Review article

MRI-Based Radiomics Methods for Predicting Ki-67 Expression in Breast Cancer: A Systematic Review and Meta-analysis

Peyman Tabnak, MD, Zanyar HajiEsmailPoor, MD, Behzad Baradaran, PhD, Fariba Pashazadeh, MS, Leili Aghebati Maleki, PhD

Rationale and Objectives: The purpose of this systematic review and meta-analysis was to assess the quality and diagnostic accuracy of MRI-based radiomics for predicting Ki-67 expression in breast cancer.

Materials and Methods: A systematic literature search was performed to find relevant studies published in different databases, including PubMed, Web of Science, and Embase up until March 10, 2023. All papers were independently evaluated for eligibility by two reviewers. Studies that matched research questions and provided sufficient data for quantitative synthesis were included in the systematic review and meta-analysis, respectively. The quality of the articles was assessed using Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) and Radiomics Quality Score (RQS) tools. The predictive value of MRI-based radiomics for Ki-67 antigen in patients with breast cancer was assessed using pooled sensitivity (SEN), specificity, and area under the curve (AUC). Meta-regression was performed to explore the cause of heterogeneity. Different covariates were used for subgroup analysis.

Results: 31 studies were included in the systematic review; among them, 21 reported sufficient data for meta-analysis. 20 training cohorts and five validation cohorts were pooled separately. The pooled sensitivity, specificity, and AUC of MRI-based radiomics for predicting Ki-67 expression in training cohorts were 0.80 [95% CI, 0.73–0.86], 0.82 [95% CI, 0.78–0.86], and 0.88 [95%CI, 0.85–0.91], respectively. The corresponding values for validation cohorts were 0.81 [95% CI, 0.72–0.87], 0.73 [95% CI, 0.62–0.82], and 0.84 [95%CI, 0.80–0.87], respectively. Based on QUADAS-2, some risks of bias were detected for reference standard and flow and timing domains. However, the quality of the included article was acceptable. The mean RQS score of the included articles was close to 6, corresponding to 16.6% of the maximum possible score. Significant heterogeneity was observed in pooled sensitivity and specificity of training cohorts ($I^2 > 75\%$). We found that using deep learning radiomic methods, magnetic field strength (3 T vs. 1.5 T), scanner manufacturer, region of interest structure (2D vs. 3D), route of tissue sampling, Ki-67 cut-off, logistic regression for model construction, and LASSO for feature reduction as well as PyRadiomics software for feature extraction had a great impact on heterogeneity according to our joint model analysis. Diagnostic performance in studies that used deep learning-based radiomics and multiple MRI sequences (e.g., DWI+DCE) was slightly higher. In addition, radiomic features derived from DWI sequences performed better than contrast-enhanced sequences in terms of specificity and sensitivity. No publication bias was found based on Deeks' funnel plot. Sensitivity analysis showed that eliminating every study one by one does not impact overall results.

Conclusion: This meta-analysis showed that MRI-based radiomics has a good diagnostic accuracy in differentiating breast cancer patients with high Ki-67 expression from low-expressing groups. However, the sensitivity and specificity of these methods still do not surpass 90%, restricting them from being used as a supplement to current pathological assessments (e.g., biopsy or surgery) to predict Ki-67 expression accurately.

Key Words: Radiomics; MRI; Artificial intelligence; Machine learning; Deep learning; Radiogenomics; Ki-67; Breast Cancer.

Acad Radiol 2024; 31:763–787

From the Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran (P.T., Z.H.); Immunology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran (P.T., Z.H., B.B., L.A.M.); Department of Immunology, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran (P.T., Z.H., B.B., L.A.M.); Research Center for Evidence-Based Medicine, Iranian Evidence-Based Medicine (EBM) Centre: A Joanna Briggs Institute (JBI) Centre of Excellence, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran (F.P.). Received September 7, 2023; revised October 1, 2023; accepted October 4, 2023. **Address correspondence to:** L.A.M e-mail: aghebatil@tbzmed.ac.ir, leili_aghebati_maleki@yahoo.com

https://doi.org/10.1016/j.acra.2023.10.010

INTRODUCTION

Breast cancer is the most common cancer in women and the fifth leading cause of cancer death, with more than 2 million new cases annually worldwide (1). Certain immunohistochemical (IHC) biomarkers, such as ER, PR, HER2, and Ki-67 have been used for the molecular classification of breast cancer for many years (2). Among them, the Ki-67 protein is a marker of cell proliferation and can be used to provide information about the proliferation of cancer cells, especially breast cancer (3). Ki-67 activity level is strongly correlated with cancer aggressiveness (4), complete pathological response to treatment (5), and prognosis of patients (6). Thus, it is considered an important factor in selecting treatment options and determining the need for follow-up in clinical settings (7). Preoperative analysis of Ki-67 expression using IHC requires tissue specimens that are typically obtained by needle biopsy. Alongside its invasiveness, this method has some disadvantages due to the heterogeneity of breast tumors and the small tissue sample obtained via this route compared to surgical excision (8). Therefore, investigating the applicability of less invasive techniques such as imaging technologies is necessary. The link between pathology and radiology has become much stronger with the rapid development of artificial intelligence (AI) in medical imaging (9). AI techniques, including machine learning (ML) and deep learning (DL), have shown promising potential in various medical applications, particularly medical imaging. These AI-driven approaches can automatically analyze complex patterns within images, offering a powerful means to extract valuable information from medical data (10). One such application is radiomics, an emerging field that involves the high-throughput extraction and analysis of quantitative features from medical images and converting these features into decision support tools. Likewise, radiogenomics is a developing branch of radiomics that involves computational methods and combines the fields of radiology and genomics. The term "radiogenomics" is a compound word created by merging "radiology" and "genomics" (11,12). Radiomics approach can utilize both hand-crafted texture analysis and ML/DL techniques to generate robust and reproducible features. Hand-crafted texture analysis involves manually extracting predefined features, such as shape, intensity, and texture, which are then used as input for ML algorithms. In contrast, DL methods, such as convolutional neural networks (CNNs), can automatically learn to identify relevant features directly from the image data without requiring manual feature extraction (13,14).

Magnetic resonance imaging (MRI) provides detailed images of soft tissues, allowing for more accurate identification of imaging features linked to molecular alterations (15) since it provides both functional and anatomical information on breast tissue, such as blood flow, tissue perfusion, and tissue composition, all of which can be related to underlying genetic and molecular dysregulation (16,17). In addition, breast MRI is also useful for predicting Ki-67 discordance between core-needle biopsy and surgical samples (18). Radiogenomic prediction of biomarkers has paved a long way to reach its current developed form (19). MRI-based radiomics, which combines the advantages of MRI with radiomics, is an attractive tool for predicting Ki-67 expression levels in solid tumors such as brain tumors (20), cholangiocarcinoma (20), soft tissue sarcoma (21), hepatocellular carcinoma (22), and bladder cancer (23). Previous meta-analyses have revealed that MRI-based radiomics approaches have an excellent diagnostic performance in predicting lymph node metastasis of breast cancer (24), pathological complete response to neoadjuvant chemotherapy (25), diagnosis of triple-negative breast cancer (26) as well as differentiation of breast cancer molecular subtypes (27). However, a systematic review and meta-analysis evaluating the diagnostic performance of MRI-based radiomics for predicting Ki-67 is still lacking. Thus, this study aims to systematically review MRI-based radiomics methods for predicting Ki-67 in breast cancer and provide a pooled result of their diagnostic accuracy.

MATERIALS AND METHODS

Statement of Design

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines were followed for conducting this systematic review (28) (Table S1). There was no overlapping systematic review related to the topic of this study in the Cochrane library.

Population, Intervention, Comparison, Outcomes (PICO)

- **Population**: Patients with breast cancer who had a preoperative MRI and subsequent histological and immunohistochemical evaluation of their Ki-67 expression following surgical resection or biopsy.
- **Intervention**: Preoperative MRI imaging of tumors was used in radiomics analysis to categorize tumors into groups based on high and low Ki-67 expression.
- **Comparison**: Assessing the discriminative ability of radiomics approaches compared to conventional molecular subtype testing (IHC).
- **Outcomes**: Evaluating the utility of radiomics to classify tumoral lesions into the correct molecular subgroups with a diagnostic test accuracy study design (e.g., providing receiver operating characteristic (ROC) curve analysis).

Search Strategy

Two independent authors (both experts in radiomics analysis for more than 2 years) electronically searched the PubMed, Web of Science, and EMBASE databases on March 10, 2023, for studies relating to our research questions. The following search terms were used: (Breast Cancer) AND

((Radiomics) OR (Machine Learning) OR (Deep Learning)) AND (Ki-67) AND (MRI). A detailed search strategy is presented in Tables S2–5 (Supplementary Materials). There was no restriction on the year of publication, and only research published in the English language was considered. All records were imported to Mendeley Reference Manager, and duplicates were removed. First, two reviewers screened the titles and abstracts of articles and removed irrelevant studies. The full text of each removed study was also reviewed by the other author. Then, a thorough full-text reading was performed by both authors to include studies relevant to our research questions. Disagreements were solved by reaching a consensus. An additional search was also conducted by another author (with more than ten years of experience in conducting systematic reviews) in Google Scholar as our supplementary database to identify missing publications.

Inclusion and Exclusion Criteria

The following inclusion criteria applied to study enrollment in qualitative synthesis: (1) studies involving patients diagnosed with breast cancer histopathologically and their Ki-67 expression status was determined by IHC staining, (2) studies that performed conventional breast MRI of tumors before surgery, (3) studies that aimed to assess diagnostic test accuracy of MRI-based radiomics for predicting Ki-67 at least by reporting values of area under the ROC curve (AUC). However, for quantitative synthesis (meta-analysis), only studies with sufficient data to reconstruct the 2 × 2 contingency table were included (at least by providing an ROC plot). The following exclusion criteria applied to study enrollment in qualitative synthesis: (1) reviews, (2) case reports, (3) animal studies, (4) comments, (5) editorials, (6) abstracts, (7) meeting reports, (8) book chapters, and (9) articles which their languages were not English. For studies with cohort overlap, the most recent study or the one with a larger sample size was included in the quantitative synthesis.

Data Extraction

Following general data were extracted from the included studies: (1) first author's name, (2) year of publication, (3) number of institutions, (4) number of lesions (for available Ki-67 status), (5) age of patients, (6) route of specimen sampling, (7) cut-off value for Ki-67, (8) scanner manufacturer, (9) scanner magnetic field, and (10) imaging sequences. Detailed extracted data from the included studies were as follows: (1) ROI structure (e.g., 3D vs. 2D), (2) ROI segmentation (e.g., automatic vs. manual), (3) number of extracted features, (4) number of selected features, (5) feature extraction software, (6) imaging features, (7) feature processing algorithm, (8) modeling algorithm, (9) type of cross-validation, and (10) use of independent validation. Following quantitative data were extracted for meta-analysis: (1) the number of positive and negative cases, (2) 2 × 2 tables with true positive, true negative, false positive, and false negative values. When comparing the diagnostic performance of various algorithms on the same sample, the algorithm that produced the best classification results was chosen. If a study only provided a ROC plot without the sensitivity (SEN) and specificity (SPEC) values in the text, we adopted the top-left method for extracting SEN and SPEC according to the ROC curve. If a study integrated MRI with another imaging modality (e.g., PET or ultrasound), the data for the MRI modality was included in the meta-analysis.

Methodological Quality Assessment

QUADAS-2

Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool for quality assessment of all studies was used with *modified* signaling questions by two independent reviewers (28). The risk of bias in each section was assessed using the following questions in the RevMan software, version 5.3.

Patient selection questions:

Were the inclusion/exclusion criteria specified?

Was the type of study specified (retrospective vs. prospective)?

Were the patients' characteristics specified?

a) **Index test questions:**

Where imaging acquisition protocols and segmentation methods detailed?

Was the image processing approach detailed?

Was a validation technique used?

b) **Reference standard question:**

Is the reference standard likely to correctly classify the target condition?

c) **Flow and timing question:**

Did all patients receive the same reference standard?

Was biopsy performed after MRI acquisition?

RQS

RQS (Radiomic Quality Score) was employed to ensure the rigor and reproducibility of radiomic studies. This metric comprises a set of criteria a radiomic study must meet to be considered high quality.

The RQS is comprised of 16 components in the radiomics workflow **(**Table S5**)**. A total score ranging from − 8–36 points is produced by rating each of the 16 components of the score, with − 8–0 points denoting 0% and 36 points representing 100% of total points (29,30). Two reviewers independently assessed the RQS of the articles, and any disagreement was solved by discussion. Table S5 (Supplementary Materials) shows each RQS item with its interpretation.

