Using artificial intelligence in the analysis of CT scans of the axillary nodes in breast cancer: a systematic review

Authors: Dr Julie Cox[1], Mr Amir Bhatti[1], Dr Amir Atapour-Abarghouei[2]

[1] University Hospital North Durham, County Durham and Darlington NHS Foundation Trust

[2] Department of Computer Science, University of Durham

Corresponding author:

Dr Julie Cox

Department of Radiology

University Hospital North Durham

Durham

DH1 5TW

Email: j.cox1@nhs.net

Telephone: 0191 333 2198

Summary

Background: Breast cancer is the most diagnosed cancer worldwide, with early diagnosis key to improved outcomes following treatment. Accurate assessment of whether breast cancer has spread to axillary lymph nodes is of vital importance in deciding on treatment paths for patients and for providing prognostic information. We therefore aimed to evaluate and synthesise findings from existing research regarding the effectiveness of using artificial intelligence (AI) to diagnose axillary lymph node metastasis on CT scans in patients who have a diagnosis of breast cancer and have had axillary surgery to establish the histological status of the axillary nodes.

Methods: The protocol was prospectively registered with PROSPERO (CRD42024598174). Four electronic databases (Embase, Medline, the Cochrane Library, and Google Scholar) were searched for relevant studies relating to different AI methods, axillary lymph nodes and breast cancer. The quality of included articles was independently appraised using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) tool. Data were narratively synthesised.

Findings: Five studies were eligible for inclusion in this review. All were single centre retrospective studies of breast cancer patients. Four studies used convolutional neural networks and one used machine learning. Results indicated that AI models were able to accurately diagnose axillary lymph node metastasis in breast cancer CT scans.

Interpretations: AI-based analyses of CT images show substantial potential to accurately localise axillary lymph nodes and predict metastasis. This method can offer an increasingly precise non-invasive diagnostic strategy for breast cancer staging. There are, however, some outstanding limitations, including limited samples and a small number of retrospective studies on this topic, which need to be overcome in future research.

Funding: No additional funding was received for this research

Research in context

Evidence before this study:

We searched Embase, Medline, the Cochrane Library, and Google Scholar from January 2010 to September 2024 using terms including “Artificial Intelligence”, “Axillary Lymph Nodes”, “CT scans”, and “Breast Cancer”. Studies were included if they evaluated AI methods to identify axillary lymph node metastases on CT scans in breast cancer patients with histologically confirmed nodal status.

Five studies met the inclusion criteria. All were retrospective, single-centre cohort studies using convolutional neural networks or machine learning models. Risk of bias was low to moderate. No prior systematic review had specifically examined AI-based analysis of axillary nodes using CT imaging.

Added value of this study

This is the first review to focus specifically on the use of AI for axillary lymph node assessment on contrast-enhanced CT in breast cancer. CNN-based models consistently achieved high accuracy, often exceeding 90%, particularly in two-stage architectures. This approach may provide a reproducible, non-invasive method of nodal staging.

Implications of all the available evidence

AI-assisted CT interpretation may enhance diagnostic accuracy and reduce the need for invasive biopsies, especially in low-resource settings where MRI and PET-CT are less accessible. Larger multicentre trials with external validation and consistent imaging protocols are now needed to support clinical adoption.

Introduction

In 2022, based on WHO statistics, approximately 2.3 million women were diagnosed with breast cancer worldwide, making it the most common cancer among women and the second most common cancer overall[1]. In the same year, breast cancer was responsible for an estimated 670,000 deaths globally.

In 2025, an analysis by the International Agency for Research on Cancer (IARC) evaluated the current and future burden of female breast cancer globally, with a detailed analysis in about 50 countries with high-quality population-level cancer data[2]. The IARC study found that on average, 1 in 20 women worldwide will be diagnosed with breast cancer in their lifetime. The findings also were that based on current rates, by 2050, there will be 3.2 million new breast cancer cases and 1.1 million breast cancer-related deaths per year. The growth and burden of the disease will disproportionately affect countries with a low Human Development Index (HDI).

