UK National Screening Committee

Automated and manual measurements of mammographic breast density in individuals undergoing breast cancer screening

External review against programme appraisal criteria for the UK National Screening Committee

Version: 1

NIHR Evidence Synthesis Group:

Aberdeen Belfast Evidence Collaboration (ABEC)

Date: February 2025

The UK National Screening Committee secretariat is hosted by the Department of Health and Social Care

Contents

About the UK National Screening Committee (UK NSC)

The UK NSC advises ministers and the NHS in the 4 UK countries about all aspects of population screening and supports implementation of screening programmes. Conditions are reviewed against evidence review criteria according to the UK NSC's evidence review process.

Read a complete list of UK NSC recommendations.

UK National Screening Committee, Southside, 39 Victoria Street, London, SW1H 0EU

www.gov.uk/uknsc

Blog: https://nationalscreening.blog.gov.uk/

For queries relating to this document, please contact: uknsc@dhsc.gov.uk.

Published Month 20XX

Plain English summary

Breast cancer is the most common cancer in UK women, with about 55,900 new cases each year, mainly in individuals aged over 50 years. Breast density, or the amount of dense tissue compared to fatty tissue in the breast, is a key risk factor for breast cancer. It is measured using mammograms and categorised into four groups (A, B, C, D) based on how dense the breast tissue is. Traditionally, radiologists assess breast density visually, known as manual measurement, but automated tools like Volpara and Quantra are now being used for more consistent results. However, it is not known whether automated tools produce similar results to manual measurement.

We looked at studies published in the literature comparing automated and manual breast density measurements. The goal was to see how well the results of automated methods match those of manual assessments. Studies were included if they focused on routine breast cancer screening (not for monitoring past cancer) and used standard 2D digital mammography.

We identified 26 studies. Most of these evaluated automated tools like Volpara and Quantra, with others assessing newer or less common software. Key findings include:

- automated and manual methods showed good overall agreement, but there were differences between versions of automated tools and between different technologies
- studies using Volpara suggested it may classify slightly more breasts as "dense" compared to manual methods, but this difference was small and varied across studies
- one study suggested that automated methods might take longer than manual assessment, but this was based on limited data.
- overall, the studies were of moderate quality. Reporting gaps made it hard to fully assess some results

The are some issues worth considering:

- Some studies did not clearly separate participants undergoing routine screening from those being monitored for past cancer, making it hard to confirm that all participants were from the general screening population.

- There was little information about how well the findings apply to minority ethnic groups or underserved populations. Similarly, there was no data about individuals who do not identify as women, which could impact the inclusivity of automated tools.
- Differences in study methods and poor reporting made it difficult to draw firm conclusions or ensure results are relevant to everyone.

Overall, our findings indicate that, in general, automated and manual methods have good agreement, with most existing evidence focusing on the Volpara and Quantra tools. They also highlight the need for more inclusive research, consistent methods and improved reporting to ensure automated breast density tools work well across all populations.

Executive summary

Purpose of the review

To determine the agreement between automated and manual measurement of mammographic breast density.

Background

In the UK, breast cancer is the most common type of cancer among women, accounting for 15% of all new cancer cases. Based on data from 2016-2018 there are around 55,900 new breast cancer cases in the UK annually, corresponding to more than 150 per day. Whilst breast cancer can occur at any age, it most commonly affects women who are over the age of 50 years and have reached menopause. The UK breast cancer screening programme currently screens all women aged 50-70 years at three-year intervals with digital mammography (images of each breast from two views). Screening allows for the early detection of breast cancer which reduces cancer-related burden and mortality. Although breast cancer screening is highly successful in preventing breast cancer mortality (20-40% reduction in risk) death due to breast cancer is still not prevented in a substantial proportion of people due to underdiagnosis.

Breasts contain glandular tissue, fibrous connective tissue, and adipose tissue. Breast density describes the relative amount of these various types of tissue as seen on a mammogram, specifically the proportion of radiologically dense fibro-glandular tissue relative to radiolucent adipose tissue on radiographic imaging. The distribution of the individual amount of fibroglandular tissue, and thus of mammographic densities across the female population, follows a typical Gaussian distribution of many biological features. In clinical practice, there are several ways of classifying breast density, including the commonly used American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) atlas, which categorises breast density into four groups with ‘A’ referring to breasts that are entirely fatty, ‘B’ referring to breasts with scattered areas of fibroglandular density, ‘C’ referring to breasts that are heterogeneously dense, and ‘D’ referring to extremely dense breasts.