Statistical Analysis

We used the MIDAS module in STATA software, version 14.2, to conduct the meta-analysis. Using bivariate random effects modeling, we calculated the pooled sensitivity and

specificity along with their 95% confidence intervals (CIs). Coupled forest plots and hierarchical summary receiver operating characteristic curves were generated to display the results visually. Meta-regression analyses were performed to further investigate the sources of study heterogeneity by including the covariates in a bivariate model. Heterogeneity across the included studies was assessed using the Cochran Q test, where a p-value < .05 indicated the presence of heterogeneity. Based on the Higgins I^2 statistic, values above 50% indicated high heterogeneity (31). A random-effects model was used to analyze pooling studies and effect size, showing that heterogeneity is considered when predicting the distribution of real effects across investigations. The threshold effects were also evaluated in Meta-disc software, version 1.4. The Deek's asymmetry test was used to establish statistical significance, and a Deek's funnel plot was made to assess publication bias (32). The trim-and-fill method was used to calculate the publication bias when a significant publication bias was found. To evaluate clinical utility, post-test probability was calculated, and a Fagan plot was generated. All p-values under 0.05 were considered significant.

RESULTS

Literature Search

Our electronic search yielded a total of 782 studies. Following the removal of 106 duplicate studies, the titles and abstracts of the remaining articles were screened, and 578 articles were excluded. Eventually, 68 articles were excluded due to a lack of DTA design or not using AI models (Fig 1). After full-text reviews, 31 articles (33–63) were deemed eligible for inclusion in the systematic review. Due to cohort overlap (36,42,46,53,61), random splitting (57), providing only AUCs (62,63), or not providing sufficient data for drawing 2 × 2 tables (58,59), 10 studies were not included in the meta-analysis. Figure 1 depicts the study selection flowchart.

Characteristics of Included Studies

Table 1 shows the general characteristics of the studies included in the systematic review. Almost 80% of studies (25/31) were published after 2020. Nearly two-thirds (21/31) of the studies were conducted in China (36–40,42–48,51–55,59–61,63), and the rest in Italy (33,35,41,56,58), Germany (34,50), South Korea (49), Turkey (57), and the USA (62). While two studies utilized multi-center data (35,63), the prediction of Ki-67 was carried out using data from a single center in both studies. Most studies had a retrospective design (28/31), while only one was prospective (59), and two had no specified study design (33,41). Only nine studies clearly mentioned that they used surgical specimen as the unique route of tissue sampling for IHC evaluation (36,38,40,43,49,54,59,60,63), while seven studies used biopsy (42,44,46,48,53,56,57), seven studies did not mention clearly which method was used (33,35,37,47,51,55,62), and eight studies used a combination of both surgical specimen and biopsy (varying between patients) (34,39,41,45,50,52,58,61). In 22 studies, the cut-off value for Ki-67 expression in IHC staining was 14% (34–38,40–42,45–49,51–53,55,58–60,62,63), and one study used two cut-off values (14% and 20%)(57). The manufacturers of the scanners were Siemens (15/31) (33–35, 41,42,44,47,48,50,51,53,54,59,61,63), GE (9/31) (36,37,43, 45,55–58,60), Philips (5/31) (36,38,40,49,52), and a combination of Siemens and GE (2/31) (39,62). The magnetic field strength of scanners was 3.0 T in 23 studies (33,35,37, 38,40–51,53–55,58,60,61,63), 1.5 T in six studies (34,36, 52,56,57,59), and a combination of both in two studies (39,62). Four studies utilized deep learning-based radiomics (38,39, 44,52). Dynamic contrast-enhanced (DCE) and T1W-contrast-enhanced MRI were commonly utilized imaging techniques for image acquisition (27/31) (33–37,39–46,48–59,62,63), followed by diffusion-weighted imaging (DWI) (16/31) (33,35, 38,40,44,45,48–52,55,57,58,60,61). Only one study used DKI and IVIM imaging methods (60). The previously mentioned methods were also commonly used together. In combination with MRI, positron emission tomography (PET)(33,50), ultrasonography (US) (39), and digital mammography (DM)/digital breast tomosynthesis (DBT) (45) were used as well. Detailed characteristics of the included studies are mentioned in Table 2. Regions of interest (ROI) structure was 3D (15/31) (33–35,40–42,47,51,55–59,62,63), 2D (7/31) (36,37,43,44, 46,49,52), a combination of both (2/31)(39,40), or not mentioned explicitly (7/31) (38,45,48,50,54,60,61). ROI delineation was performed manually in almost half of the studies (15/31) (35,36,38,40,43–46,52,54,55,57,60,61,63), then semi-automatically (10/31) (33,37,41,46,50,51,56,58,59,62), and automatically (6/31) (39,42,47,48,53,62). PyRadiomics was the most frequently used software (8/31) (34,35,40,44,45,55–57) for feature extraction, followed by in-house MATLAB (5/31) (36,41,46,58,62). For feature processing/reduction, LASSO was a commonly utilized algorithm (12/31) (36,37,40, 45–48,50,54,55,57,63). Similarly, logistic regression (LR) was frequently used for model construction (18/31) (34,36,40–49, 51,54,59–62), followed by random forest (RF) (10/31) (33–35,40,49,53,56,58,62,63), support vector machine (SVM) (9/31) (35,37,38,40,44,49,50,55,63), and naive Bayes (NB) (5/31) (34,35,37,49,63). Cross-validation was also utilized for model construction in more than two-thirds of the studies (23/31) (33–35,37,40,42,44,46–58,61–63). Five studies (42,44,48, 53,61) used data from the First Affiliated Hospital of Zhejiang Chinese Medical University. Among them, one was older (42); therefore, an updated radiomic study that used a larger overlapping cohort (48) was included in the meta-analysis. Two studies (53,61) did not specify the period of patient enrollment, and they were not included in the meta-analysis. Therefore, only two studies from this cohort (44,48) were included in the meta-analysis as the periods that patients enrolled in the study were different (2007–2011 vs. 2013–2017). Two studies used data from Shengjing Hospital of China Medical University (46,55), and the one with higher classification ability (55) was included in the meta-analysis.

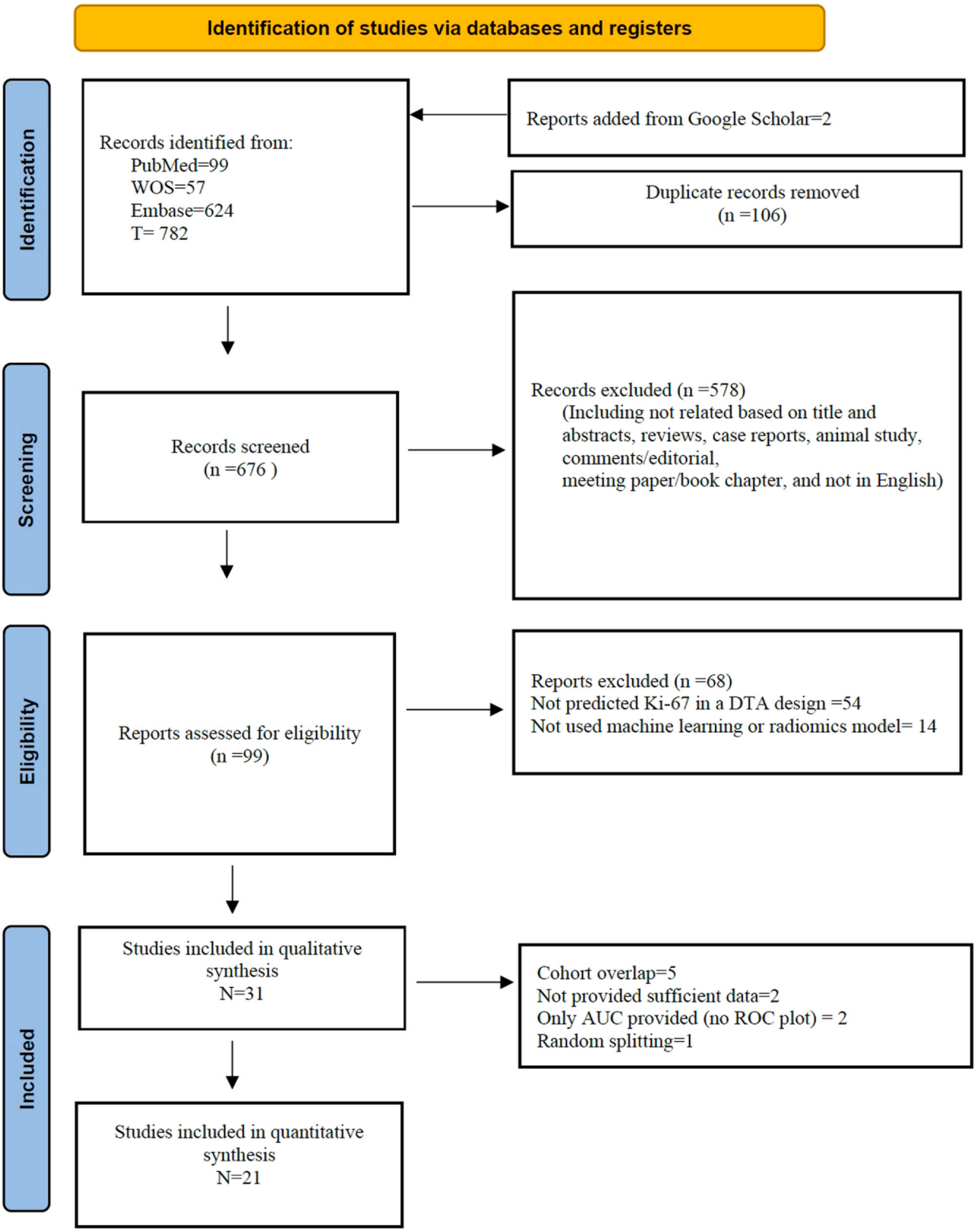

Figure 1. PRISMA flow chart of the study selection procedure for this systematic review and meta-analysis.

TABLE 1. The Baseline Characteristic of Included Studies

Author/Year	Country	Study Design	No of Centers	Number of Lesions*	Age*	Type of Specimen Sampling	Ki-67 cut-off	Scanner Manufacturer	Tesla	MRI Technique/ Sequence
Castaldo et al. 2022 (33	Italy	NM	Single	24	57 (82–35)	NM	30	Siemens	3 T	**T1W (post contrast)[a], T2W, ADC (DWI), and PET**
Demircioglu et al. 2020 (34)	Germany	Retro	Single	80	NM	Surgical resection or core needle biopsy	14	Siemens	1.5 T	**T2w, and T1w (contrast)**
Brancato et al. 2023 (35)	Italy	Retro	Two	78	NM	NM	14	Siemens	3 T	**DCE**, TSE T2, **DWI** ax, and HR Vibe T1-w fat sat
Liang et al. 2018 (36)	China	Retro	Single	318	51.81 ± 10.1	Surgical resection	14	Philips	1.5 T	**T2W** and T1W (contrast)
Ma et al. 2018 (37)	China	Retro	Single	159	29–72	NM	14	GE	3 T	**DCE-MRI**
Y.-Z. Li et al. 2022 (38)	China	Retro	Single	164	NM	Surgical resection	14	Philips	3 T	**DWI**
Qiao et al. 2022 (39)	China	Retro	Single	502	NM	Core needle biopsy or surgical pathology	20	Siemens and GE	3 T & 1.5 T	**DCE-MRI and ultrasound**
L Zhang et al. 2023 (40)	China	Retro	Single	477	51 ± 7	Surgical pathology	14	Philips	3 T	**DCE-MRI and NME-DWI (IVIM and DKI)**
Monti et al. 2018 (41)	Italy	NM	Single	49	NM	Core needle biopsies or mastectomy	14	Siemens	3 T	DCE-MRI VIBE
M Fan et al. 2018 (42)	China	Retro	Single	77	53.3(27–71)	Biopsy	14	Siemens	3.0 T	DCE-MR
He et al. 2021 (43)	China	Retro	Single	88	30–74; 49	Surgical specimens	20	GE	3 T	DCE-MRI
Ming Fan et al. 2020 (44)	China	Retro	Single	322	52.37 (22–84)	Biopsy	14	Siemens	3 T	**DCE-MR and DWI**
Jiang et al. 2022 (45)	China	Retro	Single	209	NM	Biopsy or resection specimen	14	GE	3 T	**DCE-MR, DWI, DBT, and DM**
C. Li, Song, and Yin 2021 (46)	China	Retro	Single	351	(49–52)	Biopsy	14	GE	3 T	DCE-MR
Y. Zhang et al. 2020 (47)	China	Retro	Single	128	52.36 ± 10.82	NM	14	Siemens	3 T	DWI

Table 1 (*Continued*)