Axillary lymph node involvement of breast cancer is a critical factor in breast cancer staging and treatment planning. A recently published retrospective cohort study over three decades showed a significantly better survival rate and less aggressive traits in patients without axillary LN metastasis than in participants with involved ones[3]. The importance of accurate preoperative assessment of the axillary lymph nodes in patients presenting with breast cancer is crucial as it significantly influences staging, prognosis, and treatment decisions regarding the precise extent of the initial surgery to the axilla[4].

Preoperative imaging techniques used to assess the axillary lymph nodes to detect metastatic disease include ultrasound, CT, PET-CT, and MRI scanning[5]. Biopsy of lymph nodes thought to be abnormal under ultrasound guidance is frequently used to preoperatively assess axillary lymph node involvement[6]; however, imaging guided biopsies are invasive and can result in complications[7]. Furthermore, ultrasound guided biopsies can be inaccurate and result in false negative results, meaning a tumour may be present in a lymph node but is not picked up on the biopsy[8]. In 2021, the James Lind Alliance, in their Priority Setting Partnerships with the Association of Breast Surgery, identified whether lymph node removal can be avoided as the number one priority for breast surgery research in the opinion of a group of patients, carers and clinicians for breast cancer research[9].

These challenges in clinical practice have led to the exploration of AI in analysis of the imaging characteristics of the axillary lymph nodes to improve diagnostic accuracy, reduce false negative results, and potentially enable axillary surgery to be more accurately tailored to the individual patient and their nodal status[10-14]. Most of the literature to date has focused on evaluating AI in combination with MRI or PET CT imaging. CT scanning with intravenous contrast provides high quality reproducible

images, which can be formatted in multiple planes. Globally, CT scanning capacity greatly exceeds MRI and PET CT capacity in terms of numbers of scanners available which may make CT scanning a more feasible imaging option for detailed evaluation rather than MRI or PET CT scanning[15].

This systematic review therefore aims to answer the following research questions:

1. Which AI models have been used for axillary lymph node analysis in breast cancer CT scans?

2. How do AI models compare to conventional diagnostic methods in terms of sensitivity, specificity, and overall accuracy including NPV, PPV, and AUC?

3. What are the clinical outcomes reported associated with AI-assisted axillary node analysis compared with standard case?

4. What are the limitations and biases present in the studies evaluating AI in evaluating axillary lymph nodes in CT scans of patients with breast cancer?

Methods

This review is reported according to the PRISMA guidelines for reporting on systematic reviews[16]. The protocol was prospectively registered with PROSPERO (CRD42024598174).

Search strategy

Four electronic databases (Embase, Medline, the Cochrane Library, and Google Scholar) were searched for relevant studies from January 2010 until September 2024. The search strategy was developed in collaboration with an expert librarian, and was based on the following keywords:

- "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Radiomics"

- "Axillary Lymph Nodes" OR "Axillary Nodes" OR "Lymph Node Analysis" OR "Lymphatic Metastasis"

- "Breast Cancer" OR "Breast Neoplasms"

A full search strategy is available in Appendix 1. The reference lists of any relevant systematic reviews were also handsearched for studies which met our inclusion criteria.

Screening

Identified references deduplicated, then imported into Rayyan for title and abstract, and full text screening. Screening was conducted independently by two members of the research team (JC, RA), with any discrepancies resolved through discussion with the wider team.

Inclusion and exclusion criteria

The inclusion and exclusion criteria are shown in Table 1 below.

