BMC Medical Imaging

SYSTEMATIC REVIEW Open Access

A Bayesian meta-analysis on MRI-based radiomics for predicting EGFR mutation in brain metastasis of lung cancer

Peyman Tabnak[1,2*], Zana Kargar[1], Mohammad Ebrahimnezhad[1] and Zanyar HajiEsmailPoor[1]

Abstract

Objectives This study aimed to investigate the diagnostic test accuracy of MRI-based radiomics studies for predicting EGFR mutation in brain metastasis originating from lung cancer.

Methods This meta-analysis, conducted following PRISMA guidelines, involved a systematic search in PubMed, Embase, and Web of Science up to November 3, 2024. Eligibility criteria followed the PICO framework, assessing population, intervention, comparison, and outcome. The RQS and QUADAS-2 tools were employed for quality assessment. A Bayesian model determined summary estimates, and statistical analysis was conducted using R and STATA software.

Results Eleven studies consisting of nine training and ten validation cohorts were included in the meta-analysis. In the training cohorts, MRI-based radiomics showed robust predictive performance for EGFR mutations in brain metastases, with an AUC of 0.90 (95% CI: 0.82–0.93), sensitivity of 0.84 (95% CI: 0.80–0.88), specificity of 0.86 (95% CI: 0.80–0.91), and a diagnostic odds ratio (DOR) of 34.17 (95% CI: 19.16–57.49). Validation cohorts confirmed strong performance, with an AUC of 0.91 (95% CI: 0.69–0.95), sensitivity of 0.79 (95% CI: 0.73–0.84), specificity of 0.88 (95% CI: 0.83–0.93), and a DOR of 31.33 (95% CI: 15.50–58.3). Subgroup analyses revealed notable trends: the T1C + T2WI sequences and 3.0 T scanners showed potential superiority, machine learning-based radiomics and manual segmentation exhibited higher diagnostic accuracy, and PyRadiomics emerged as the preferred feature extraction software.

Conclusion This meta-analysis suggests that MRI-based radiomics holds promise for the non-invasive prediction of EGFR mutations in brain metastases of lung cancer.

Keywords MRI, Radiomics, EGFR, Brain metastasis, NSCLC, Artificial intelligence, Machine learning, Deep learning

*Correspondence:
Peyman Tabnak
drtabnak@gmail.com; tabnakp@tbzmed.ac.ir
[1]Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Immunology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Introduction

In recent decades, lung cancer has stood out among the predominant cancer-related fatalities globally [1], comprising approximately 1.8 million deaths annually [2]. Lung cancer is the most common type of cancer that metastasizes to the brain, affecting approximately 7–10% of non-small cell lung cancer (NSCLC) patients at the time of diagnosis. Additionally, 20–40% of patients with NSCLC may develop brain metastases (BM) at later stages of the disease [3]. A substantial rise in the incidence of BMs has been observed in recent years, possibly due to prolonged patient survival. Current BM patients benefit from more efficient therapeutic methods and advanced imaging modalities, thereby improving BM detection compared to earlier methods [4]. Although there have been specific advances in sophisticated treatments and augmented survival rates, BM is still a notable cause of morbidity linked with progressive neurological deficits [5]. Analyzing tumor molecular subtypes utilizing gene expression profiling could offer deeper insight into their biology and pave the way for personalized BM therapy. In this regard, mutations in epidermal growth factor receptor (EGFR) are found in 10–60% of NSCLC cases and contribute to unfavorable survival outcomes [6]. Stimulation of EGFR through ligand binding triggers receptor tyros in kinase function, leading to cell proliferation and metastasis [7]. On the other hand, the emerging evidence indicates that EGFR-positive lung cancer patients suffering from brain metastasis experience enhanced survival compared to those lacking these mutations. This improvement is accompanied by elevated response to whole-brain radiotherapy and the development of novel chemotherapy drugs, EGFR tyrosine kinase inhibitors (TKI) [8–10]. Therefore, early and precise discrimination between patients with EGFR mutation and those with EGFR wild type is a crucial prerequisite for effective therapeutic decision-making [11]. The clinical assessment of EGFR mutations is contingent on obtaining biopsy tissue samples and blood specimens. Blood analyses encounter limitations, comprising substandard sample quality, high costs, and a heightened incidence of false positives [12]. Nevertheless, brain metastases are typically diminutive and can be dispersed throughout the brain, making it impractical and often unfeasible to perform invasive biopsies or surgical resections for molecular testing purposes. Consequently, the majority of metastatic brain lesions are detected through magnetic resonance (MR) imaging without confirming the presence of pathological tissue. Hence, it is prudent to explore non-invasive imaging-based techniques for detecting the mutation status of lung cancer-mediated brain metastasis patients [13, 14]. In this regard, radiomics has been the center of debates as a methodology for converting medical images into consistent and quantitative data to facilitate clinical decision support. Radiomics is a process of employing mathematical, machine learning, or deep learning algorithms to extract quantitative features from diagnostic images. These methods unveil the hidden characteristics of tumor images that are not discernable to the naked eye, offering the prediction of intended outcomes [15]. This process initiates with **Data Acquisition** and contributes to the gathering of digital format images, such as magnetic resonance (MR), ultrasound (US), and CT scans. Standard protocols are applied to provide reliability, reproducibility, and comparability in radiomics results. In the **Segmentation Step**, the region of interest (ROI) is delineated either manually, automatically, or semi-automatically. Despite the precision of manual segmentation, it is time-consuming and requires a large amount of data. Conversely, automatic segmentation lacks human input but can be prone to errors. Semiautomatic techniques integrate both strategies, requiring manual correction of automatically generated ROIs. Following **Feature Extraction**, machine learning or deep learning approaches are used to extract quantitative radiomics features from the ROI. Upon the **Modeling Process**, the selected radiomics features, either alone or combined with closely related genetic, biological, and clinical data, are recruited for model development. Eventually, the model's accuracy is internally and externally evaluated to ascertain its predictive efficacy beyond the primary population used for training, which is referred to as the **Validation Step** [16]. Radiomics has proven highly effective in assessing gene mutation status and devising personalized treatment approaches in oncology, relying on the analysis of medical imaging data [17]. Numerous prior studies have investigated the application of imaging-based radiomics for evaluating EGFR mutation status in lung cancer. The majority of these studies have concentrated on brain metastases [18, 19].

However, despite the promising potential of these methods, the diagnostic accuracy and overall performance of MRI-based radiomics and deep learning models remain unclear. Given the growing body of literature on this topic, we aimed to conduct a meta-analysis to evaluate the diagnostic performance of MRI-based radiomics and deep learning models for predicting EGFR mutation status in brain metastases from lung cancer. This meta-analysis is needed to consolidate the current evidence, identify potential factors influencing model performance, and provide insights into the clinical applicability of these approaches. In this study, we will first outline the systematic methodology employed in the study, including eligibility criteria, data extraction, and statistical analysis methods, in the Materials and Methods section. We will then present the diagnostic performance results of MRI-based radiomics for predicting EGFR mutations, along with heterogeneity and subgroup analyses, in the

Results section. In the Discussion, we will interpret these findings, compare them with existing literature, address the study's limitations, and propose future research directions. Finally, we will conclude by summarizing the study's contributions and its potential implications for clinical practice and future investigations.

Contributions of the study

1. This is the first meta-analysis to specifically focus on the diagnostic accuracy of MRI-based radiomics and deep learning models for EGFR prediction in lung cancer brain metastases, filling a critical research gap.
2. It highlights the potential of non-invasive radiomics approaches to complement or replace invasive biopsy techniques, offering a safer alternative for patients.
3. It identifies optimal radiomics methodologies, including segmentation techniques, imaging sequences, and modeling algorithms, providing insights for future clinical applications.

Materials and methods

This meta-analysis was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines [20]. An institutional review board (IRB) was not obtained because of the nature of the study. The protocol for this systematic review and meta-analysis was registered with the International Prospective Register of Systematic Reviews (PROSPERO) under the registration number CRD42024609642.

Search strategy

A systematic literature search was conducted through PubMed, Embase, and Web of Science to identify relevant studies from inception to October 14, 2023, and updated on November 3, 2024. The search was limited to studies in the English language, using the following keywords and relevant free text terms: ("Radiomics" OR "Machine learning" OR "Deep learning") AND "EGFR" AND "Brain Metastasis" AND "Lung Cancer" AND "MRI".

Eligibility criteria

The literature acquired through database retrieval was imported into Mendeley software. Following removing duplicate publications, a complete examination of titles and abstracts was undertaken to exclude literature that did not adhere to the specified inclusion criteria. Subsequently, the full texts of the remaining studies underwent careful scrutiny for their definitive inclusion. The PICO framework, encompassing population, intervention, comparison, and outcome, guided the identification of relevant studies with the following parameters: (1) Population: Individuals diagnosed with lung cancer and brain metastasis, with EGFR status assessment in primary tumors; (2) Intervention: Utilization of MRI-based radiomics; (3) Comparison: Evaluation of MRI radiomics for EGFR detection in contrast to pathological approaches; and (4) Outcome: Provision of comprehensive diagnostic test accuracy indicators, including sensitivity and specificity.

Inclusion criteria comprised (a) the use of radiomics to predict EGFR status in brain metastases of lung cancer patients, (b) all participants possessing documented pathological EGFR status, and (c) the availability of sufficient data for constructing a 2×2 contingency table containing true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Exclusion criteria included (a) review papers, case reports, meetings, letters, abstracts, editorials, comments, posters, and guidelines; (b) studies not employing radiomics for EGFR prediction; (c) literature published in languages other than English, (d) and studies with cohort overlap.