Author/Year	Country	Study Design	No of Centers	Number of Lesions*	Age*	Type of Specimen Sampling	Ki-67 cut-off	Scanner Manufacturer	Tesla	MRI Technique/ Sequence
M Fan et al. 2020 (48)	China	Retro	Single	144	52 years 28–83	Biopsy	14	Siemens	3 T	**DCE-MRI and DWI**
Song et al. 2022 (49)	South Korea	Retro	Single	300	56.6 ± 12.0; Range: 32–93	Surgical specimens	14	Philips	3 T	**T2WI, DWI, and T1w (contrast)**
Umutlu et al. 2021 (50)	Germany	Retro	Single	124	54 years (range 31–86 y)	Surgical specimens and biopsy	15[b]	Siemens	3 T	**DCE, T2W-TSE, T1w, ADC (DWI),** and PET
Ni et al. 2020 (51)	China	Retro	Single	112	46.5 (25 - 72)	NM	14	Siemens	3 T	**DWI,** DCE T1, T1WI, and T2WI
W. Liu et al. 2021 (52)	China	Retro	Single	328	50.32 ± 10	Surgical resection or biopsy	14	Philips	1.5 T	**T2WI, DWI, T1 + C (DCE), and mp-MRI**
L Zhang et al. 2022 (53)	China	Retro	Single	208	52.5 (29–84)	Biopsy	14	Siemens	3 T	DCE-MRI
Zhou et al. 2020 (54)	China	Retro	Single	126	NM	Surgical resection	20	Siemens	3 T	Contrast-enhanced T1WI (Pk-**DCE-MRI**)
Feng and Yin 2022 (55)	China	Retro	Single	205	50.56 ± 9.7	NM	14	GE	3 T	DCE-MRI and DWI
Pesapane et al. 2021 (56)	Italy	Retro	Single	81	47.3 (41.1–53.7)	CNB	20	GE	1.5 T	Contrast-enhanced T1-w
Kayadibi et al. 2022 (57)	Turkey	Retro	Single	159	49.9 ± 9.9	Biopsy	**14** and 20	GE	1.5 T	Contrast-enhanced T1w and **ADC** (DWI)
Santucci et al. 2021 (58)	Italy	Retro	Single	164	54.86 (30–84)	Biopsy or surgery	14	GE	3 T	T1W, DCE **-T2WI,** and DWI
Sun et al. 2020 (59)	China	Pro	Single	167	55.0 ± 11.9 (25–80)	Surgical specimens	14	Siemens	1.5 T	CDTV-**DCE-MRI** and T2W fast spin echo
Wang et al. 2022 (60)	China	Retro	Single	227	50.8 ± 10.3 (27–86)	Surgically excised	14	GE	3 T	**IVIM, DKI,** and DWI
Ming Fan et al. 2017 (61)	China	Retro	Single	82	51.65 (27–71)	Breast biopsy or surgery	14	Siemens	3 T	**DWI (ADC)**

Table 1 (*Continued*)

Author/Year	Country	Study Design	No of Centers	Number of Lesions*	Age*	Type of Specimen Sampling	Ki-67 cut-off	Scanner Manufacturer	Tesla	MRI Technique/ Sequence
Saha et al. 2018 (62)	USA	Retro	Single	922	52.25 (21.75–89.49)	NM	14	Siemens and GE	1.5 T and 3 T	DCE-MRI
Ming et al. 2022 (63)	China	NM	Multi	246	NM	Surgical specimens	20	Siemens	3 T	T1-weighted DCE-MRI

NM; not mentioned
[a] Selected for meta-analysis based on the bolded sequence(s).
[b] Considered as 14 in meta-analysis.

Two additional studies, both conducted at Guangdong Provincial People's Hospital, raised concerns about potential cohort overlap (36,52). From these, the latest study (52), distinguished by superior classification and a larger sample size, was chosen for inclusion in the meta-analysis.

The study by Santucci et al. (58) analyzed data using a cut-off of 14% for Ki-67 but reported the number of positives and negatives with a cut-off of 20%. Therefore, it was omitted from the meta-analysis. In addition, the study by Sun et al. (59) was also not included in the meta-analysis due to a lack of reporting the number of positive and negative Ki-67 lesions. Likewise, the study by Kayadibi et al. (57) was not considered in the meta-analysis due to randomly splitting training and test groups without reporting the number of positives and negatives in each set. One study (44) only provided data for the validation cohort. Tables S6 and S7 provide detailed quantitative data extracted from the training and validation cohorts.

Quality Assessment

QUADAS-2

Based on QUADAS-2, the risk of bias and applicability concerns for the included studies are shown in Figure 2. In the patient selection domain, the risk of bias was low overall and high only in a few studies due to not mentioning study design (retrospective vs. prospective) or inclusion/exclusion criteria (Fig 2a). Also, overall applicability concern was low in this domain (Fig 2b). Only one study included some patients receiving chemotherapy (Kayadibi et al.), and another (Ma et al.) did not clearly mention inclusion and exclusion criteria. Similarly, the overall risk of bias for the index test was lower than 25% and only high in several studies due to not using any validation method. Two studies that combined MRI with other imaging modalities did not report separate results for their radiomics analysis, leading to some applicability concerns in the index test domain for matching with review questions. Major concerns were raised in the reference standards section since one-fourth of studies used biopsy as the only route of tissue sampling, and almost half of the studies either used both biopsy and surgical specimens (different per patient) or did not mention the sampling route. Since all studies matched with the review questions for the reference standard section (histopathological evaluation), low applicability concerns were detected for this domain. Similarly, a high or unclear risk of bias was detected in the flow and timing domain as some studies did perform biopsies before MRI acquisition or used different types of sampling per patient (surgical specimens or biopsy) for histopathological evaluation of Ki-67. Taken together, the quality of the included articles was almost acceptable according to the QUADAS-2 assessment.

RQS

The 31 studies had a mean RQS of 5.90, corresponding to 16.6% of the total score, a median of 5, and a range of − 1

TABLE 2. Detailed Characteristics of Studies

Study	ROI Structure	ROI Segmentation	No. of Extracted Features	No. of Selected Feature (s)	Feature Extraction Software	Imaging Features	Feature Processing Algorithm	Modeling Algorithm	Cross Validation
Castaldo et al. 2022 (33)	3D	Semiautomatic	5 (ADC, T2w, and T1w-PC) 19 (T2w and T1w-PC)	-	PMOD	First order (mean, variance, skewness, kurtosis, and energy), and second order (energy, contrast, entropy, homogeneity, correlation, sum average, variance, dissimilarity, and autocorrelation.)	PCA	**LDA**[a], RF, and LogitBoost	3-fold stratified cross-validation
Demircioglu et al. 2020 (34)	3D	Sub automatic	13118	8 features were used for Ki-67 analysis	PyRadiomics	Shape, first order, GLCM, GLDM, GLRLM, and NGTDM	randomized LR, chi-square, f-score, t-score, and mutual information	Naive Bayes, RF, and **LR**	5-fold stratified cross-validation
Brancato et al. 2023 (35)	3D	Manual	Above 1000	2 (Post contrast)	PyRadiomics	Shape, first-order, GLCM, GLSZM, NGTDM, GLDM, GLRLM, wavelet, and Gaussian-filtered features	**$Chi2_{(1st)}$**, Fisher Score, Gini Index, mutual information, ReliefF, LR-RFE (LASSO), and **$CFS_{(2nd)}$**	KNN, NB, SVM, DT, **MLP**, and RF	10-fold cross-validation and Leave-one-out cross-validation (LOOCV)
Liang et al. 2018 (36)	2D	Manual	10207	16 (T2W) 14 (T1 + C) features	in-house MATLAB	Intensity, shape, texture, and wavelet	Intraclass correlation coefficients (ICCs), z-score, and LASSO	LR	-
Ma et al. 2018 (37)	2D	Semiautomatic	56	13 (Postcontrast)	NM	Morphological features, GSS, texture features (GLCM, GLDS and Tamura)	LASSO	kNN, **NB**, and SVM	10-fold cross-validation
Y.-Z. Li et al. 2022 (38)	NM	Manual	-	-	-	-	Deep learning algorithms	**CNN+SVM**	-

Table 2 (*Continued*)

Study	ROI Structure	ROI Segmentation	No. of Extracted Features	No. of Selected Feature (s)	Feature Extraction Software	Imaging Features	Feature Processing Algorithm	Modeling Algorithm	Cross Validation
Qiao et al. 2022 (39)	2D &3D	Automatic	-	-	3D ResNet-18 structure	-	discriminative feature learning method based on a min-max feature loss training	Feature fusion module using MUM-Net	-
(L Zhang et al. 2023 (40)	3D	Manual	742	-	PyRadiomics	Morphology, histogram, and GLCM	RFE	RF, **AB**, SVM, LDA, and LR	5-fold cross validation
Monti et al. 2018 (41)	3D	Semiautomatic	163	25	in-house MATLAB	First- and second-order features, Shape features, *R*1, *k*trans, *k*ep, Ve, iAUC maps, TIRM, postC, and GLCM	Stepwise forward feature selection, maximized Spearman's rank correlation, and maximal information coefficient	LR	-
M Fan et al. 2018 (42)	3D	Automatic	18	-	-	Texture features (GLCM)	removing features with high similarity (maximized correlation coefficient)	LR	LOOCV
He et al. 2021 (43)	2D	Manual	-	-	FIREVOXEL	Texture (GLCM)	-	Multivariate logistic regression	-
Ming Fan et al. 2020 (44)	2D	Manual	107	-	PyRadiomics	Statistic, morphological and texture (GLCM, GLRLM, GLSZM, GLDM, NGTDM) features from super-resolution ADC images.	Evolutionary algorithm-based optimization method	LR and deep learning (SRGAN, EDSR, and bicubic)	10-fold cross-validation
Jiang et al. 2022 (45)	NM	Manual	1967	2 DCE-MRI, and 3 DW-MRI	PyRadiomics	First-order, shape, texture, and filtered features	LASSO	LR	-

Table 2 (*Continued*)

Study	ROI Structure	ROI Segmentation	No. of Extracted Features	No. of Selected Feature (s)	Feature Extraction Software	Imaging Features	Feature Processing Algorithm	Modeling Algorithm	Cross Validation
C. Li, Song, and Yin 2021 (46)	2D	Semi-automatically	2474	17 (intratumoral), 20 (peritumoral), and 20 (combined)	MATLAB	First-order statistics, shape and texture (GLCM, Laws, and Gabor)	WLCX, MRMR, and LASSO	LR	10-fold cross-validation
Y. Zhang et al. 2020 (47)	3D	Automatic	1029	11	Radcloud	First order statistic, shape, and texture (GLCM, GLRLM, and GLSZM)	LASSO	LR	5-fold cross validation
M Fan et al. 2020 (48)	NM	Automatic	97	At least 11	NM	Morphologic, texture, statistics, and dynamic features	LASSO and RFE	**Multitask learning**, LR, SVM, and MTC	LOOCV
Song et al. 2022 (49)	2D	Manual	-	-	-	Morphology, kinetic features, and apparent diffusion coefficient (ADC) at DWI	-	RF, AB, DT, kNN, SVM NB, LDA, and **LR**	10-fold cross-validation
Umutlu et al. 2021 (50)	NM	Semiautomatically	101	Maximum 6 features selected for each model	ITK-SNAP	First order, GLCM, RLM, and NGTDM	LASSO	SVM	5-fold cross validation
Ni et al. 2020 (51)	3D	Semiautomatic	396	84	Artificial Intelligent Kit (A.K.)	Texture parameters, GLSZM, GLCM, form factor parameters, RLM and histogram	Removing outlier variables, Pearson correlation analysis, mean center and standard deviation scale, and linear smoothing filtering	Fisher discriminant analysis (backward selection), and LR	LOOCV

Table 2 (*Continued*)

Study	ROI Structure	ROI Segmentation	No. of Extracted Features	No. of Selected Feature (s)	Feature Extraction Software	Imaging Features	Feature Processing Algorithm	Modeling Algorithm	Cross Validation
W. Liu et al. 2021 (52)	2D	Manual	1472	100(T2), 180 (T1 + C), 160 (DWI), and 420 (mp-MRI)	VGG16 model (CNN)	Deep learning features	minimum-maximum normalization method, Sobel filter for edge detection, and CLAHE method for local contrast enhancement	MLP classifier	5-fold cross validation
L Zhang et al. 2022 (53)	2D and 3D	Automatic	104	at least 10	PyRadiomics	Shape, statistics, and texture features (GLCM, GLDM, GLRLM, GLSZM, NGTDM) from tumor subregion-based Ktrans map	CAM-CM method	Multivariate RF	5-fold cross-validation
Zhou et al. 2020 (54)	NM	Manual	386	4	A.K. software	Histogram, geometry and texture (GLCM, GLRLM, and GLSZM) features from Pk map	LASSO	Multivariate and univariate LR	5-fold cross-validation
Feng and Yin 2022 (55)	3D	Manual	946	5 (wash-out), 5 (wash-in), 6 (SER), 1 (ADC maps), 14 (DCE-MRI parameter maps), and 15 (DCE-MRI combined with ADC maps)	PyRadiomics	First order, shape, and texture (GLCM, GLSZM, GLRLM, and GLDM)	Backward selection approach, ICC, normalizing with Z-score, and LASSO	SVM	10-fold cross-validation
Pesapane et al. 2021 (56)	3D	Semiautomatic	1037	136	PyRadiomics	morphological, histogram-based, and textural descriptors (GLCM, GLRLM, GLSZM, NGTDM, and GLDM).	Spearman's coefficient, Rank Sum test, and Wilcoxon	RF	repeated k-fold cross validation (5 × 10 CV)

