. Beets-Tan, and M. Maas, "Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review," *Clinical Colorectal Cancer* 20, no. 1 (2021): 52–71, https://doi.org/10.1016/j.clcc.2020.11.001.

13. G. M. Walls, S. O. S. Osman, K. H. Brown, et al., "Radiomics for Predicting Lung Cancer Outcomes Following Radiotherapy: A Systematic Review," *Clinical Oncology* 34, no. 3 (2022): e107–e122, https://doi.org/10.1016/j.clon.2021.10.006.

14. A. Rahmim, P. Huang, N. Shenkov, et al., "Improved Prediction of Outcome in Parkinson's Disease Using Radiomics Analysis of Longitudinal DAT SPECT Images," *NeuroImage: Clinical* 16 (2017): 539–544, https://doi.org/10.1016/j.nicl.2017.08.021.

15. A. Bettinelli, M. Branchini, F. De Monte, A. Scaggion, and M. Paiusco, "Technical Note: An IBEX Adaption Toward Image Biomarker Standardization," *Medical Physics* 47, no. 3 (2020): 1167–1173, https://doi.org/10.1002/mp.13956.

16. J. D. Shur, S. J. Doran, S. Kumar, et al., "Radiomics in Oncology: A Practical Guide," *RadioGraphics* 41, no. 6 (2021): 1717–1732, https://doi.org/10.1148/rg.2021210037.

17. D. Kawahara, X. Tang, C. K. Lee, Y. Nagata, and Y. Watanabe, "Predicting the Local Response of Metastatic Brain Tumor to Gamma Knife Radiosurgery by Radiomics With a Machine Learning Method," *Frontiers in Oncology* 10 (2021): 569461, https://doi.org/10.3389/fonc.2020.569461.

18. J. Choe, S. M. Lee, K. H. Do, et al., "Outcome Prediction in Resectable Lung Adenocarcinoma Patients: Value of CT Radiomics," *European Radiology* 30, no. 9 (2020): 4952–4963, https://doi.org/10.1007/s00330-020-06872-z.

19. X. Fave, L. Zhang, J. Yang, et al., "Delta-Radiomics Features for the Prediction of Patient Outcomes in Non-Small Cell Lung Cancer," *Scientific Reports* 7, no. 1 (2017): 588, https://doi.org/10.1038/s41598-017-00665-z.

20. Y. Huang, Z. Liu, L. He, et al., "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, no. 3 (2016): 947–957, https://doi.org/10.1148/radiol.2016152234.

21. G.-W. Ji, F.-P. Zhu, Y.-D. Zhang, et al., "A Radiomics Approach to Predict Lymph Node Metastasis and Clinical Outcome of Intrahepatic Cholangiocarcinoma," *European Radiology* 29, no. 7 (2019): 3725–3735, https://doi.org/10.1007/s00330-019-06142-7.

22. Z. Jiang, B. Wang, X. Han, et al., "Multimodality MRI-Based Radiomics Approach to Predict the Posttreatment Response of Lung Cancer Brain Metastases to Gamma Knife Radiosurgery," *European Radiology* 32, no. 4 (2022): 2266–2276, https://doi.org/10.1007/s00330-021-08368-w.

23. A. Mouraviev, J. Detsky, A. Sahgal, et al., "Use of Radiomics for the Prediction of Local Control of Brain Metastases After Stereotactic Radiosurgery," *Neuro-Oncology* 22, no. 6 (2020): 797–805, https://doi.org/10.1093/neuonc/noaa007.

24. F. Khalvati, Y. Zhang, A. Wong, and M. A. Haider, "Radiomics," *Encyclopedia of Biomedical Engineering* 3, (2019): 597–603, https://doi.org/10.1016/b978-0-12-801238-3.99964-1.

25. W. Yu, C. Tang, B. P. Hobbs, et al., "Development and Validation of a Predictive Radiomics Model for Clinical Outcomes in Stage I Non-Small Cell Lung Cancer," *International Journal of Radiation Oncology, Biology, Physics* 102, no. 4 (2018): 1090–1097, https://doi.org/10.1016/j.ijrobp.2017.10.046.