Data extraction

Two authors independently carried out the extraction of data and the assessment of study quality. The fundamental data extracted encompassed details such as the first author's name and year of publication, the country where the study was conducted, study design, number of centers, type of primary lung cancer (if provided), number of lesions/patients with brain metastases, utilization of separate validation cohorts for model validation, scanner manufacturer, magnetic field strength (Tesla), evaluated MRI sequence, type of radiomics approach (deep learning vs. machine learning), and the integration of clinical factors with radiomic features for model construction. Additionally, comprehensive information was extracted, covering aspects such as ROI dimension, ROI segmentation software, feature extraction software, number of extracted/selected features, radiomics feature types, interclass correlation coefficient assessment of features, feature reduction algorithm, and model construction algorithm.

Quality assessment

For the evaluation of included articles, a rigorous assessment of methodological quality was undertaken using two specialized tools: the Radiomics Quality Score (RQS) and the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). The RQS, a 16-item scoring system designed for radiomics studies, was employed to gauge the methodological robustness and reporting completeness of the included studies, with assessments conducted independently by two authors to ensure reliability [21]. Concurrently, a modified version of the QUADAS-2 tool, specifically tailored for diagnostic accuracy radiomics studies, was utilized to assess the risk of bias

and concerns about applicability in key domains such as patient selection, index test, reference standard, and flow and timing. This dual-tool approach provided a comprehensive evaluation covering both radiomics and diagnostic accuracy aspects [22, 23]. To enhance reliability, the assessments using both tools were conducted independently by two authors, with any discrepancies resolved through consensus discussions. The following questions were used in the modified version of QUADAS-2:

Patient selection domain

- Were the inclusion/exclusion criteria specified?
- Was the type of study specified (retrospective vs. prospective)?
- Were the patients' characteristics specified?

Index test

- Were imaging acquisition protocols and segmentation methods detailed?
- Was the image processing approach detailed?
- Was a validation technique used?

Reference standard

- Was the reference standard likely to classify the target condition correctly?
- Was a biopsy taken from brain metastasis?

Flow and timing

- Did all patients receive the same reference standard?

Statistical analysis

The summary estimates for sensitivity (SENS) and specificity (SPEC), combining these metrics along with the diagnostic odds ratio (DOR) across all included studies with 95% confidence intervals (CIs) were illustrated using forest plots. A summary receiver operating characteristic curve was constructed, and the meta-analysis was conducted using the "meta4diag" and "INLA" packages in the R language, as well as the "midas" module in STATA version 14.2 [24, 25]. Summary receiver operating characteristic (SROC) curves were constructed in meta4diag package. In Bayesian statistics, uncertainty is often characterized by probability distributions. In this context, the uncertainty in the estimated area under the curve (AUC) is captured through the Bayesian framework by sampling from the posterior distributions of sensitivity and specificity conditional on the given data. The process involves creating multiple groups of sensitivity and specificity values by drawing samples from the posterior distributions. Each set of sampled values leads to the generation of a different SROC curve and, consequently, a different estimated AUC. By repeating this process multiple times (1000 bootstrap), a distribution of estimated AUC values is obtained. From this distribution, statistical summaries such as the mean, standard deviation, and quantiles can be calculated to provide a more comprehensive characterization of the uncertainty associated with the estimated AUC. Diagnostic accuracy performance was generally assessed using the AUC, categorized into low (0.5–0.7), moderate (0.7–0.8), good (0.8–0.9), and excellent (0.9-1.0) discriminatory power levels. Heterogeneity among studies was estimated through Cochran's Q test and Higgins' I^2 statistic, with I^2 values above 50% indicating significant heterogeneity. The threshold effect was evaluated using Spearman's correlation coefficient between sensitivity and false-positive rate in MetaDiSc software. Subgroup analyses comparing pooled AUCs, sensitivity, specificity, and DOR among different subgroups were conducted using the "meta4diag" package. Publication bias was explored through a Deeks funnel plot, and significance was determined using the Deeks asymmetry test, with a threshold set at p-values below 0.05 for significance.

Results

Literature search

According to Fig. 1, illustrating the study selection process, a total of 171 studies were identified through the literature search, among which 79 records were duplicates. Consequently, 92 remaining studies underwent screening based on their titles and abstracts to determine their relevance to the research question. At this stage, 73 studies were excluded due to a lack of relevance regarding the titles/abstracts. Therefore, 19 studies were considered eligible for an in-depth full-text review. However, seven of these studies did not predict EGFR mutation or predicted T790M mutation across EGFR-positive patients. Consequently, 11 eligible studies for the meta-analysis were detected [18, 26–35].

Characteristics of the included studies

Table 1 shows the basic characteristics of the included studies. Eleven studies containing ten validation and nine training cohorts, with available data for extraction, were included in the meta-analysis, with a total number of 1634 patients. Six studies had separate validation cohorts [18, 26, 28–30, 32–34], and among them, two had two cohorts (internal and external) [28, 34], and one did not provide data for calculation 2×2 table in its training cohort [26, 33]. All studies were retrospectively designed. Six studies were conducted in China [18, 28–30, 32, 34], and the rest of the studies were in South Korea [26, 27], India [31], Israel [33], and the United States [35]. Eight studies were single-center [18, 26, 27, 30–32, 35], and two

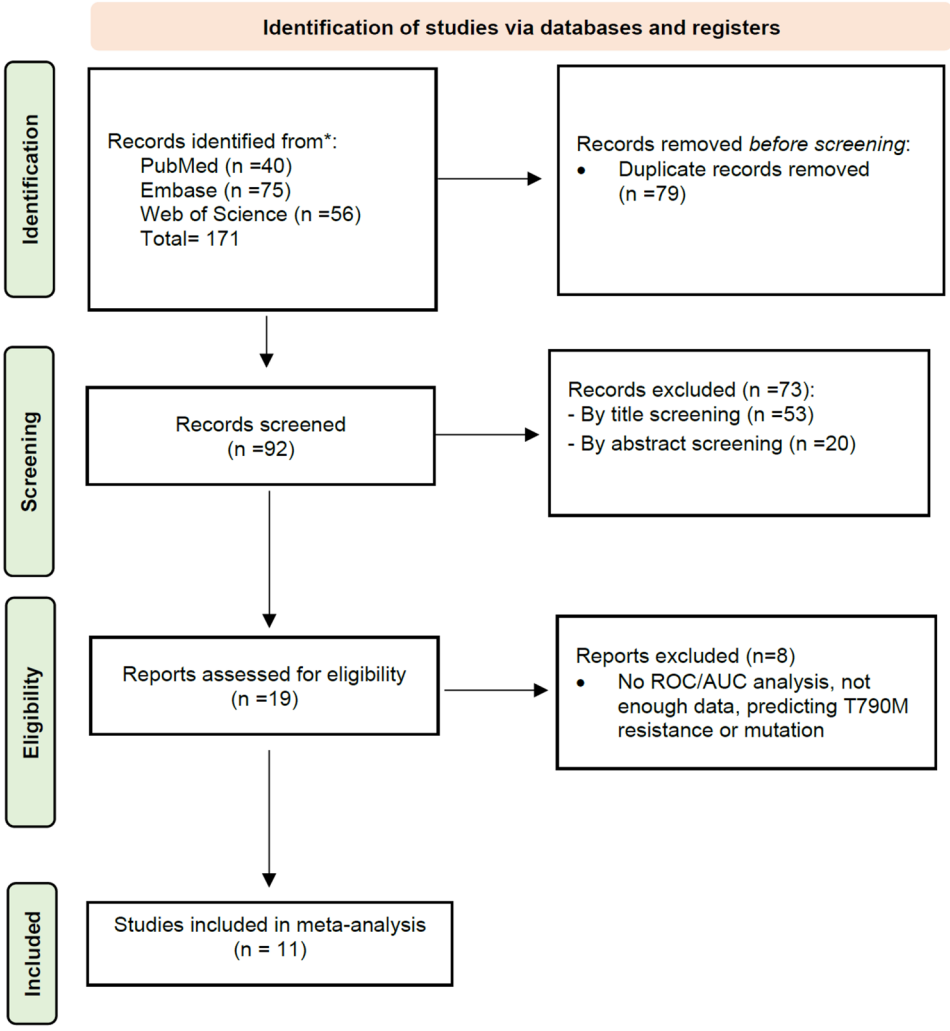

Fig. 1 PRISMA flowchart of the study

were multicenter [28, 29, 33, 34]. Scanner manufacturers were Philips [18, 26], GE [28], Siemens [28, 29, 32, 34, 35], a combination of GE/Philips [30], and not mentioned in two studies [27, 33]. Magnetic field strength was 3.0 T [18, 26, 28, 29, 34, 35], 1.5T [31], and a combination of both [30, 32] and not mentioned in two studies [27, 33]. Three studies combined clinical factors with radiomics signatures [30, 31, 35]. Two studies used deep learning-based radiomics methods [29, 31, 33], and the rest used conventional machine learning radiomics methods [18, 26–28, 30, 32, 34, 35]. Different MRI sequences were used to extract radiomics features, and T1-CE was chosen frequently [18, 27, 29, 31–35]. Table 2 shows the detailed characteristics of the radiomics models. ROI was delineated manually in nine studies, and two study used semi-automatic ROI delineation [26, 35]. Three studies used a 3D ROI structure [23, 27, 32], while most ROI structures were 2D [18, 24–26, 28–31]. ITK-SNAP was the most frequently used software for ROI segmentation in seven studies [18, 28, 29, 31, 32, 34, 35], followed by 3D Slicer [26, 30] and AnalyzeDirect [33]. Likewise, PyRadiomics was the most frequently applied software for extracting radiomics features [18, 26, 28, 29, 32, 34, 35]. The extracted features ranged from 3,934 to 107 across the studies. Various algorithms were used for feature reduction across the studies, and LASSO was used in one-third of the included studies [18, 28, 29, 32, 34].

