OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Artificial Intelligence

Artificial Intelligence for Prediction of Breast Cancer–Related Lymphedema: A Systematic Review

Hatan Mortada, MBBS*†
Mustafa Qais Muhsin Al-Khafaji, MD‡
Zainalabden E. Jefri, MBBS§
Faris Sultan, MBBS¶
Saad Alrobaiea, MBBS, MSc, SB-PLAST‖
Feras Alshomer, MBBS, MSc, SB-PLAST**

Background: Breast cancer–related lymphedema (BCRL) impairs quality of life and function, and early detection is crucial for preventive intervention. AI models, particularly machine learning and deep learning, hold promise for improving BCRL prediction and personalized risk stratification. This systematic review evaluates AI applications for predicting BCRL and, where available, summarizes how their reported performance compared with traditional diagnostic or non-AI methods.

Methods: This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. A literature search in MEDLINE, EMBASE, Cochrane Library, and Google Scholar was conducted in July 2024 using keywords related to breast cancer, lymphedema, and AI. Relevant studies were screened and selected based on predefined inclusion and exclusion criteria, with performance metrics such as accuracy, sensitivity, and specificity extracted for analysis.

Results: Seven studies from 2018 to 2024 were included, with 5284 patients. Among the 5 studies reporting BCRL classification performance, AI models achieved accuracies of 81%–93.75% (mean 87.2%), mean sensitivity 85.5%, and mean specificity 87.6%. Commonly reported predictors included higher body mass index, greater lymph node burden, and receipt of chemotherapy or radiotherapy. Younger age and neoadjuvant chemotherapy were associated with higher predicted risk in some studies.

Conclusions: AI shows promise for early detection and risk stratification in BCRL, but current models remain exploratory rather than clinically established. Small sample sizes, data heterogeneity, absent external validation, and inconsistent calibration highlight the need for larger, multi-institutional datasets, standardized evaluation, and prospective validation before clinical integration. *(Plast Reconstr Surg Glob Open 2026;14:e8132; doi: 10.1097/GOX.0000000000008132; Published online 6 October 2026.)*

*From the *Division of Plastic Surgery, Department of Surgery, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia; †Department of Plastic Surgery and Burn Unit, King Saud Medical City, Riyadh First Health Cluster, Riyadh, Saudi Arabia; ‡Department of Plastic Surgery, Norfolk and Norwich University Hospital, Norwich, United Kingdom; §Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia; ¶Division of Plastic and Reconstructive Surgery, Department of Surgery, Dalhousie University, Halifax, NS, Canada; ‖Department of Plastic Surgery and Burn Unit, Security Forces Hospital, Riyadh, Saudi Arabia; and **Division of Plastic Surgery, Department of Surgery, King Saud bin Abdulaziz University for Health and Sciences, Riyadh, Saudi Arabia.*

Received for publication December 20, 2025; accepted July 29, 2026.

All extracted data are included in the manuscript and supplemental files.

Copyright © 2026 The Authors. Published by Wolters Kluwer Health, LLC. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/GOX.0000000000008132

INTRODUCTION

Lymphedema in patients with breast cancer is a serious condition characterized by abnormal accumulation of fluid in tissues that normally drain via the lymphatic system as a result of compromised lymphatic drainage; and is usually caused by surgical intervention (such as mastectomies) and/or radiation treatment for breast cancer. Breast cancer–related lymphedema (BCRL) often manifests through severe swelling, skin changes, pain, and reduced mobility of the affected limb.[1] The burden of BCRL is considerable; the cumulative prevalence of BCRL increases over time, 6.8% at baseline, increasing to 23.8% after 7 years, and has been reported to occur at varying rates in different populations, ranging from 10.1% to 42.2% based on body mass index (BMI) and surgery type.[2–4] Known risk factors for developing BCRL are higher BMI values, increased numbers of lymph nodes dissected, taxanes used in chemotherapy, total mastectomies, regional nodal irradiation, and use of conventionally fractionated radiotherapy as opposed to hypofractionated radiotherapy.[3] Traditionally, clinicians have used clinical assessment,

Disclosure statements are at the end of this article, following the correspondence information.

measuring arm circumference, volumetric assessments, lymphoscintigraphy, and bioimpedance analysis to diagnose BCRL[5–7]; however, recent advances in AI and machine learning offer an exciting opportunity to identify patterns in larger datasets to predict and prevent BCRL more accurately. Although there have been many advances in AI and machine learning applications in healthcare, AI's utility in predicting BCRL is underdeveloped due to many barriers including data privacy concerns and potential biases in algorithms.[8] This systematic review aims to evaluate the performance of AI-based approaches for BCRL prediction and to identify the most effective techniques for early and accurate detection of lymphedema in breast cancer patients. Where available, reported comparisons with traditional or non-AI methods are also summarized. The goal of this review is to improve diagnostic accuracy and patient outcomes in breast cancer care.