26. S. Pang, M. Field, J. Dowling, S. Vinod, L. Holloway, and A. Sowmya, "Training Radiomics-Based CNNS for Clinical Outcome Prediction: Challenges, Strategies and Findings," *Artificial Intelligence in Medicine* 123 (2022): 102230, https://doi.org/10.1016/j.artmed.2021.102230.

27. A. Stanzione, M. Gambardella, R. Cuocolo, A. Ponsiglione, V. Romeo, and M. Imbriaco, "Prostate MRI Radiomics: A Systematic Review and Radiomic Quality Score Assessment," *European Journal of Radiology* 129 (2020): 109095, https://doi.org/10.1016/j.ejrad.2020.109095.

28. G. Spadarella, G. Calareso, E. Garanzini, L. Ugga, A. Cuocolo, and R. Cuocolo, "MRI Based Radiomics in Nasopharyngeal Cancer: Systematic Review and Perspectives Using Radiomic Quality Score (RQS) Assessment," *European Journal of Radiology* 140 (2021): 109744, https://doi.org/10.1016/j.ejrad.2021.109744.

29. Y. Zheng, D. Geng, T. Yu, et al., "Prognostic Value of Pretreatment MRI Texture Features in Breast Cancer Brain Metastasis Treated With Gamma Knife Radiosurgery," *Acta Radiologica* 62, no. 9 (2021): 1208–1216, https://doi.org/10.1177/0284185120956296.

30. E. Chang, M. Z. Joel, H. Y. Chang, et al., "Comparison of Radiomic Feature Aggregation Methods for Patients With Multiple Tumors," *Scientific Reports* 11, no. 1 (2021): 9758, https://doi.org/10.1038/s41598-021-89114-6.

31. J. H. Park, B. S. Choi, J. H. Han, et al., "MRI Texture Analysis for the Prediction of Stereotactic Radiosurgery Outcomes in Brain Metastases From Lung Cancer," *Journal of Clinical Medicine* 10, no. 2 (2021): 237, https://doi.org/10.3390/jcm10020237.

32. E. Karami, M. Ruschin, H. Soliman, A. Sahgal, G. J. Stanisz, and A. Sadeghi-Naini, "An MR Radiomics Framework for Predicting the Outcome of Stereotactic Radiation Therapy in Brain Metastasis," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2019), 1022–1025, https://doi.org/10.1109/embc.2019.8856558.

33. H. Wang, J. Xue, T. Qu, et al., "Predicting Local Failure of Brain Metastases After Stereotactic Radiosurgery With Radiomics on Planning MR Images and Dose Maps," *Medical Physics* 48, no. 9 (2021): 5522–5530, https://doi.org/10.1002/mp.15110.

34. D. Kawahara, A. Jensen, J. Yuan, Y. Nagata, and Y. Watanabe, "Predicting the BRAF Mutation With Pretreatment MRI Radiomics Features for Melanoma Brain Metastases Receiving Gamma Knife Radiosurgery," *Clinical Radiology* 78, no. 12 (2023): e934–e940, https://doi.org/10.1016/j.crad.2023.08.012.

35. P. Du, X. Liu, R. Xiang, et al., "Development and Validation of a Radiomics-Based Prediction Pipeline for the Response to Stereotactic Radiosurgery Therapy in Brain Metastases," *European Radiology* 33, no. 12 (2023): 8925–8935, https://doi.org/10.1007/s00330-023-09930-4.

36. S. Ali Jalalifar, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, "Predicting the Outcome of Radiotherapy in Brain Metastasis by Integrating the Clinical and MRI-Based Deep Learning Features," *Medical Physics* 49, no. 11 (2022): 7167–7178, https://doi.org/10.1002/mp.15814.

37. G. Carloni, C. Garibaldi, G. Marvaso, et al., "Brain Metastases From NSCLC Treated With Stereotactic Radiotherapy: Prediction Mismatch Between Two Different Radiomic Platforms," *Radiotherapy & Oncology* 178 (2023): 109424, https://doi.org/10.1016/j.radonc.2022.11.013.

38. C.-Y. Liao, C.-C. Lee, H.-C. Yang, et al., "Predicting Survival After Radiosurgery in Patients With Lung Cancer Brain Metastases Using Deep Learning of Radiomics and EGFR Status," *Physical and Engineering Sciences in Medicine* 46, no. 2 (2023): 585–596, https://doi.org/10.1007/s13246-023-01234-7.