Table 2 (*Continued*)

Study	ROI Structure	ROI Segmentation	No. of Extracted Features	No. of Selected Feature (s)	Feature Extraction Software	Imaging Features	Feature Processing Algorithm	Modeling Algorithm	Cross Validation
Kayadibi et al. 2022 (57)	3D	Manual	2212	15 (ADC), 2 (C+MRI), and 10 (combination of both MRI sequence)	PyRadiomics	Features extracted by PyRadiomics (first order statistics, shape, and texture (GLCM GLRLM GLSZM NGTDM GLDM))	LASSO, removing multi collinear features, and lambda optimization	GLM	10-fold cross-validation
Santucci et al. 2021 (58)	3D	Semiautomatic	253	3(First-order), 12(GLCM 3D), and 12 (TOP-LBP)	MATLAB	First-order, Second-order TOP-LBP, and second-order GLCM	Best firsts wrapper	RF	10-fold cross-validation
Sun et al. 2020 (59)	3D	Semiautomatic	-	7 (histogram-based statistical) and 4 (texture-based) features	MR multipara-metric analysis software	Histogram-based (Ktrans, kep, Ve, and IAUC60) and texture-based features	-	LR	-
Wang et al. 2022 (60)	NM	Manual	-	-	-	functional parameters of DWI, IVIM, and DKI	-	LR	-
Ming Fan et al. 2017 (61)	NM	Manual	48 statistical features and 120 ratios of ADC features	13 features	-	Statistical features and ratios of ADC features	Evolutionary algorithm (EA)-based optimization method	LR	LOOCV
Saha et al. 2018 (62)	3D	Automatic	529	4 features (with AUC > 0.5)	MATLAB	Enhancement and kinetic, texture features, and spatial heterogeneity of tumor	Highest cross-validation AUC and removing features with high correlation with other features	RF and LR	K-fold cross-validation

Table 2 (*Continued*)

Study	ROI Structure	ROI Segmentation	No. of Extracted Features	No. of Selected Feature (s)	Feature Extraction Software	Imaging Features	Feature Processing Algorithm	Modeling Algorithm	Cross Validation
Ming et al. 2022 (63)	3D	Manual	15,494	-	PyRadiomics	Histogram features, texture features, Laplacian of Gaussian (LoG) filter features, and wavelet features	LASSO	ENR, SVM, RF, and NB	5-fold cross-validation

AB; adaboost, DR; decision tree, EDSR; enhanced deep super-resolution network, ENR; elastic net regression, GLM; general linear model, KNN; k-nearest neighbors, LASSO; least absolute shrinkage and selection operator, LDA; linear discriminant analysis, LR; logistic regression, MLP; multi-layer perceptron, MRMR; minimum redundancy maximum relevance, MUM-Ne; multi-scale U-shape chained M-shape convolution network, NM; not mentioned, PCA; principal component analysis, RF; random forest, RFE; recursive feature elimination, SRGAN; super-resolution generative adversarial networks, SVM; support vector machine, WLCX; Wilcoxon rank-sum test.

[a] Selected data for meta-analysis was based on the bolded modeling algorithm.

(0%) to 12 (33%). Approximately 40 percent of the studies received a score below 10% (Fig 3). "Well-documented image acquisition protocols", "discrimination statistics", and "biological correlation" were performed for all the included studies. "Multiple segmentation" and "feature reduction" were also performed in more than 75% and 90% of the studies, respectively. One-fourth of studies provided cut-off points for their models. None of the included studies considered "phantom study" and "cost-effectiveness analysis". Only one study received the score for "imaging at multiple time points" due to using a *temporal* validation cohort (52). In addition, the items including "potential clinical application", "open science reporting", and "multivariable analysis" received scores in one (45), two (56,63), and two (58,62) studies, respectively. Likewise, calibration statics were performed in three studies (45,56,60), and only one (59) had a prospective design. The overall low RQS score was due to not using independent validation cohorts by most studies (20/31), leading to a loss of 5 points in the "validation" item. Regarding the "comparison to gold standard" item, a full score was given only to studies that used surgery as the unique route of tissue sampling for histopathological assessment. Detailed scoring for all of the studies is provided in Table S8 (Supplementary Materials).

Meta-analysis

Diagnostic Test Accuracy Analysis

20 training cohorts and five independent validation cohorts were pooled separately. In training and validation cohorts, pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) were 0.80 [95% CI, 0.73–0.86] vs. 0.81 [95% CI, 0.72–0.87], 0.82 [95% CI, 0.78–0.86] vs. 0.73 [95% CI, 0.62–0.82], 4.6 [95% CI, 3.4–6.1] vs. 3.0 [95% CI, 2.1–4.2], 0.24 [95% CI, 0.17–0.34] vs. 0.26 [95% CI, 0.18–0.37], and 19 [95% CI, 10–35] vs. 11 [95% CI, 7–19]. As expected, the area under the curve (AUC) for the summary ROC plot of validation cohorts 0.84 [95% CI, 0.80–0.87] was slightly lower than training cohorts 0.88 [95% CI, 0.85–0.91]. The coupled forest plots and SROC curves are represented in Figures 4–6.

Heterogeneity Analysis

I^2 statistics did not reveal significant heterogeneity for sensitivity (I^2 =44.2%) (p-value = 0.13) and specificity (I^2 = 18.49%) (p-value = 0.30) in validation cohorts. However, the heterogeneity for sensitivity (I^2 = 86.58%) and specificity (I^2 = 81.89%) in training cohorts was severe (p-values = 0.00), making subgroup analysis necessary. In addition, the threshold analyses for both sets were performed in Meta-DiSc software to calculate Spearman's correlation coefficient between the sensitivity logit and the specificity logit. However, for both cohorts, the estimated correlation was not significant (p-value > 0.05).

Figure 2. Risk of bias and applicability concerns according to Quality Assessment of Diagnostic Accuracy Studies-2 tool: (**a**) per study assessment (**b**) per domain summary.

Heterogeneity Exploration and Meta-regression

We used numerous covariates to investigate their contribution to study heterogeneity (Table 3). We found out that the following factors caused significant heterogeneity in the meta-analysis according to the joint model analysis using meta-regression: magnetic field strength of scanner (3.0 T vs. 1.5 T) (p-value = 0.00), scanner manufacturer (p-value = 0.00), ROI structure (2D vs. 3D) (p-value = 0.00), MRI imaging method (contrast-enhanced vs. diffusion-weighted imaging) (p-value = 0.00), reference standard for measuring Ki-67 (biopsy vs. surgery) (p-value = 0.00), Ki-67 cut-off (≤ 14 vs. > 14) (p-value = 0.04), type of radiomics algorithm (deep learning vs. conventional) (p-value = 0.02), type of modeling algorithm (logistic regression vs. other) (p-value = 0.05), feature extraction software (PyRadiomics vs. other) (p-value = 0.00), and feature reduction algorithm (LASSO vs. other) (p-value = 0.00).

Subgroup Analysis

According to the subgroup analysis, studies with sample sizes smaller than 150 had a higher pooled sensitivity (85% vs. 78%)(p-value = 0.20, not significant) and specificity (86% vs. 80%) (p-value = 0.00). In addition, in studies in which we used the top left method for calculating the 2 * 2 table, pooled specificity (86% vs. 81%) (p-value = 0.06) was higher. In comparison, sensitivity was significantly lower (78% vs. 81%) (p-value = 0.02) than those studies that reported enough data. Our subgroup analysis also revealed interesting findings. We found that studies that used Philips scanners had a significantly higher specificity compared to other scanners (87% vs. 80%) (p-value = 0.05). In addition, combining features derived from contrast-enhanced MRI with DWI could increase sensitivity (from 76% to 84%) (p-value = 0.18, not significant) and specificity (from 82% to 83%) (p-value = 0.05). Likewise, studies that used LR for model construction had a significantly lower sensitivity (75% vs. 84%) (p-value = 0.00) and specificity (76% vs. 84%) (p-value = 0.00) than those that used other algorithms (e.g., LDA, MLP, SVM, AB, or RF instead). Similarly, those studies that used LASSO for feature reduction had a significantly lower specificity (78% vs. 86%) (p-value = 0.00). Studies that tissue sampling was performed only via the surgical route had a lower sensitivity (77% vs. 81%) (p-value = 0.12, not significant) and specificity (83% vs. 84%) (p-value = 0.00) compared to those that used biopsy alone or in combination with surgery. Although deep learning-based radiomics had a higher pooled sensitivity (93% vs. 77%) (p-value = 0.83) and specificity (90% vs. 80%)(p-value = 0.10) compared to conventional radiomics methods, due to the small number of deep learning-based studies ($n = 3$ vs. $n = 17$), these differences were not statistically significant (p-value > 0.05). The results of the subgroup analysis are summarized in Table 3.

Sensitivity Analysis

We removed every study individually and pooled the remaining studies to investigate whether it could impact meta-

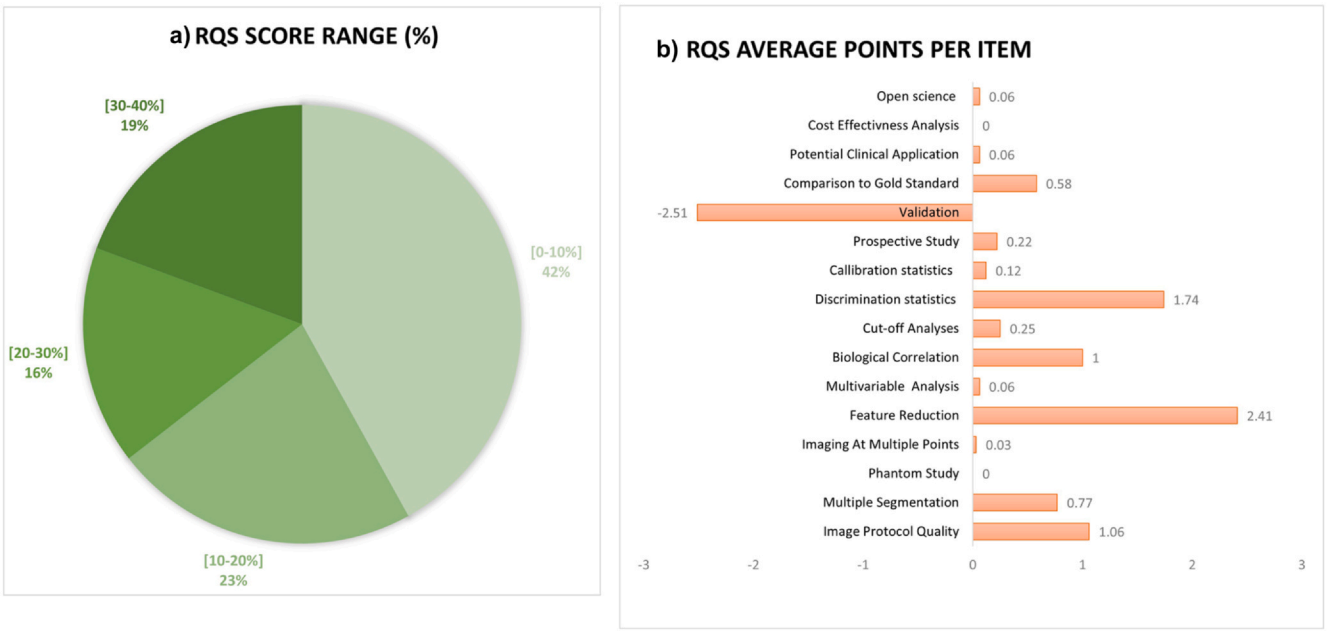

Figure 3. Methodological quality assessment based on the RQS tool. (**a**) Proportion of studies per different RQS range percentages. (**b**) Average points per each RQS item. RQS, radiomics quality score.