Takeaways

Question: Can AI models improve early prediction and risk stratification of breast cancer–related lymphedema (BCRL) compared with traditional diagnostic approaches?

Findings: This systematic review of 7 studies (2018–2024; 5284 patients) evaluated machine learning and deep learning models for BCRL prediction. Across the 5 studies reporting classification performance, accuracies ranged from 81% to 93.75%, with a mean sensitivity of 85.5% and a specificity of 87.6%. Higher body mass index, lymph node burden, and chemotherapy or radiotherapy were recurrent predictors.

Meaning: AI models show promise for earlier, personalized BCRL risk stratification, but they remain exploratory until externally validated in prospective multicenter cohorts.

MATERIALS AND METHODS

Literature Search

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines[9] and Cochrane review methods[10] and was registered in PROSPERO (ID: CRD42024561004). Literature searches were conducted in July 2024 across MEDLINE, EMBASE, Cochrane Library, and Google Scholar, without time-frame limitation. Search terms included "breast cancer," "breast neoplasm," "breast carcinoma," "lymphedema," "lymphatic obstruction," "artificial intelligence," "machine learning," "deep learning," "neural network," "predictive model," "prediction," "early detection," and "early diagnosis."

Study Selection

The population, intervention, comparison, outcome, and study design framework guiding study selection is presented in Table 1. Articles were extracted from the databases and screened using Rayyan.[11] Four independent reviewers screened titles and abstracts, with full texts reviewed when the title and abstract provided insufficient information. Reference lists of included studies were also screened to identify additional relevant articles. Disagreements were resolved by an additional reviewer. The inclusion criteria were (1) studies focusing on breast cancer patients, published in English up to July 2024, without time-frame restriction; (2) studies utilizing AI, machine learning, or deep learning for prediction or risk stratification of lymphedema; (3) studies applying AI to BCRL classification, risk stratification, or structured data extraction relevant to BCRL identification, with quantitative model performance reported where applicable; and (4) prospective cohort, retrospective, cross-sectional, randomized controlled, case report, or case series designs. Direct comparison with traditional diagnostic or non-AI methods was not required, but comparator data were extracted when available.

Exclusion criteria were (1) nonoriginal study designs (systematic reviews, narrative reviews, editorials, opinion pieces, economic analyses, animal, in vitro, or cadaver studies); (2) no application of AI to BCRL classification, risk stratification, or data extraction; and (3) duplicate publications or overlapping populations, with only the most comprehensive study retained.

Screening and Data Extraction

Full-text reports were independently assessed by 2 reviewers, who screened studies against the inclusion criteria and extracted data including: study characteristics (first author, publication year, country, study design, sample size, follow-up duration), patient demographics and clinical features (age, sex, comorbidities, breast cancer stage, treatments

Table 1. Key Items for Framing the Aim, Search Strategy, and Study Inclusion and Exclusion Criteria for the Systematic Review

Item	Definition
Population	Patients with breast cancer at risk or diagnosed with lymphedema
Intervention	Use of AI modalities such as machine learning and deep learning for prediction or risk stratification of lymphedema
Comparison	Traditional diagnostic methods for lymphedema (eg, clinical evaluation, arm circumference measurement, lymphoscintigraphy, bioimpedance analysis) or non-AI prediction approaches, when reported
Outcome	Accuracy, sensitivity, specificity of AI-based models for predicting lymphedema; risk stratification and early detection outcomes
Study designs	Cross-sectional, retrospective, prospective, and cohort studies
Setting	Hospitals, breast cancer treatment centers, and research facilities
Timing	Post-breast cancer treatment follow-up ranging from months to years

received), AI methodology (AI type and technique, data sources, key model features, software used), comparator methods, and AI model performance metrics (accuracy, sensitivity, specificity, area under the receiver operating characteristic curve [AUC], and F1 score). Disagreements were resolved by an additional reviewer.