For the modeling algorithm, logistic regression (LR) was adopted frequently [18, 27, 28, 30, 32, 34], followed by random forest (RF) [26, 27, 35] and convolutional neural network (CNN) [31, 33] models.

Quality assessment

QUADAS-2

The results of the modified QUADAS-2 assessment for the nine included studies are illustrated in Fig. 2. In the patient selection domain, a study [31] was deemed to have a high risk of bias due to including patients receiving

Table 1 Basic characteristics of the included studies

Author/Year	Country	Study design	Centers	Cancer Type	Number of lesions/patients	Validation cohort	Scanner manufacturer	Tesla	MRI technique/ sequence	Radiomics Type	Combined
Ahn et al. 2020 [27]	South Korea	Retro	1	LUAD, SCLC	61P	No	NM	NM	T1-CE	ML	No
Chen et al.2020 [35]	USA	Retro	1	LUAD	110P	No	Siemens	3T	T1-CE and FLAIR	ML	Yes
Wang et al. 2021 [18]	China	Retro	1	LUAD	52P	Yes	Philips	3T	T1-CE, **T2-FLAIR**, T2WI, and DWI	ML	No
Fan et al. 2022 [34]	China	Retro	2	NSCLC	310P	Yes	Siemens	3T	T1-CE and T2WI	ML	No
Haim et al. 2022 [33]	Israel	Retro	2	NSCLC	293P	Yes	NM	NM	T1-CE	DL	No
Zheng et al. 2022 [32]	China	Retro	1	-	162P	Yes	Siemens	3T/1.5T	T1CE, T2WI, T2 FLAIR	ML	No
Mahajan et al. 2023 [31]	India	Retro	1	NSCLC	117P	No	GE	1.5T	T1WI, T2WI, T1-CE, and FLAIR	DL	Yes
Huang et al. 2023 [30]	China	Retro	1	LUAD	58P	Yes	GE and Philips	3T/1.5T	FLAIR, T1W, T2W.	ML	Yes
Cao et al. 2024 [29]	China	Retro	2	NSCLC	232P	Yes	Siemens	3.0 T	T1-C and T2-W MRI	DL	No
Cao et al. 2022 [28]	China	Retro	2	NSCLC	188P	Yes	Siemens	3.0 T	T1CE and T2W	ML	No
Park et al. 2021 [26]	South Korea	Retro	1	NSCLC	51P	Yes	Philips	3.0 T	T1C and DTI	ML	No

Abbreviations: T1-CE, contrast-enhanced T1-weighted; FLAIR, fluid-attenuated inversion recovery; T2WI, T2-weighted imaging, DWI, diffusion-weighted imaging; ML, machine learning; DL, deep learning

systemic chemotherapy, and an unclear risk of bias was considered for one study [18] due to not mentioning the exact exclusion and inclusion criteria. In the index test domain, three studies were considered to have a high risk of bias due to not using any validation technique [31] or poor image protocol quality [27, 28]. For the flow and timing domain, one study [31] was deemed to have an unclear risk of bias due to uncertainty in receiving the same reference standard across the participants, and one considered to have a high risk of bias due to using biopsy or surgery, different per patients [28]. More than half of the studies (5/11) [18, 28, 30, 34, 35] were deemed to have a high risk of bias in the reference standard section since the EGFR mutation assessment was performed on primary lung lesions rather than brain metastasis. In addition, there were unclear risks of bias in the three studies as the source of biopsy for EGFR evaluation was not mentioned [27, 28, 31]. However, no high applicability concern was detected in almost all studies, indicating that the included articles matched the review questions.

RQS score

Table 3 presents individual and overall RQS scores for the included studies. The average RQS score for the nine studies was 10.27 (28.5%, ranging from 8.3 to 41.6%), with one study scoring below 10%. About three-quarters of the studies (8/11) achieved scores between 11 and 13 points, corresponding to 30–36% of the total possible points. None of the studies used phantom study, imaging at multiple time points, prospective design, decision curve analysis (potential clinical application), and cost-effectiveness analysis. In contrast, biological correlation, feature reduction, discrimination statistics, and comparison to the gold standard were performed in all studies. Imaging protocol quality was satisfying in more than 63% of the studies (7/11) and poor in four studies [27–29, 31]. Multiple segmentations (by different radiologists/software) were performed in about three-quarters of the studies (8/11) [18, 26, 30–35]. Multivariable (combined model) was conducted in four studies [28, 30, 31, 35]. Cut-off analysis was only provided in one study [30].

Meta-analysis

Diagnostic accuracy test

In the training cohorts, the diagnostic indicators, AUC (0.90 [0.82–0.93]), SENS (0.84 [0.80–0.88]), SPEC (0.86 [0.81–0.90]), PLR (6.14 [4.1–9.34]), NLR (0.19 [0.13–0.24]), and DOR (34.17 [19.16–57.49]), were pooled, and

Table 2 Detailed characteristics of the included studies

Study	ROI Dimension	ROI Segmentation	ROI segmentation Software	Feature extraction Software	Selected features(*n*)/ extracted (*n*)	Radiomics feature types	ICC	Feature reduction algorithm	Modeling algorithm
Ahn et al. 2020 [27]	2D	Manually	NM	MATLAB	22 / 1209	First-order, second-order, higher-order	-	RF, norm minimization, concave minimization, MRMR, Relief, Laplacian	**RF**, SVM, AdaBoost, LR
Chen et al.2020 [35]	3D	Semi-automated	ITK-SNAP	Pyradiomics	50 / 2520	Intensity-based, shape-based, textural features	ICC > 0.8	ICC > 0.8, MRMR, RF	RF
Wang et al. 2021 [18]	2D	Manually	ITK-SNAP	Pyradiomics	9 / 438	First-order, texture features, shape, wavelet	-	LASSO	LR
Fan et al. 2022 [34]	2D	Manually	ITK-SNAP	Pyradiomics	8 / 1967	First-order, shape, texture features	ICC > 0.85	Mann–Whitney U test ICC > 0.85, LASSO	LR
Haim et al. 2022 [33]	2D	Manually	AnalyzeDirect	Fast.ai framework	-	-	-	ResNet-50	CNN
Zheng et al. 2022 [32]	2D	Manually	ITK-SNAP	Pyradiomics	26 / 1470	First-order, shape, texture features	ICC > 0.7	Univariate analysis, LASSO	LR
Huang et al. 2023 [30]	3D	Manually	3D slicer	FeAture Explorer software	9/107	First order, intensity, shape, and textural features	ICC of > 0.75	RFE and LR	LR
Mahajan et al. 2023 [31]	2D	Manually	ITK snap	Tensorflow	-	Semantic features	-	Deep learning methods	CNN
Cao et al. 2024 [29]	2D	Manual	ITK-Snap	PyRadiomics	1976 extracted	First-order, shape-based, and texture features	ICC > 0.85	LASSO	MSF-Net
Cao et al. 2022 [28]	2D	Manual	ITK-Snap	PyRadiomics	11/3,934	Shape-based, first-order, texture features	ICC > 0.85	Mann-Whitney U test, LASSO regression, logistic regression with AIC	LR
Park et al. 2021 [26]	3D	Semi-automatic	3D Slicer	PyRadiomics	5/526	First-order, shape-based, and second-order texture features	ICC > 0.75	Tree-based feature selection	SVM, RF, LDA

Abbreviations: NM, not mentioned, RF, random forest; SVM, support vector machine; LR, logistic regression; MRMR, minimum redundancy maximum relevance; LASSO, least absolute shrinkage and selection operator; RFE, recursive feature elimination; CNN, convolutional neural network; CV, cross-validation

their respective 95% confidence intervals were determined. In the validation cohorts, the values for diagnostic indicators were as follows: AUC (0.91 [0.84–0.94]), SENS (0.79 [0.73–0.84]), SPEC (0.88 [0.83–0.93]), PLR (7.15 [4.39–11.23]), NLR (0.24 [0.17–0.31]), and DOR (31.33 [15.50–58.3]). The coupled forest plot showing sensitivity and specificity in training and validation cohorts is shown in Fig. 3. The forest plot showing DOR in training and validation groups is shown in Fig. 4. In addition, the SROC shows the estimated AUC with summary points, credible regions, and prediction regions in training and validation cohorts in Fig. 5.

