# RESEARCH



# Deep learning-based breast MRI for predicting axillary lymph node metastasis: a systematic review and meta-analysis

Chia-Fen Lee<sup>1,2</sup>, Joseph Lin<sup>3,4,5</sup>, Yu-Len Huang<sup>6</sup>, Shou-Tung Chen<sup>3,4</sup>, Chen-Te Chou<sup>1</sup>, Dar-Ren Chen<sup>3,4</sup> and Wen-Pei Wu<sup>1,2,7\*</sup>

# Abstract

**Background** To perform a systematic review and meta-analysis that assesses the diagnostic performance of deep learning algorithms applied to breast MRI for predicting axillary lymph nodes metastases in patients of breast cancer.

**Methods** A systematic literature search in PubMed, MEDLINE, and Embase databases for articles published from January 2004 to February 2025. Inclusion criteria were: patients with breast cancer; deep learning using MRI images was applied to predict axillary lymph nodes metastases; sufficient data were present; original research articles. Quality Assessment of Diagnostic Accuracy Studies-AI and Checklist for Artificial Intelligence in Medical Imaging was used to assess the quality. Statistical analysis included pooling of diagnostic accuracy and investigating between-study heterogeneity. A summary receiver operating characteristic curve (SROC) was performed. R statistical software (version 4.4.0) was used for statistical analyses.

**Results** A total of 10 studies were included. The pooled sensitivity and specificity were 0.76 (95% Cl, 0.67–0.83) and 0.81 (95% Cl, 0.74–0.87), respectively, with both measures having moderate between-study heterogeneity ( $l^2 = 61\%$  and 60%, respectively; p < 0.01). The SROC analysis yielded a weighted AUC of 0.788.

**Conclusion** This meta-analysis demonstrates that deep learning algorithms applied to breast MRI offer promising diagnostic performance for predicting axillary lymph node metastases in breast cancer patients. Incorporating deep learning into clinical practice may enhance decision-making by providing a non-invasive method to more accurately predict lymph node involvement, potentially reducing unnecessary surgeries.

\*Correspondence:

Wen-Pei Wu

wenpei1220@gmail.com

- <sup>1</sup> Department of Radiology, Changhua Christian Hospital, Changhua, Taiwan
- <sup>2</sup> Department of Biomedical Imaging and Radiological Sciences, National Yang Ming Chiao Tung University, Taipei, Taiwan
- <sup>3</sup> Division of General Surgery, Changhua Christian Hospital, Changhua, Taiwan

<sup>5</sup> Division of Breast Surgery, Yuanlin Christian Hospital, Yuanlin, Taiwan

<sup>7</sup> Department of Medical Imaging, Changhua Christian Hospital, 135 Nanxiao Street, Changhua 500, Taiwan

# Introduction

Breast cancer is a significant global health concern, being the most commonly diagnosed cancer and the leading cause of cancer-related death among women worldwide. In 2020, there were an estimated 2.3 million new cases of female breast cancer, representing 11.7% of all cancer cases, with 685,000 deaths recorded [1]. Staging of the axillary lymph nodes (ALNs) is crucial in managing breast cancer, as it guides the clinical stages, treatment planning, and prognosis. The most precise method to assess the status of ALN in breast cancer patients is by performing a surgical lymph node biopsy, which can be



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.gr/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.gr/licenses/by/4.0/. The Creative Commons Public Domain and redit line to the data.

<sup>&</sup>lt;sup>4</sup> Comprehensive Breast Cancer Center, Changhua Christian Hospital, Changhua, Taiwan

<sup>&</sup>lt;sup>6</sup> Department of Computer Science, Tunghai University, Taichung, Taiwan

performed by axillary lymph node dissection (ALND) or sentinel lymph node biopsy (SLNB) [2, 3].

The use of surgical axillary lymph node staging among older women with early breast cancer remains controversial. Based on the findings of the ongoing randomized controlled trials such as SOUND, INSEMA, BOOG 2013-08, and NAUTILUS, the omission of SLNB was safe in selected patients with small breast cancers, favorable hormone receptor-positive status, and clinically negative axillary lymph nodes [4–11]. These findings indicate that axillary surgery can be safely avoided in such patients when the absence of pathological information does not affect the postoperative treatment plan. However, despite this evidence, concerns remain among surgeons regarding these recommendations. Given the lack of survival benefit from axillary clearance in SLN-positive patients, the key question is: Should SLNB be omitted in selected low-risk patients with early breast cancer, or does it still provide enough therapeutic value to justify its routine use?