In clinical practice, breast density assessment has traditionally been conducted by subjective manual measurement where the radiologist visually inspects

mammograms to categorise the density of the breast. More recently, automated and semi-automated quantitative methods of breast density assessments have been developed to improve the reproducibility of breast density assessment, and they may improve workflow efficiency.

Automated software such as Volpara and Quantra provide density assessments that correspond to the four BI-RADS density categories, and there is evidence that they correlate well with the categories. However, a review by Patterson and colleagues (2019) for the UK National Screening Committee (UK NSC) found that, while the test-retest reliability of automated methods was good, and reliability was better than human readers, there was a paucity of high-quality evidence. The authors found that the concordance between automated methods was variable, and they concluded that automated methods cannot be used interchangeably to measure breast density. Since the concordance between automated and manual methods has yet to be established, this review sought to evaluate the agreement (concordance) between automated and manual methods of measuring mammographic breast density.

Focus of the review

Comprehensive search strategies were developed by an information scientist and the following databases were searched from 2014 onwards: MEDLINE, Embase, Cochrane Database of Systematic Reviews, CENTRAL, Scopus, and Web of Science. There were no restrictions on study type or language. Full-text articles of published studies were eligible for inclusion if they reported the agreement between automated or semi-automated measurement of breast density with manual (visual) measurement of breast density using the BI-RADS breast density scoring system (editions 3, 4 or 5) for 2D digital mammography or the resources required to measure breast density for the two methods, and included at least 60% of participants in their sample who underwent mammography for routine breast cancer screening and had no prior history of breast cancer. Studies of synthetic or spectral mammography were not deemed eligible for inclusion.

Results

The review included 26 articles. Eight studies evaluated Volpara software, two studies evaluated Quantra, two evaluated Volpara and Quantra in the same study,

three studies evaluated Volpara and other automated algorithms (Cumulus Hand Delineation and ImageJ software, EfficientNetB0 deep learning software and AI-CAD Lunit INSIGHT MMG). The remaining 11 studies evaluated various other individual automated software.

Most studies reported substantial agreement between automated and manual measurements of breast density, indicating that there is good concordance between automated and manual measurements of breast density. Nevertheless, there was considerable variation both between automated technologies and within different versions of automated software. Robust conclusions are difficult to draw due to the small number of studies evaluating similar versions of automated software using comparable BI-RADS editions.

The meta-analysis comparing the density classification of the studies that evaluated Volpara software showed a slightly higher categorisation as dense than non-dense from Volpara in comparison to manual classification (risk difference 0.03, 95% confidence interval -0.03, 0.10). There are no dominant or small studies in the meta-analysis with weights between 7.38% and 9.53% but the I^2 statistic value of 97.92% indicates considerable heterogeneity between the studies.

One study indicated that automated density assessment with Volpara may take longer than manual assessment, although the results were derived from a small study that evaluated 250 mammograms. It is, therefore, difficult to draw firm conclusions on any differences for the time requirements of manual and automated density assessment.

Overall, the included studies were of moderate quality even though many items of the ReBIP checklist items were rated as unclear due to insufficient reporting in the full-text publications.

Limitations

While we have made every effort to ensure that the included studies are representative of the UK general breast cancer screening population, including contacting authors to clarify the composition of their study samples, confirming the eligibility of study populations has been challenging. This difficulty arises because the term ‘screening’ is often used interchangeably by study authors to refer to imaging for

breast cancer detection in the general screening population as well as imaging for surveillance to detect recurrent or second primary breast cancer. As a result, we cannot rule out with certainty the possibility that some included studies may have study populations with fewer than 60% general screening participants. It is also possible that some relevant studies have been excluded because we were unable to establish the composition of their populations.

The generalisability of our findings for minority ethnic and underserved groups remains uncertain due to the poor reporting of participants' ethnic and socioeconomic characteristics in the included studies. Moreover, none of the studies reported data on the inclusion of individuals who do not identify as women. This lack of information could be problematic if the automated technologies were trained on datasets that exclude select groups, potentially introducing bias into their application. It is, therefore, unclear whether our findings are truly representative of the broader screening population, highlighting the need for more inclusive research and transparent reporting.

Evidence uncertainties

Overall, there is a paucity of evidence evaluating similar versions of automated software against manual breast density measurement using similar editions of the BI-RADS atlas. This makes it difficult to draw firm conclusions on the concordance of automated software with manual measurement, particularly for newer versions or less common software. There is a need for more consistent research with diverse participant populations to ensure results are representative of the wider screening population.