Quality Assessment and Risk of Bias

Risk of bias was assessed using the Prediction model Risk of Bias Assessment Tool (PROBAST), designed specifically for diagnostic and prognostic prediction model studies.[12] PROBAST evaluates 4 domains, participants, predictors, outcome, and analysis, through 20 signaling questions, each answered as yes, probably yes, probably no, no, or no information. Domain-level judgments were classified as low, high, and unclear risk of bias. Overall risk of bias was rated as low when all domains were low risk, high when at least 1 domain was high risk, and unclear when no domain was high risk but one or more were unclear. The level of evidence was classified according to the American Society of Plastic Surgeons evidence rating scale, as described by Sullivan et al.[13]

Analysis

Descriptive statistical analysis was performed using Microsoft Excel. Categorical variables (eg, study design, AI modalities, and treatment types) were summarized using frequencies and percentages. Continuous variables (eg, age, BMI, sample size) were reported as means and SDs for normally distributed data, or medians and interquartile ranges for nonnormally distributed data. AI model performance metrics, including accuracy, sensitivity, and specificity, were summarized using means and ranges. Pooled performance metrics were calculated only for the 5 studies reporting BCRL classification performance; Carleton et al and Nascimben et al were excluded from quantitative pooling, as they reported data extraction accuracy and cluster-label classification accuracy, respectively. Each study contributed its author-designated best-performing model. Due to heterogeneity in study designs, outcome definitions, and inconsistent reporting of comparator methods, comparisons between AI-based models and traditional or non-AI approaches were synthesized narratively rather than by formal quantitative pooling.

RESULTS

Study Selection

A total of 1990 studies were identified from 4 databases: MEDLINE (592), EMBASE (984), Cochrane Library (214), and Google Scholar (200). After removing duplicates, 1625 studies remained for eligibility screening. After excluding 1602 records during title and abstract screening, 23 articles were assessed in full text. Of these, 16 were excluded: 6 for wrong study design, 5 as duplicate publications, 3 for outcome relevance, 1 as a non-English study, and 1 for inaccessible full text. Two of the outcome-relevance exclusions were studies by Yang et al[14] and Xu et al,[15] which predicted lymphovascular invasion, a tumor pathological characteristic, rather than BCRL. This resulted in 7 studies (2018–2024) being included in the qualitative synthesis.[16–22] The selection process is detailed in the PRISMA flow diagram (Fig. 1).

Study Characteristics

Among the 7 included studies, 3 were conducted in China (42.9%), 2 in the United States (28.6%), 1 in Italy (14.3%), and 1 in Korea (14.3%). Five studies (71.4%) used a cross-sectional or observational design, with 2 retrospective studies (28.6%) (Table 2). The primary aim of most studies was to develop or evaluate AI models for BCRL prediction; Carleton et al instead applied natural language understanding to extract clinical data in the context of axillary staging de-escalation. Most studies developed supervised machine learning models for BCRL prediction using clinical, symptom, or laboratory data; the remaining studies applied unsupervised machine learning for risk stratification and natural language understanding for clinical data extraction.

Patient Demographics and Clinical Characteristics

Sample sizes ranged from 294 to 2137, with a combined total of 5284 patients. Two studies reported follow-up duration: 1 reported 5–8 years (median 5.5 y) and the other a mean of 28.7 months. Across the 5 studies reporting a numeric mean age, mean ages ranged from 49.9 to 59.8 years; Carleton et al reported a median age of 67 years, and Fu et al reported age bands. Three studies reported comorbidities, noting small percentages of smoking (4%) and alcohol use (3.8%), with hypertension more prevalent in the lymphedema group (21.8%) compared to nonlymphedema patients (14.3%). Breast cancer stage was reported in 4 studies, with most patients in stage II (52.8%–68.1%).

Treatment Details

Surgical interventions were reported in 7 studies. Sentinel lymph node biopsy (SLNB) was documented in 5 studies, whereas mastectomy was the most commonly reported surgery type. Modified radical mastectomy was performed in 55.7% of participants in 1 study, and 76% of patients had a mastectomy in another without specifying the type of mastectomy. Radiation therapy rates were reported in 4 studies, ranging from 32.2% to 67.2% of patients. Any-chemotherapy rates were reported in 4 studies, ranging from 61.4% to 86.5% of patients, administered before or after surgery.

