



(REVIEW ARTICLE)



DIAGNOSTIC ACCURACY OF ARTIFICIAL INTELLIGENCE IN THE UPSTAGING PREDICTION OF DUCTAL CARCINOMA IN SITU: A SYSTEMATIC REVIEW AND META-ANALYSIS

Titania Sophia Qelmi ¹, Adrian Faza Raditya ¹, Kiandra Annisa Nandini ² and Citrawati Dyah Kencono Wungu ^{3,*}

¹ Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia.

² Department of Medicine, Faculty of Medicine and Health, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

³ Department of Physiology and Medical Biochemistry, Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia.

* Corresponding Author

ORCID Details

Titania Sophia Qelmi: <https://orcid.org/0009-0003-9088-7193>

Adrian Faza Raditya: <https://orcid.org/0009-0009-5160-4529>

Kiandra Annisa Nandini: <https://orcid.org/0009-0001-9478-0938>

Citrawati Dyah Kencono Wungu: <https://orcid.org/0000-0001-5180-957X>

World Journal of Advanced Research and Reviews, 2026, 32(01), 802–811

Publication history: Received on 31 August 2026; revised on 10 October 2026; accepted on 12 October 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.32.1.2580>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Background: Ductal carcinoma in situ (DCIS) may be upstaged to invasive carcinoma after surgical excision. Predicting upstaging before surgery may help guide treatment planning. Artificial intelligence (AI)-based imaging has been investigated for this purpose, but its overall performance remains unclear.

Objective: To evaluate the diagnostic accuracy of AI-based systems for predicting DCIS upstaging to invasive carcinoma.

Materials and Methods: A systematic review and meta-analysis was conducted following PRISMA 2020 and registered in PROSPERO (CRD420261516994). PubMed, Scopus, EBSCOhost, ProQuest, and Web of Science were searched from inception to September 2026. Studies evaluating AI-based imaging models for DCIS upstaging were included. Pooled sensitivity and specificity were calculated using a bivariate random-effects model. Risk of bias was assessed using QUADAS-2.

Results: Six studies involving 701 participants were included. Pooled sensitivity was 0.77 (95% CI, 0.57–0.89) and specificity was 0.78 (95% CI, 0.64–0.87), with substantial heterogeneity ($I^2 = 84.0\%$ and 90.6% , respectively). The pooled log diagnostic odds ratio (DOR) was 2.15 (95% CI, 1.64–2.65). The SROC analysis showed moderate overall discriminative performance. No significant differences were observed across imaging modalities, while AI model type significantly affected specificity but not sensitivity. Deeks' funnel plot suggested possible publication bias ($p = 0.046$).

Conclusion: AI-based imaging showed moderate accuracy for predicting DCIS upstaging. However, substantial heterogeneity and methodological limitations limit confidence in these findings. Further prospective studies with external validation are needed.

Keywords: Artificial-Intelligence; Breast Imaging; DCIS; Diagnostic accuracy; Upstaging

1 INTRODUCTION

Ductal carcinoma in situ (DCIS) is a non-invasive breast neoplasm in which malignant epithelial cells remain confined to the ductal system [1]. Its detection has increased with the widespread use of mammographic screening [2], accounting for approximately 16–28% of screen-detected breast cancers [3]. DCIS most commonly appears as microcalcifications on mammography and non-mass enhancement on magnetic resonance imaging (MRI), although some cases may be mammographically occult [1,4]. Its biological heterogeneity and variable potential for progression to invasive carcinoma raise concerns about overdiagnosis and overtreatment [1,5]. Moreover, 17–26% of biopsy-diagnosed DCIS cases may be upstaged to invasive carcinoma following surgical excision [6,7], affecting treatment planning, including decisions regarding sentinel lymph node biopsy [8].

Radiological assessment of DCIS remains challenging because its imaging features may overlap with benign lesions [1], and interpretation of microcalcifications shows considerable inter-reader variability [9]. Artificial intelligence (AI), particularly machine learning and deep learning, has therefore been investigated to improve the assessment of breast imaging [10]. In DCIS, AI applications include diagnosis, prediction of occult invasive disease (upstaging), and classification. Previous studies have reported promising performance across different imaging modalities, with AUC values ranging from approximately 0.70 to 0.98 [11,12]. However, performance varies across clinical tasks, imaging modalities, and AI approaches.

Previous systematic reviews have evaluated AI mainly in breast cancer screening [13], while a recent review specifically examined radiological AI in DCIS [14]. Although the latter included 46 studies, substantial heterogeneity limited its findings to narrative synthesis, without pooled estimates of sensitivity and specificity [14]. Therefore, this systematic review and meta-analysis aims to evaluate the diagnostic accuracy of AI-based systems in predicting upstaging from biopsy-diagnosed DCIS to invasive carcinoma.