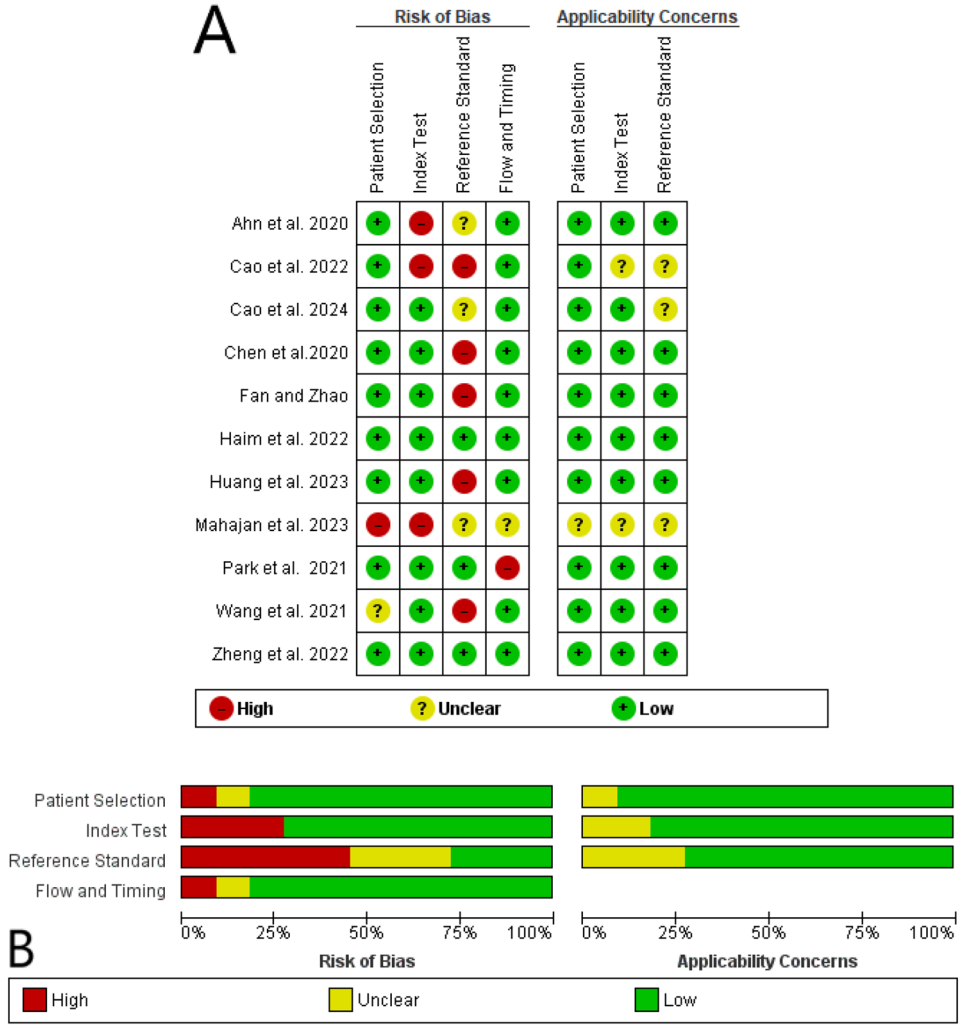

Fig. 2 Risk of bias assessment (left) and applicability concerns (right) in each study (**A**) and overall (**B**)

Heterogeneity test

In the training cohorts, considerable heterogeneities were observed in sensitivity ($I^2 = 79.83\%$) and specificity ($I^2 = 77.92\%$) values, as evidenced by p-value < 0.05 for Cochran's Q test. However, in the validation cohorts, moderate heterogeneity was observed in the pooled specificity value ($I^2 = 33.45\%$, p-value = 0.14), while Higgins' I^2 was near 0 for the pooled sensitivity (p-value = 0.84). Spearman's correlation coefficient did not show a significant threshold effect in both training (p-value = 0.3) and validation (p-value = 0.4) cohorts.

Subgroup analysis

Subgroup analysis was performed based on different cofactors to compare their diagnostic performance. The diagnostic indicators, including SENS, SPEC, and AUCs for each subgroup, are shown in Table 4. Herein, we discuss the differences in each subgroup accordingly:

Regional differences

In training models, six cohorts were investigated in China, whose pooled AUC was higher than other countries (AUC = 0.91) (Table 4). In validation cohorts, eight cohorts belonged to China (AUC = 0.83), one to Israel (AUC = 0.91) and one to South Korea (AUC = 0.91).

Comparison of MRI sequences

Across the training cohorts, different MRI sequences were used, and their comparison can be difficult due to the small number of cohorts in each subgroup. Based on our Bayesian subgroup analysis, we found that using features from the T1C sequence combined with T2W *might* outperform other sequences in terms of AUC (0.99). These results were also derived from validation cohorts, indicating that the T1C + T2W sequences might have a higher diagnostic performance compared to other sequences (AUC = 0.95). However, drawing this conclusion needs further investigation.

Table 3 Radiomics quality score (RQS) assessment based on 16 questions

Study	Radiomics Score																
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Score
Ahn	0	0	0	0	3	0	1	0	2	0	0	-5	2	0	0	0	3
Chen	1	1	0	0	3	1	1	0	1	0	0	-5	2	0	0	0	5
Wang	1	1	0	0	3	0	1	0	2	0	0	2	2	0	0	1	13
Fan	1	1	0	0	3	0	1	0	2	0	0	3	2	0	0	0	13
Haim	1	0	0	0	3	0	1	0	1	0	0	3	2	0	0	0	11
Zheng	1	1	0	0	3	0	1	0	1	0	0	2	2	0	0	1	12
Mahajan	0	0	0	0	3	1	1	0	2	0	0	-5	2	0	0	0	4
Huang	1	1	0	0	3	1	1	1	2	1	0	2	2	0	0	0	15
Cao 2024	0	1	0	0	3	0	1	0	2	0	0	3	2	0	0	0	12
Cao 2022	0	1	0	0	3	1	1	0	2	0	0	3	2	0	0	0	13
Park	1	1	0	0	3	0	1	0	2	0	0	2	2	0	0	0	12
Score	0.63	0.73	0	0	3	0.36	1	0.09	1.72	0.09	0	0.45	2	0	0	0	10.27

Questions: Q1: Image Protocol Quality, Q2: Multiple Segmentation, Q3: Phantom Study, Q4: Imaging at Multiple Points, Q5: Feature Reduction, Q6: Multivariable Analyses, Q7: Biological Correlation, Q8: Cut-off Analyses, Q9: Discrimination Statics, Q10: Calibration Statics, Q11: Prospective Study, Q12: Validation, Q13: Comparison to Gold Standard, Q14: Potential Clinical Application, Q15: Cost Effectiveness Analyses, Q16: Open science and Data

Scanner magnetic field strength

As expected in both training and validation cohorts, the diagnostic accuracy of 3.0 T scanners was higher (AUC = 0.90 in training) compared to 1.5T scanners or cohorts with both 3.0 T and 1.5 T scanners. In addition, the sensitivity and specificity of 3.0 T scanners were higher, ranging from 91 to 100% in training and 93–100% in validation cohorts.

ROI structure

The comparison between 2D and 3D ROI structures in the meta-analysis indicates distinct performance differences in training and validation cohorts. In the training cohort, 2D ROI achieved a slightly higher mean AUC (0.89 vs. 0.86) and sensitivity (0.86 vs. 0.82) compared to 3D ROI, while 3D ROI demonstrated marginally better specificity (0.87 vs. 0.85). In the validation cohort, 3D ROI outperformed 2D ROI in both AUC (0.85 vs. 0.80) and sensitivity (0.82 vs. 0.77) but had slightly lower specificity (0.86 vs. 0.90). Additionally, 2D ROI methods were used in more studies (6 in both training and validation) compared to 3D ROI (3 in training and 4 in validation), which could influence the generalizability of the results for the 3D approach.

ROI segmentation method

Only one study employed semiautomatic segmentation in the training cohorts, yielding a mean AUC of 0.79, which was lower than the 0.89 achieved by manual segmentation. Notably, the specificity for semiautomatic segmentation was slightly higher (0.90 vs. 0.85). However, because only one study used semiautomatic ROI segmentation ($n = 1$), further research is needed to validate these findings. In contrast, the pooled AUC for semiautomatic segmentation was higher than for manual segmentation (0.84 vs. 0.81). Overall, the limited number of studies on semiautomatic segmentation makes it impossible to draw a definitive conclusion at this point.

ROI segmentation software

Among different software used for ROI segmentation, ITK-SNAP was used most frequently in training ($n = 8$) and validation cohorts ($n = 7$) and also had a higher diagnostic accuracy compared to 3DSlicer in training (0.92 vs. 0.81) and validation (0.86 vs. 0.81) cohorts. In validation cohorts, a study used Analyze Direct software, and its AUC was slightly lower than ITK-SNAP (0.81 vs. 0.84).