Introduction and approach

Background

In the UK, breast cancer is the most common type of cancer among women, accounting for 15% of all new cancer cases. Based on data from 2016-2018 there are around 55,900 new breast cancer cases in the UK annually, corresponding to more than 150 per day.[1] Whilst breast cancer can occur at any age, it most commonly affects women who are over the age of 50 years and have reached menopause.

The UK breast cancer screening programme currently screens all women aged 50-70 years at three-year intervals with mammography (images of each breast from two views). Screening allows for the early detection of breast cancer which reduces cancer-related burden and mortality.[2-4] Although breast cancer screening is highly successful in preventing breast cancer mortality (20-40% reduction in risk)[3-5] death due to breast cancer is still not prevented in a substantial proportion of people due to underdiagnosis.[6]

Breasts contain glandular tissue, fibrous connective tissue, and adipose tissue. Breast density describes the relative amount of these various types of tissue as seen on a mammogram, specifically the proportion of radiologically dense fibro-glandular tissue relative to radiolucent adipose tissue on radiographic imaging. The distribution of the individual amount of fibroglandular tissue, and thus of mammographic densities across the female population, follows a typical Gaussian distribution of many biological features.[7] In clinical practice, there are several ways of classifying breast density, including the commonly used American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) atlas,[8] which categorises breast density into four groups,with ‘A’ referring to breasts that are entirely fatty, ‘B’ referring to breasts with scattered areas of fibroglandular density, ‘C’ referring to breasts that are heterogeneously dense, and ‘D’ referring to extremely dense breasts. In clinical practice, breast density assessment has traditionally been conducted by subjective manual measurement where the radiologist visually inspects mammograms to categorise the density of the breast.

In the context of screening, breast density is of concern for two reasons: women with high breast density have an increased risk of breast cancer than those with low

breast density,[8] and the sensitivity of mammography screening is lower in women with more dense breasts.[9] Breast density in women also appears to vary by ethnicity. Whilst breast cancer incidence rates have been traditionally lower in Eastern countries, rates are rapidly rising, potentially attributable to increased obesity, reduced physical activity, and decreased reproduction.[10] Breast cancer is the highest incidence of malignancy in Japanese women. The Japanese Breast Screening Programme includes women in their 40s and has no upper age limit. Although breast cancer mortality rates have been declining in developed Western countries in the early 1990s attributable to screening programmes, the implementation of screening programmes in Japan has not been associated with a reduction in mortality rates. This could be due to underdiagnosis with standard mammography as Japanese women typically have more dense breasts than women in Western countries.[10, 11]

Women with extremely dense breasts (BI-RADS group D) and those with moderately dense breasts (BI-RADS group C) are, therefore, at risk of underdiagnosis. Together these two groups (BI-RADS C and D) may account for almost half of the screening population.[12] Earlier identification of breast cancer through supplemental screening modalities for women with dense breasts would allow for earlier intervention and better clinical outcomes. A risk-adapted screening protocol, wherein women with dense breasts are offered supplemental or enhanced screening modalities, is increasingly being considered.

Aside from standard mammography, several other imaging modalities may be used to detect breast cancer in women with breast density. These include magnetic resonance imaging (MRI), ultrasonography (using either hand-held or automated modalities), contrast-enhanced mammography (CEM), and digital breast tomosynthesis (DBT).

In 2019, the UK National Screening Committee (UK NSC) commissioned a report on whether additional screening with ultrasound after negative standard mammography in women with breast density would be beneficial.[7] The field of breast imaging is a rapidly evolving area, and in the UK, the multicentre Breast Screening - Risk Adapted Imaging for Density (BRAID) study (due for publication in early 2025) sought to determine if supplemental abbreviated MRI, CEM and automated ultrasonography (ABUS) better detected cancer in women with dense breasts.[13]

On this basis, the UK NSC commissioned an updated evidence review to evaluate supplemental imaging modalities for breast cancer screening to detect breast cancer in women with dense breasts to support decision-making about the current UK breast screening programme.

Objectives

Specifically, this review had three objectives:

1. Objective 1: To determine the agreement (concordance) between automated and manual measurement of mammographic breast density
2. Objective 2: To determine the effect of an additional imaging modality to supplement standard mammography compared with standard mammography alone for identifying breast cancer in women with dense breasts.
3. Objective 3: To review evidence on existing economic models assessing the costs and consequences of enhanced mammographic screening for women with breast density.