2 MATERIALS AND METHODS

2.1 Study Design and Protocol Registration

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [15]. The study protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) under registration number CRD420261516994 [16].

2.2 Search Strategy

A systematic literature search was conducted in PubMed, SCOPUS, EBSCOhost, ProQuest, and Web of Science from inception to September 2026. The search strategy incorporated relevant keywords, Medical Subject Headings (MeSH), and Boolean operators. The complete search strategies for each database are provided in Supplementary File S1.

2.3 Eligibility Criteria

According to Supplementary File S2, studies were eligible if they included adults (≥ 18 years) undergoing breast imaging for suspected or screen-detected breast lesions, diagnostic or screening mammography, or suspected/confirmed ductal carcinoma in situ (DCIS) or other non-invasive breast lesions; asymptomatic screening populations were also eligible when DCIS detection was evaluated. Eligible interventions comprised artificial intelligence (AI)-based radiologic diagnostic systems, including machine learning (ML), deep learning (DL), convolutional neural networks (CNNs), computer-aided detection/diagnosis (AI-based CAD), and other AI/ML algorithms used to detect or classify DCIS on breast imaging. No comparator was required. Primary outcomes included true-positive, false-positive, true-negative, and false-negative results, with diagnostic accuracy assessed using sensitivity, specificity, the summary receiver operating characteristic (SROC) curve, and log diagnostic odds ratio (DOR). Secondary outcomes included diagnostic accuracy according to imaging modality, AI model, geographical region, and study type.

2.4 Study Selection

All retrieved records were imported into rayyan.ai to identify and remove duplicates [17]. Two independent reviewers screened titles and abstracts, followed by full-text assessment of potentially eligible studies. Disagreements were resolved through discussion or consultation with a third reviewer. The selection process is illustrated in a PRISMA flow diagram.

2.5 Data Extraction

Two independent reviewers extracted data using a standardized data extraction form. Extracted information included the first author and year of publication, country and study design, sample size and participant characteristics, intervention and comparator characteristics, follow-up duration, primary and secondary outcomes, and relevant effect estimates with measures of variability. Any discrepancies between reviewers were resolved through consensus or consultation with a third reviewer.

2.6 Risk-of-Bias Assessment

The methodological quality and risk of bias of included studies were independently assessed by two reviewers using QUADAS-2 [18]. Disagreements were resolved through discussion or consultation with a third reviewer. The results were summarized using ROBVIS [19].

2.7 Statistical Analysis

Statistical analyses were conducted using Rstudio software version 2026.01.1+403 (RStudio, PBC, Boston, MA, USA) with the mada, meta, metafor, and dplyr packages [20]. A bivariate random-effects model based on the Reitsma approach was used to jointly synthesize sensitivity and specificity, construct summary receiver operating characteristic (SROC) curves for evaluating the overall diagnostic performance of AI-based models for sepsis diagnosis. Separate pooled estimates of sensitivity and specificity were additionally calculated using generalized linear mixed-effects models (GLMM) with logit transformation. Pooled diagnostic odds ratios (DOR) were synthesized using a random-effects model with restricted maximum likelihood (REML) estimation. Between-study heterogeneity was assessed using Cochran's Q test and the I^2 statistic. Potential sources of heterogeneity were explored through subgroup and sensitivity analyses according to imaging modality, AI model, geographical region, and study type. Publication bias was assessed using Deeks' Funnel Plot. A p-value of <0.05 was considered statistically significant. The complete Rstudio code is shown in Supplementary File S3.

2.8 Certainty of Evidence

The certainty of evidence for each relevant outcome was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, considering risk of bias, inconsistency, indirectness, imprecision, and publication bias [21].

3 RESULTS AND DISCUSSION

3.1 Study Selection

The initial search identified 1000 records across the selected databases. After removing 393 duplicates, 607 records were screened by title and abstract. Subsequently, 51 full-text articles were assessed for eligibility, of which 6 studies met the inclusion criteria and were included in the systematic review and meta-analysis. The selection process is presented in Supplementary File S4.

3.2 Characteristics of Included Studies

A total of 6 studies involving 701 participants were included [22–27]. The studies were conducted in China, Japan, South Korea, Turkey, and the USA. Studies were published between 2022 and 2026. The principal characteristics of the included studies are summarized in Supplementary File S5.

3.3 Risk-of-Bias Assessment

The risk-of-bias assessment demonstrated low risk of bias for patient selection and reference standard across all included studies, while the index test domain showed high risk of bias in three studies and some concerns in two studies. The flow and timing domain was judged as low risk in three studies and raised some concerns in three studies. Overall, three studies were judged as having high risk of bias, while three had some concerns. The detailed assessments are presented in Supplementary File S6.

