INPLASY

Artificial Intelligence for Breast Cancer Molecular Subtype Prediction from Medical Imaging: a Systematic Review of Literature

INPLASY202650096
doi: 10.37766/inplasy2026.5.0096
Received: 17 May 2026
Published: 17 May 2026

Morales-Duran, JC; Osorno-Castillo, K; Diaz, GM.

Corresponding author:
Gloria Mercedes Diaz Cabrera

gloriadiaz@itm.edu.co

Author Affiliation:
Departamento de Sistemas, Institucion Universitaria ITM, Medellin, Colombia.

ADMINISTRATIVE INFORMATION

Support - This work was supported by Institucion Universitaria ITM, young researchers program 2025.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202650096

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 17 May 2026 and was last updated on 17 May 2026.

INTRODUCTION

Review question / Objective This systematic review aims to evaluate the diagnostic performance (AUC, accuracy, sensitivity, and specificity) of artificial intelligence methods applied to medical imaging for the non-invasive classification of breast cancer molecular subtypes (Luminal A, Luminal B, HER2-positive, and triple-negative) regranding to the immunohistochemical analysis as the reference standard, regranding to the reported performance metrics (AUC, accuracy, sensitivity, and specificity). And evaluate the methodological quality and risk of bias of included original peer-reviewed studies using an AI-adapted QUADAS-2 framework.

Rationale Breast cancer is the second most common malignancy worldwide, with over 2.3 million new cases in 2022. Its biological heterogeneity is captured through four major molecular subtypes (Luminal A, Luminal B, HER2-positive, and triple-negative) each requiring distinct therapeutic approaches. Currently, subtype determination relies on immunohistochemical analysis of biopsy specimens, a method that is invasive, subject to inter-observer variability, and often associated with delays in clinical decision-making. These limitations have motivated growing interest in non-invasive alternatives based on artificial intelligence applied to medical imaging.
Several systematic reviews addressing related topics have been published in recent years. One meta-analysis focused exclusively on ultrasound radiomics for molecular subtype prediction, reporting pooled AUCs exceeding 0.85 for specific subtypes. Another catalogued 81 publications using mammography and tomosynthesis for subtyping, staging, and prognosis. A third examined radiogenomics associations across multiple modalities. However, these prior works present important limitations that leave key questions unanswered.
First, existing reviews are restricted to a single imaging modality or a narrowly defined imaging-

genomics framework, and none encompasses the full spectrum of breast imaging modalities relevant to molecular subtyping, including DCE-MRI, ultrasound, mammography, contrast-enhanced spectral mammography, and digital breast tomosynthesis. Second, and critically, none of these prior reviews applied a formal risk-of-bias assessment instrument adapted to the specific methodological challenges of AI-based diagnostic studies. Without such an assessment, it is impossible to distinguish technically sound findings from those derived from studies with inflated performance due to data leakage, image-level data splitting, or inadequate external validation. Third, the absence of quality filtering means that reported performance metrics across the existing literature cannot be reliably interpreted or compared.

This systematic review was therefore conducted to address these gaps. By covering all relevant imaging modalities and applying a rigorous AI-adapted QUADAS-2 framework, this review provides a more complete and methodologically sound synthesis of the current evidence. The need for this new review is further supported by the rapid growth of publications in this field between 2020 and 2025, which requires an updated and critically appraised assessment to inform both researchers and clinicians about the true state of evidence and the barriers that remain for clinical translation.

Condition being studied The condition being studied is breast cancer, specifically its molecular subtype classification (Luminal A, Luminal B, HER2-positive, and triple-negative) through medical imaging. Breast cancer is the second most common malignancy worldwide, characterized by significant biological heterogeneity with distinct prognostic and therapeutic implications across subtypes. This review addresses the capacity of artificial intelligence methods applied to medical imaging modalities (including DCE-MRI, ultrasound, mammography, contrast-enhanced spectral mammography, and digital breast tomosynthesis) to non-invasively identify these molecular subtypes, which are currently determined through invasive biopsy and immunohistochemical analysis.