Pre-operative axillary ultrasound has a variable sensitivity, with a standard identification rate of metastatic axillary lymph nodes at 40–50% in breast cancer patients prior to SLNB [11, 12]. However, it is operator-dependent, and scanning deep areas is difficult, especially in obese patients [13, 14]. Given the MRI's ability to evaluate all axillary LNs regardless of the depth of nodes and body status of patients, preoperative breast MRI has emerged as a valuable tool for assessing axillary lymph node metastases in breast cancer patients. At the lymph node level, MRI demonstrated the highest sensitivity at 0.70 (95% CI: 0.58–0.80) and the highest specificity at 0.89 (95% CI: 0.85–0.92) compared to ultrasound or PET-CT [15–18].

Over the past years, significant advancements in AI algorithms have transformed the medical imaging landscape, particularly in enhancing diagnostic precision and efficiency. Deep learning, distinct from traditional machine learning and radiomics, operates as an end-toend model that automatically learns and extracts intrinsic features from medical images, eliminating human intervention and bypassing traditional, complex processes [19]. Deep learning models, particularly convolutional neural networks (CNNs), making them ideal for complex tasks like predicting axillary lymph node metastases from breast MRI [20-31]. Given the rapid advancement of deep learning applications, different AI algorithms and the target of prediction, whether the focus is on the tumor itself or the axillary lymph nodes, leading to varying accuracy.

While previous sparse systematic reviews/meta-analysis has covered AI techniques, including both machine learning and deep learning [32, 33], their findings remain sparse and lack a focused evaluation of deep learning's potential. Therefore, this study aimed to conduct a systematic review and meta-analysis to evaluate the diagnostic accuracy of deep learning-based MRI in predicting lymph node metastases in breast cancer, offering a comprehensive assessment of its clinical efficacy.

## **Materials and methods**

#### Study design and eligibility criteria for study selection

This study was conducted in adherence to the Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA) 2020 guidelines [34]. The PRISMA checklists are provided in supplementary material (Tables S1 and S2). The study protocol was registered with the International Platform of Registered Systematic Review and Meta-analysis Protocols (INPLASY) under the registration number INPLASY202520053.

# Search strategy and literature screening

Two authors, C.F.L. and W.P.W., with 6 and 10 years of experience in breast imaging, respectively, conducted a comprehensive literature search to gather relevant research data. We searched PubMed, MEDLINE and Embase databases for English-language articles published from January 2004 to February 2025. The predefined search terms included "artificial intelligence," "breast," "breast cancer," "lymph nodes," and "magnetic resonance imaging." The two authors independently screened titles and abstracts for eligibility and reviewed the full text of potentially relevant articles. The detailed search strategy is presented in Supplementary Text S1.

Articles were included based on the satisfaction of all the following criteria: (I) inclusion of patients diagnosed with breast cancers; (II) deep learning using breast MRI images were applied to predict axillary lymph nodes metastases; (III) sufficient data were present in terms of predictive performance of the deep learning algorithms; (IV) original research articles. Studies without enough information to calculate true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values were excluded.

#### Study selection, data collection, and quality assessment

Data from relevant studies were extracted and recorded in a predefined Microsoft Excel spreadsheet. For each analyzed study, the following information was collected: first author, publication year, number of patients in each dataset, specific AI algorithm model, MRI sequence used, reference standard, diagnostic accuracy, validation method, and AUC value. Diagnostic accuracy metrics, including TP, FP, TN, and FN, were either manually derived or reconstructed. These values were directly



Fig. 1 A flowchart of the literature review and study selection

entered into contingency tables and used to calculate sensitivity and specificity. In cases where external validation was conducted, this data was prioritized. For studies that examined multiple algorithms, data from the algorithm demonstrating the best performance were extracted.

Finally, study quality assessment was performed by the same two readers according to the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) and Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-AI [35, 36]. Discrepancies were resolved in consensus.