3.4 Primary Outcome

3.4.1 Pooled Sensitivity and Specificity

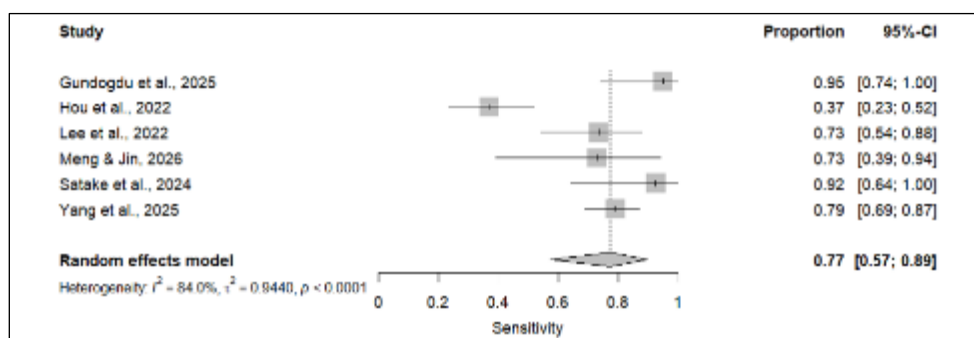


Figure 1: Forest plot of pooled sensitivity of AI-based models for predicting DCIS upstaging.

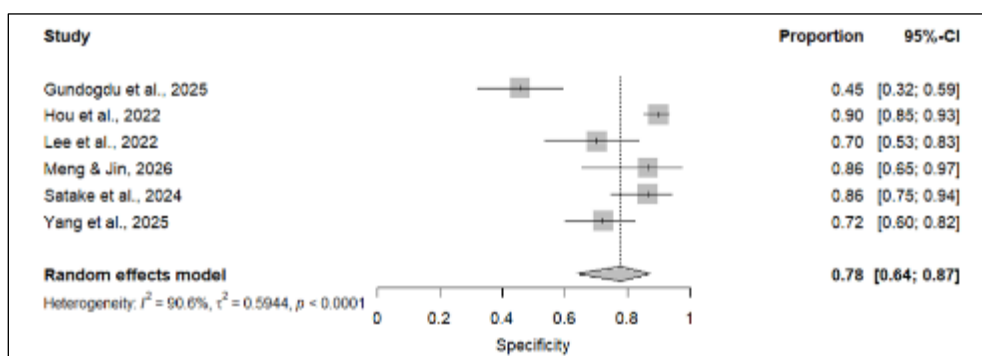


Figure 2: Forest plot of pooled specificity of AI-based models for predicting DCIS upstaging.

Based on Figure 1 and Figure 2, the pooled diagnostic sensitivity of AI for DCIS upstaging prediction across six studies was 0.77 (95% CI 0.57–0.89), with substantial heterogeneity ($I^2 = 84.0\%$, $\tau^2 = 0.9440$, $p < 0.0001$). Individual study sensitivities ranged from 0.37 to 0.95. The pooled specificity was 0.78 (95% CI 0.64–0.87), also demonstrating substantial heterogeneity ($I^2 = 90.6\%$, $\tau^2 = 0.5944$, $p < 0.0001$), with individual estimates ranging from 0.45 to 0.90. These findings indicate considerable variability in diagnostic performance across the included studies.

3.4.2 SROC Curve

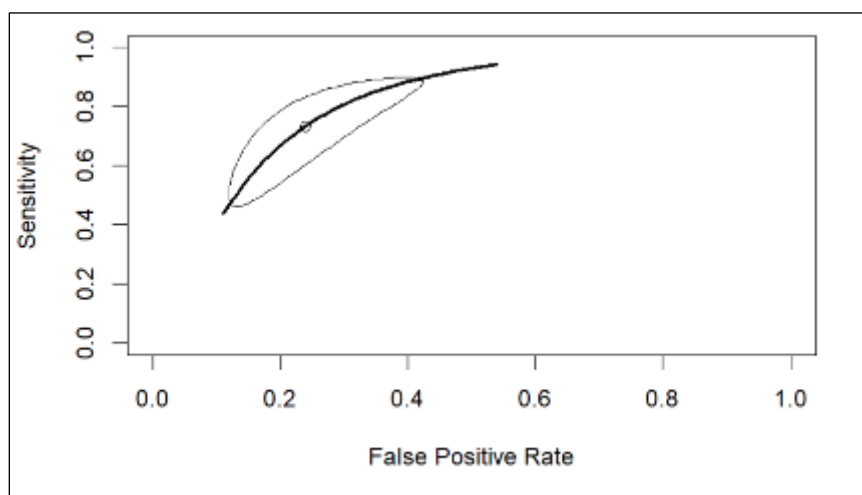


Figure 3: Summary receiver operating characteristic (SROC) curve of artificial intelligence–based models for predicting pathological upstaging of ductal carcinoma in situ.