METHODS

Search strategy The literature search was conducted across four electronic databases: Scopus, Web of Science, ScienceDirect, and PubMed. The search was limited to English-language, peer-reviewed original articles published between January 1, 2020, and March 25, 2025.

The following search string was applied across all databases, combining keywords and Boolean operators:

("breast cancer") AND ("subtype") AND ("image" OR "MRI" OR "ultrasound" OR "mammography" OR "CESM") AND ("prediction" OR "classification" OR "analysis") AND ("radiomics" OR "machine learning" OR "deep learning").

Participant or population The population includes patients with breast cancer enrolled in studies that applied artificial intelligence to medical imaging for molecular subtype prediction. No restrictions were applied based on age, ethnicity, or geographic origin. Studies were eligible regardless of imaging modality (DCE-MRI, ultrasound, mammography, contrast-enhanced spectral mammography, or digital breast tomosynthesis), provided that molecular subtype classification was reported as the primary outcome.

Intervention The intervention of interest is the application of artificial intelligence techniques to medical imaging for the non-invasive prediction of breast cancer molecular subtypes. This encompasses two main categories of approaches: radiomics-based machine learning models (including support vector machines, logistic regression, random forests, and k-nearest neighbors) and deep learning architectures such as convolutional neural networks and deep neural networks. Eligible interventions must be applied to in vivo medical images. Studies relying exclusively on conventional statistical analysis without an artificial intelligence component, or those using non-in vivo images such as histological specimens, were not eligible.

Comparator The comparator is the standard immunohistochemical analysis of tissue samples obtained through biopsy, which evaluates the expression of estrogen receptor, progesterone receptor, HER2, and Ki-67 to determine breast cancer molecular subtype. This method serves as the reference standard against which the diagnostic performance of artificial intelligence models was assessed. Studies without an explicit human or clinical comparator were also eligible, provided that immunohistochemical-based molecular subtype labels were used as ground truth for model training and evaluation.

Study designs to be included Prospective and retrospective diagnostic accuracy studies, single-center and multicenter observational studies, and cross-sectional studies reporting the performance of artificial intelligence models applied to medical imaging for breast cancer molecular subtype

classification will be included. Only original, peer-reviewed research articles available in full text and published in English will be considered eligible.

Eligibility criteria Eligible studies must apply artificial intelligence techniques (including radiomics-based machine learning or deep learning) to in vivo medical imaging for the prediction or classification of at least one breast cancer molecular subtype (Luminal A, Luminal B, HER2-positive, or triple-negative), using immunohistochemical analysis as the reference standard. Studies must be original, peer-reviewed research articles available in full text, published in English between January 1, 2020, and March 25, 2025.
Studies will be excluded if they focus exclusively on outcomes other than molecular subtype classification, such as neoadjuvant chemotherapy response, pathological complete response, or disease-free survival; rely solely on conventional statistical analysis without an artificial intelligence component; use non-in vivo images such as histological or ex vivo specimens; or correspond to non-original publications including systematic reviews, narrative reviews, meta-analyses, editorials, letters, commentaries, or conference abstracts and proceedings not available as full-text articles.

Information sources Information sources will include PubMed, Scopus, Web of Science, and ScienceDirect. Additional records will be identified by manually screening reference lists of eligible studies and relevant systematic reviews on related topics. No grey literature, dissertations, theses, or conference abstracts were considered as additional sources. The search was restricted to English-language, peer-reviewed original articles published between January 1, 2020, and March 25, 2025.

Main outcome(s) The main outcomes are the diagnostic performance metrics of artificial intelligence models applied to medical imaging for breast cancer molecular subtype classification, including area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity. Performance metrics were extracted for each reported molecular subtype classification task (Luminal A, Luminal B, HER2-positive, and triple-negative breast cancer). For studies reporting multiple classification configurations, metrics were extracted for the best-performing model per task. A secondary outcome is the methodological quality and risk of bias of included studies, assessed across four domains(patient selection, index test, reference standard, and flow and timing) using an AI-adapted QUADAS-2 framework proposed for the authors.