AI Modalities and Techniques

Machine learning was the most commonly used modality, present in 5 (71.4%) studies, followed by deep learning in 1 (14.3%) study and natural language understanding in 1 (14.3%) study. Logistic regression and support vector machine were the most frequent AI techniques. Neural networks (artificial neural network, multilayer perceptron), Random Forest, and Decision Trees appeared across multiple studies. Data sources included clinical records, imaging data, and

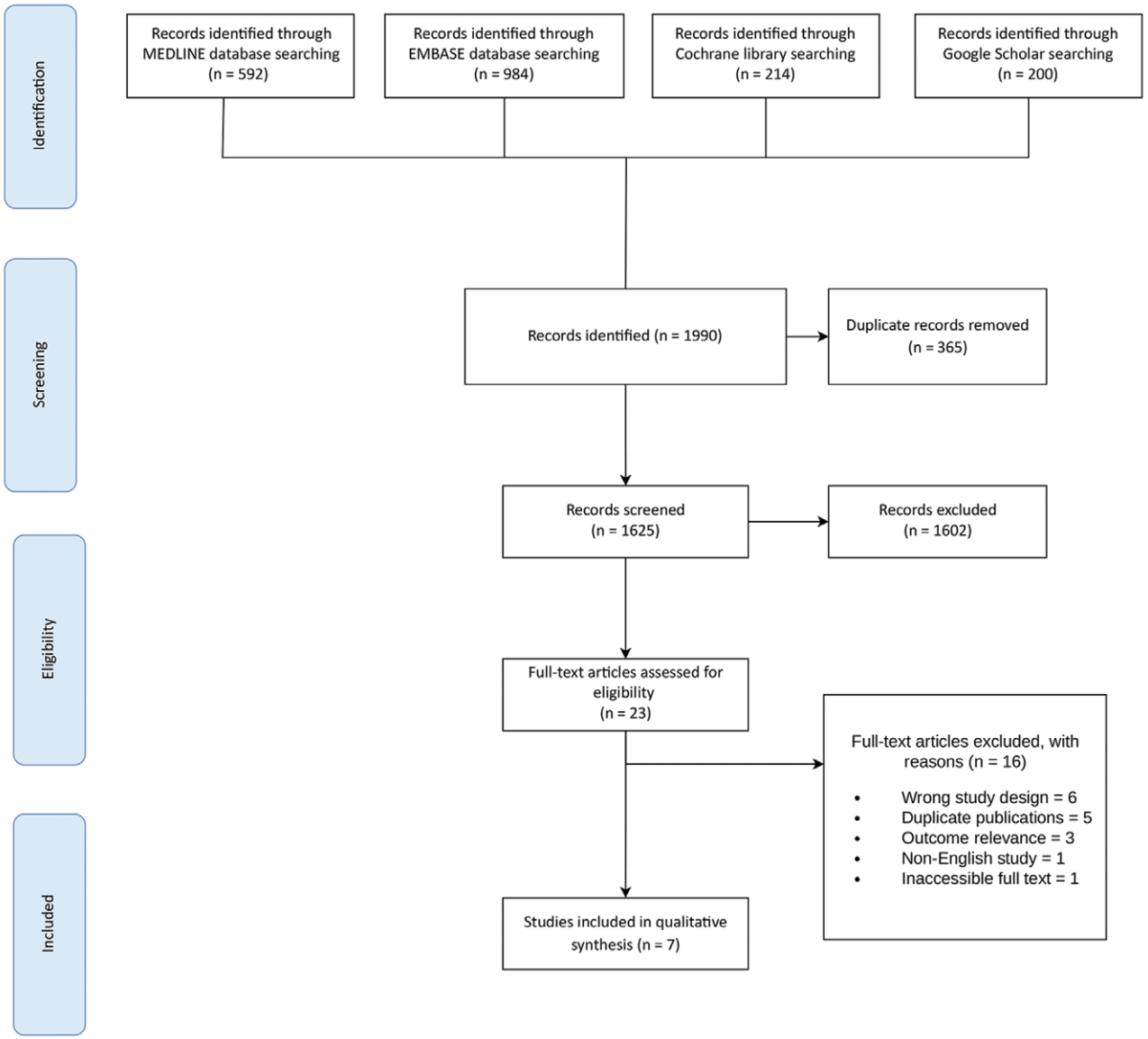


Fig. 1. PRISMA flowchart illustrating the study selection process.