According to Figure 3, the SROC curve illustrates the overall diagnostic performance of AI-based approaches for predicting DCIS upstaging, with the summary operating point reflecting the pooled sensitivity and false-positive rate.

The SROC curve demonstrated favorable overall diagnostic performance, with an AUC of 0.812, indicating good discriminative ability of AI-based models for predicting pathological upstaging of ductal carcinoma in situ (DCIS).

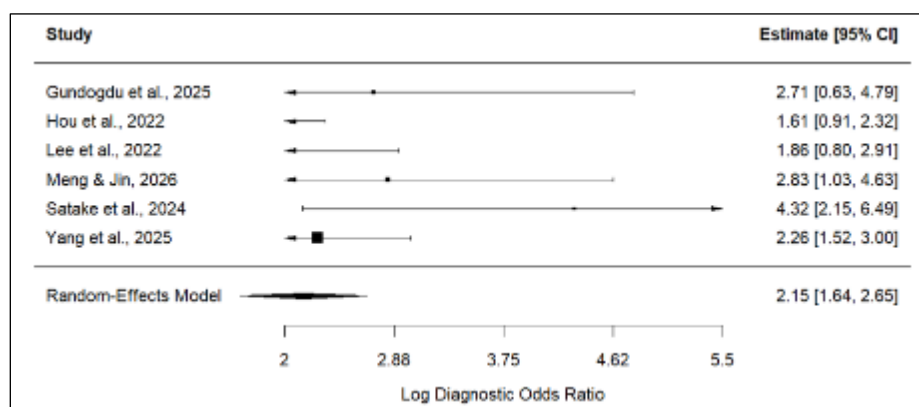


Figure 4: Forest plot of log diagnostic odds ratios (DORs) of AI models for DCIS upstaging prediction.

The forest plot of diagnostic odds ratios (DORs) showed variation in diagnostic performance across the six included studies, with individual log DOR estimates ranging from 1.61 to 4.32. The random-effects model yielded a pooled log DOR of 2.15 (95% CI 1.64–2.65), indicating an overall association between AI-based assessment and DCIS upstaging prediction. The forest plot is shown in Figure 4.

3.5 Subgroup Analyses

3.5.1 Subgroup by Imaging Modality

Subgroup analysis by imaging modality demonstrated pooled sensitivities of 0.75 (95% CI 0.18–0.97) for mammography, 0.78 (95% CI 0.65–0.87) for breast MRI, and 0.79 (95% CI 0.69–0.87) for ultrasound. The corresponding pooled specificities were 0.73 (95% CI 0.35–0.93), 0.81 (95% CI 0.70–0.88), and 0.72 (95% CI 0.60–0.82), respectively. Substantial heterogeneity was observed in the mammography subgroup for both sensitivity ($I^2 = 90.3\%$) and specificity ($I^2 = 97.9\%$), while the breast MRI subgroup demonstrated no significant heterogeneity in sensitivity ($I^2 = 0\%$) and moderate heterogeneity in specificity ($I^2 = 54.1\%$). The tests for subgroup differences were not statistically significant for either sensitivity ($p = 0.9769$) or specificity ($p = 0.4422$), indicating no significant differences in diagnostic performance across imaging modalities. The forest plots of subgroup analysis by imaging modality are shown in Supplementary File S7 and Supplementary File S8.

3.5.2 Subgroup by AI Model

Subgroup analysis by AI model demonstrated a pooled sensitivity of 0.95 (95% CI 0.74–1.00) for deep learning, 0.72 (95% CI 0.39–0.91) for radiomics, and 0.73 (95% CI 0.58–0.84) for radiomics combined with machine learning. Regarding specificity, radiomics achieved a pooled estimate of 0.84 (95% CI 0.74–0.91), while radiomics combined with machine learning demonstrated a pooled specificity of 0.76 (95% CI 0.64–0.85). The test for subgroup differences was statistically significant for specificity ($p < 0.0001$), but not for sensitivity ($p = 0.2094$), indicating variation in specificity across AI model categories without statistically significant differences in sensitivity. The forest plots of subgroup analysis by AI model are shown in Supplementary File S9 and Supplementary File S10.

3.5.3 Subgroup by Geographical Region

Subgroup analysis by geographical region revealed a pooled sensitivity of 0.80 (95% CI 0.73–0.86) among studies conducted in Asia, with low heterogeneity ($I^2 = 9.1\%$), whereas the single North American study reported a sensitivity of 0.37 (95% CI 0.23–0.52). For specificity, the pooled estimate in Asia was 0.73 (95% CI 0.58–0.84), with substantial heterogeneity ($I^2 = 83.1\%$), while the North American study reported a specificity of 0.90 (95% CI 0.85–0.93). The tests for subgroup differences were statistically significant for both sensitivity ($p < 0.0001$) and specificity ($p = 0.0043$). However, the North American subgroup consisted of only one study, limiting the interpretation of these geographical differences. The forest plots of subgroup analysis by geographical region are shown in Supplementary File S11 and Supplementary File S12.