This document addresses Objective 1 and complies with UK NSC criterion 4 for a population screening programme, which requires a simple, safe precise and validated screening test.

Methods

General

This systematic review was commissioned by the UK NSC and was conducted in accordance with the recommendations of the Cochrane Handbook for Systematic Reviews of Interventions[14] and in adherence with the Preferred Reporting Items for Systematic Reviews (PRISMA) guidelines.[15] The methods were pre-specified in a protocol and registered with the PROSPERO International Prospective Register of Systematic Reviews, available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42024550250.

Patient and public involvement (PPI)

Two PPI partners were part of the study Advisory Group, which also included academic and clinical experts. One PPI partner has lived experience of undergoing mammography for routine breast screening and the other has lived experience of breast cancer. PPI partners participated in regular Advisory Group meetings, where

they contributed to discussions and made recommendations at each stage of the project.

Language and inclusivity statement

Most people who use the UK's breast screening programme identify as women, though not all do. While using exclusively gender-neutral language can enhance inclusivity, it may also reduce clarity. None of the studies included in our review reported data on non-binary participants. We have, therefore, chosen to use both 'women' and gender-neutral language where appropriate. We acknowledge that this is a compromise; however, when we refer to 'women', we ask the readers to interpret this as including all individuals who use the breast screening service, not only those who identify as women.

Role of the funding source

The NIHR Aberdeen-Belfast Evidence Collaboration (ABEC) was funded by the NIHR Evidence Synthesis Programme to conduct this review (project no. NIHR164221). The funder of the study and the UK NSC contributed to the conceptualisation of the research question and study design, but had no role in data collection, data analysis, data interpretation, or writing of the report.

Rationale for the review

Criterion 4 — There should be a simple, safe, precise and validated screening test

Question — Does the automated measurement of mammographic breast density provide similar results to manual measurement in individuals undergoing breast cancer screening?

In clinical practice, breast density assessment has traditionally been conducted by subjective manual measurement where the radiologist visually inspects mammograms to categorise the density of the breast.[6] More recently, automated and semi-automated quantitative methods of breast density assessments have been developed to improve the reproducibility of breast density assessment and may improve workflow efficiency.

Automated software such as Volpara and Quantra provide density assessments that correspond to the four BI-RADS density categories and there is evidence that they correlate well with the categories.[6, 16, 17] However, a review by Patterson and colleagues (2019) for the National Screening Committee found that, while the test-retest reliability of automated methods was good, and reliability was better than human readers, there was a paucity of high-quality evidence.[18] The authors found that the concordance between automated methods was variable, and they concluded that automated methods cannot be used interchangeably to measure breast density. Since the concordance between automated and manual methods has yet to be established, this review sought to evaluate the agreement between automated and manual methods of measuring breast density.

Eligibility for inclusion in the review

Search strategy

Comprehensive search strategies were developed by an information scientist with input from our expert advisors to identify studies of any design that compared manual and automated measurement of breast density. The databases searched were MEDLINE, Embase, Cochrane Database of Systematic Reviews, CENTRAL, Scopus, and Web of Science. There were no restrictions on study type or language at the search stage, but results were limited to 2014 onwards. All references were

exported to Endnote for recording and deduplication. The reference lists of all articles selected for full text appraisal were screened for additional studies. Details of the search strategies are reported in Appendix 1.

Study selection

Full-text articles of published studies were eligible for inclusion if they reported the agreement between automated or semi-automated measurement of breast density with manual (visual) measurement of breast density using the BI-RADS breast density scoring system[8] (editions 3, 4 or 5)[9, 19, 20] for 2D digital mammography or the resources required to measure breast density for the two methods, and included at least 60% of participants in their sample who underwent mammography for routine breast cancer screening and had no prior history of breast cancer. Studies of synthetic or spectral mammography were not deemed eligible for inclusion. The rationale for our eligibility criteria was that 2D digital mammography is the standard imaging modality used in the UK National Breast Screening Programme. We also aimed to include mammograms from participants that are representative of the UK general screening population, rather than those undergoing mammography for diagnostic indications or for surveillance for second primary or recurrent breast cancer. Case-control studies, systematic reviews, editorials, letters and opinion articles were considered ineligible. Conference abstracts were excluded but we attempted to investigate whether fuller information was available from another source. Details of the review eligibility criteria are provided in Table 1.