Table 2. Study and Patient Population Characteristics

Study ID	Country	Study Design	n	Follow-up	Mean Age, y	Cancer Stage
Carleton et al[16]	USA	Observational	925	5–8 y	67*	Early stage ER+, cN0
Nascimben et al[17]	Italy	Cross-sectional	294	N/A	59.8 ± 12.9	Stage I–IV
Wu et al[18]	China	Retrospective cohort	370	28.7 months	49.9 ± 7.5	Stage I–IV
Wei et al[19]	China	Cross-sectional	533	N/A	58 ± 11.3	N/A
Fu et al[20]	USA	Cross-sectional	355	N/A	40–59 (53%)	N/A
Du et al[21]	China	Observational	670	N/A	50.4 ± 10	Stage 0–IV
Trinh et al[22]	Korea	Retrospective	2137	N/A	55.9 ± 11.4	Not specified

*Carleton et al age is a median; remaining values are means.
±, mean ± SD; cN0: clinically node-negative ; ER+: estrogen receptor positive; n, sample size; N/A: not available.

patient-reported outcomes, with most studies relying on electronic medical records. Common AI model features included age, BMI, breast cancer stage, lymph node involvement, and treatment details such as chemotherapy and radiotherapy.

Diagnostic Methods and Outcomes

Pooled performance was assessed for the 5 studies reporting BCRL classification metrics; Carleton et al and Nascimben et al were excluded, as they reported data extraction accuracy and cluster-label accuracy,

respectively, rather than BCRL classification. Across these 5 studies, AI models achieved accuracies ranging from 81% to 93.75%, with a mean of 87.2% (±5.6%). Sensitivity ranged from 77.1% to 95.65% (mean 85.5%, ±7.6%), whereas specificity, reported in 4 studies, ranged from 82.0% to 91.03% (mean 87.6%, ±3.9%). AUC values, reported in 4 studies, ranged from 0.870 to 0.965, with a mean of 0.914 (±0.043). F1 scores, reported in 3 studies, ranged from 0.68 to 0.904, with a mean of 0.816 (±0.122). With regard to model validation, all 5 studies relied on internal validation (cross-validation, a single random train-test split, or both), and none performed external validation on a separately recruited, independent cohort. The per-study performance and validation approach are detailed in Table 3.

Model-identified Predictors and Associated Characteristics

Frequently reported predictors or associated characteristics in AI-based BCRL models included a higher number of positive lymph nodes, higher BMI, postoperative chemotherapy, and radiotherapy. In some studies, younger age, neoadjuvant chemotherapy, advanced N-stage disease, and mastectomy were associated with higher predicted BCRL risk. Lifestyle-related variables, including higher meat/fried-food consumption and lower exercise frequency, were also reported as associated features in some models (Table 4).

Level of Evidence and Bias Assessment

Risk of bias results are summarized in Table 5. Most studies were rated as low risk in the participants and outcome domains, although some lacked clear lymphedema outcome definitions. The predictors domain showed low or unclear risk across most studies. The analysis domain was the most common source of high risk of bias, driven by inadequate sample sizes relative to candidate predictors, inconsistent calibration assessment, absence of external validation, and limited overfitting prevention measures. Overall, 4 studies were rated as high risk of bias, 1 as unclear, and 2 as low risk.

Table 3. AI Modalities, Software Platforms, and Performance Metrics

Study ID	AI Modality	Best Model	Software/Platform	Accuracy	Sensitivity	Specificity	AUC	F1	Validation
Carleton et al[16]	NLU	NLU pipeline (LLM + ML + rules)	Custom NLU pipeline	93%*	NR	NR	NR	NR	Internal (manual verification, 50% subset)
Nascimben et al[17]	Unsupervised ML	Random Forest on cluster labels	Python (scikit-learn, UMAP-learn)	NR†	NR	NR	NR	NR	Internal (repeated 5-fold CV)
Wu et al[18]	Machine learning	Logistic regression	Python (scikit-learn)	91.6%	91.4%	NR	0.965	0.865	Internal (single 80:20 split)
Wei et al[19]	Machine learning	Logistic regression	Python, R (web app)	82.5%	77.1%	88.3%	0.889	NR	Internal (7:3 split, 10-fold CV)
Fu et al[20]	Deep learning	Artificial neural network	MATLAB	93.75%	95.65%	91.03%	NR	NR	Internal (5-fold CV)
Du et al[21]	Machine learning	Logistic regression	Python (scikit-learn)	81%	79%	82%	0.87	0.68	Internal (7:3 split, 10-fold CV)
Trinh et al[22]	Machine learning	Random Forest	R (caret, randomForest)	87.0%	84.3%	89.1%	0.931	0.904	Internal (80:20 hold-out, 3 repeats, 10-fold CV)

Values reflect each study's author-designated best-performing model.
*Carleton et al reported a per-patient data extraction accuracy, not BCRL classification performance.
†Nascimben et al reported balanced accuracy for cluster-label assignment (up to 99.4%), not BCRL classification. Carleton et al and Nascimben et al are therefore excluded from pooled performance estimates. Wu et al reported precision and recall in place of specificity. Fu et al reported an AUC of 0.751 for a non-AI symptom-count comparator, not for the ANN.
ANN, artificial neural network; CV, cross-validation; LLM, large language model; ML, machine learning; NLU, natural language understanding; NR, not reported.