Table 1: inclusion and exclusion criteria

	Inclusion	Exclusion
Population	Breast cancer patients of any age, as diagnosed using any recognised diagnostic criteria.	Studies involving non-human participants. Studies focusing on patients receiving neoadjuvant chemotherapy. Studies not focussed on axillary nodes. Studies based on images of slides. Studies about radiotherapy modelling of fields with there is no histological outcome.
Intervention	AI-based analysis of axillary lymph nodes. Studies involving CT scans.	Studies involving other imaging methods (e.g. MRI). Studies involving radiomics only. Conventional diagnostic methods (e.g. imaging, imaging guided biopsy, sentinel lymph node biopsy).
Comparator	Conventional diagnostic methods (e.g. ultrasound, MRI, biopsy).	Studies with no comparison group.
Outcome	Studies exploring the effectiveness of AI in identifying axillary lymph node metastasis.	Studies which do not contain any of the specified outcomes.

	Studies comparing AI and traditional diagnostic methods in identifying axillary lymph node metastasis. Studies identifying clinical outcomes for patients with AI-assisted axillary node analysis compared with standard case. Studies identifying any biases and limitations of AI-assisted axillary node analysis.	
Study design	Any peer reviewed cohort or case control study Studies published in English Studies published between 2010 – 2024	Study protocols or protocol papers Case studies Conference abstracts Review articles Commentaries or editorials Qualitative studies Articles where the full text is not available Studies published in other languages where an English translation has not been provided Studies published before 2010

Data extraction

A standardised data extraction form was created using Microsoft Excel to collect information from each study, including:

- Study characteristics (authors, year, country, sample size, setting)

- Type of AI model used (e.g., CNN, SVM, Random Forest)

- Data sources (e.g., imaging modalities, histopathology)

- Performance metrics (sensitivity, specificity, accuracy, etc.)

- Study outcomes (grouped by research question)

This was first piloted on a small number of included studies and any necessary amendments made before extracting all data. Data was extracted independently by two reviewers (JC, RA), with any discrepancies resolved through discussion.

Quality appraisal

The quality of included articles was appraised using an amended version of the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) tool[17] independently by two reviewers (JC, RA), with any discrepancies resolved through discussion. This tool was amended to remove questions about the timing of the index test and reference standard, as these were not relevant to the retrospective studies in this review.

Evidence synthesis

A narrative synthesis was conducted to summarise the findings across studies, highlighting key patterns, strengths, and weaknesses of the AI models. We followed ESRC guidelines to conduct a narrative synthesis[18], involving four stages: developing a theory; developing a preliminary synthesis; exploring the relationships in the data and assessing the robustness of the synthesis. This narrative synthesis was undertaken by two members of the review team (JC, RA), and reviewed by the other members.

A quantitative synthesis of included studies was considered; however, this was not conducted due to the small number of included papers and heterogeneity of outcomes reported.

Role of the funding source

There was no funding source for this study.

Results

Study selection

After removing duplicates, 472 references were screened for eligibility by their title and abstract. Eleven were included for full text screening, of which five papers were found to be eligible for inclusion in this review. Figure 1 shows a full breakdown of the study selection and screening process. Full details of reasons for exclusion at full text screening are presented in Appendix 2.