3.5.4 Subgroup by Study Type

Subgroup analysis by study design demonstrated pooled sensitivities of 0.75 (95% CI 0.18–0.97) for retrospective cohort studies, 0.78 (95% CI 0.65–0.87) for retrospective studies, and 0.79 (95% CI 0.69–0.87) for retrospective multicentre studies. The corresponding pooled specificities were 0.73 (95% CI 0.35–0.93), 0.81 (95% CI 0.70–0.88), and

0.72 (95% CI 0.60–0.82), respectively. Heterogeneity was substantial in the retrospective cohort subgroup for both sensitivity ($I^2 = 90.3\%$) and specificity ($I^2 = 97.9\%$), while the retrospective study subgroup showed no significant heterogeneity in sensitivity ($I^2 = 0\%$) and moderate heterogeneity in specificity ($I^2 = 54.1\%$). The tests for subgroup differences were not statistically significant for sensitivity ($p = 0.9769$) or specificity ($p = 0.4422$), suggesting no statistically significant variation in diagnostic performance according to study design. The forest plots of subgroup analysis by study design are shown in Supplementary File S13 and Supplementary File S14.

3.5.5 Sensitivity Analyses

The leave-one-out sensitivity analysis demonstrated that the pooled sensitivity remained relatively stable following the sequential exclusion of individual studies. According to Supplementary File S15, the pooled sensitivity across the five remaining studies ranged from 0.784 to 0.836, compared with the overall pooled sensitivity of 0.804. Excluding Gundogdu et al. (2025) resulted in the lowest pooled sensitivity (0.784), whereas excluding Yang et al. (2025) yielded the highest estimate (0.836). These findings suggest that the pooled sensitivity was not substantially influenced by the exclusion of any single study.

Based on Supplementary File S16, the leave-one-out specificity analysis showed that the pooled specificity varied from 0.685 to 0.782 following the sequential exclusion of individual studies, compared with the overall pooled specificity of 0.733. Excluding Satake et al. (2024) yielded the lowest pooled specificity (0.685), whereas excluding Gundogdu et al. (2025) resulted in the highest estimate (0.782). The variation in pooled estimates indicates some influence of individual studies on the overall specificity, although the estimates remained within a relatively consistent range.

3.6 Publication Bias

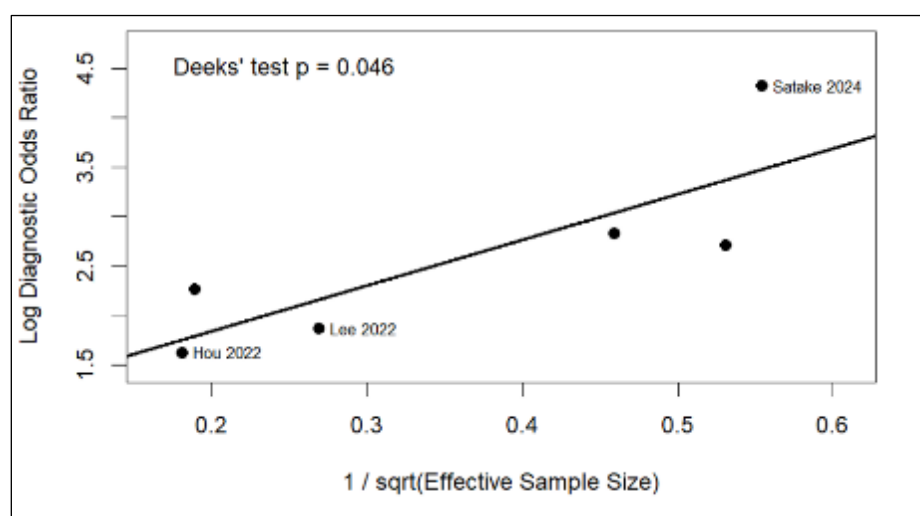


Figure 5: Deeks' funnel plot for assessment of potential publication bias in studies evaluating artificial intelligence-based prediction of pathological upstaging of ductal carcinoma in situ.

The Deeks' funnel plot demonstrated potential publication bias among the included studies, with a statistically significant Deeks' test ($p = 0.046$). The Deek's funnel plot is shown in Figure 5.

3.7 Certainty of Evidence

The certainty of evidence for the diagnostic accuracy estimates was assessed using the GRADE framework. Two reviewers independently evaluated the certainty of evidence across five domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Absolute effects were presented per 1,000 patients tested across pre-test probabilities of 5%, 20%, and 50%. Certainty was assessed separately for true positives, false negatives, true negatives, and false positives. According to Table 1, overall certainty was rated as very low for all outcomes, with downgrading primarily due to risk of bias, substantial heterogeneity, imprecision, and suspected publication bias. Heterogeneity was substantial for both sensitivity- and specificity-related outcomes ($I^2 = 84.0\%$ and 90.6% , respectively), while Deeks' funnel plot demonstrated statistically significant asymmetry ($p = 0.046$).