Table 4. Identified High-risk Characteristics for BCRL

Study ID	Identified High-risk Characteristics
Carleton et al[16]	N/A (compared nodal positivity and lymphedema rates by age and T stage; no predictor ranking)
Nascimben et al[17]	Higher number of metastatic lymph nodes, tumor grade, hormone therapy, age group, trastuzumab, HER2, taxane-based chemotherapy; also N stage, higher BMI, lymphovascular invasion, comorbidities
Wu et al[18]	No. positive lymph nodes, BCRL on the same side as surgery, history of SLNB, meat/fried-food dietary preference, exercise frequency under 3 times per week, months since surgery
Wei et al[19]	Symptom-based model: skin toughness/thickness, tingling, arm limitation, arm swelling, blistering, arm tightness, arm heaviness
Fu et al[20]	N/A (symptom-based detection using 26 self-reported lymphedema symptoms; no clinical risk-factor ranking)
Du et al[21]	BMI (highest importance), number of positive lymph nodes, surgical side, N stage, neoadjuvant chemotherapy, type of lymph node surgery, postoperative radiotherapy, type of surgery, clinical staging
Trinh et al[22]	No. lymph nodes harvested (highest importance), taxane-based chemotherapy, age, radiation

N/A, no clinical risk-factor ranking produced.

Table 5. Risk of Bias Assessment of Included Studies Using the PROBAST

Study	Participants	Predictors	Outcome	Analysis	Overall Risk of Bias
Carleton et al[16]	Low	Low	Low	High	High
Nascimben et al[17]	Low	Low	Low	High	High
Wu et al[18]	Low	Low	Low	High	High
Wei et al[19]	Low	Low	Low	Low	Low
Fu et al[20]	Low	Low	High	Low	High
Du et al[21]	Low	Low	Unclear	Low	Unclear
Trinh et al[22]	Low	Low	Low	Low	Low

DISCUSSION

This systematic review included 7 studies applying AI to predict BCRL. Machine learning was the most commonly used modality (71.4%), followed by deep learning (14.3%) and natural language understanding (14.3%). Across the 5 studies reporting BCRL classification metrics, AI models demonstrated promising performance, with accuracy ranging from 81% to 93.75%, mean sensitivity of 85.5%, and mean specificity of 87.6%. However, these estimates warrant cautious interpretation given the retrospective nature of much of the evidence, heterogeneity in study design and outcome definitions, variable sample sizes, and limited external validation. Models utilized diverse data sources, including clinical records, imaging data, and patient-reported outcomes, with BMI, breast cancer stage, lymph node involvement, and treatment details emerging as recurrent predictive features. Frequently identified high-risk subgroups included patients with higher numbers of positive lymph nodes, those receiving chemotherapy or radiotherapy, younger women, and patients undergoing neoadjuvant chemotherapy. Lifestyle factors such as high meat/fried-food consumption and low exercise frequency were also reported as associated features. Importantly, because the included studies were observational and prediction-focused, these variables should be interpreted as predictive associations rather than causal determinants of BCRL. Although comparison with traditional diagnostic or non-AI prediction approaches was a review objective, such comparator data were inconsistently reported and could only be interpreted narratively.

These findings align with previous research. Du et al reported that logistic regression achieved an accuracy of 81%, an AUC of 0.87, and a sensitivity of 79% using similar predictive features.[21] Lin et al evaluated 17 models with AUC values of 0.680 to 0.908; the 4 AUC values available among our included studies fell within a narrower range of 0.870 to 0.965, whereas that review similarly noted high risk of bias across most studies.[23] Ozmen and Schwarz reported a higher prediction accuracy of 95.25% using a deep learning model trained on synthetic data generated by generative adversarial networks, suggesting that synthetic data augmentation may enhance AI performance in this domain.[24] However, this study was excluded from our review, as it was available only as a conference abstract without full peer-reviewed data.