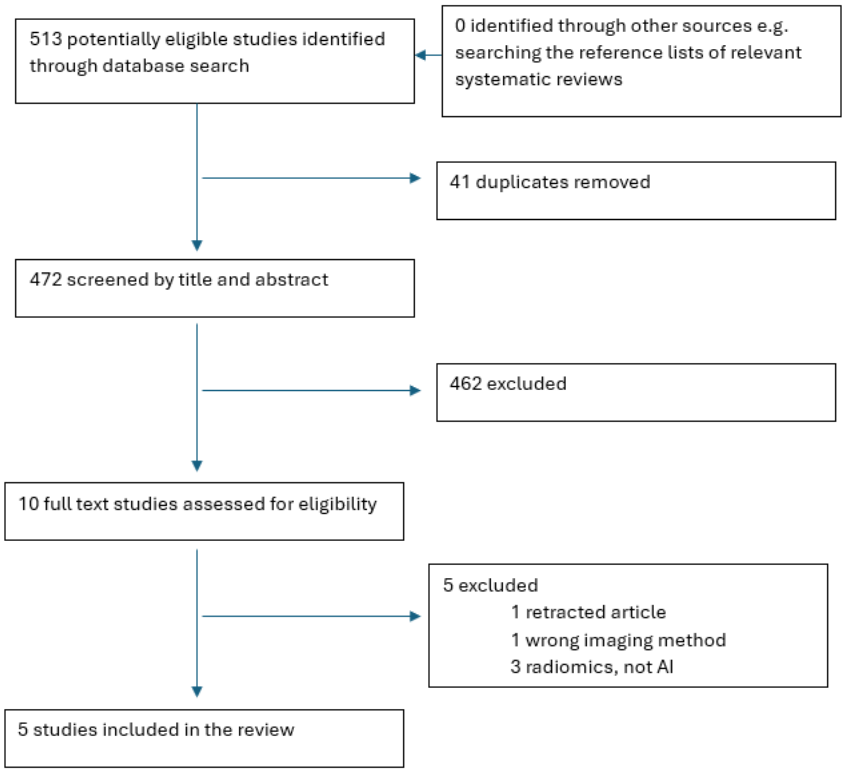

Figure 1: Study selection

Quality of included studies

The QUADAS-2 tool indicated that three studies were viewed having overall low risk, with low risk or unclear risk identified across all but one of the domains. Two studies were indicated to have higher risk of bias, with high risk recorded in three out of the seven domains. The domain which was identified as having the highest risk of bias across all studies was domain 2a, regarding the index test and how its findings were interpreted, with all studies having an unclear or high risk of bias. Full details of the QUADAS-2 ratings for each study are available in Supplementary Appendix 3.

Study characteristics

Of the five included studies, three were conducted in China[10,11,14] and two in South Korea[12,13]. All had a retrospective cohort study design and all used data from a single centre. Two of the papers[10,11] used the same cohort of breast cancer patients. Full details of included studies are shown in Table 2.

Table 2. Characteristics of included studies

Author (date)	**Country**	**Study design**	**Aim of study**	**N participants**	**Population** (e.g. newly diagnosed breast cancer patients)	**Participant demographics** Age: mean (range)	**Setting** (e.g. single or multi-centre)
Liu et al., (2021)[10]	China	Retrospective cohort study	To exploit efficient deep learning schemes to predict the metastasis of ALN automatically via CECT images	401	Pathologically confirmed diagnosis of breast cancer	Not reported	Single centre
Liu et al., (2022)[11]	China	Retrospective cohort study	To develop non-invasive and automatic schemes for ALN localization and metastasis prediction via CECT image	401	Pathologically confirmed diagnosis of breast cancer	Not reported	Single centre
Park et al., (2023)[12]	South Korea	Retrospective cohort study	To evaluate the performance of the ML algorithm for predicting ALN metastasis with chest CECT	266	pathologically confirmed diagnosis of breast cancer	Age: 50 (23-91)	single centre
Park et al., (2024)[13]	South Korea	Retrospective cohort study	To develop a deep learning	523 (303 malignant and	Histologically confirmed	Not reported	Single centre

			model using clinical implication-applied pre-processed CT images to enhance the prediction of ALN metastasis in breast cancer patients.	220 benign lymph nodes)	breast cancer patients		
Yang et al., (2019)[14]	China	Retrospective cohort study	To evaluate the non-invasive predictive performance of deep learning features based on staging CT for sentinel lymph node metastasis of breast cancer.	348	Histologically confirmed primary invasive breast cancer patients	Age: Primary cohort SLN-negative 50.00 +- 10.66, SLN positive 51.98 +-9.43 Validation cohort SLN-negative 52.73 +- 9.84, SLN positive 50.17 +-9.69	Single centre

Data synthesis

Liu et al.[10] reported an AUC of 0.945 and an accuracy of 90.1% for their convolutional neural networks (CNN)-based model, with sensitivity and specificity of 87.5% and 91.2%, respectively. Their follow-up study[11] further improved upon this performance using a two-stage spatial-information-fusion CNN, with an AUC of 0.961 and an overall accuracy of 92.6%.