# Statistical analysis

Statistical analysis was performed by the author (C.F.L.) using package meta (v 8.0-2), package lme4 (v 1.1-36), package mada (v 0.5.11), package emmeans (v 1.10.1), package rje (v 1.12.1) in R version 4.4.0 (RStudio, Boston, MA). The pooled proportion analysis of detectability estimates with 95% confidence intervals (CI) was performed using a random-effects model. This model assumes significant diversity among different studies and accounts for both intra-study sampling errors and inter-study variances. It typically provides wider confidence intervals than the fixed effects model. The predictive accuracy was quantified using pooled sensitivity, specificity with 95% CI. A *p*-value < 0.10 indicates the presence of heterogeneity. To quantify the effect of heterogeneity, the inconsistency index  $(I^2)$  was used.  $I^2$  values of 25%, 50%, and 75% are interpreted as low, moderate, and high, respectively [37]. A cross-hairs plot was also generated to better display the variability in sensitivity and specificity estimates [38].

Summary receiver-operating characteristic curves using the bivariate method were constructed to display the mean value. The area under the curve (AUC) and the partial area under the curve (pAUC), which is calculated by restricting the computation of the AUC to the observed false positive rates, were also calculated. To ensure a more accurate summary estimation, weighting based on study sample size was applied, giving larger studies greater influence on the final SROC curve. An AUC of 0.9~1 is considered excellent, 0.8~0.9 very good, 0.7~0.8 good, 0.6~0.7 sufficient, and 0.5~0.6 bad in terms of diagnostic accuracy [38].

#### Results

## Literature search

Figure 1 summarizes the search and screening results for the relevant studies. Initially, we identified a total of 2667 articles through the literature review, with 524 articles found in PubMed, 205 articles from MEDLINE and 1938 articles from Embase. After removing 638 duplicate articles, 2029 unique studies remained. These records underwent screening by title and abstract, resulting in the exclusion of 1981 studies that did not meet the inclusion criteria. A total of 48 full-text articles were assessed for eligibility. During full-text reading, we excluded 38 articles because their study structures or incomplete data were not relevant to our study. Finally, 10 studies were included for meta-analysis.

# **Quality assessment**

The QUADAS-AI tool consists of four domains, including patient selection, index test, reference standard, and

Studies	N of Patients	Al Algorithm	MRI Sequences	Reference Standard	Validation method	External Validation	AUC	₽	£	FN
Nguyen, 2020 [31]	357	CNN	DCE-MRI	Histopathology	Nested, stratified group tenfold cross-validation	No	0.71	10	10	4
Ren, 2020 [ <mark>25</mark> ]	66	CNN	T1WI (post contrast)	<sup>18</sup> FDG-PET	fivefold cross-validation	No	0.91	20	9	2
Ren, 2022 [ <mark>26</mark> ]	56	CNN	T1WI+T2WI	<sup>18</sup> FDG-PET	threefold cross-validation	No	0.88	20	4	S
Santucci, 2022 [ <mark>24</mark> ]	153	CNN	DCE-MRI	Histopathology	tenfold cross-validation	No	0.77	4	2	-
Zhang, 2022 [ <mark>30</mark> ]	252	ResNet50	Multiparametric MRI model	Histopathology	Split training-test	No	0.91	17	m	m
Gao, 2023 [ <mark>23</mark> ]	941	ResNet, RCNet, SVM	DCE-MRI	Histopathology	fivefold cross-validation	Yes	0.85	38	10	10
Guo, 2024 [ <mark>28</mark> ]	2063	ConvRNN, CNN	DCE-MRI	Histopathology	Split training-test	Yes	0.81	45	11	35
Polat, 2024 [ <mark>27</mark> ]	350	CNN	DCE-MRI	MRI, axillary US, US-guided needle biopsy, Histopathology	Nested fivefold cross-validation	No	0.87	27	Ś	ŝ
Zhou, 2024 [ <mark>29</mark> ]	1259	DLWPS (ResNet101, ResNeXt101, DenseNet)	DCE-MRI	Histopathology	fivefold cross validation	Yes	0.92	29	;-	6
AUC area under the rect attention module, 2D C enhanced magnetic res	eiver operating cur onvRNN convolutio onance imaging, T	ves, SEN sensitivity, SPE speci nal neural network and recur 1W/ T1-weighted imaging, 72	ficity, <i>PPV</i> positive predictiv rent neural network, <i>DLWP</i> : <i>W</i> /T2-weighted imaging, <i>D</i>	e value, <i>NPV</i> negative predictive valu 5 deep learning-based whole proces <i>WI</i> diffusion-weighted imaging	ie, CNN convolutional neural network, <i>R</i> s system, <i>ResNeXt101</i> modified version	<i>CNet</i> ResNet and of ResNet101, <i>D</i> 0	convolut E-MRI dyı	ional bl amic co	ock ontrast-	

studies	
ncluded	
of the ii	
Characteristics	
-	
Table	

flow and timing [39]. For individual studies, each domain was assessed for applicability concerns, categorized as high, low, or unclear. Two reviewers independently conducted the quality assessments, and any discrepancies were resolved by a third reviewer to achieve consensus. Most studies were rated as the unclear risk. The detailed risk of bias information on each enrolled study can be found in supplemental Fig. 1.