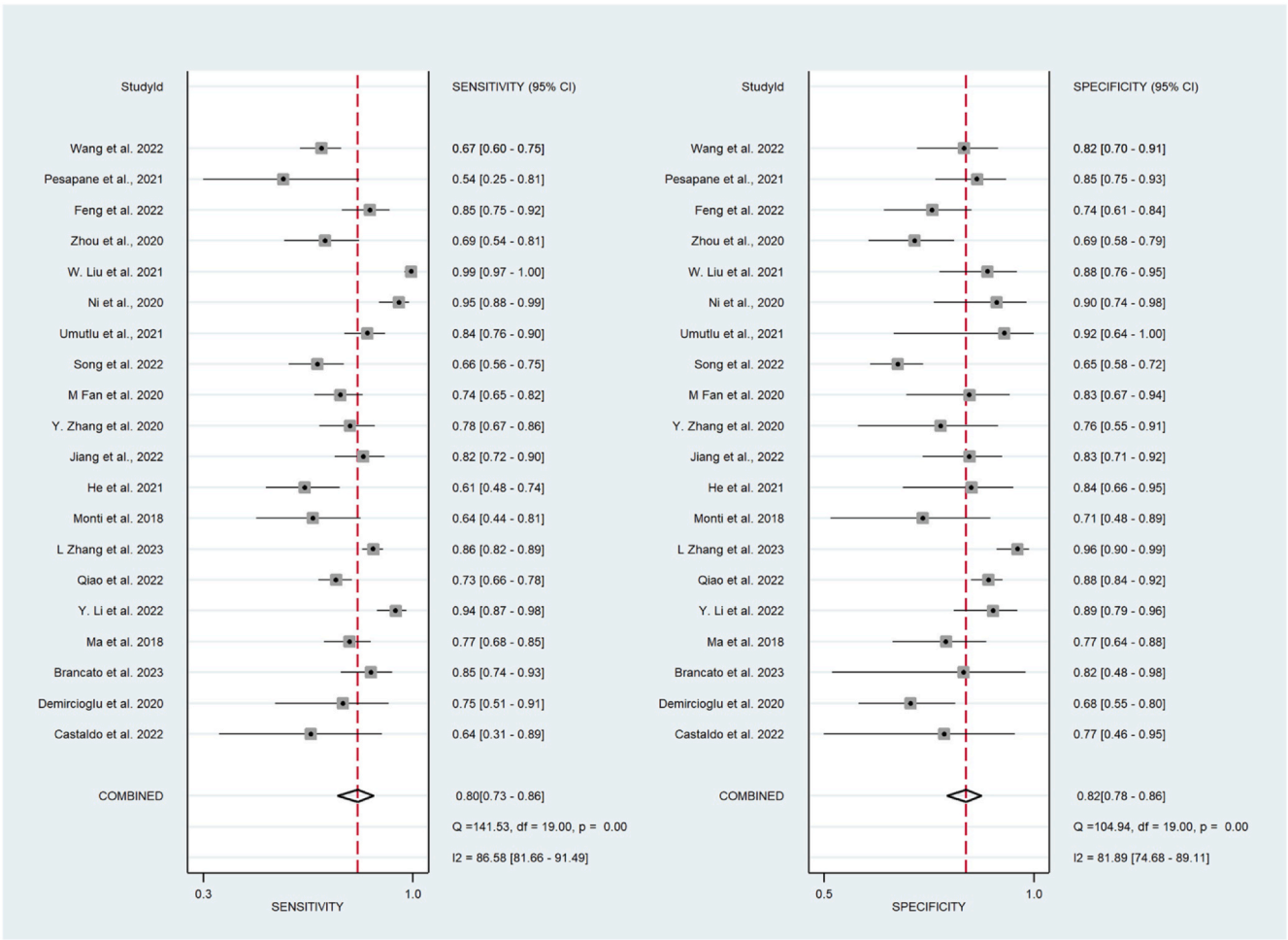

Figure 4. Forest plots of the sensitivity and specificity of MRI radiomics for prediction of Ki-67 in breast cancer (training cohorts).

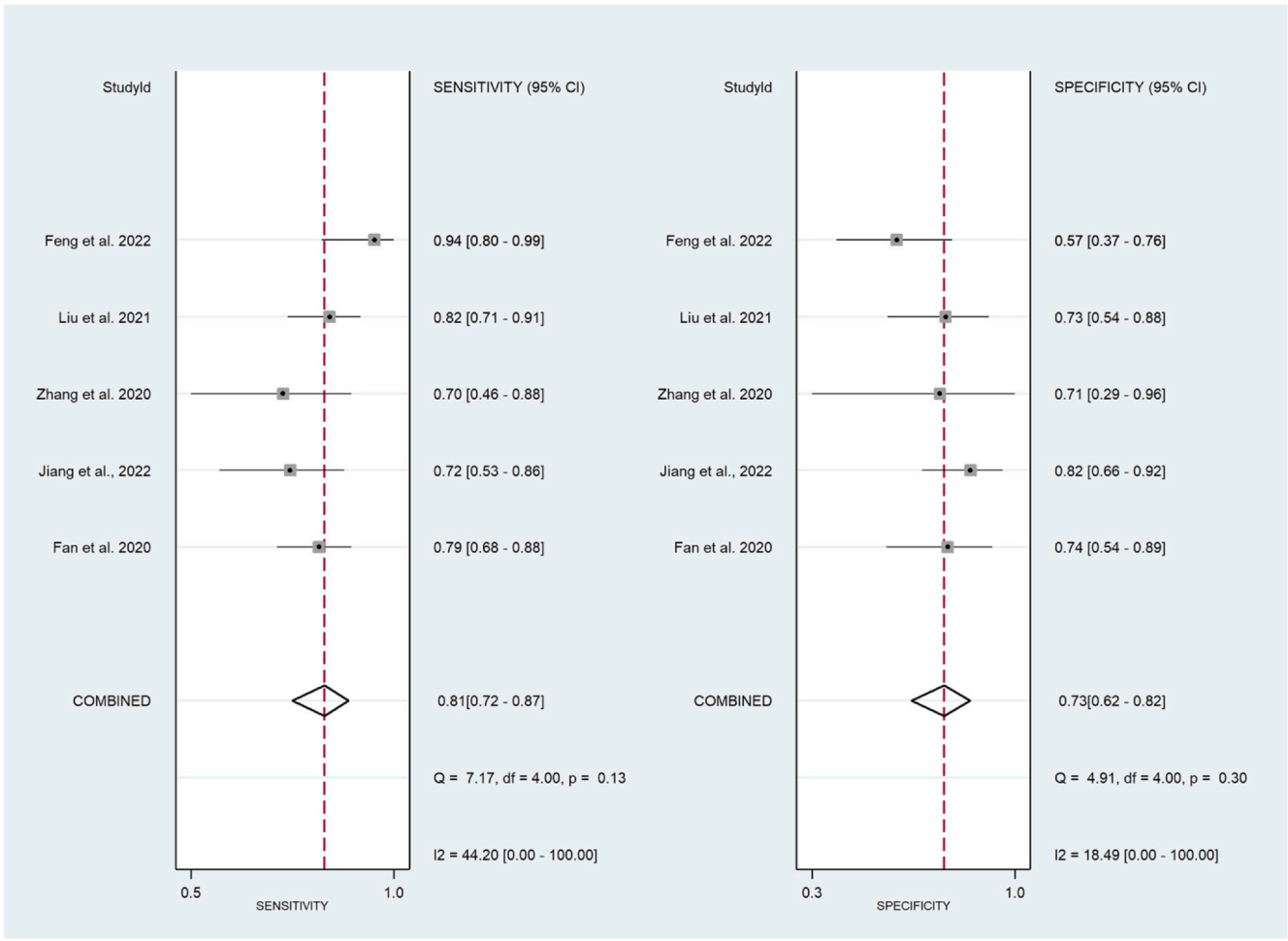

Figure 5. Forest plots of the sensitivity and specificity of MRI radiomics for prediction of Ki-67 in breast cancer (validation cohorts).

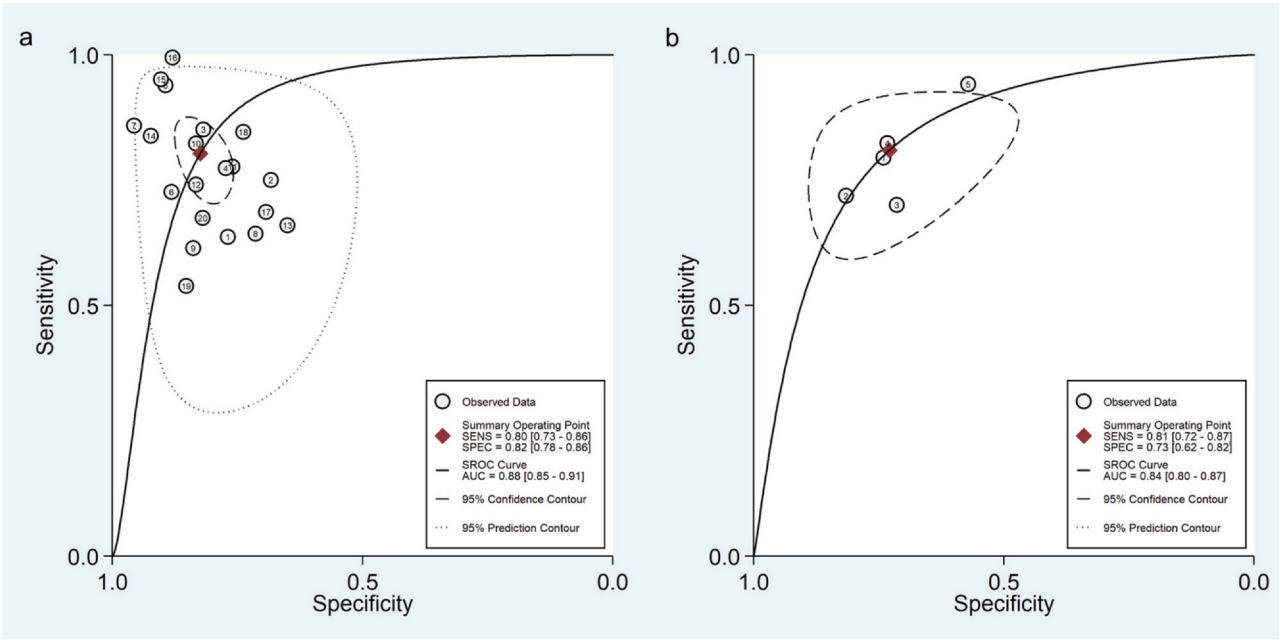

Figure 6. Summary receiver operating characteristic curves of MRI radiomics for predicting Ki-67 in training (**a**) cohorts and validation (**b**) cohorts.

TABLE 3. Investigation of Heterogeneity Using Meta-regression and Subgroup Analysis