In contrast, Park et al.[12] applied traditional machine learning models based on radiomic features, reporting an AUC of 0.863 and an accuracy of 83.5%, with sensitivity and specificity of 84.5% and 81.5%, respectively. Park et al.[13] extended this work by incorporating clinical implications into pre-processed deep learning models, leading to a notable improvement with an AUC of 0.931, a sensitivity of 88.9% and a specificity of 89.2%. Similarly, Yang et al.[14] employed ResNet-based architectures for sentinel lymph node metastasis prediction, achieving an AUC of 0.858 and an accuracy of 82.3%, with sensitivity ranging from 80.6% to 83.9% and specificity between 83.9% and 85.1% across primary and validation cohorts.

Findings across these studies suggest that deep learning systems can reliably capture morphological signatures correlated with lymph node metastasis, with CNN-based models frequently exceeding 90% accuracy[12,14]. Notably, classification and localisation steps were often integrated into two-stage frameworks, where initial detection or regional proposal of the lymph node area was followed by metastatic involvement prediction. These two-stage approaches have consistently demonstrated superior performance, particularly in terms of negative predictive value and specificity, when compared to conventional single-stage strategies.

Discussion

Summary of findings

Across studies examining the use of artificial intelligence (AI) in contrast-enhanced computed tomography (CECT) for axillary lymph node (ALN) analysis in breast cancer, deep-learning-based models have demonstrated encouraging potential. Early radiomics approaches relied on hand-crafted features and machine learning algorithms (e.g., SVM, logistic regression) but were constrained by relatively narrow feature sets or complex data preprocessing[12,13]. Subsequent studies introduced CNNs that exploited CECT image textures directly to achieve substantial diagnostic accuracy for ALN metastasis classification and detection[10]. These CNN models, often pre-trained on large-scale image datasets and further fine-tuned on CECT data, were shown to outperform classic machine learning methods. In particular, advanced architectures, such as multi-stage or spatial-

information-fusion-based networks, localised ALNs accurately despite their typically small size, thus improving the precision of metastasis prediction[11].

Overall, the evidence supports the adoption of deep learning for ALN analysis in CECT imaging, with the potential to enhance diagnostic confidence and streamline clinical workflows. However, variations in dataset characteristics, model architectures and preprocessing techniques highlight the need for further multi-centre validation before widespread clinical implementation.

A central strength of AI-based methods is their capacity to automatically learn complex image features that may be imperceptible to the human eye, even if expertly trained, especially relevant for detecting small or subtle ALNs[10,11]. This holds particular importance in scenarios where conventional manual interpretations fall short due to fatigue or observer variability. Additionally, two-stage CNN architectures leverage existing pre-training on large-scale natural images, which minimises the negative impact of relatively small local datasets. The improved performance metrics, such as AUC, negative predictive value and specificity, have direct clinical implications and enable more accurate preoperative staging and sparing patients from unnecessary invasive procedures[14].

Nevertheless, limitations persist. Despite the incorporation of advanced modules (e.g., spatial attention blocks), limited generalisation can emerge from single-centre or homogenous data[12]. Furthermore, label noise may exist if pathologic nodal findings do not align perfectly with image slices, especially in cases with multiple lymph nodes in proximity[11]. Image preprocessing routines also vary (e.g., window levels, contrast timing), which might hinder direct model comparison or replication. Finally, these methods currently focus on binary classification (metastasis vs. non-metastasis) and less on quantitative nodal burden or borderline nodal changes over time. Addressing these limitations requires prospective, multi-centre trials incorporating standardised imaging protocols, consistent labelling and robust external validation with real-world data.