The meta-analysis of the quality assessment, based on the CLAIM criteria and presented in Table S3, revealed an average total score of 34.1 across these studies, with a standard deviation of 3.67 [40]. This indicates a moderate level of variability in the overall quality of the included research.

## Basic characteristics of the included literature

The characteristics of each eligible study are detailed in Table 1. All studies used a retrospective design, six involving internal validation [24–27, 30, 31] and four with external validation [23, 28, 29, 41].

Most studies used the primary tumor as the target lesion for predicting axillary lymph node metastases. However, three studies included both the primary tumor and axillary lymph nodes as targets [23, 25, 26].

Seven of these studies reported the best model performance using DCE-MRI images [23–25, 27–29, 31, 41]. In 10 eligible studies, different AI algorithms were used for modeling. Two of these studies incorporated clinical information in addition to images [27, 31], and one used ensemble learning [30]. Most studies identified lymph node metastases through pathological examination, including surgical resection or needle biopsy. However, one study relied on clinical node status [31] and two studies employed <sup>18</sup>FDG-PET for indirect assessment [25, 26].

#### Diagnostic accuracy and heterogeneity

Figure 2 presents forest plots illustrating sensitivities (Fig. 2a) and specificities (Fig. 2b), each with the corresponding 95% confidence intervals. The overall pooled sensitivity and specificity were 0.76 (95% CI, 0.67–0.83) and 0.81 (95% CI, 0.74–0.87), respectively, with both measures having moderate between-study heterogeneity ( $I^2$ =61% and 60%; p<0.01 and p<0.01, respectively). In the subgroup analysis, studies using only the primary tumor as the target reported an overall pooled sensitivity of 0.73 (95% CI, 0.61–0.82) and specificity of 0.78 (95% CI, 0.69–0.86). In contrast, studies using both the axillary lymph nodes and the primary tumor as targets showed slightly higher overall pooled sensitivity and specificity of 0.81 (95% CI, 0.72–0.88) and 0.86 (95% CI, 0.77–0.92), respectively, as illustrated in Fig. 3.

Figure 4 depicts a cross-hair plot that visualizes the sensitivity and specificity estimates for each study included in the meta-analysis.

The pooled SROC curve by the bivariate approach showed good diagnostic accuracy with a weighted AUC of 0.79 and a weighted pAUC value of 0.76. The Diagnostic Odds Ratio (DOR) across studies was 14.7 (95% CI 7.6–28.3), with heterogeneity observed at 64.1% ( $I^2$ =64.%, *p*=0.003), suggesting moderate to high heterogeneity (Fig. 5).

## Discussion

The transition from traditional machine learning to deep learning marks a major milestone in the evolution of artificial intelligence, enabling more complex data-driven decision-making and predictive analytics across various domains, particularly in medical imaging [42]. Deep learning leverages deep neural networks with multiple layers to automatically analyze complex and diverse data [43]. This advancement has driven significant progress in areas such as image processing and predictive modeling, highlighting deep learning's pivotal role in personalized treatments and diagnostic tools to identify diseases more accurately and quickly.

This meta-analysis is to evaluate the diagnostic performance of deep learning-based algorithms applied to breast MRI for predicting lymph node metastasis in breast cancer patients. Breast MRI is increasingly recognized as a crucial imaging modality for pre-operative staging [44]. However, variations in image quality and the subjectivity of radiological interpretation can lead to false positives and false negatives, potentially resulting in unnecessary surgeries for breast cancer patients. From the radiologists' perspective, the role of breast MRI in determining axillary lymph nodes metastases has shown moderate sensitivity and negative predictive value, but only low-to medium specificity [45]. This can lead to either overestimation or underestimation of nodal staging (N staging). Therefore, a non-invasive and highly accurate method to predict axillary lymph node metastasis before surgery is essential to guide clinical decisionmaking effectively.