Clinically, if prospectively validated, these AI models could support earlier risk stratification by identifying high-risk patients for closer surveillance and timely interventions such as physical therapy or lifestyle modifications. With a mean sensitivity of 85.5%, these models could theoretically identify approximately 86 of every 100 at-risk patients, whereas a mean specificity of 87.6% suggests that approximately 88 of every 100 lower-risk patients may be correctly classified, potentially reducing unnecessary monitoring and referral. In practice, such stratification may guide earlier referral to lymphedema prevention services and more targeted patient education. However, these sensitivity and specificity estimates should be interpreted cautiously, as positive and negative predictive values, which depend on disease prevalence, were not consistently reported across the included studies and are essential for translating model performance into clinically meaningful decision-making.

This review has several strengths, including the use of PROBAST for bias assessment, incorporation of studies utilizing diverse data sources and AI methodologies, transparent reporting of 2 lymphovascular invasion-focused studies[14,15] excluded at full-text screening, and a clearly delineated analytic approach in which two studies applying AI to data extraction and unsupervised risk stratification were retained for qualitative synthesis but excluded from pooled performance estimates. However, several limitations must be acknowledged. First, substantial methodological heterogeneity across the included studies, in AI algorithms, predictor types, lymphedema definitions, and validation strategies, precluded quantitative meta-analysis and limited direct comparability of model performance. Second, many studies had relatively small sample sizes, and none underwent external validation in an independent cohort, restricting generalizability. Third, direct head-to-head comparisons between AI models and traditional diagnostic methods were not available in all studies. Fourth, the geographic concentration of included studies, 3 of 7 (42.9%) from China, with no representation from the Middle East, Africa, or South America, limits the applicability of findings across diverse patient demographics, healthcare systems, and surgical practices. Fifth, calibration was inconsistently assessed across the included studies. Only a minority reported formal calibration (Wei et al[19] reported Brier scores, and Wu et al[18] presented calibration plots), and where calibration was examined the methods were heterogeneous, which prevents any pooled appraisal of how well predicted probabilities aligned with observed outcomes. This omission has direct clinical implications. A model may achieve strong discrimination (a high AUC) while remaining poorly calibrated, systematically over- or underestimating an individual patient's absolute risk. Because BCRL risk-based decisions, including the intensity of surveillance, the timing of referral to lymphedema

prevention services, and the framing of patient counseling, depend on the reliability of the predicted probability itself rather than on rank ordering alone, a miscalibrated model could lead to overtreatment and unnecessary monitoring of lower-risk patients, or false reassurance of higher-risk patients. The absence of consistent calibration data therefore limits the safe definition of clinical risk thresholds and meaningful decision-curve or net-benefit analysis, reinforcing that these models are not yet suitable for individual-level clinical decision-making. As summarized in Table 3, all studies reporting BCRL classification performance relied exclusively on internal validation, and none was externally validated in a separately recruited, independent cohort. This substantially limits confidence in how the reported performance would generalize to new patient populations, healthcare systems, and surgical practices. Future research should prioritize larger, geographically diverse, multi-institutional cohorts with standardized lymphedema outcome definitions and consistent model development frameworks. Greater emphasis on external validation, calibration assessment, and uniform performance metric reporting would improve cross-study comparability and facilitate more robust evidence synthesis. Incorporating real-time patient-reported outcomes and synthetic data augmentation also hold promise. Overall, current AI models for BCRL prediction should be viewed as exploratory decision-support candidates rather than clinically established tools.

CONCLUSIONS

This systematic review highlights the emerging potential of AI, particularly machine learning and deep learning, in predicting BCRL, with models demonstrating promising accuracy, sensitivity, and specificity. Recurrent predictive variables identified across studies may support early risk stratification and personalized monitoring, though these reflect predictive associations rather than established causal risk factors. Limitations including small sample sizes, methodological heterogeneity, and absent external validation underscore that these models remain exploratory. Future prospective, multicenter validation studies with standardized outcome definitions and larger, geographically diverse cohorts are needed before AI-based prediction tools can be considered for routine clinical use.

Hatan Mortada, MBBS
Division of Plastic Surgery, Department of Surgery
King Saud University Medical City, King Saud University
Riyadh, Saudi Arabia
Department of Plastic Surgery and Burn Unit
King Saud Medical City, Riyadh First Health Cluster
Riyadh, Saudi Arabia
E-mail: Hatanmortada@gmail.com
X: @HatanMortada
Instagram: @HatanMortada

DISCLOSURE

The authors have no financial interest to declare in relation to the content of this article.