Machine learning-based radiomics vs. deep learning radiomics

The comparison between ML-based and DL-based methods in the meta-analysis reveals that ML generally outperforms DL in terms of AUC and sensitivity across both training and validation phases. Specifically, ML achieved

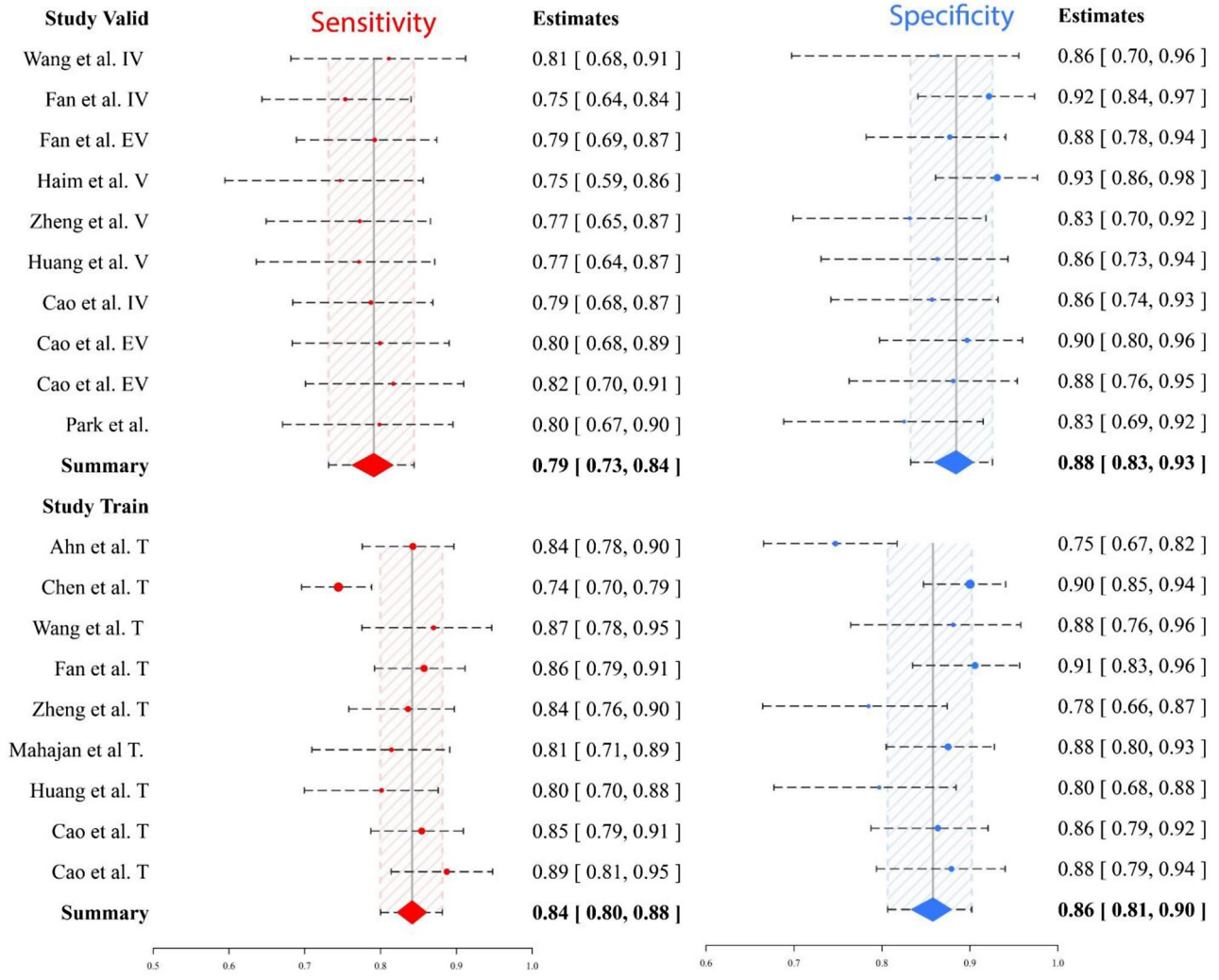

Fig. 3 Forest plots of the sensitivity and specificity in validation and training cohorts

a higher mean AUC (0.89 vs. 0.88 in training, 0.82 vs. 0.79 in validation) and sensitivity (0.86 vs. 0.83 in training, 0.80 vs. 0.75 in validation). However, DL methods demonstrated superior specificity, particularly in the validation phase (0.92 vs. 0.87). It is also notable that ML was used in a greater number of cohorts (7 in training and 8 in validation) compared to DL (2 in both phases), which may affect the robustness and generalizability of the findings for DL-based approaches.

Combined models vs. radiomics-only models

In the training cohorts, the radiomics-only category comprised five cohorts with a pooled AUC of 0.88. In contrast, the combined model subgroups (integrating radiomics with potential clinical factors) consisted of two cohorts, with an AUC of 0.78. Because the validation cohorts did not use combined models, a subgroup analysis was not performed; consequently, the pooled AUC for radiomics-only models was 0.85. While radiomics models demonstrated higher pooled sensitivity compared to combined models, their pooled specificity was lower.

Feature extraction software

PyRadiomics was the most frequently used software for feature extraction, and its diagnostic performance, based on mean AUCs, was higher than that of other software in training (0.93) and validation (0.84) cohorts.

Feature reduction algorithm

Among all feature-reduction approaches, LASSO was employed most frequently, yielding robust diagnostic performance in both the training (AUC = 0.92) and validation (AUC = 0.83) cohorts. By comparison, random forest (RF)–based methods were used in two training cohorts and showed high AUC values (0.95), but were not applied in any validation cohorts. Deep learning–based and recursive feature elimination (RFE) approaches had modest performance in the training cohorts (AUC = 0.84

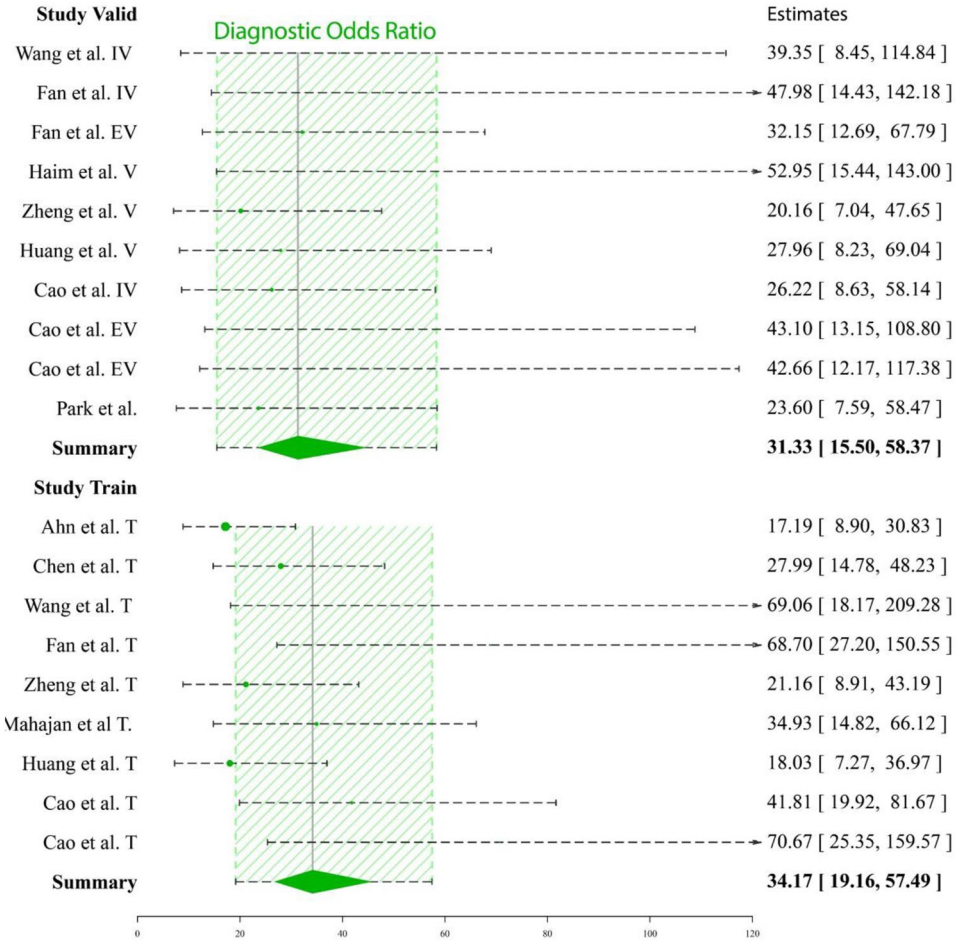

Fig. 4 Diagnostic odds ratio across training and validation cohorts

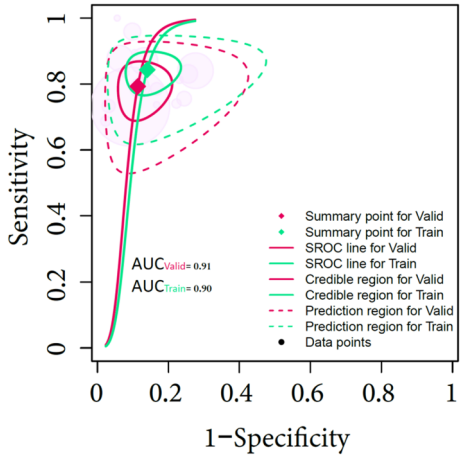

Fig. 5 SROC curves showing the estimated AUC with summary points, credible regions, and prediction regions in training and validation cohorts

and 0.79, respectively) and slightly lower or comparable results in validation cohorts.

Modeling algorithm

In the training cohorts, LR achieved the highest mean AUC (0.91), outperforming both random forest (RF, 0.83) and deep learning (DL, 0.88). In the validation cohorts, LR continued to show robust performance (AUC = 0.83), which was higher than the DL-based methods (AUC = 0.80) but lower than linear discriminant analysis (LDA, AUC = 0.85).