Our meta-analysis demonstrated promising results for axillary lymph nodes metastases classification, with a pooled sensitivity and specificity of 76% and 81%, respectively. The diagnostic performance by radiologists reported sensitivities of 77% and 55% and specificities of 90% and 86%, as demonstrated in the meta-analysis by Zhou et al. [17](2018) and Zhang et al. [46](2020). Meanwhile, machine learning-based models have shown significant potential with pooled sensitivities and specificities ranging from 76 to 85% and 81% to 83% in the meta-analysis by Zhang et al. [47](2022), Chen et al. [33]

Study	Events	Total						Proportion	95%-CI	Weight (common)	Weight (random)
Nguyen, 2020	10	14		-		+1	-	0.71	[0.42, 0.92]	5.8%	10.1%
Ren. 2020	20	22				++	*	0.91	[0.71: 0.99]	3.7%	7.8%
Ren, 2022	20	25			_	-	-	0.80	[0.59; 0.93]	8.1%	11.7%
Santucci, 2022	4	5		_		++-	_	0.80	[0.28; 0.99]	1.6%	4.4%
Zhang, 2022	17	20			_	+ +		0.85	[0.62; 0.97]	5.2%	9.5%
Gao, 2023	38	48			-	- 00		0.79	[0.65: 0.90]	16.1%	14.8%
Guo, 2024	45	80		_	101			0.56	[0.45: 0.67]	40.0%	17.7%
Polat, 2024	27	30						0.90	[0.73; 0.98]	5.5%	9.8%
Zhou, 2024	29	38			_	1		0.76	[0.60; 0.89]	14.0%	14.3%
Common effect model		282				4		0.72	[0.66: 0.78]	100.0%	
Random effects model Heterogeneity: t <sup>2</sup> = 62%, p	< 0.01			-	-	\$	_	0.78	[0.68; 0.86]		100.0%
		0	0.2	0.4	0.6	0.8	1				
				Sens	vitiviti						

Fig. 2 Forest plots of pooled sensitivity (a) and specificity(b) for the deep learning model to detect axillary lymph node metastasis in breast cancer



Fig. 3 Forest plots of the sensitivity (a) and specificity (b) for subgroup analysis according to the target of assessment (the primary tumor only or the axillary lymph nodes and the primary tumor)

(2021), and Liu et al. [32] (2023). These findings indicate that deep learning-based models offer comparable, if not superior, diagnostic performance compared to traditional radiologist interpretations and other machine learning models to detect axillary lymph node metastasis (ALNM) in breast cancer patients.

To date, many meta-analyses have been performed to evaluate the diagnostic performance of breast MRI for ALNM. Chen et al. found that MRI sequences and algorithms as key factors influencing the diagnostic accuracy of machine learning-assisted MRI [33]. Our subgroup analysis revealed that studies targeting both the primary tumor and axillary lymph nodes demonstrated better diagnostic performance compared to those focusing solely on the primary tumor. Factors such as differences in study populations, imaging protocols, image segmentation methods (eg, included breast lesion or lymph nodes) and model architectures may contribute to this heterogeneity and should be carefully addressed in future research.

A deep learning-based algorithm can enhance the speed of image interpretation for detecting lymph node metastases in breast cancer and is likely to be integrated alongside clinical judgment to optimize decision-making. However, challenges and limitations remain. With a pooled sensitivity and specificity of approximately 0.80, the algorithm could potentially miss 1 in 5 women with lymph node metastases, which may affect critical



Fig. 4 Cross-hair plot of studies included in the meta-analysis



Fig. 5 Summary receiver operating characteristic (SROC) curves based on the bivariate approach

treatment decisions, such as the need for surgery or neoadjuvant chemotherapy. On the other hand, the model correctly identifies most patients without axillary lymph node metastases, helping to minimize unnecessary interventions, such as surgical lymph node dissection or overtreatment.

Moving forward, while the findings of this meta-analvsis are promising, several considerations should be taken into account. First, the heterogeneity observed in both sensitivity and specificity underscores the need for further investigation into the sources of variability across studies. The integration of deep learning in medical imaging brings several ethical challenges that must be addressed to ensure safe and equitable use. Deep learning models trained on limited or region-specific datasets has the possibility to produce biased results, leading to reduced diagnostic accuracy in underrepresented populations. There were relatively few deep learning studies eligible for inclusion. Due to incomplete reporting of results in few studies, estimates of diagnostic performance were derived from a limited data set. Finally, the diversity in scanner types, imaging protocols, and criteria for defining lymph node metastasis across studies may have influenced the accuracy of the results.