Covariates		*n*	Sensitivity	P_1	Specificity	P_2	Joint Model Analysis		
							I^2	LRT chi^2	P Value
Sample size	< 150	7	0.85 [0.76–0.93]	0.20	0.86 [0.80–0.91]	0.00	7	2.15	0.34
	> 150	13	0.78 [0.69–0.86]		0.80 [0.74–0.86]				
Top left method	Yes	5	0.78 [0.64–0.91]	0.06	0.86 [0.78–0.93]	0.02	0	1.94	0.38
	No	15	0.81 [0.74–0.88]		0.81 [0.76–0.86]				
Country	China	13	0.83 [0.77–0.90]	0.43	0.76 [0.68–0.85]	0.02	44	3.60	0.17
	Others	7	0.73 [0.60–0.86]		0.76 [0.68–0.85]				
Segmentation	Manual	10	0.83 [0.76–0.91]	0.18	0.83 [0.77–0.89]	0.00	0	1.30	0.52
	Other (automatic and semiautomatic)	10	0.77 [0.66–0.87]		0.82 [0.76–0.89]				
Magnetic Field Strength	3.0 T	16	0.79 [0.72–0.86]	0.04	0.82 [0.77–0.87]	0.02	100	20.10	**0.00**
	1.5 T	3	0.89 [0.78–1.00]		0.82 [0.72–0.93]				
Manufacturer[1]	Siemens	8	0.79 [0.68–0.90]	0.04	0.79 [0.71–0.87]	0.00	93	30.42	**0.00**
	Others	10	0.82 [0.74–0.91]		0.84 [0.78–0.89]				
Manufacturer[2]	GE	6	0.74 [0.60–0.88]	0.01	0.81 [0.73–0.89]	0.00	94	31.65	**0.00**
	Other	12	0.84 [0.77–0.91]		0.82 [0.77–0.88]				
Manufacturer[3]	Philips	4	0.91 [0.84–0.98]	0.89	0.87 [0.80–0.94]	0.05	94	34.37	**0.00**
	Others	14	0.77 [0.69–0.85]		0.80 [0.75–0.86]				
ROI structure	3D	9	0.80 [0.68–0.92]	0.18	0.82 [0.76–0.89]	0.04	98	112.74	**0.00**
	2D	3	0.88 [0.75–1.00]		0.84 [0.73–0.94]				
MRI Sequence	Contrast-enhanced + DWI MRI	9	0.84 [0.77–0.92]	0.18	0.83 [0.77–0.90]	0.00	0	1.60	0.45
	Single-sequence MRI (e.g., DWI or DCE)	11	0.76 [0.67–0.86]		0.82 [0.76–0.87]				
MRI Sequence	Contrast-enhanced	7	0.69 [0.62–0.77]	0.00	0.79 [0.73–0.86]	0.00	99	146.37	0.00
	DWI	3	0.90 [0.85–0.95]		0.87 [0.78–0.95]				
Reference Standard	Surgical specimens	6	0.77 [0.63–0.90]	0.12	0.83 [0.75–0.91]	0.01	97	65.42	**0.00**
	Biopsy or combination	8	0.81 [0.70–0.92]		0.84 [0.77–0.91]				
Ki-67 cut-off	≤ 14	15	0.84 [0.78–0.89]	0.92	0.82 [0.77–0.87]	0.00	68	6.32	**0.04**
	> 14	5	0.65 [0.49–0.82]		0.82 [0.74–0.90]				
Radiomics Algorithm	Deep learning	3	0.93 [0.87–0.99]	0.83	0.90 [0.83–0.96]	0.10	74	7.55	**0.02**
	Conventional	17	0.77 [0.70–0.84]		0.80 [0.76–0.85]				
Modeling Algorithm	LR	9	0.75 [0.65–0.85]	0.00	0.76 [0.70–0.83]	0.00	5.87	5.87	**0.05**
	Others	11	0.84 [0.77–0.91]		0.86 [0.82–0.91]				
Feature Extraction Software	PyRadiomics	6	0.80 [0.67–0.93]	0.10	0.83 [0.76–0.91]	0.00	95	44.07	**0.00**
	Others	11	0.82 [0.74–0.91]		0.84 [0.78–0.89]				
Feature Reduction Algorithm	LASSO	7	0.81 [0.70–0.91]	0.06	0.78 [0.71–0.85]	0.00	91	22.73	**0.00**
	Others	12	0.81 [0.73–0.89]		0.86 [0.82–0.90]				
RQS	≥ 6	6	0.84 [0.75–0.94]	0.20	0.80 [0.72–0.88]	0.00	34	3.04	0.22
	< 6	14	0.78 [0.71–0.86]		0.84 [0.79–0.89]				

analysis. However, no significant change in AUC, sensitivity, specificity, PLR, NLR, and DOR was observed when each study was excluded.

Publication Bias

Deeks's test was used to investigate potential publication bias in the included studies; however, the funnel plot asymmetry test did not reveal a significant publication bias in both validation (Fig7a, p-value = 0.50) or training cohorts (Fig 7b, P-value = 0.40).

Clinical Utility

According to the Fagan plot for training cohorts (Fig 8a), when the pretest was positive, MRI-based radiomics could boost the post-test probability to 53% from 20% with a PLR of 5 for predicting Ki-67. When the pretest was negative, it could reduce the post-test probability to 6% with an NLR of 0.24. Similar results were observed in the validation cohorts (Fig 8b).

DISCUSSION

Ki-67, serving as a proliferation index, provides valuable information regarding tumor aggressiveness. Therefore, accurate estimation of Ki-67 is essential for both tumor classification and prognosis prediction (3). At present, conventional core needle biopsies are employed to acquire breast tissues necessary for immunohistochemical analysis (8). While Ki-67 heterogeneity exists in breast cancer, ranging from 1% to 90% within different intratumoral regions, it is crucial to assess the entire tumor for a comprehensive analysis (36,64). Imaging approaches can be beneficial as they provide much more information about tumor heterogeneity. In addition, recent studies reveal that pre-operative core needle biopsy can significantly elevate the Ki-67 index (65–67), subsequently worsening disease outcomes in HER2-negative patients (66,67). Thus, utilizing less invasive methods such as MRI appears promising. Another advantage of imaging over the biopsy is its ability to evaluate surrounding tumor environments (peri-tumoral regions) that reflect lymphatic invasion and infiltration (45).

Radiomics is a specialized field in medical imaging that entails extracting and analyzing quantitative features from medical images. This process holds the potential to offer supplementary insights into Ki-67 expression in breast cancer by scrutinizing the distinctive characteristics of breast tumors depicted in medical images. (44). Constructing models using machine learning (60) or deep learning (36) algorithms can provide key information about Ki-67 expression in breast cancer. This could complement traditional pathology methods, potentially reduce the need for invasive biopsies, and guide treatment decisions.

This systematic review and meta-analysis explored whether MRI-based radiomics can predict Ki-67 expression in patients with breast cancer. By pooling the results of 20 training cohorts, we observed that MRI-based radiomics methods are promising for Ki-67 prediction, with a pooled sensitivity, specificity, and AUC of 0.80, 0.82, and 0.88, respectively. The corresponding values for five independent validation cohorts were 0.81, 0.73, and 0.84, respectively.

In training cohorts, significant heterogeneity was observed, which made heterogeneity exploration necessary. Initially, we assessed Spearman's correlation coefficient to eliminate the possibility of a threshold effect. The results indicated that a threshold effect is unlikely to cause heterogeneity (p-value > 0.05). Then, we investigated other potential sources of heterogeneity using univariable meta-regression and identified numerous implicated variables.

As mentioned in previous systematic reviews (68,69), magnetic field strength and manufacturer of scanners are two crucial factors that influence algorithm performance, two factors that also contributed to the heterogeneity of our results based on meta-regression. We found that using deep learning radiomic methods is a great cause of between-study heterogeneity. Recently, in a meta-analysis by Liang et al. (70), the performance of deep learning-based MRI radiomics was higher than conventional radiomcis with machine learning for prediction of response to neoadjuvant chemotherapy in breast cancer. In contrast, a meta-analysis by Zhang et al. (71) showed that conventional MRI radiomcis with machine learning methods could act slightly better than deep learning based-studies for predicting axillary lymph node metastasis. Similar to our study, the number of deep learning-based articles was few in both meta-analyses. Therefore, whether deep learning radiomics are superior to conventional radiomics for predicting Ki-67 expression requires more investigation.

Different ROI segmentation methods (2D vs. 3D) were also detected as a possible source of heterogeneity. Unfortunately, some included studies did not clearly mention ROI structure, forcing us to exclude them while doing subgroup analysis. Whole tumor analysis is only possible when 3D tumor segmentation is applied. This method also has the advantage of assessing tumor heterogeneity compared to 2D segmentation (72). In our subgroup analysis, only three studies used 2D segmentation, and surprisingly, their pooled specificity was significantly higher than 3D methods. Thus, evaluating the performance of different segmentation structures should be investigated in future studies.

LR is a widely applied machine-learning modeling algorithm for constructing radiomics signatures. However, it has disadvantages like lower classification accuracy (73). This factor contributed to interstudy heterogeneity, and based on subgroup analysis, we found that studies that used LR for model construction had a significantly lower sensitivity and specificity than newer algorithms such as SVM and AdaBoost. A similar finding was observed for studies that also used the LASSO algorithm for feature reduction. These findings must be taken into consideration in future studies for a more accurate classification ability of radiomics signatures.

Single-parameter MRI features derived from ADC maps (e.g., ADC mean) and DCE-MRI parametric maps (e.g.,

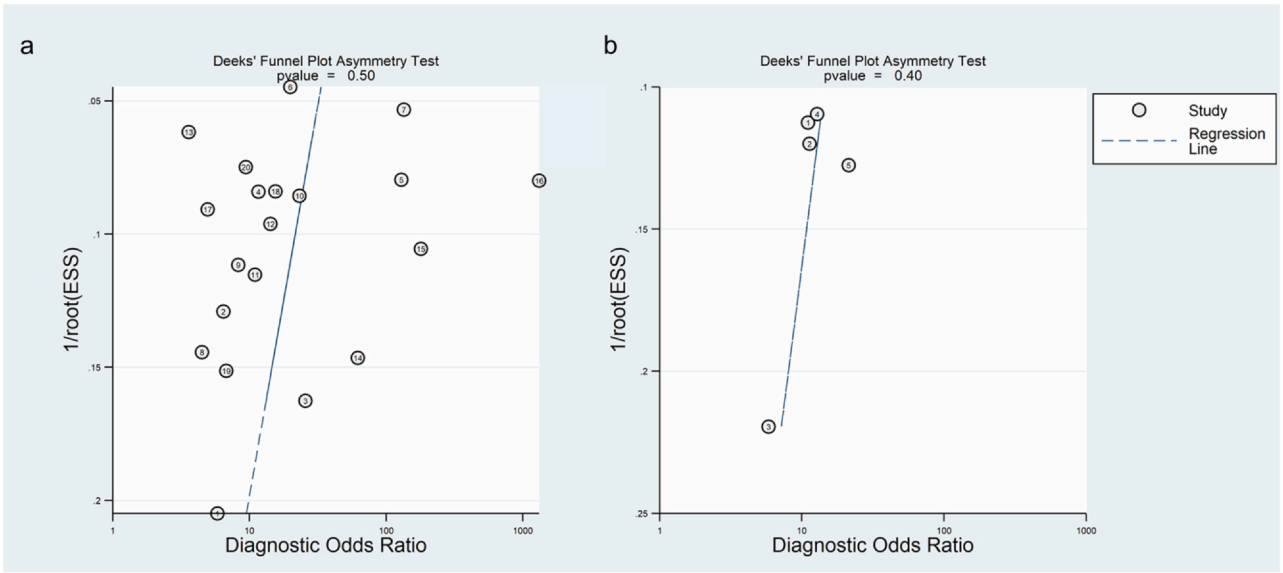

Figure 7. Deeks' funnel plot with superimposed regression line. In both training (**a**) and validation cohorts (**b**), the funnel plot asymmetry test revealed no publication bias (P-values > 0.10).

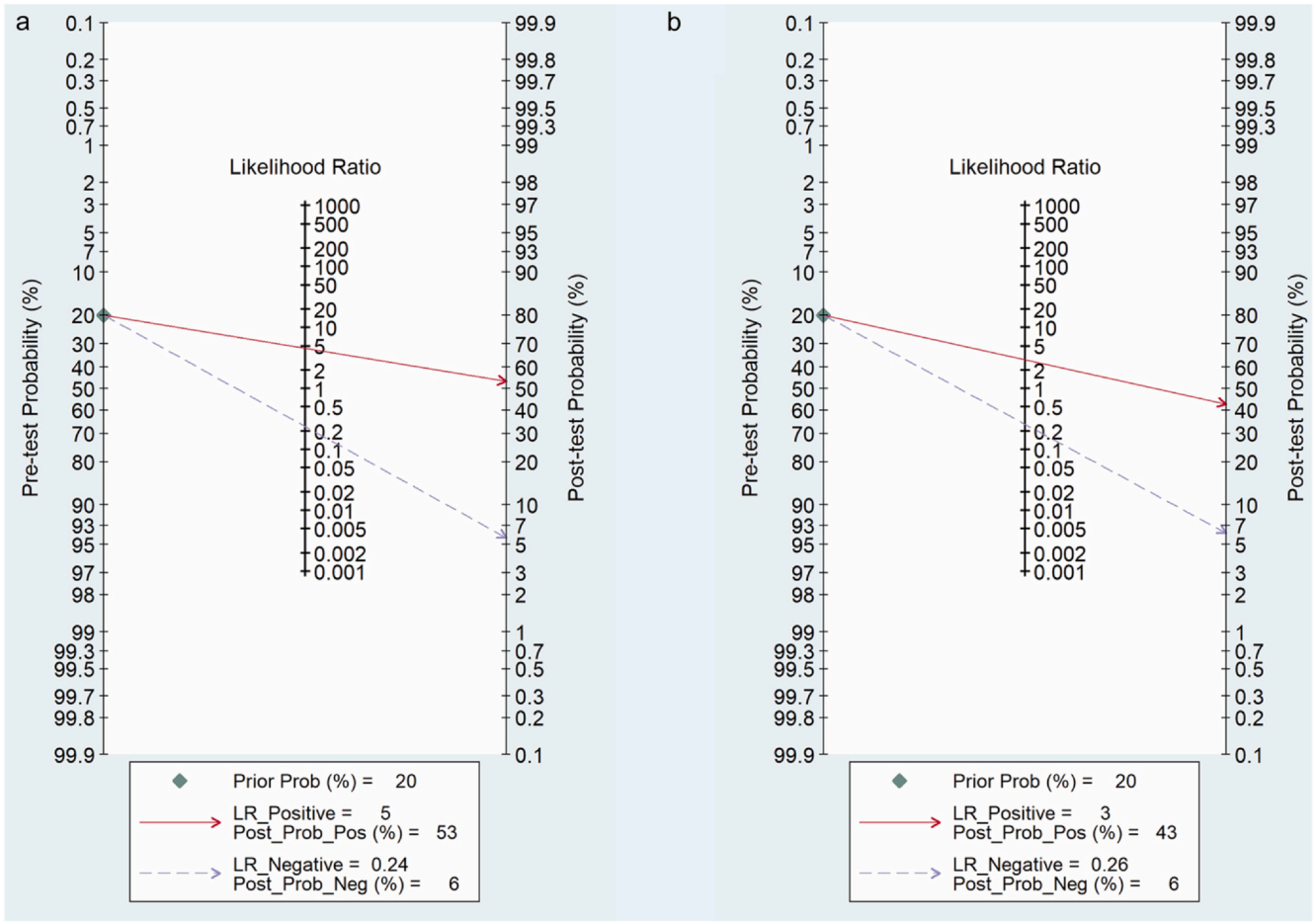

Figure 8. Fagan plots for assessing the clinical utility in training (**a**) and validation (**b**) cohorts.