Implications for future research

Although current algorithms report favourable outcomes in classifying and localising ALNs, challenges remain. First, obtaining robust performance in heterogeneous imaging conditions with differing CT scanners, scan protocols or patient anatomy necessitates more extensive multi-institutional data pooling[13]. The limited scale of most datasets underscores the need for larger, standardised repositories to mitigate overfitting and to facilitate external validation. Second, while many models emphasise classification accuracy, optimising clinical utility requires real-time deployment and interpretability: deeper exploration of explainable AI - e.g., heatmaps revealing salient features[19], occlusion sensitivity analysis to assess regional impact[20] and class-discriminative

activation mapping for enhanced feature attribution[21]- could strengthen clinician trust and acceptance. Moreover, refining multi-modal integration, combining CT with MRI, ultrasound or biopsy data, holds promise for improved confidence in borderline imaging findings. Third, to maximise clinical impact, future work should evaluate the effect of AI-driven CECT classification on treatment decisions (e.g., reduced surgical interventions or accelerated care pathways). Prospective randomised studies to investigate outcomes, including mortality and cost-effectiveness, would be highly valuable.

Strengths and limitations

To our knowledge, this is the first review to synthesise studies regarding the effectiveness of using AI to diagnose axillary lymph node metastasis in breast cancer CT scans. We followed robust methodological guidelines in the conduct and write up of this review, and all stages of study screening, quality assessment and data extraction were conducted independently by two members of the review team.

However, there are some limitations to note when interpreting the findings. Firstly, this field is in its infancy, and we only identified five papers which met our inclusion criteria. These studies were all conducted in single centres, limiting the generalisability of findings. Two studies were also judged to have a high risk of bias. Due to the heterogeneity of included studies, we were unable to perform a quantitative synthesis to examine the overall effect of AI screening across studies.

Equity and Global Health Relevance

CT-based AI solutions offer the potential for a scalable and cost-effective approach to improving axillary lymph node assessment in breast cancer, particularly in low- and middle-income countries, where access to MRI and PET-CT is often limited. Given the greater availability and affordability of CT scanners globally, integrating AI into routine CT imaging workflows could help reduce diagnostic inequities by providing accurate and non-invasive nodal staging in settings with constrained radiological expertise. This aligns with the objectives of the WHO Global Breast Cancer Initiative, which advocates for equitable access to timely diagnosis and treatment across all health systems, particularly in low- and middle-income countries[22].

Conclusion

Artificial intelligence-based analyses of contrast-enhanced CT images show substantial potential to accurately localise ALNs and predict metastasis, and as such can offer an increasingly precise non-invasive diagnostic strategy for breast cancer staging. Enhanced architectures incorporating attention mechanisms and multi-scale feature fusion have

moved beyond conventional radiomics to deliver improved sensitivity and specificity. Nonetheless, efforts to standardise data acquisition, expand external validation and link detection to downstream clinical decision-making remain vital next steps.

Contributors

JC (Dr J Cox) conceived the study, designed the review protocol, led the systematic search and data extraction, and wrote the first draft of the manuscript.

AA (Dr Amir Atapour-Abarghouei) provided technical expertise on artificial intelligence methodologies, contributed to data interpretation, and revised the manuscript for important intellectual content.

AB (Mr A Bhatti) supported drafting of the results and discussion sections.

All authors critically reviewed and approved the final manuscript. JC is the guarantor of the study.

Declarations of Interest

The authors declare no conflicts of interest.

Data Sharing Statement

This study is a systematic review of previously published literature. All data analysed during this review are available in the published articles included in the review and their supplementary materials. No individual-level or unpublished raw data were generated or used. Extracted data tables and the QUADAS-2 assessments used in the synthesis are available from the corresponding author upon reasonable request.

Acknowledgments

The authors would like to acknowledge the assistance of Sarah-Jane Raymond, Librarian employed by CDDFT, for her assistance with the literature searches which formed the basis of the quantitative synthesis and Dr Rebecca Appleton, for methodical advice about systematic reviews.