Publication bias

The analysis conducted using Deeks' asymmetry test did not reveal any substantial publication bias within the training (p-value of 0.33) and validation (p-value of 0.98) cohorts incorporated in the investigation, as illustrated in Fig. 6. This suggests that there is no notable skewness in the distribution of published studies, indicating a relatively unbiased representation of the available research on the topic. The absence of notable publication

Table 4 Subgroup analyses in training and validation cohorts

Subgroups		Training ($n=9$)				Validation ($n=10$)			
		N	AUC (mean)	SENS	SPEC	N	AUC (mean)	SENS	SPEC
Country	South Korea	1	0.86 (0.57–0.96)	0.82 (0.59–0.95)	0.71 (0.43–0.90)	1	0.85 (0.65–0.87)	0.81(0.61–0.94)	0.77 (0.55–0.91)
	USA	1	0.80 (0.44–0.94)	0.72 (0.43–0.91)	0.90 (0.73–0.97)	0	-	-	-
	China	6	0.91 (0.84–0.96)	0.88 (0.81–0.94)	0.87 (0.79–0.92)	8	0.83 (0.81, 0.92)	0.79 (0.74, 0.84)	0.88 (0.82, 0.92)
	India	1	0.84 (0.48–0.97)	0.77 (0.48–0.94)	0.87 (0.67–0.97)	0	-	-	-
	Israel	0	-	-	-	1	0.91 (0.77,1.00)	0.69 (0.46, 0.86)	0.98 (0.91, 1.00)
MRI Sequence	T1C	1	1.00 (0.40–0.99)	1.00 (0.98–1.00)	1.00 (1.00–1.00)	1	0.99 (0.58–0.98)	0.96 (0.87–0.99)	1.00 (1.00–1.00)
	T1C + FLAIR	1	0.83 (0.15–0.98)	0.70 (0.26–0.95)	0.90 (0.73–0.97)	0	-	-	-
	FLAIR	2	0.93 (0.66–0.99)	0.88 (0.66–0.98)	0.83 (0.67–0.95)	2	0.85 (0.70–0.95)	0.81 (0.65–0.91)	0.83 (0.65–0.94)
	T1C + T2W	3	0.99 (0.74–0.98)	0.97 (0.83–1.00)	0.96 (0.86–0.99)	5	0.95 (0.80–0.94)	0.92 (0.82–0.97)	0.97 (0.91–0.99)
	T1C + T2WI + FLAIR	1	0.86 (0.33–0.99)	0.79 (0.39–0.98)	0.73 (0.44–0.92)	1	0.78 (0.59–0.92)	0.74 (0.55–0.88)	0.77 (0.55–0.92)
	DTI + T1C	0	-	-	-	1	0.85 (0.62–0.96)	0.81 (0.60–0.94)	0.77 (0.54–0.91)
	T1C + T1W + T2W + FLAIR	1	0.87 (0.26–0.99)	0.75 (0.32–0.97)	0.87 (0.67–0.97)	0	-	-	-
Magnetic Field	3T	5	0.99 (0.81–0.96)	0.97 (0.91–0.99)	0.97 (0.94–0.98)	7	0.95 (0.82–0.93)	0.92 (0.86–0.96)	0.97 (0.93–0.99)
	3T/1.5T	2	0.85 (0.53–0.94)	0.79 (0.58–0.92)	0.75 (0.65–0.83)	2	0.78 (0.66–0.91)	0.74 (0.61–0.85)	0.80 (0.66–0.91)
	1.5T	1	0.91 (0.37–0.96)	0.77 (0.45–0.95)	0.88 (0.80–0.94)	0	-	-	-
Radiomics method	ML	7	0.89 (0.81–0.94)	0.86 (0.79–0.91)	0.85 (0.79–0.91)	8	0.82 (0.81–0.93)	0.80 (0.74–0.85)	0.87 (0.81–0.92)
	DL	2	0.88 (0.69–0.95)	0.83 (0.67–0.93)	0.87 (0.75–0.95)	2	0.79 (0.75–0.96)	0.75 (0.63–0.85)	0.92 (0.82–0.97)
Combined Radiomics	Only Radiomics	5	0.88 (0.85–0.95)	0.87 (0.83–0.90)	0.84 (0.78–0.90)	10	0.81 (0.83–0.93)	0.79 (0.74–0.83)	0.89 (0.83–0.93)
	Combined	2	0.78 (0.69–0.95)	0.75 (0.65–0.84)	0.89 (0.79–0.95)	0	-	-	-
ROI structure	2D	6	0.89 (0.82–0.94)	0.86 (0.80–0.92)	0.85 (0.78–0.91)	6	0.80 (0.81–0.94)	0.77 (0.70–0.83)	0.90 (0.83–0.94)
	3D	3	0.86 (0.75–0.94)	0.82 (0.71–0.91)	0.87 (0.76–0.94)	4	0.85 (0.81–0.95)	0.82 (0.74–0.89)	0.86 (0.76–0.93)
Segmentation	Manual	8	0.89 (0.84–0.94)	0.86 (0.82–0.91)	0.85 (0.79–0.90)	9	0.81 (0.83–0.93)	0.79 (0.73–0.83)	0.90 (0.85–0.94)
	Semiautomatic	1	0.79 (0.56–0.93)	0.72 (0.52–0.87)	0.90 (0.76–0.97)	1	0.84 (0.64–0.96)	0.81 (0.61–0.94)	0.76 (0.50–0.92)
ROI environment	ITK-SNAP	8	0.92 (0.82–0.94)	0.87 (0.80–0.92)	0.89 (0.85–0.92)	7	0.84 (0.80–0.94)	0.80 (0.74–0.85)	0.88 (0.83–0.93)
	3Dslicer	1	0.81 (0.42–0.94)	0.74 (0.46–0.92)	0.75 (0.56–0.89)	2	0.81 (0.65–0.93)	0.77 (0.63–0.88)	0.80 (0.65–0.90)
	Analyze Direct	0	-	-	-	1	0.81 (0.57–0.99)	0.69 (0.45–0.87)	0.98 (0.91–1.00)
Feature extraction software	MATLAB	1	0.87 (0.45–0.97)	0.82 (0.52–0.96)	0.72 (0.51–0.87)	0	-	-	-
	PyRadiomics	6	0.93 (0.81–0.96)	0.88 (0.80–0.94)	0.89 (0.84–0.93)	8	0.84 (0.78–0.94)	0.80 (0.74–0.85)	0.87 (0.81–0.92)
	TensorFlow	1	0.88 (0.38–0.97)	0.77 (0.42–0.96)	0.87 (0.73–0.96)	0	-	-	-
	Feature Explorer	1	0.81 (0.35–0.96)	0.73 (0.38–0.94)	0.75 (0.52–0.90)	1	0.79 (0.54–0.96)	0.74 (0.52–0.89)	0.83 (0.59–0.96)
	Fast.ai.framework	0	-	-	-	1	0.80 (0.59–1.00)	0.69 (0.45–0.87)	0.98 (0.91–1.00)

Table 4 (continued)

Subgroups		Training ($n=9$)				Validation ($n=10$)			
		N	AUC (mean)	SENS	SPEC	N	AUC (mean)	SENS	SPEC
Feature Reduction Methods	RF	2	0.95 (0.69–0.93)	0.91 (0.77–0.98)	0.92 (0.76–0.99)	0	-	-	-
	LASSO	5	0.92 (0.88–0.97)	0.90 (0.84–0.94)	0.88 (0.81–0.94)	4	0.83 (0.80–0.93)	0.80 (0.74–0.85)	0.88 (0.82–0.93)
	Tree Based	-	-	-	-	1	0.85 (0.63–0.96)	0.81 (0.60–0.94)	0.76 (0.52–0.92)
	DL	1	0.84 (0.59–0.96)	0.78 (0.55–0.93)	0.87 (0.67–0.97)	1	0.79 (0.63–0.99)	0.69 (0.45–0.87)	0.98 (0.91–1.00)
	RFE	1	0.79 (0.54–0.94)	0.75 (0.52–0.90)	0.74 (0.46–0.92)	1	0.78 (0.60–0.95)	0.74 (0.52–0.89)	0.83 (0.59–0.96)
Modeling Algorithm	RF	3	0.83 (0.61–0.94)	0.78 (0.61–0.90)	0.82 (0.67–0.93)	-	-	-	-
	LR	5	0.91 (0.84–0.96)	0.88 (0.81–0.94)	0.87 (0.78–0.93)	7	0.83 (0.79–0.95)	0.80 (0.73–0.85)	0.89 (0.82–0.94)
	DL	2	0.88 (0.66–0.96)	0.83 (0.67–0.93)	0.87 (0.73–0.95)	2	0.80 (0.71–0.98)	0.75 (0.62–0.86)	0.92 (0.81–0.98)
	LDA	-	-	-	-	1	0.85 (0.60–0.97)	0.81 (0.59–0.94)	0.76 (0.46–0.93)