Table 1: GRADE assessment result

Outcome Assessed	Number of Studies (Number of Patients)	Certainty
True positives	6 studies	Very low
False negatives	207 patients	
True negatives	6 studies	Very low
False positives	494 patients	

4 DISCUSSION

The present systematic review and meta-analysis evaluated the accuracy of imaging-based AI models for predicting pathological upstaging in patients with biopsy-confirmed ductal carcinoma in situ (DCIS). Pooled sensitivity was 0.77 (95% CI 0.57–0.89) and specificity was 0.78 (95% CI 0.64–0.87), indicating moderate performance. However, considerable heterogeneity in sensitivity ($I^2 = 84.0\%$) and specificity ($I^2 = 90.6\%$), along with high overall risk of the studies bias limits the applicability of these summary estimates across clinical settings.

These findings broadly align with previous reviews identifying potential applications of AI in DCIS assessment alongside substantial methodological variability. A systematic review of AI-assisted histological DCIS diagnosis reported sensitivity of 83.88% and specificity of 85.49%; however, direct comparison is inappropriate because histological classification and preoperative upstaging prediction involve different inputs and target conditions. Similarly, Wu et al. highlighted heterogeneous performance, predominantly retrospective evidence, and limited external validation in radiological DCIS assessment [14,28]. Imaging-derived texture and enhancement features may capture characteristics associated with occult invasion, although these associations do not establish a biological mechanism [29].

Subgroup analysis showed the highest sensitivity estimates of 0.95 (95% CI 0.74–1.00) for deep learning, while a noticeable lower result for radiomics (0.72 (95% CI 0.39–0.91)) and radiomics combined with machine learning (0.73 (95% CI 0.58–0.84)). The deep-learning estimate originated from one study and was accompanied by specificity of approximately 0.45, precluding conclusions about superiority. A study by Shi et al., (2018) also demonstrated that deep-learning approaches do not uniformly produce superior discrimination [12]. Conversely, Hou et al. (2022) reported sensitivity of 37% at a high-specificity operating point of 90%, illustrating how threshold selection can produce contrasting sensitivity–specificity profiles [23].

Sensitivity heterogeneity was lower among Asian studies (I^2 approximately 9%) and breast MRI studies ($I^2 = 0\%$). These findings suggest greater consistency within selected subsets, but the small subgroup sizes and overlapping study characteristics prevent attributing this reduction independently to geography or imaging modality. Furthermore, the pooled log DOR of 2.15 (95% CI 1.64–2.65) corresponds to a DOR of 8.58 (95% CI 5.16–14.15), supporting discriminatory ability without establishing sufficient accuracy for a particular clinical decision [30].

Clinically, these findings support further evaluation of AI as an adjunct to preoperative risk assessment. A study by Slavkova et al. (2025) found that combining MRI radiomic features with clinical and qualitative imaging information improved upstaging prediction [31]. Separately, Meurs et al. (2023) externally validated a clinicopathological upstaging model, reporting a decision-curve evidence of potential net benefit for selecting patients for sentinel lymph-node biopsy [32]. Together, these studies provide a rationale for evaluating whether imaging-based AI adds clinically useful information to existing risk assessment. However, neither establishes that AI-guided management reduces unnecessary procedures or safely avoids axillary staging. Given the moderate pooled accuracy and substantial variability observed in this review, prospective clinical-impact studies are needed before predictions are used to guide treatment de-escalation. The current evidence therefore does not establish that AI-guided decisions improve outcomes or safely support treatment de-escalation.

4.1 Strengths

This study possesses several strengths. Quantitative synthesis across several imaging modalities, structured assessment of risk of bias and certainty, leave-one-out sensitivity analyses, and examination of funnel-plot asymmetry was done thoroughly. Multiple subgroup analyses explored potential sources of heterogeneity. These approaches improve transparency and characterise uncertainty, although they cannot eliminate limitations in the underlying studies.

4.2 Limitations

However, several limitations persist. The evidence was limited by six studies, retrospective designs, and small cohorts. Four of the six quantitatively synthesised studies were judged at high overall risk of bias, primarily concerning the index test, needing further caution about possible optimism in reported performance. Deeks' test indicated statistically significant funnel-plot asymmetry ($p = 0.046$), raising concern about small-study effects and possible publication bias. The small subgroups limit comparisons between models and modalities with low subgroup I^2 values should not be interpreted as proof of reproducibility.