References

1. World Health Organization. Breast cancer: facts and figures. 2023. https://www.who.int/news-room/fact-sheets/detail/breast-cancer [Accessed 17/04/25]

2. International Agency for Research on Cancer (IARC). Global cancer observatory. 2025. https://gco.iarc.fr/ [Accessed 17/04/2025]

3. Mousavi-Kiasary SMS, Bayat M, Abbasvandi F, Khoundabi B, Mousavi F, Akbari A, Bagherian M, Zandi A, Honarvar B, Akbari ME. Tumor characteristics and survival rate of axillary metastatic breast cancer patients: a three decades retrospective cohort study. Scientific Reports. 2025;15(1):4571. doi:10.1038/s41598-024-84115-7

4. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines.

5. Esserman LJ, Lin N, Traina TA, Morrow M, Hudis CA, Winer EP. Imaging the axilla in patients with breast cancer: A call to action. JAMA Oncol. 2017;3(4):550–551. doi:10.1001/jamaoncol.2016.5171

6. Hieken TJ, Boughey JC, Jones KN, Degnim AC, Jakub JW, McLaughlin SA, et al. Accuracy of ultrasound-guided axillary lymph node biopsy after neoadjuvant chemotherapy in patients with node-positive breast cancer. J Clin Oncol. 2013;31(36):4311–4316. doi:10.1200/JCO.2013.49.2295

7. Pilewskie M, Morrow M. Complications of image-guided breast biopsy. Breast J. 2015;21(5):522–528. doi:10.1111/tbj.12454.

8. Bedi DG, Krishnamurthy R, Krishnamurthy S, Edeiken BS, Le-Petross H, Fornage BD, Bassett RL Jr, Hunt KK. Cortical morphologic features of axillary lymph nodes as a predictor of metastasis in breast cancer: in vitro sonographic study. AJR Am J Roentgenol. 2008;191(3):646–652. doi:10.2214/AJR.07.2460

9. James Lind Alliance. Priority Setting Partnership for breast cancer surgery. 2021.

10. Liu Z, Ni S, Yang C, Sun W, Huang D, Su H, Shu J, Qin N. Axillary lymph node metastasis prediction by contrast-enhanced computed tomography images for breast cancer patients based on deep learning. Comput Biol Med. 2021 Sep;136:104715. doi: 10.1016/j.compbiomed.2021.104715.

11. Liu Z, Huang D, Yang C, Shu J, Li J, Qin N. Efficient axillary lymph node detection via two-stage spatial-information-fusion-based CNN. Computer Methods and Programs in Biomedicine. 2022 Aug 1;223:106953.

12. Park S, Kim JH, Cha YK, Chung MJ, Woo JH, Park S. Application of Machine Learning Algorithm in Predicting Axillary Lymph Node Metastasis from Breast Cancer on Preoperative Chest CT. Diagnostics. 2023 Sep 14;13(18):2953.

13. Park TY, Kwon LM, Hyeon J, Cho BJ, Kim BJ. Deep Learning Prediction of Axillary Lymph Node Metastasis in Breast Cancer Patients Using Clinical Implication-Applied Preprocessed CT Images. Current Oncology. 2024 Apr 18;31(4):2278-88.

14. Yang X, Wu L, Ye W, Zhao K, Wang Y, Liu W, Li J, Li H, Liu Z, Liang C. Deep learning signature based on staging CT for preoperative prediction of sentinel lymph node metastasis in breast cancer. Academic Radiology. 2020 Sep 1;27(9):1226-33.

15. OECD. Health at a Glance 2023: CT and MRI units. https://www.oecd.org [Accessed 17/04/2025]

16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

17. 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.

18. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N, Roen K, Duffy S. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version. 2006 Apr 1;1(1):b92.

19. Brima Y, Atemkeng M. Saliency-driven explainable deep learning in medical imaging: bridging visual explainability and statistical quantitative analysis. BioData Mining. 2024 Jun 22;17(1):18.

20. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. InComputer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13 2014 (pp. 818-833). Springer International Publishing.

21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. InProceedings of the IEEE international conference on computer vision 2017 (pp. 618-626).