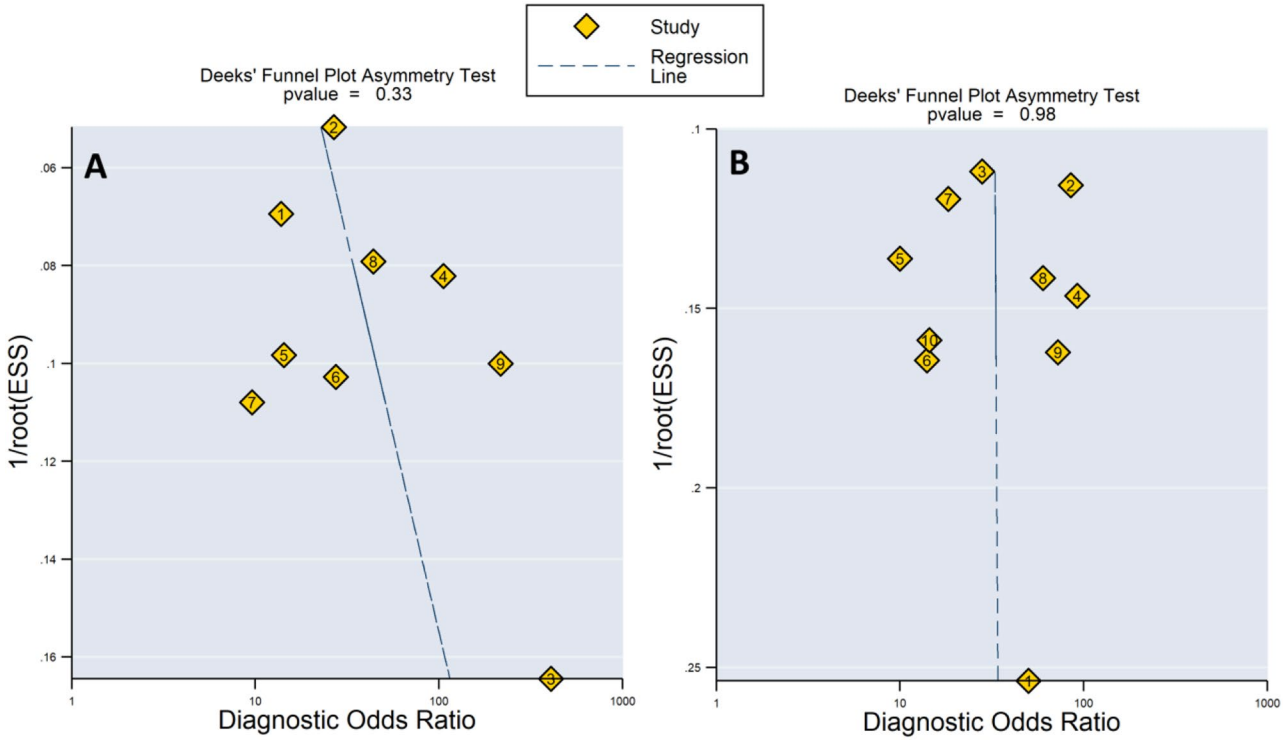

Fig. 6 -Deeks' funnel plot for testing publication bias in training (**A**) and validation (**B**) cohorts

bias bolsters the trustworthiness and rigor of the overall findings, reinforcing the credibility of the synthesized conclusions.

Discussion

Brain metastases are a frequent complication of lung cancer, with an incidence that varies depending on factors such as cancer type and stage. Lung cancer, recognized for its propensity to metastasize, is estimated to exhibit brain metastases in approximately 20% of patients at the time of diagnosis. Certain genetic mutations, such as EGFR and ALK, further increase the risk of brain involvement in NSCLC [37]. Predicting EGFR mutations in BMs of lung cancer is crucial for guiding treatment decisions and optimizing therapeutic outcomes. EGFR mutations influence the choice of targeted therapies, with certain EGFR TKIs, such as osimertinib, showing efficacy in addressing both systemic disease and brain lesions. The determination of EGFR mutation status provides valuable prognostic information, aiding in personalized treatment

planning. This information helps in avoiding unnecessary treatments with limited efficacy and potential side effects [38]. Biopsy procedures to determine EGFR mutations in brain metastases face challenges due to the dispersed and small nature of these lesions across the brain. Invasive techniques like biopsies or surgeries may not be practical, especially when dealing with numerous or hard-to-reach metastases. As a result, identifying brain metastases often relies on non-invasive imaging methods such as MRI, preceding the need for pathological confirmation. Recognizing these obstacles, there is an increasing need to develop non-invasive imaging-based approaches for assessing the mutation status of brain metastases in lung cancer patients. Non-invasive methods offer a more viable and less risky way to gather crucial genetic information, facilitating more effective and personalized treatment decisions for individuals with brain metastases [13, 35]. Thus, in this meta-analysis, we assessed the diagnostic performance of MRI-based radiomics methods as a novel approach for non-invasively predicting EGFR mutations in patients with BMs originating from lung cancers. The diagnostic performance of the radiomics models for predicting EGFR mutations in brain metastases of lung cancer is promising. The pooled Bayesian model demonstrated *excellent* discriminative ability with high AUC values in nine training (AUC: 0.90) and ten validation (AUC: 0.91) cohorts. Sensitivity and specificity were well-balanced, with SENS of 0.81 and SPEC of 0.86 in the training cohort and SENS of 0.74 and SPEC of 0.91 in the validation cohort. Nguyen et al. [39] have previously investigated the diagnostic performance of radiomics methods based on different imaging modalities in primary lesions of NSCLC, and the results showed a pooled AUC of 0.789, representing a moderate discriminatory power. However, it should be noted that most of the included studies in their meta-analysis were based on CT-scan images, which generally perform inferior to MRI for predicting biomarkers [40].

To ensure the quality of the articles included in this meta-analysis, two different tools, namely QUADAS-2 and RQS, were utilized for quality assessment. The RQS tool revealed a mean score of 10.2, indicating a low to moderate adherence to available guidelines for radiomics studies among the included studies. Conversely, while there were no applicability concerns for the studies included in the meta-analysis according to the QUADAS-2 tool, a major bias risk was identified in the reference standard domain. Many studies did not determine EGFR mutation through biopsy/surgery from the BMs, relying solely on histopathological examination of primary lesions. This introduces a significant risk of bias in the reference standard domain. Similar to other diagnostic test accuracy meta-analyses, this study demonstrated high heterogeneity in pooled sensitivity and specificity values for both types of cohorts. The heterogeneity was particularly pronounced for pooled specificity compared to pooled sensitivity, suggesting that the results across the studies were more consistent in terms of sensitivity. The presence of heterogeneity necessitates the conduct of subgroup analysis, which we performed accordingly. Most of the included articles in this meta-analysis were from China, which may introduce regional bias. As we observed in the subgroup analysis, the overall diagnostic performance of Chinese articles was higher compared to other countries. Radiomics researchers often combine clinical factors with radiomics features to create more robust and clinically relevant models. While radiomics provides quantitative insights from medical images, integrating clinical data enhances predictive accuracy by improving sensitivity and specificity. However, in contrast to previous meta-analyses [41, 42], we observed that the predicted performance of combined models is even lower than radiomics signatures. It appears that the combined model, incorporating both radiomics and potentially other factors, did not result in a noticeable improvement in diagnostic accuracy compared to using only radiomics. Hence, additional research is warranted to integrate clinical variables that exhibit a stronger correlation with EGFR mutation. In addition, some differences related to the MRI imaging technique and methods were observed: First, images derived from 3.0 T scanners performed much better than 1.5 T ones or those consisting of both. Second, we observed that the T1C sequence combined with the T2WI imaging sequence might perform better than other techniques. However, due to the small number of the included cohorts in each subgroup, drawing a definite conclusion is not possible and requires more original studies in the future. Likewise, ROI structure (2D vs. 3D), segmentation (semiauto vs. manual), and ROI delineation software also impact the final model's diagnostic performance. Our findings suggest that using manual segmentation, and specific ROI segmentation software (such as ITK-SNAP) might lead to better diagnostic performance for predicting EGFR mutation in brain metastasis. Lastly, LASSO and LR are widely used for feature selection and model building in radiomics studies, and their performance was better overall compared to other models. However, in previous meta-analyses, these models were shown to be inferior to modern modeling algorithms, such as SVM, for predicting Ki-67 expression in breast cancer by MRI-radiomics [22].

Limitations

The meta-analysis faces several limitations that warrant consideration in interpreting its findings. Firstly, the exclusion of eight studies due to potential cohort overlap, while essential for preserving data independence, introduces a risk of selection bias, as these excluded

studies may have offered unique perspectives. Secondly, the inclusion of a small number of studies ($n = 11$) raises concerns about statistical power and generalizability. Thirdly, some studies lack a separate validation cohort or missed data for the training cohort, potentially compromising the reliability of the reported predictive models. The limited ability to generalize subgroup analysis results due to small subgroup sizes and diverse categorizations further complicates the interpretation. Lastly, a notable limitation involves the risk of bias in the reference standard domain, as reliance on primary lung lesions for EGFR mutation detection may introduce diagnostic uncertainty, emphasizing the need for caution when extrapolating results to brain metastases directly. Transparent acknowledgment of these limitations is crucial for a nuanced understanding of the meta-analysis outcomes and informs directions for future research and clinical applications.

Conclusion

In summary, this meta-analysis investigated the diagnostic performance of MRI-based radiomics methods for predicting EGFR mutations in brain metastases originating from lung cancers. The results indicate a promising performance, with the pooled Bayesian model demonstrating good discriminative ability in both training and validation cohorts. Specifically, the AUC values for the ROC analysis were high, with an AUC of 0.90 in the training cohort and 0.91 in the validation cohort. These values, exceeding the threshold of excellence (0.9), highlights the effectiveness of MRI-based radiomics in distinguishing EGFR mutation status in BMs. The use of non-invasive imaging methods, particularly MRI-based radiomics, is emphasized as a valuable and less risky approach for obtaining crucial genetic information, facilitating more effective and personalized treatment decisions for individuals with brain metastases.

Despite some limitations and potential biases identified in the reference standard domain, this study highlights the need for continued research in the field of radiomics, particularly in refining MRI sequences, segmentation methods, and feature extraction techniques. The observed high diagnostic performance in studies from China, as well as the superior results from 3.0 T MRI scanners, further emphasize the importance of optimizing imaging protocols to enhance diagnostic accuracy.

Acknowledgements
The authors express their gratitude to Dr. Ramin Shahidi and Dr. Farzaneh Shojaeshafiei for their invaluable assistance during the revision process.