4.3 Implications for future research

Larger prospective, multicentre studies with independent external validation are warranted. Future studies should prespecify models and clinically relevant thresholds, standardise upstaging definitions, and report calibration alongside discrimination. Comparisons with established clinicopathological assessment and prospective evaluation of clinical net benefit are needed to determine whether AI improves surgical planning without increasing missed invasion or unnecessary procedures.

5 CONCLUSION

Imaging-based AI models demonstrated moderate pooled sensitivity of 0.77 (95% CI 0.57–0.89) and specificity of 0.78 (95% CI 0.64–0.87) for predicting DCIS upstaging. Considerable heterogeneity, frequent high risk of bias, and possible publication bias limit confidence in clinical applicability. Artificial-intelligence warrants further investigation as an adjunct to preoperative assessment, but prospective external validation and evidence of clinical benefit are required before routine implementation.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declared no potential competing interest with respect to the research, authorship, and/or publication of this article.

Author contributions

TSQ: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Visualization, Writing – Original Draft, Writing – Review & Editing, Project Administration. AFR: Investigation, Writing – Original Draft, Writing – Review & Editing. KAN: Investigation, Writing – Original Draft, Writing – Review & Editing. CDKW: Validation, Supervision. All authors reviewed and approved the final manuscript.

Data availability

All data analyzed in this study were extracted from previously published studies and are available in the article and its supplementary materials.

Supplementary Material

Supplementary materials are published as provided by the authors and can be accessed at below link. <https://archive.org/details/supplementary-material-wjarr-2026-2580>

REFERENCES

- [1] WHO Classification of Tumours Editorial Board. Breast tumours. 5th ed. Lyon (France): International Agency for Research on Cancer; 2019. (WHO classification of tumours; vol. 2).
- [2] Virnig BA, Tuttle TM, Shamlivan T, Kane RL. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. *J Natl Cancer Inst.* 2010;102(3):170-8. doi:10.1093/jnci/djp482
- [3] Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst.* 2002;94(20):1546-54. doi:10.1093/jnci/94.20.1546
- [4] Bragg A, Candelaria R, Adrada B, Huang M, Rauch G, Santiago L, et al. Imaging of noncalcified ductal carcinoma in situ. *J Clin Imaging Sci.* 2021;11:34. doi:10.25259/JCIS_48_2021

- [5] Hwang ES, Hyslop T, Lynch T, Ryser MD, Weiss A, Wolf A, et al. Active monitoring with or without endocrine therapy for low-risk ductal carcinoma in situ: the COMET randomized clinical trial. *JAMA*. 2025;333(11):972-80. doi:10.1001/jama.2024.26698
- [6] Brennan ME, Turner RM, Ciatto S, et al. Ductal carcinoma in situ at core-needle biopsy: meta-analysis of underestimation and predictors of invasive breast cancer. *Radiology*. 2011;260(1):119-28. doi:10.1148/radiol.11102368
- [7] Grimm LJ, Ryser MD, Partridge AH, Thompson AM, Thomas JS, Wesseling J, et al. Surgical upstaging rates for vacuum assisted biopsy proven DCIS: implications for active surveillance trials. *Ann Surg Oncol*. 2017;24(12):3534-40. doi:10.1245/s10434-017-6018-9
- [8] Shin YD, Lee HM, Choi YJ. Necessity of sentinel lymph node biopsy in ductal carcinoma in situ patients: a retrospective analysis. *BMC Surg*. 2021;21:159. doi:10.1186/s12893-021-01170-x
- [9] Lazarus E, Mainiero M, Schepps B, et al. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology*. 2006;239(2):385-91. doi:10.1148/radiol.2392042127
- [10] McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94. doi:10.1038/s41586-019-1799-6
- [11] Yin H, Jiang Y, Xu Z, Huang W, Chen T, Lin G. Apparent diffusion coefficient-based convolutional neural network model can be better than sole diffusion-weighted magnetic resonance imaging to improve the differentiation of invasive breast cancer from breast ductal carcinoma in situ. *Front Oncol*. 2022;11:805911. doi:10.3389/fonc.2021.805911
- [12] Shi B, Grimm LJ, Mazurowski MA, Baker JA, Marks JR, King LM, et al. Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features. *J Am Coll Radiol*. 2018;15(3 Pt B):527-34.
- [13] Freeman K, Geppert J, Stinton C, et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ*. 2021;374:n1872. doi:10.1136/bmj.n1872
- [14] Wu C, Bartak A, Sharma R. Artificial intelligence in the radiologic assessment of ductal carcinoma in situ: a systematic review. *Front Oncol*. 2026;16:1870768. doi:10.3389/fonc.2026.1870768
- [15] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow C D, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*. 2021;372(71). doi:10.1136/bmj.n71
- [16] Schiavo JH. PROSPERO: An International Register of Systematic Review Protocols. *Medical Reference Services Quarterly*. 2019 Apr 3;38(2):171–80. doi:10.1080/02763869.2019.1588072
- [17] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a Web and Mobile App for Systematic Reviews. *Systematic Reviews*. 2016 Dec;5(1). doi:10.1186/s13643-016-0384-4
- [18] Whiting PF. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine*. 2011 Oct 18;155(8):529. doi:10.7326/0003-4819-155-8-201110180-00009
- [19] McGuinness LA, Higgins JPT. Risk - of - bias VISualization (robvis): An R package and Shiny web app for visualizing risk - of - bias assessments. *Research Synthesis Methods*. 2020 May 6;12(1). doi:10.1002/jrsm.1411
- [20] R Core Team. R Foundation for Statistical Computing [Internet]. 2025 [cited 2026 Oct 7]. R: A Language and Environment for Statistical Computing. Available from: <https://www.r-project.org/>
- [21] Guyatt GH, Oxman A D, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008 Apr 24;336(7650):924–6. doi:10.1136/bmj.39489.470347.ad
- [22] Gundogdu A, Wetherilt CS, Alpar A, Abdullah S, Yilmaz OC, Celik L. Invasion prediction with artificial intelligence in ductal carcinoma in situ (DCIS) patients: a proof-of-concept study. *BMC Surgery*. 2025 Oct 21;25(1). doi:10.1186/s12893-025-03252-6
- [23] Hou R, Grimm LJ, Mazurowski MA, Marks JR, King LM, Maley CC, et al. Prediction of Upstaging in Ductal Carcinoma in Situ Based on Mammographic Radiomic Features. *Radiology*. 2022 Apr 1;303(1):54–62. doi:10.1148/radiol.210407
- [24] Lee H, Park JH, Nguyen AT, Do LN, Park MH, Lee JS, et al. Prediction of the histologic upgrade of ductal carcinoma in situ using a combined radiomics and machine learning approach based on breast dynamic contrast-enhanced magnetic resonance imaging. *Frontiers in Oncology*. 2022 Nov 2;12. doi:10.3389/fonc.2022.1032809