ACKNOWLEDGMENT

This work was supported by the College of Medicine Research Center, Deanship of Scientific Research, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia.

REFERENCES

1. Larocque G, McDiarmid S. The legacy of lymphedema: impact on nursing practice and vascular access. *Can Oncol Nurs J.* 2019;29:194–203.
2. Ren Y, Kebede MA, Ogunleye AA, et al. Burden of lymphedema in long-term breast cancer survivors by race and age. *Cancer.* 2022;128:4119–4128.
3. Byun HK, Chang JS, Im SH, et al. Risk of lymphedema following contemporary treatment for breast cancer: an analysis of 7617 consecutive patients from a multidisciplinary perspective. *Ann Surg.* 2021;274:170–178.
4. Chuang XE, Chia CLK. Breast cancer-related lymphedema (BCRL): should we be doing more or less for the axilla? *Ann Acad Med Singap.* 2024;53:67–68.
5. Pagliara D, Grieco F, Rampazzo S, et al. Prevention of breast cancer-related lymphedema: an up-to-date systematic review of different surgical approaches. *J Clin Med.* 2024;13:555.
6. Choi JE, Chang MC. Management of lymphedema is really a matter in patients with breast cancer. *World J Clin Cases.* 2024;12:2482–2486.
7. Cornish BH, Chapman M, Hirst C, et al. Early diagnosis of lymphedema using multiple frequency bioimpedance. *Lymphology.* 2001;34:2–11.
8. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nat Med.* 2022;28:31–38.
9. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.
10. Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0.* The Cochrane Collaboration; 2011. Available at: https://training.cochrane.org/handbook. Accessed June 1, 2021.
11. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan: a web and mobile app for systematic reviews. *Syst Rev.* 2016;5:210.
12. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170:51–58.
13. Sullivan D, Chung KC, Eaves FF, 3rd, et al. The level of evidence pyramid: indicating levels of evidence in plastic and reconstructive surgery articles. *Plast Reconstr Surg.* 2011;128:311–314.
14. Yang X, Fan X, Lin S, et al. Assessment of lymphovascular invasion in breast cancer using a combined MRI morphological features, radiomics, and deep learning approach based on dynamic contrast-enhanced MRI. *J Magn Reson Imaging.* 2024;59:2238–2249.
15. Xu Z, Xie Y, Wu L, et al. Using machine learning methods to assess lymphovascular invasion and survival in breast cancer: performance of combining preoperative clinical and MRI characteristics. *J Magn Reson Imaging.* 2023;58:1580–1589.
16. Carleton N, Saadawi G, McAuliffe PF, et al. Use of natural language understanding to facilitate surgical de-escalation of axillary staging in patients with breast cancer. *JCO Clin Cancer Inform.* 2024;8:e2300177.
17. Nascimben M, Lippi L, de Sire A, et al. Algorithm-based risk identification in patients with breast cancer-related lymphedema: a cross-sectional study. *Cancers (Basel).* 2023;15:336.
18. Wu X, Guan Q, Cheng ASK, et al. Comparison of machine learning models for predicting the risk of breast cancer-related lymphedema in Chinese women. *Asia Pac J Oncol Nurs.* 2022;9:100101.

19. Wei X, Lu Q, Jin S, et al. Developing and validating a prediction model for lymphedema detection in breast cancer survivors. *Eur J Oncol Nurs*. 2021;54:102023.
20. Fu MR, Wang Y, Li C, et al. Machine learning for detection of lymphedema among breast cancer survivors. *Mhealth*. 2018;4:17.
21. Du J, Yang J, Yang Q, et al. Comparison of machine learning models to predict the risk of breast cancer-related lymphedema among breast cancer survivors: a cross-sectional study in China. *Front Oncol*. 2024;14:1334082.
22. Trinh XT, Chien PN, Long NV, et al. Development of predictive models for lymphedema by using blood tests and therapy data. *Sci Rep*. 2023;13:19720.
23. Lin Q, Yang T, Yongmei J, et al. Prediction models for breast cancer-related lymphedema: a systematic review and critical appraisal. *Syst Rev*. 2022;11:217.
24. Ozmen BB, Schwarz G. Predicting breast cancer related lymphedema after immediate lymphatic reconstruction: an artificial intelligence approach with synthetic data. *Plast Reconstr Surg Glob Open*. 2024;12(1 Suppl 1):65.