Author contributions
PT designed the search strategy. PT and ZK searched the database and selected the studies. ZK and PT extracted the data and evaluated the studies quality. PT did the analysis. PT wrote the manuscript and edited it with the assistance of ME and ZHE. All authors have read and approved the content and contributed to the study conception and design.

Funding
No fund was received for this study.

Data availability
The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Declarations

Ethical approval
Not applicable.

Informed consent
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 10 July 2024 / Accepted: 22 January 2025
Published online: 10 February 2025

References

1. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol. 2023;20:624–39.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
3. Ali A, Goffin JR, Arnold A, Ellis PM. Survival of patients with non-small-cell lung cancer after a diagnosis of brain metastases. Curr Oncol. 2013;20:300–6.
4. Derks SHAE, van der Veldt AAM, Smits M. Brain metastases: the role of clinical imaging. Br J Radiol. 2022;95:20210944.
5. Kancharla P, Ivanov A, Chan S, Ashamalla H, Huang RY, Yanagihara TK. The effect of brain metastasis location on clinical outcomes: a review of the literature. Neuro-Oncology Adv. 2019;1:vdz017.
6. Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). Am J Cancer Res. 2015;5:2892.
7. Koulouris A, Tsagkaris C, Corriero AC, Metro G, Mountzios G. Resistance to TKIs in EGFR-mutated non-small cell lung cancer: from mechanisms to new therapeutic strategies. Cancers (Basel). 2022;14:3337.
8. Jiang T, Min W, Li Y, Yue Z, Wu C, Zhou C. Radiotherapy plus EGFR TKIs in non-small cell lung cancer patients with brain metastases: an update meta-analysis. Cancer Med. 2016;5:1055–65.
9. Lee CK, Wu Y-L, Ding PN, Lord SJ, Inoue A, Zhou C, et al. Impact of specific epidermal growth factor receptor (EGFR) mutations and clinical characteristics on outcomes after treatment with EGFR tyrosine kinase inhibitors versus chemotherapy in EGFR-mutant lung cancer: a meta-analysis. J Clin Oncol. 2015;33:1958–65.
10. Fan Y, Huang Z, Fang L, Miu L, Lin N, Gong L et al. Chemotherapy and EGFR tyrosine kinase inhibitors for treatment of brain metastases from non-small-cell lung cancer: survival analysis in 210 patients. Onco Targets Ther 2013:1789–803.
11. Yoon H-Y, Ryu J-S, Sim YS, Kim D, Lee SY, Choi J, et al. Clinical significance of EGFR mutation types in lung adenocarcinoma: a multi-centre Korean study. PLoS ONE. 2020;15:e0228925.
12. Biaoxue R, Shuanying Y. Tissue or blood: which is more suitable for detection of EGFR mutations in non-small cell lung cancer? Int J Biol Markers. 2018;33:40–8.
13. Di Lorenzo R, Ahluwalia MS. Targeted therapy of brain metastases: latest evidence and clinical implications. Ther Adv Med Oncol. 2017;9:781–96.
14. Rubino S, Oliver DE, Tran ND, Vogelbaum MA, Forsyth PA, Yu H-HM, et al. Improving brain metastases outcomes through therapeutic synergy

between stereotactic radiosurgery and targeted cancer therapies. Front Oncol. 2022;12:854402.
15. McCague C, Ramlee S, Reinius M, Selby I, Hulse D, Piyatissa P, et al. Introduction to radiomics for a clinical audience. Clin Radiol. 2023;78:83–98. https://doi.org/10.1016/j.crad.2022.08.149.
16. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. J Nucl Med. 2020;61:488–95. https://doi.org/10.2967/jnumed.118.222893.
17. Qi Y, Zhao T, Han M. The application of radiomics in predicting gene mutations in cancer. Eur Radiol. 2022;32:4014–24.
18. Wang G, Wang B, Wang Z, Li W, Xiu J, Liu Z, et al. Radiomics signature of brain metastasis: prediction of EGFR mutation status. Eur Radiol 2021;31:4538–47. https://doi.org/10.1007/s00330-020-07614-x.
19. Chen BT, Jin T, Ye N, Mambetsariev I, Wang T, Wong CW, et al. Predicting Survival Duration With MRI Radiomics of Brain Metastases From Non-small Cell Lung Cancer. Front Oncol 2021;11. https://doi.org/10.3389/fonc.2021.621088.
20. Salameh J-P, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. Bmj 2020;370.
21. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749–62. https://doi.org/10.1038/nrclinonc.2017.141.
22. Tabnak P, HajiEsmailPoor Z, Baradaran B, Pashazadeh F, Maleki LA. MRI-Based Radiomics methods for Predicting Ki-67 expression in breast Cancer: a systematic review and Meta-analysis. Acad Radiol; 2023.
23. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36.
24. Guo J, Riebler A. meta4diag: bayesian bivariate meta-analysis of diagnostic test studies for routine practice. ArXiv Prepr ArXiv151206220 2015.
25. Shin JJ, Zurakowski D. Null hypotheses, interval estimation, and bayesian analysis. Otolaryngol Neck Surg. 2017;157:919–20.
26. Park YW, An C, Lee JS, Han K, Choi D, Ahn SS et al. Diffusion tensor and postcontrast T1-weighted imaging radiomics to differentiate the epidermal growth factor receptor mutation status of brain metastases from non-small cell lung cancer. Neuroradiology. 2021;63:343–52. https://doi.org/10.1007/s00234-020-02529-2
27. Ahn SJ, Kwon H, Yang J-J, Park M, Cha YJ, Suh SH, et al. Contrast-enhanced T1-weighted image radiomics of brain metastases may predict EGFR mutation status in primary lung cancer. Sci Rep 2020;10:8905. https://doi.org/10.1038/s41598-020-65470-7.
28. Cao R, Pang Z, Wang X, Du Z, Chen H, Liu J, et al. Radiomics evaluates the EGFR mutation status from the brain metastasis: a multi-center study. Phys Med Biol 2022;67.https://doi.org/10.1088/1361-6560/ac7192.
29. Cao R, Fu L, Huang B, Liu Y, Wang X, Liu J, et al. Brain metastasis magnetic resonance imaging-based deep learning for predicting epidermal growth factor receptor (EGFR) mutation and subtypes in metastatic non-small cell lung cancer. Quant Imaging Med Surg. 2024;14:4749.
30. Huang Z, Tu X, Yu T, Zhan Z, Lin Q, Huang X. Peritumoural MRI radiomics signature of brain metastases can predict epidermal growth factor receptor mutation status in lung adenocarcinoma. Clin Radiol 2023.
31. Mahajan A, B G, Wadhwa S, Agarwal U, Baid U, Talbar S, et al. Deep learning based automated epidermal growth factor receptor and anaplastic lymphoma kinase status prediction of brain metastasis in non-small cell lung cancer. Explor Target Anti-Tumor Ther. 2023;4:657–68. https://doi.org/10.37349/etat.2023.00158.
32. Zheng L, Xie H, Luo X, Yang Y, Zhang Y, Li Y, et al. Radiomic Signatures for Predicting EGFR Mutation Status in Lung Cancer Brain Metastases. Front Oncol 2022;12. https://doi.org/10.3389/fonc.2022.931812.
33. Haim O, Abramov S, Shofty B, Fanizzi C, DiMeco F, Avisdris N, et al. Predicting EGFR mutation status by a deep learning approach in patients with non-small cell lung cancer brain metastases. J Neurooncol 2022;157:63–9. https://doi.org/10.1007/s11060-022-03946-4.
34. Fan Y, Zhao Z, Wang X, Ai H, Yang C, Luo Y, et al. Radiomics for prediction of response to EGFR-TKI based on metastasis/brain parenchyma (M/BP)-interface. Radiol Med 2022;127:1342–54. https://doi.org/10.1007/s11547-022-01569-3.
35. Chen BT, Jin T, Ye N, Mambetsariev I, Daniel E, Wang T, et al. Radiomic prediction of mutation status based on MR imaging of lung cancer brain metastases. Magn Reson Imaging 2020;69:49–56. https://doi.org/10.1016/j.mri.2020.03.002.
37. Myall NJ, Yu H, Soltys SG, Wakelee HA, Pollom E. Management of brain metastases in lung cancer: evolving roles for radiation and systemic treatment in the era of targeted and immune therapies. Neuro-Oncology Adv. 2021;3:v52–62.
38. Kelly WJ, Shah NJ, Subramaniam DS. Management of brain metastases in epidermal growth factor receptor mutant non-small-cell lung cancer. Front Oncol. 2018;8:208.
39. Nguyen HS, Ho DKN, Nguyen NN, Tran HM, Tam K-W, Le NQK. Predicting EGFR Mutation Status in non–small cell Lung Cancer using Artificial Intelligence: a systematic review and Meta-analysis. Acad Radiol; 2023.
40. Qin Q, Deng LP, Chen J, Ye Z, Wu YY, Yuan Y et al. The value of MRI in predicting hepatocellular carcinoma with cytokeratin 19 expression: a systematic review and meta-analysis. Clin Radiol 2023.
41. HajiEsmailPoor Z, Tabnak P, Baradaran B, Pashazadeh F, Aghebati Maleki L. Diagnostic performance of CT-Scan based Radiomics for Prediction of Lymph Node Metastasis in Gastric Cancer: a systematic review and Meta-analysis. Front Oncol n d;13:1185663.
42. HajiEsmailPoor Z, Kargar Z, Tabnak P. Radiomics diagnostic performance in predicting lymph node metastasis of papillary thyroid carcinoma: a systematic review and meta-analysis. Eur J Radiol 2023:111129.

Publisher's note