22. World Health Organization. Global Breast Cancer Initiative: reducing global breast cancer mortality through timely diagnosis and treatment. Geneva: WHO; 2021. https://www.who.int/initiatives/global-breast-cancer-initiative [Accessed 17/04/2025]

Appendix 1: Example search strategy

Search ID#	Search Terms	Search Options	Last Run Via	Results
S22	S9 AND S12 AND S17	Limiters - Publication Date: 20140101-20241231 Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	107
S21	S9 AND S12 AND S17	Limiters - Publication Date: 20190101-20241231 Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	95
S20	S9 AND S12 AND S17	Limiters - Publication Date: 20180101-20241231 Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	100
S19	S9 AND S12 AND S17	Limiters - Publication Date: 20140101-20241231 Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen -	107

			Advanced Search Database - MEDLINE Complete	
S18	S9 AND S12 AND S17	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	136
S17	S13 OR S14 OR S15 OR S16	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	487,129
S16	TI breast N3 cancer OR AB breast N3 cancer	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	386,537
S15	TI breast N3 tumo#r OR AB breast N3 tumo#r	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases	29,674

			Search Screen - Advanced Search Database - MEDLINE Complete	
S14	TI breast N3 neoplasm* OR AB breast N3 neoplasm*	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	2,605
S13	(MH "Breast Neoplasms")	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	343,558
S12	S10 OR S11	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	286,499

S11	TI lymph node* OR AB lymph node*	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	253,464
S10	(MH "Lymph Nodes")	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	95,509
S9	S7 AND S8	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	69,455
S8	TI analys* OR AB analys*	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database -	6,674,738

			MEDLINE Complete	
S7	S1 OR S2 OR S3 OR S4 OR S5 OR S6	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	197,625
S6	TI LLM OR AB LLM	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	1,451
S5	TI large language model* OR AB large language model*	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	2,171
S4	TI machine learning OR AB machine learning	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search	105,945

			Database - MEDLINE Complete	
S3	TI AI OR AB AI	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	55,360
S2	TI Artificial Intelligence OR AB Artificial intelligence	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	46,640
S1	(MH "Artificial Intelligence")	Expanders - Apply equivalent subjects Search modes - Proximity	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - MEDLINE Complete	47,634

Appendix 2: Reasons for exclusion

Authors	Year	Journal	Reason for exclusion
Ashokkumar,N. and Meera,S. and Anandan,P. and Murthy,M. Y. B. and Kalaivani,K. S. and Alahmadi,T. A. and Alharbi,S. A. and Raghavan,S. S. and Jayadhas,S. A.	2022	BioMed research international	retracted article
Lai,J. and Chen,Z. and Liu,J. and Zhu,C. and Huang,H. and Yi,Y. and Cai,G. and Liao,N.	2024	International journal of surgery (London, England)	wrong imaging method
Tang,Y. and Che,X. and Wang,W. and Su,S. and Nie,Y. and Yang,C.	2022	Medical physics	radiomics, not AI
Zhang,J. and Cao,G. and Pang,H. and Li,J. and Yao,X.	2021	Biomolecules & biomedicine	radiomics, not AI
Zhang,Jieqiu and Yin,Wei and Yang,Lu and Yao,Xiaopeng	2024	Molecular imaging and biology	radiomics, not AI
Zhu,Y. and Ma,Y. and Zhai,Z. and Liu,A. and Wang,Y. and Zhang,Y. and Li,H. and Zhao,M. and Han,P. and Yin,L. and He,N. and Wu,Y. and Sechopoulos,I. and Ye,Z. and Caballo,M.	2024	European radiology	radiomics, not AI

Appendix 3: Quality assessment scores

Author (date)	Domain 1a	Domain 1b	Domain 2a	Domain 2b	Domain 3a	Domain 3b	Domain 4
Liu et al., 2021							
Liu et al., 2022							
Park et al., 2023							
Park et al., 2024							
Yang et al., 2019