Additional outcome(s) Not applicable.

Data management Two reviewers independently screened titles and abstracts using the Rayyan web-based platform for systematic reviews. Potentially eligible articles were subsequently retrieved in full text for detailed eligibility assessment and risk-of-bias evaluation. Any discrepancies between reviewers at any stage were resolved through discussion and consensus. Data extraction was performed independently by both reviewers using a standardized extraction form, collecting the following variables for each included study: publication year, study design, sample size and class distribution, data source, imaging modality, molecular subtype classification task, AI model architecture, and reported performance metrics including AUC, accuracy, sensitivity, and specificity. Statistical analyses and tabulation of results were performed using standard spreadsheet software.

Quality assessment / Risk of bias analysis Methodological quality was assessed using an AI-adapted QUADAS-2 framework for diagnostic accuracy studies. Two reviewers independently evaluated risk of bias and applicability concerns across four domains: patient selection, index test, reference standard, and flow and timing. Additional AI-specific signaling questions were incorporated into each domain to address methodological aspects critical to AI-based studies, including reporting of imaging acquisition protocols and scanner information, preprocessing strategies, rationale and breakdown of training and test sets, patient-level data partitioning, class distribution stratification, external validation, hyperparameter tuning, appropriateness of evaluation metrics, and independence between molecular subtype labels and AI model outputs. Each domain was rated as low, high, or unclear risk of bias. Disagreements between reviewers were resolved through discussion and consensus. Studies presenting high or unclear risk of bias in at least one domain were excluded from the final qualitative synthesis.

Strategy of data synthesis Due to the substantial heterogeneity across included studies in terms of imaging modalities, acquisition protocols, molecular subtype classification tasks, AI model architectures, and reported performance metrics, a quantitative synthesis through meta-analysis was not feasible. Data synthesis was therefore conducted through a qualitative narrative

approach. For each included study, diagnostic performance metrics (including AUC, accuracy, sensitivity, and specificity) were extracted and tabulated according to imaging modality, AI approach (machine learning vs. deep learning), and classification task. Results were organized and described separately by imaging modality to avoid conflating modality-specific performance. Risk-of-bias ratings across the four QUADAS-2 domains were summarized for all assessed studies. Only studies with low risk of bias across all domains were included in the final qualitative synthesis.

Subgroup analysis Not applicable. Given the qualitative nature of the data synthesis, formal subgroup analyses were not performed.

Sensitivity analysis The primary sensitivity analysis consisted of restricting the final qualitative synthesis to studies rated as low risk of bias across all four QUADAS-2 domains (patient selection, index test, reference standard, and flow and timing). Studies presenting high or unclear risk of bias in at least one domain were excluded to assess the robustness of findings under strict methodological quality criteria. An additional sensitivity analysis was performed by restricting inclusion to studies with external validation on an independent dataset, to evaluate whether diagnostic performance estimates remained consistent when assessed under more rigorous generalizability conditions.

Language restriction English.

Country(ies) involved Colombia.

Other relevant information This systematic review was conducted in accordance with the PRISMA 2020 guidelines. The review has been completed and submitted for publication. This protocol is being registered retrospectively, as the review was conducted prior to registration due to time constraints related to the submission process.

Keywords Breast Cancer, Medical Imaging, Molecular subtype, Artificial intelligence, QUADAS.

Dissemination plans The results of this systematic review will be disseminated through publication in a peer-reviewed scientific journal.

Contributions of each author

Author 1 - Juan Camilo Morales Duran.
Email: juanmorales288981@correo.itm.edu.co
Author 2 - Kevin Osorno Castillo.
Email: kevinosorno@itm.edu.co
Author 3 - Gloria Mercedes Diaz Cabrera.
Email: gloriadiaz@itm.edu.co