# Conclusions

In conclusion, our meta-analysis demonstrates that deep learning-based MRI offers promising diagnostic performance for predicting axillary lymph node metastases in breast cancer. These results suggest that deep learning models have the potential to complement radiologists, enhance diagnostic accuracy, and reduce unnecessary surgical interventions. However, variability among studies highlights the need for standardized imaging protocols and larger, multicenter studies to further validate their clinical utility. Future research should focus on integrating deep learning with clinical and genomic data to enable more personalized and precise treatment planning, ultimately improving patient outcomes.

# **Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s40644-025-00863-3.

Additional file 1 Additional file 2 Additional file 3

#### Acknowledgements

No acknowledgements to declare.

#### Authors' contributions

Conceptualization: Wen-Pei Wu. Data curation: Chia-Fen Lee, Wen-Pei Wu, Joseph Lin, Yu-Len Huang, Chen-Te Chou, Dar-Ren Chen. Formal analysis: Wen-Pei Wu. Funding acquisition: N/A. Investigation: Chia-Fen Lee, Wen-Pei Wu, Dar-Ren Chen. Methodology: Wen-Pei Wu, Joseph Lin, Shou-Tung Chen, Chen-Te Chou, Dar-Ren Chen. Project administration: Dar-Ren Che. Resources: Shou-Tung Chen, Dar-Ren Chen. Software: Chia-Fen Lee. Supervision: Shou-Tung Chen, Chen-Te Chou, Dar-Ren Chen. Validation: Chia-Fen Lee, Wen-Pei Wu, Dar-Ren Chen. Visualization: Shou-Tung Chen, Dar-Ren Chen. Writingoriginal draft: Chia-Fen Lee, Wen-Pei Wu. Writing-review & editing: Wen-Pei Wu, Joseph Lin, Dar-Ren Chen. Approval of final manuscript: All authors.

#### Funding

All authors declare that they received no funding for this work.

#### Data availability

No datasets were generated or analysed during the current study.

#### Declarations

**Ethics approval and consent to participate** Not applicable.

- - ---

**Consent for publication** Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 22 September 2024 Accepted: 12 March 2025 Published online: 31 March 2025

#### References

- Sedeta ET, Jobre B, Avezbakiyev B. Breast cancer: global patterns of incidence, mortality, and trends. J Clin Oncol. 2023;41(16\_suppl):10528-.
- Moore MP, Kinne DW. Axillary lymphadenectomy: a diagnostic and therapeutic procedure. J Surg Oncol. 1997;66(1):2–6.
- Kelley MC, Hansen N, McMasters KM. Lymphatic mapping and sentinel lymphadenectomy for breast cancer. Am J Surg. 2004;188(1):49–61.
- Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of axillary dissection vs no axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: the ACOSOG Z0011 (Alliance) randomized clinical trial. JAMA. 2017;318(10):918–26.
- Krag D, Harlow S, Julian T. Breast cancer and the NSABP-B32 sentinel node trial. Breast Cancer. 2004;11(3):221–4 discussion 64–6.
- Carleton N, Zou J, Fang Y, Koscumb SE, Shah OS, Chen F, et al. Outcomes after sentinel lymph node biopsy and radiotherapy in older women with early-stage, estrogen receptor-positive breast cancer. JAMA Netw Open. 2021;4(4): e216322.
- Kunkler IH, Williams LJ, Jack WJ, Cameron DA, Dixon JM, investigators PI. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. Lancet Oncol. 2015;16(3):266–73.
- Castelo M, Sutradhar R, Faught N, Mata D, Hahn E, Nguyen L, et al. The association between surgical axillary staging, adjuvant treatment use and survival in older women with early stage breast cancer: a populationbased study. Ann Surg Oncol. 2023;30(7):3901–12.
- Jung JG, Ahn SH, Lee S, Kim EK, Ryu JM, Park S, et al. No axillary surgical treatment for lymph node-negative patients after ultra-sonography [NAUTILUS]: protocol of a prospective randomized clinical trial. BMC Cancer. 2022;22(1):189.
- Reimer T, Stachs A, Nekljudova V, Loibl S, Hartmann S, Wolter K, et al. Restricted axillary staging in clinically and sonographically node-negative early invasive breast cancer (c/iT1-2) in the Context of Breast Conserving Therapy: First Results Following Commencement of the Intergroup-Sentinel-Mamma (INSEMA) Trial. Geburtshilfe Frauenheilkd. 2017;77(2):149–57.
- Houssami N, Diepstraten SC, Cody HS 3rd, Turner RM, Sever AR. Clinical utility of ultrasound-needle biopsy for preoperative staging of the axilla in invasive breast cancer. Anticancer Res. 2014;34(3):1087–97.