Ktrans, Kep, Ve, and Vp) are strongly correlated with Ki-67 expression in breast cancer (73). Therefore, these features have frequently been utilized in radiomics studies to detect breast tumors with high Ki-67 expression (41,47,54). Previous meta-analyses have shown that combined DCE-MRI and DWI have superior diagnostic accuracy than either DCE-MRI or DWI alone for differentiating malignant and benign breast lesions (74). In line with this, our subgroup analysis also yielded similar results for combined DCE-MRI and DWI radiomics analysis for predicting Ki-67 in breast cancer. Our investigation comparing DWI and DCE-MRI for distinguishing highly expressing Ki-67 lesions yielded noteworthy findings. Notably, ADC demonstrated superior efficacy in this context, suggesting that it could stand alone as a reliable modality when circumstances limit the use of multiple MRI techniques. The observed advantage of ADC may be attributed to its sensitivity to tissue microstructure and cellularity, aligning with the characteristics of highly expressing Ki-67 lesions. While acknowledging the study's limitations, including small number of studies using DWI alone compared to DCE-MRI alone, our results imply that, under certain conditions, prioritizing ADC in clinical protocols may offer a pragmatic and practical approach. These findings contribute valuable insights for researchers, opening avenues for further investigations into refining imaging strategies for lesions with high Ki-67 expression.

We highlighted the role of various feature extraction software, including PyRadiomics, in contributing to the observed heterogeneity of results. The variability stems from diverse algorithmic approaches and parameter settings across different software. This diversity, rather than being inherently detrimental, underscores the nuanced impact of the choice of feature extraction software on study outcomes. Recognizing this variability, we emphasize the importance of understanding and transparently reporting the software employed, as it significantly influences the radiomic landscape.

Our meta-regression analysis underscores the multifaceted nature of inter-study heterogeneity, revealing influential factors such as the route of tissue sampling (biopsy vs. surgery) and the choice of Ki-67 expression cut-off. Importantly, existing evidence suggests that the biopsy process may dynamically impact Ki-67 expression in breast lesions over time. Simultaneously, the variability introduced by employing a specific Ki-67 cut-off emphasizes the importance of standardization in defining high expression. The interplay of these factors significantly contributes to the observed heterogeneity in radiomic analyses. Acknowledging these influences, our findings highlight the nuanced considerations associated with both sample acquisition methods and the chosen Ki-67 cut-off. Transparent reporting and thoughtful consideration of these variables are essential steps toward enhancing result interpretation and comparability across radiomics studies.

We also evaluated the quality of the included studies in this review using QUADAS-2 and RQS, which are routinely used for quality assessment in diagnostic test accuracy and radiomic studies, respectively. According to QUADAS-2, there were some risks of bias in reference standards as well as flow and timing domains. First, one-fourth of studies used biopsy as the method of tissue sampling, and more than half of the studies either did not mention the sampling method or used biopsy or surgery, different per patient. Second, the overall risk of bias in the flow and timing domain remained unclear as many studies did not mention whether MRI was performed before biopsy. In addition, if all patients do not receive a standard reference method for IHC evaluation (e.g., biopsy used in one and surgery in another), a bias occurs in the flow and timing domain. For this reason, some studies were detected with high risks of bias in the flow of timing domain. Overall, low applicability concern was detected in patient selection, index test, and reference standard domains, making the quality of the included articles acceptable. Unfortunately, the mean RQS point of the included studies was near 6, equivalent to 16.6% of the total score (6/36). Generally, most systematic review articles receive more overall RQS scores than this value. One cause of this significant difference was that in many of the included studies in our review, an independent validation cohort was not used for testing models' reproducibility, leading to a loss of at least 5 points in this section as the minimum score of this item in studies that have a validation cohort is + 2. According to the RQS checklist, validation should be performed without retraining; therefore, studies that only use cross-validation methods lost seven points compared to studies with an independent validation cohort.

Compared to cross-validation methods, employing an independent validation cohort reduces the likelihood of overfitting, provides a more precise estimation of real-world performance, allows a more thorough assessment of the model's performance, and offers more generalizability. An independent validation cohort ensures that the model is assessed using entirely new data, which is more unbiased and enables a more precise estimation of the model's performance. That is why the "validation" item of the RQS tool is misinterpreted in systematic review articles (75). Our investigation did not find a significant link between RQS (Radiomics Quality Score) and the heterogeneity of results. While acknowledging the importance of study quality metrics, our study suggests that factors other than RQS played more substantial roles in shaping the observed variability across studies. This highlights the complexity of sources contributing to the heterogeneity of the results. Although RQS is adopted frequently in systematic review articles of radiomic studies, and it has been about five years since this tool was introduced, many newer checklists are emerging, indicating that the RQS might not be a completely ideal tool for assessing the quality of radiomics studies.

Limitations

This systematic review and meta-analysis has several limitations that are necessary to mention:

a) Almost all of the included studies were retrospectively designed. Prospective studies are generally regarded as superior to retrospective studies due to standardized imaging protocols, timely and relevant radiomic feature extraction, standardized and blinded data collection, and optimized study design. These factors can all enhance the quality and relevance of the study results.
b) Unfortunately, some studies were excluded from the meta-analysis due to probable or definitive cohort overlap and insufficient data reporting. Cohort overlap can lead to biased and inaccurate estimates in meta-analysis and should be carefully considered during study selection and analysis to ensure the validity and accuracy of the results.
c) The lack of validation cohorts in most of the included studies caused a significant drop in the overall RQS score and hindered subgroup analysis in validation cohorts due to the small number of included studies.
d) Many of the included studies did not mention the route of tissue sampling that subsequently caused unclear risks of bias in reference standard and flow and timing sections of the QUADAS-2 tool.
e) Only a small number of studies used deep learning methods for model establishment. DL and CNN offer a more automated and efficient approach to feature extraction, allowing for extracting high-level features from images. This makes them particularly suitable for radiomics studies, where the complexity and volume of medical imaging data can be high.

Future Perspectives

In the future, researchers should consider these points:

a) Further studies should use independent validation cohorts, especially external validation, as it is an essential step in developing radiomics models and helps to ensure their generalizability, robustness, clinical applicability, and regulatory approval.
b) It is recommended to conduct an MRI before any intervention, as there is evidence indicating an increase in Ki-67 levels following core needle biopsy. This alteration in Ki-67 levels has been shown to impact patients' prognosis.
c) Authors conducting radiomics studies are encouraged to design their research protocol according to current standards for scientific rigor, consistency, transparency, and clinical relevance. This can help increase the reliability and generalizability of the findings, ultimately leading to better patient care.
d) Combining radiomic features with non-radiomic features (e.g., age and other clinical indicators) can lead to more comprehensive and accurate diagnostic performance.
e) Transparent reporting of the diagnostic performance of radiomics models is essential since these data are necessary for conducting a DTA-type meta-analysis.
f) Making code and data publicly available can significantly improve knowledge transfer and reproducibility of radiomics studies according to RQS guidelines. Thus, radiomics researchers are encouraged to share their code and scans.

CONCLUSION

This systematic review and meta-analysis showed that MRI-based radiomics have a robust potential utility for predicting Ki-67 in breast cancer. However, in line with a previous meta-analysis that evaluated the diagnostic performance of radiomics for differentiating breast cancer subtypes (27), the pooled AUCs of existing studies in both validation and training cohorts were less than 0.90, indicating that these studies are currently in their early phases and restricting the possibility of using these tools as a supplement to current pathological assessments (e.g., biopsy or surgery) to predict Ki-67 expression very accurately. Therefore, more research is required to improve the diagnostic ability of radiomics for predicting Ki-67 in breast cancer.

ETHICS APPROVAL

All procedures were carried out with the approval of the Ethics Committee of Tabriz University of Medical Sciences (IR.TBZMED.REC.1401.013).

FUNDING

This study was supported by the Research Vice-Chancellor at Tabriz University of Medical Sciences, Iran [Grant No. 69395].

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT by OpenAI to improve paper readability. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

None.

AVAILABILITY OF DATA AND MATERIAL

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

AUTHORS CONTRIBUTION

PT and ZH designed the study strategy. PT, ZHE, FP searched the database and selected the studies. ZH and PT extracted the data and evaluated the studies quality. PT did the analysis. PT wrote the manuscript and edited with the assistance of LAM, BB and ZHE. LAM and BB supervised the work. All authors have read and approved the content.

INFORMED CONSENT

Not applicable.

APPENDIX A. SUPPORTING INFORMATION

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.acra.2023.10.010.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71:209–249.
2. Eliyatkın N, Yalçın E, Zengel B, et al. Molecular classification of breast carcinoma: from traditional, old-fashioned way to a new age, and a new way. J Breast Heal 2015; 11:59.
3. Inwald EC, Klinkhammer-Schalke M, Hofstädter F, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. Breast Cancer Res Treat 2013; 139:539–552.
4. Tagliafico AS, Bignotti B, Rossi F, et al. Breast cancer Ki-67 expression prediction by digital breast tomosynthesis radiomics features. Eur Radiol Exp 2019; 3:1–6.
5. Tao M, Chen S, Zhang X, et al. Ki-67 labeling index is a predictive marker for a pathological complete response to neoadjuvant chemotherapy in breast cancer: a meta-analysis. Med (Baltimore) 2017; 96.
6. Nielsen TO, Leung SCY, Rimm DL, et al. Assessment of Ki67 in breast cancer: updated recommendations from the international Ki67 in breast cancer working group. J Natl Cancer Inst 2021; 113:808–819.
7. Łukasiewicz S, Czeczelewski M, Forma A, et al. Breast cancer—-epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. Cancers (Basel) 2021; 13:4287.
8. Kalvala J, Parks RM, Green AR, et al. Concordance between core needle biopsy and surgical excision specimens for Ki-67 in breast cancer–a systematic review of the literature. Histopathology 2022; 80:468–484.
9. Banna GL, Olivier T, Rundo F, et al. The promise of digital biopsy for the prediction of tumor molecular features and clinical outcomes associated with immunotherapy. Front Med 2019; 6:172.
10. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. Nat Rev Cancer 2018; 18:500–510.
11. Shui L, Ren H, Yang X, et al. The era of radiogenomics in precision medicine: an emerging approach to support diagnosis, treatment decisions, and prognostication in oncology. Front Oncol 2021; 10:570465.
12. Rosenstein BS, West CM, Bentzen SM, et al. Radiogenomics: radiobiology enters the era of big data and team science. Int J Radiat Oncol Biol Phys 2014; 89:709–713.
13. Rogers W, Thulasi Seetha S, Refaee TAG, et al. Radiomics: from qualitative to quantitative imaging. Br J Radiol 2020; 93:20190948.
14. Avanzo M, Wei L, Stancanello J, et al. Machine and deep learning methods for radiomics. Med Phys 2020; 47:e185–e202.
15. Peeken JC, Spraker MB, Knebel C, et al. Tumor grading of soft tissue sarcomas using MRI-based radiomics. EBioMedicine 2019; 48:332–340.
16. Fields BKK, Demirjian NL, Hwang DH, et al. Whole-tumor 3D volumetric MRI-based radiomics approach for distinguishing between benign and malignant soft tissue tumors. Eur Radiol 2021; 31:8522–8535.
17. Liu F, Wang M, Li H. Role of perfusion parameters on DCE-MRI and ADC values on DWMRI for invasive ductal carcinoma at 3.0 Tesla. World J Surg Oncol 2018:16. https://doi.org/10.1186/s12957-018-1538-8
18. Wei W, Ji Y, Tang Z, et al. Breast magnetic resonance imaging can predict Ki67 discordance between core needle biopsy and surgical samples. J Magn Reson Imaging 2023; 57:85–94.
19. Lo Gullo R, Daimiel I, Morris EA, et al. Combining molecular and imaging metrics in cancer: radiogenomics. Insights Imaging 2020; 11:1–17.
20. Zhang L, Liu X, Xu X, et al. An integrative non-invasive malignant brain tumors classification and Ki-67 labeling index prediction pipeline with radiomics approach. Eur J Radiol 2023; 158:110639.
21. Yang Y, Zhang L, Wang T, et al. MRI fat-saturated T2–weighted radiomics model for identifying the Ki-67 index of soft tissue sarcomas. J Magn Reson Imaging 2022; 58:534–545. https://doi.org/10.1002/jmri.28518
22. Hu X, Zhou J, Li Y, et al. Added value of viscoelasticity for MRI-based prediction of Ki-67 expression of Hepatocellular Carcinoma using a deep learning combined radiomics (DLCR) model. Cancers (Basel) 2022; 14:2575.
23. Zheng Z, Gu Z, Xu F, et al. Magnetic resonance imaging-based radiomics signature for preoperative prediction of Ki67 expression in bladder cancer. Cancer Imaging 2021; 21:1–14.
24. Gong X, Guo Y, Zhu T, et al. Diagnostic performance of radiomics in predicting axillary lymph node metastasis in breast cancer: A systematic review and meta-analysis. Front Oncol 2022; 12.
25. O'Donnell JPM, Gasior SA, Davey MG, et al. The accuracy of breast MRI radiomic methodologies in predicting pathological complete response to neoadjuvant chemotherapy: a systematic review and network meta-analysis. Eur J Radiol 2022:110561.
26. Sha YS, Chen JF. MRI-based radiomics for the diagnosis of triple-negative breast cancer: a meta-analysis. Clin Radiol 2022; 77:655–663.
27. Davey MG, Davey MS, Boland MR, et al. Radiomic differentiation of breast cancer molecular subtypes using pre-operative breast imaging–a systematic review and meta-analysis. Eur J Radiol 2021; 144:109996.
28. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155:529–536.
29. Park JE, Kim D, Kim HS, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. Eur Radiol 2020; 30:523–536.
30. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017; 14:749–762.
31. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557–560.
32. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005; 58:882–893.
33. Castaldo R, Garbino N, Cavaliere C, et al. A complex radiomic signature in luminal breast cancer from a weighted statistical framework: a pilot study. Diagnostics 2022; 12. https://doi.org/10.3390/diagnostics12020499
34. Demircioglu A, Grueneisen J, Ingenwerth M, et al. A rapid volume of interest-based approach of radiomics analysis of breast MRI for tumor decoding and phenotyping of breast cancer. PLoS One 2020; 15:e0234871. https://doi.org/10.1371/journal.pone.0234871
35. Brancato V, Brancati N, Esposito G, et al. A two-step feature selection radiomic approach to predict molecular outcomes in breast cancer. Sensors (Basel) 2023; 23. https://doi.org/10.3390/s23031552
36. Liang C, Cheng Z, Huang Y, et al. An MRI-based radiomics classifier for preoperative prediction of Ki-67 status in breast cancer. Acad Radiol 2018; 25:1111–1117. https://doi.org/10.1016/j.acra.2018.01.006