- [25] Meng Y, Jin H. MRI-based radiomics with SHAP interpretation for preoperative prediction of upstaging in ductal carcinoma in situ: a comparative study of intratumoral, peritumoral, habitat, and fusion models. *Translational Cancer Research*. 2026 Aug;15(8):649–9. doi:10.21037/tcr-2026-1601
- [26] Satake H, Kinoshita F, Ishigaki S, Kato K, Jo Y, Shimada S, et al. Predictive Performance of Radiomic Features Extracted from Breast MR Imaging in Postoperative Upgrading of Ductal Carcinoma in Situ to Invasive Carcinoma. *Magnetic Resonance in Medical Sciences*. 2025;24(4):n/a. doi:10.2463/mrms.mp.2023-0168
- [27] Yang L, Li X, Wang Z, Li Q, Fu J, Zou X, et al. Super - Resolution Ultrasound Radiomics Can Predict the Upstaging of Ductal Carcinoma In Situ. *Cancer Medicine*. 2025 Aug;14(15). doi:10.1002/cam4.71155
- [28] Co M, Lau YCC, Qian YXY, Chan MCR, Wong DK, Lui KH, et al. Artificial Intelligence in Histologic Diagnosis of Ductal Carcinoma In Situ. *Mayo Clinic Proceedings: Digital Health*. 2023 Sep;1(3):267–75. doi:10.1016/j.mcpdig.2023.05.008
- [29] Harowicz MR, Saha A, Grimm LJ, Marcom PK, Marks JR, Hwang ES, et al. Can algorithmically assessed MRI features predict which patients with a preoperative diagnosis of ductal carcinoma in situ are upstaged to invasive breast cancer? *Journal of Magnetic Resonance Imaging*. 2017 Feb 9;46(5):1332–40. doi:10.1002/jmri.25655
- [30] Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*. 2003 Nov;56(11):1129–35. doi:10.1016/s0895-4356(03)00177-x
- [31] Slavkova KP, Kang R, Kazerouni AS, Biswas D, Belenky V, Chitalia R, et al. MRI-based Radiomic Features for Risk Stratification of Ductal Carcinoma in Situ in a Multicenter Setting (ECOG-ACRIN E4112 Trial). *Radiology*. 2025 Apr 1;315(1). doi:10.1148/radiol.241628
- [32] Meurs CJC, van Bekkum S, van Rosmalen J, Menke-Pluijmers MBE, Siesling S, Westenend PJ. Validation and Clinical Utility of a Prediction Model for the Risk of Upstaging to Invasive Breast Cancer After a Biopsy Diagnosis Ductal Carcinoma In Situ. *Annals of Surgical Oncology*. 2023 Aug 4;30(12):7069–80. doi:10.1245/s10434-023-13929-y