- Huang Y, Zheng S, Lin Y. Accuracy and utility of preoperative ultrasound-guided axillary lymph node biopsy for invasive breast cancer: a systematic review and meta-analysis. Comput Intell Neurosci. 2022;2022:3307627.
- Rauch GM, Kuerer HM, Jochelson MS. To look or not to look? Yes to Nodal Ultrasound! J Breast Imaging. 2021;3(6):659–65.
- Marino MA, Avendano D, Zapata P, Riedl CC, Pinker K. Lymph node imaging in patients with primary breast cancer: concurrent diagnostic tools. Oncologist. 2020;25(2):e231–42.
- Li Z, Gao Y, Gong H, Feng W, Ma Q, Li J, et al. Different imaging modalities for the diagnosis of axillary lymph node metastases in breast cancer: a systematic review and network meta-analysis of diagnostic test accuracy. J Magn Reson Imaging. 2023;57(5):1392–403.
- Harnan SE, Cooper KL, Meng Y, Ward SE, Fitzgerald P, Papaioannou D, et al. Magnetic resonance for assessment of axillary lymph node status in early breast cancer: a systematic review and meta-analysis. Eur J Surg Oncol. 2011;37(11):928–36.
- Zhou P, Wei Y, Chen G, Guo L, Yan D, Wang Y. Axillary lymph node metastasis detection by magnetic resonance imaging in patients with breast cancer: A meta-analysis. Thorac Cancer. 2018;9(8):989–96.
- Chen ST, Lai HW, Chang JH, Liao CY, Wen TC, Wu WP, et al. Diagnostic accuracy of pre-operative breast magnetic resonance imaging (MRI) in predicting axillary lymph node metastasis: variations in intrinsic subtypes, and strategy to improve negative predictive value-an analysis of 2473 invasive breast cancer patients. Breast Cancer. 2023;30(6):976–85.
- 19. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
- Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. BMC Cancer. 2021;21(1):1058.
- 21. Rokhshad R, Salehi SN, Yavari A, Shobeiri P, Esmaeili M, Manila N, et al. Deep learning for diagnosis of head and neck cancers through radiographic data: a systematic review and meta-analysis. Oral Radiol. 2024;40(1):1–20.
- Forte GC, Altmayer S, Silva RF, Stefani MT, Libermann LL, Cavion CC, et al. Deep learning algorithms for diagnosis of lung cancer: a systematic review and meta-analysis. Cancers (Basel). 2022;14(16):3856–66. https:// doi.org/10.3390/cancers14163856.
- Gao J, Zhong X, Li W, Li Q, Shao H, Wang Z, et al. Attention-based deep learning for the preoperative differentiation of axillary lymph node metastasis in breast cancer on DCE-MRI. J Magn Reson Imaging. 2023;57(6):1842–53.
- Santucci D, Faiella E, Gravina M, Cordelli E, de Felice C, Beomonte Zobel B, Iannello G, Sansone C, Soda P. CNN-Based Approaches with Different Tumor Bounding Options for Lymph Node Status Prediction in Breast DCE-MRI. Cancers (Basel). 2022;14(19):4574–92. https://doi.org/10.3390/ cancers14194574.
- 25. Ren T, Cattell R, Duanmu H, Huang P, Li H, Vanguri R, et al. Convolutional neural network detection of axillary lymph node metastasis using standard clinical breast MRI. Clin Breast Cancer. 2020;20(3):e301–8.
- Ren T, Lin S, Huang P, Duong TQ. Convolutional neural network of multiparametric MRI accurately detects axillary lymph node metastasis in breast cancer patients with pre neoadjuvant chemotherapy. Clin Breast Cancer. 2022;22(2):170–7.
- Polat DS, Nguyen S, Karbasi P, Hulsey K, Cobanoglu MC, Wang L, et al. Machine learning prediction of lymph node metastasis in breast cancer: performance of a multi-institutional MRI-based 4D convolutional neural network. Radiol Imaging Cancer. 2024;6(3): e230107.
- Guo YJ, Yin R, Zhang Q, Han JQ, Dou ZX, Wang PB, Lu H, Liu PF, Chen JJ, Ma WJ. MRI-Based Kinetic Heterogeneity Evaluation in the Accurate Access of Axillary Lymph Node Status in Breast Cancer Using a Hybrid CNN-RNN Model. J Magn Reson Imaging. 2024;60(4):1352–64. https://doi. org/10.1002/jmri.29225.
- Zhou H, Hua Z, Gao J, Lin F, Chen Y, Zhang S, et al. Multitask deep learning-based whole-process system for automatic diagnosis of breast lesions and axillary lymph node metastasis discrimination from dynamic contrast-enhanced-MRI: a multicenter study. J Magn Reson Imaging. 2024;59(5):1710–22.