39. P. Du, X. Liu, L. Shen, et al., "Prediction of Treatment Response in Patients With Brain Metastasis Receiving Stereotactic Radiosurgery Based on Pre-Treatment Multimodal MRI Radiomics and Clinical Risk Factors: A Machine Learning Model," *Frontiers in Oncology* 13 (2023): 1114194, https://doi.org/10.3389/fonc.2023.1114194.

40. P. Lambin, R. T. H. Leijenaar, T. M. Deist, et al., "Radiomics: The Bridge Between Medical Imaging and Personalized Medicine," *Nature Reviews Clinical Oncology* 14, no. 12 (2017): 749–762, https://doi.org/10.1038/nrclinonc.2017.141.

41. R. W. Y. Granzier, T. J. A. van Nijnatten, H. C. Woodruff, M. L. Smidt, and M. B. I. Lobbes, "Exploring Breast Cancer Response Prediction to Neoadjuvant Systemic Therapy Using MRI-Based Radiomics: A Systematic Review," *European Journal of Radiology* 121 (2019): 108736, https://doi.org/10.1016/j.ejrad.2019.108736.

42. Q. Wang, C. Li, J. Zhang, et al., "Radiomics Models for Predicting Microvascular Invasion in Hepatocellular Carcinoma: A Systematic Review and Radiomics Quality Score Assessment," *Cancers* 13, no. 22 (2021): 5864, https://doi.org/10.3390/cancers13225864.

43. A. Ponsiglione, A. Stanzione, R. Cuocolo, et al., "Cardiac CT and MRI Radiomics: Systematic Review of the Literature and Radiomics Quality Score Assessment," *European Radiology* 32, no. 4 (2022): 2629–2638, https://doi.org/10.1007/s00330-021-08375-x.

44. T. Akinci D'Antonoli, A. U. Cavallo, F. Vernuccio, et al., "Reproducibility of Radiomics Quality Score: An Intra- and Inter-Rater Reliability Study," *European Radiology* 34, no. 4 (2024): 2791–2804, https://doi.org/10.1007/s00330-023-10217-x.

45. M. Ismail, M. A. M. Hanifa, E. I. M. Mahidin, H. A. Manan, and N. Yahya, "Cone Beam Computed Tomography (CBCT) and Megavoltage Computed Tomography (MVCT)-Based Radiomics in Head and Neck Cancers: A Systematic Review and Radiomics Quality Score Assessment," *Quantitative Imaging in Medicine and Surgery* 14, no. 9 (2024): 6963–6977, https://doi.org/10.21037/qims-24-334.

46. Q. Zhang, J. Cao, J. Zhang, et al., "Differentiation of Recurrence From Radiation Necrosis in Gliomas Based on the Radiomics of Combinational Features and Multimodality MRI Images," *Computational and Mathematical Methods in Medicine*, 2019 no. 1 (2019): 2893043, https://doi.org/10.1155/2019/2893043.

47. Y. Cui, X. Yang, Z. Shi, et al., "Radiomics Analysis of Multiparametric MRI for Prediction of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer," *European Radiology* 29, no. 3 (2019): 1211–1220, https://doi.org/10.1007/s00330-018-5683-9.

48. S. A. Jalalifar, H. Soliman, A. Sahgal, and A. Sadeghi-Naini, "Impact of Tumour Segmentation Accuracy on Efficacy of Quantitative MRI Biomarkers of Radiotherapy Outcome in Brain Metastasis," *Cancers* 14, no. 20 (2022): 5133, https://doi.org/10.3390/cancers14205133.

49. I. Fornacon-Wood, H. Mistry, C. J. Ackermann, et al., "Reliability and Prognostic Value of Radiomic Features Are Highly Dependent on Choice of Feature Extraction Platform," *European Radiology* 30, no. 11 (2020): 6241–6250, https://doi.org/10.1007/s00330-020-06957-9.

50. J. J. Foy, K. R. Robinson, H. Li, M. L. Giger, H. Al-Hallaq, and S. G. Armato III, "Variation in Algorithm Implementation Across Radiomics Software," *Journal of Medical Imaging* 5, no. 4 (2018): 044505, https://doi.org/10.1117/1.JMI.5.4.044505.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.