37. Ma W, Ji Y, Qi L, et al. Breast cancer Ki67 expression prediction by DCE-MRI radiomics features. Clin Radiol 2018; 73(909):e1–909. https://doi.org/10.1016/j.crad.2018.05.027. e5.
38. Li Y-Z, Huang Y-H, Su X-Y, et al. Breast MRI segmentation and Ki-67 high- and low-expression prediction algorithm based on deep learning. Comput Math Methods Med 2022; 2022. https://doi.org/10.1155/2022/1770531
39. Qiao M, Liu C, Li Z, et al. Breast tumor classification based on MRI-US images by disentangling modality features. IEEE J Biomed Heal Informatics 2022; 26:3059–3067. https://doi.org/10.1109/JBHI.2022.3140236
40. Zhang L, Zhou X-X, Liu L, et al. Comparison of dynamic contrast-enhanced MRI and non-mono-exponential model-based diffusion-weighted imaging for the prediction of prognostic biomarkers and molecular subtypes of breast cancer based on radiomics. J Magn Reson Imaging 2023; 58:1590–1602. https://doi.org/10.1002/jmri.28611
41. Monti S, Aiello M, Incoronato M, et al. DCE-MRI pharmacokinetic-based phenotyping of invasive ductal carcinoma: A radiomic study for prediction of histological outcomes. Contrast Media Mol Imaging 2018; 2018. https://doi.org/10.1155/2018/5076269
42. Fan M, Cheng H, Zhang P, et al. DCE-MRI texture analysis with tumor subregion partitioning for predicting Ki-67 status of estrogen receptor-positive breast cancers. J Magn Reson Imaging 2018; 48:237–247. https://doi.org/10.1002/jmri.25921
43. He X, Zhou J, Ye S, et al. Differences in tumour heterogeneity based on dynamic contrast-enhanced MRI between tumour and peritumoural stroma for predicting Ki-67 status of invasive ductal carcinoma. Clin Radiol 2021; 76(470):e13–e470. https://doi.org/10.1016/j.crad.2020.12.008. e22.
44. Fan M, Liu Z, Xu M, et al. Generative adversarial network-based super-resolution of diffusion-weighted imaging: Application to tumour radiomics in breast cancer. NMR Biomed 2020; 33:.e4345 https://doi.org/10.1002/nbm.4345
45. Jiang T, Song J, Wang X, et al. Intratumoral and peritumoral analysis of mammography, tomosynthesis, and multiparametric MRI for predicting Ki-67 level in breast cancer: a radiomics-based study. Mol Imaging Biol 2022; 24:550–559. https://doi.org/10.1007/s11307-021-01695-w
46. Li C, Song L. Yin J. intratumoral and peritumoral radiomics based on functional parametric maps from breast DCE-MRI for prediction of HER-2 and Ki-67 status. J Magn Reson Imaging 2021; 54:703–714. https://doi.org/10.1002/jmri.27651
47. Zhang Y, Zhu Y, Zhang K, et al. Invasive ductal breast cancer: preoperative predict Ki-67 index based on radiomics of ADC maps. Radiol Med 2020; 125:109–116. https://doi.org/10.1007/s11547-019-01100-1
48. Fan M, Yuan W, Zhao W, et al. Joint prediction of breast cancer histological grade and Ki-67 expression level based on DCE-MRI and DWI radiomics. IEEE J Biomed Heal Informatics 2020; 24:1632–1642. https://doi.org/10.1109/JBHI.2019.2956351
49. Song SE, Cho KR, Cho Y, et al. Machine learning with multiparametric breast MRI for prediction of Ki-67 and histologic grade in early-stage luminal breast cancer. Eur Radiol 2022; 32:853–863. https://doi.org/10.1007/s00330-021-08127-x
50. Umutlu L, Kirchner J, Bruckmann NM, et al. Multiparametric integrated (18)F-FDG PET/MRI-based radiomics for breast cancer phenotyping and tumor decoding. Cancers (Basel) 2021; 13. https://doi.org/10.3390/cancers13122928
51. Ni M, Zhou X, Liu J, et al. Prediction of the clinicopathological subtypes of breast cancer using a fisher discriminant analysis model based on radiomic features of diffusion-weighted MRI. BMC Cancer 2020; 20:1073. https://doi.org/10.1186/s12885-020-07557-y
52. Liu W, Cheng Y, Liu Z, et al. Preoperative prediction of Ki-67 status in breast cancer with multiparametric mri using transfer learning. Acad Radiol 2021; 28:e44–e53. https://doi.org/10.1016/j.acra.2020.02.006
53. Zhang L, Fan M, Wang S, et al. Radiomic analysis of pharmacokinetic heterogeneity within tumor based on the unsupervised decomposition of dynamic contrast-enhanced MRI for predicting histological characteristics of breast cancer. J Magn Reson Imaging 2022; 55:1636–1647. https://doi.org/10.1002/jmri.27993
54. Zhou X, Gao F, Duan S, et al. Radiomic features of Pk-DCE MRI parameters based on the extensive Tofts model in application of breast cancer. Phys Eng Sci Med 2020; 43:517–524. https://doi.org/10.1007/s13246-020-00852-9
55. Feng S, Yin J. Radiomics of dynamic contrast-enhanced magnetic resonance imaging parametric maps and apparent diffusion coefficient maps to predict Ki-67 status in breast cancer. Front Oncol 2022; 12. https://doi.org/10.3389/fonc.2022.847880
56. Pesapane F, Rotili A, Botta F, et al. Radiomics of MRI for the prediction of the pathological response to neoadjuvant chemotherapy in breast cancer patients: a single referral centre analysis. Cancers (Basel) 2021; 13. https://doi.org/10.3390/cancers13174271
57. Kayadibi Y, Kocak B, Ucar N, et al. Radioproteomics in breast cancer: prediction of Ki-67 expression with MRI-based radiomic models. Acad Radiol 2022; 29:S116–S125. https://doi.org/10.1016/j.acra.2021.02.001
58. Santucci D, Faiella E, Cordelli E, et al. The impact of tumor edema on T2-weighted 3T-MRI invasive breast cancer histological characterization: a pilot radiomics study. Cancers (Basel) 2021; 13. https://doi.org/10.3390/cancers13184635
59. Sun K, Zhu H, Chai W, et al. Whole-lesion histogram and texture analyses of breast lesions on inline quantitative DCE mapping with CAIPIRINHA-Dixon-TWIST-VIBE. Eur Radiol 2020; 30:57–65. https://doi.org/10.1007/s00330-019-06365-8
60. Wang W, Zhang X, Zhu L, et al. Prediction of prognostic factors and genotypes in patients with breast cancer using multiple mathematical models of MR diffusion imaging. Front Oncol 2022; 12:825264.
61. Fan M, He T, Zhang P, et al. Heterogeneity of diffusion-weighted imaging in tumours and the surrounding stroma for prediction of Ki-67 proliferation status in breast cancer. Sci Rep 2017; 7:1–9.
62. Saha A, Harowicz MR, Grimm LJ, et al. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. Br J Cancer 2018; 119:508–516.
63. Ming W, Zhu Y, Bai Y, et al. Radiogenomics analysis reveals the associations of dynamic contrast-enhanced–MRI features with gene expression characteristics, PAM50 subtypes, and prognosis of breast cancer. Front Oncol 2022; 12. https://doi.org/10.3389/fonc.2022.943326
64. Boros M, Moncea D, Moldovan C, et al. Intratumoral heterogeneity for Ki-67 index in invasive breast carcinoma: a study on 131 consecutive cases. Appl Immunohistochem Mol Morphol 2017; 25:338–340.
65. Chen X, Zhu S, Fei X, et al. Surgery time interval and molecular subtype may influence Ki67 change after core needle biopsy in breast cancer patients. BMC Cancer 2015; 15:1–10.
66. Li S, Chen X, Shen K. Association of Ki-67 change pattern after core needle biopsy and prognosis in HR+/HER2– early breast cancer patients. Front Surg 2022; 9.
67. Tong Y, Dai J, Huang J, et al. Ki67 increase after core needle biopsy associated with worse disease outcome in HER2-negative breast cancer patients. Sci Rep 2023; 13:2489.
68. Rouvière O, Jaouen T, Baseilhac P, et al. Artificial intelligence algorithms aimed at characterizing or detecting prostate cancer on MRI: How accurate are they when tested on independent cohorts?–A systematic review. Diagn Interv Imaging 2022; 104:221–234.
69. Kim HY, Cho SJ, Sunwoo L, et al. Classification of true progression after radiotherapy of brain metastasis on MRI using artificial intelligence: a systematic review and meta-analysis. Neuro-Oncology Adv 2021; 3:vdab080.
70. Liang X, Yu X, Gao T. Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: a systematic review and meta-analysis. Eur J Radiol 2022; 150:110247.
71. Zhang J, Li L, Zhe X, et al. The diagnostic performance of machine learning-based radiomics of DCE-MRI in predicting axillary lymph node metastasis in breast cancer: a meta-analysis. Front Oncol 2022; 12:799209.
72. Granzier RWY, van Nijnatten TJA, Woodruff HC, et al. Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: a systematic review. Eur J Radiol 2019; 121:108736.
73. Shin JK, Kim JY. Dynamic contrast-enhanced and diffusion-weighted MRI of estrogen receptor-positive invasive breast cancers: Associations

between quantitative MR parameters and Ki-67 proliferation status. J Magn Reson Imaging 2017; 45:94–102. https://doi.org/10.1002/jmri.25348

74. Zhang L, Tang M, Min Z, et al. Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis. Acta Radiol 2016; 57:651–660.

75. Jia L-L, Zhao J-X, Zhao L-P, et al. Current status and quality of radiomic studies for predicting KRAS mutations in colorectal cancer patients: A systematic review and meta-analysis. Eur J Radiol 2022; 158:110640.