- Zhang X, Liu M, Ren W, Sun J, Wang K, Xi X, et al. Predicting of axillary lymph node metastasis in invasive breast cancer using multiparametric MRI dataset based on CNN model. Front Oncol. 2022;12: 1069733.
- Nguyen S, Polat D, Karbasi P, Moser D, Wang L, Hulsey K, et al. Preoperative prediction of lymph node metastasis from clinical DCE MRI of the primary breast tumor using a 4D CNN. Med Image Comput Comput Assist Interv. 2020;12262:326–34.
- Liu CJ, Zhang L, Sun Y, Geng L, Wang R, Shi KM, et al. Application of CT and MRI images based on an artificial intelligence algorithm for predicting lymph node metastasis in breast cancer patients: a meta-analysis. BMC Cancer. 2023;23(1):1134.
- Chen C, Qin Y, Chen H, Zhu D, Gao F, Zhou X. A meta-analysis of the diagnostic performance of machine learning-based MRI in the prediction of axillary lymph node metastasis in breast cancer patients. Insights Imaging. 2021;12(1):156.
- Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. BMJ. 2020;370: m2632.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
- Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. Nat Med. 2021;27(10):1663–5.
- 37. Lee YH. Overview of the process of conducting meta-analyses of the diagnostic test accuracy. J Rheum Dis. 2018;25(1):3.
- Schlattmann P. Tutorial: statistical methods for the meta-analysis of diagnostic test accuracy studies. Clin Chem Lab Med. 2023;61(5):777–94.
- Guni A, Sounderajah V, Whiting P, Bossuyt P, Darzi A, Ashrafian H. Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies Using AI (QUADAS-AI): Protocol for a Qualitative Study. JMIR Res Protoc. 2024;13: e58202.
- Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2(2):e200029.
- 41. Tang X, Zhang H, Mao R, Zhang Y, Jiang X, Lin M, et al. Preoperative prediction of axillary lymph node metastasis in patients with breast cancer through multimodal deep learning based on ultrasound and magnetic resonance imaging images. Acad Radiol. 2025;32(1):1–11.
- Rane NL, Mallick SK, Kaya O, Rane J. Machine learning and deep learning architectures and trends: A review. In Applied Machine Learning and Deep Learning: Architectures and Techniques. Deep Science Publishing: 2024. p. 1–38. https://doi.org/10.70593/978-81-981271-4-3\_1.
- Shafik W. Deep learning impacts in the field of artificial intelligence. Book Deep Learning Concepts in Operations Research. Auerbach Publications: 2024. p. 18. http://dx.doi.org/10.1201/9781003433309-2.
- Aktaş A, Gürleyik MG, Aydın Aksu S, Aker F, Güngör S. Diagnostic value of axillary ultrasound, MRI, and (18)F-FDG-PET/ CT in determining axillary lymph node status in breast cancer patients. Eur J Breast Health. 2022;18(1):37–47.
- Chayakulkheeree J, Pungrassami D, Prueksadee J. Performance of breast magnetic resonance imaging in axillary nodal staging in newly diagnosed breast cancer patients. Pol J Radiol. 2019;84:e413–8.
- Zhang X, Liu Y, Luo H, Zhang J. PET/CT and MRI for identifying axillary lymph node metastases in breast cancer patients: systematic review and meta-analysis. J Magn Reson Imaging. 2020;52(6):1840–51.
- Zhang J, Li L, Zhe X, Tang M, Zhang X, Lei 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.

# **Publisher's Note**

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