At the start of the study selection process, two reviewers (CR and SD) independently screened 20% of the titles and abstracts to ensure consistency by comparing their results. The remaining citations were screened by a single reviewer (CR). All potentially relevant articles were retrieved in full and assessed for inclusion by one reviewer (CR), with a second reviewer (MB, DC and SD) checking all articles labelled as unclear (20%). Any disagreements were resolved through discussion between reviewers.

We attempted to contact the corresponding authors of studies where details of the study population were unclear to determine whether the proportions of participants in their samples met our eligibility criteria, provided the study included at least 200 participants (or 200 mammograms if the number of participants was not reported).

Due to time constraints, we excluded studies with unclear population details that involved fewer than 200 participants or 200 mammograms without attempting to contact the study authors.

Table 1: Inclusion criteria for the key question

Key question	Inclusion criteria					
	Population	**Target condition**	**Intervention**	**Comparator**	**Outcome**	**Study type**
Does the automated measurement of mammographic breast density provide similar results to manual measurement in individuals undergoing breast cancer screening?	Individuals between 40 and 70 years of age undergoing breast cancer screening	Breast density	Semi-automated and fully automated methods. Methods of measuring risk from AI technologies applied to mammograms that are not based on breast density were excluded. Semi-automated methods may include Cumulus, ImageJ-based method or DM-scan. Fully automated methods may include Densitas, DM-scan, LIBRA, Quantra, SXA, or Volpara.	Manual (visual) measurement of breast density (% density or BI-RADS classification edition 3, 4 or 5).	Agreement between manual and automated methods for measuring breast density Resources needed to measure breast density (number and experience of health professionals performing the measurement)	Studies of any design published in English in the last 10 years that assessed the agreement between measurements obtained using a semi-automated or fully automated method with those obtained from a manual measurement. Only studies published in full-text articles were considered eligible for inclusion. Conference abstracts were excluded but we investigated whether fuller information was available from another source.

Data extraction and risk of bias assessment

A single reviewer (CR) conducted data extraction using a prespecified data extraction form that was developed with input from the Advisory Group and in accordance with guidance from the PRO-EDI[21] initiative for considering equality, diversity and inclusion of participant characteristics in evidence syntheses. The same reviewer conducted risk of bias assessment using an adapted version of the Review Body for Interventional Procedures () quality assessment tool for non-randomised comparative and case series studies. The ReBIP 18-item checklist was originally developed for NICE and was adapted from several quality assessment checklists and guidance documents, including the NHS Centre for Reviews and Dissemination's guidance, Verhagen and colleagues, Downs and Black and the Generic Appraisal Tool for Epidemiology (GATE).[22-25] The tool assesses bias and generalisability, sample definition and selection, description of the intervention, outcome assessment, adequacy of follow-up, and performance of the analysis. Individual ReBIP question items 1 to 12 were rated as 'yes', 'no' or 'unclear'. A rating of 'yes' denoted the optimal rating for methodological quality. Items 13 to 18 of the checklist were considered unsuitable for the scope of the current review.

Data analysis

Synthesising agreement data across studies proved challenging due to the inconsistency in how agreement was measured and reported. Investigators used a range of statistical methods to assess agreement between manual and automated measurement including the area under the curve, measure of diagnostic accuracy such as sensitivity or specificity, Spearman's rank correlation coefficient, Pearson's correlation coefficient and the kappa statistics. The choice of methods was at the discretion of each study's authors. Among studies reporting the kappa statistics, there was variation in whether the kappa was weighted and, if so, whether linear or quadratic weights were applied. In certain cases, only the kappa value was provided with no accompanying indication of precision. Furthermore, the basis of agreement differed across studies. In some, the agreement was assessed for binary classifications (dense versus non-dense), while in others related to the four density categories.

Where studies reported the proportions of participants classified into breast density categories by both automated and manual measurements, we analysed these proportions for studies using similar automated software to determine the ability of the software to consistently classify participants as having dense or non-dense breasts compared with manual measurement. For studies reporting numerical data for both automated and manual measurements, we conducted a meta-analysis to compare the proportions classified as dense and non-dense. This analysis was possible for studies comparing Volpara to manual measurement. A random-effects meta-analysis was performed to compare the dense proportions. Among the Volpara studies, there were four multi-arm studies. To avoid potential bias, we did not split the Volpara group when there were multiple control groups, nor did we combine control groups when these represented different manual measurements of the same participants. Therefore, for the main meta-analysis we selected the more recent version of Volpara when two were available (e.g., Gemici 2020);[26] breast imaging experts in Eom (2018),[27] randomly selected observer 1 in Singh (2016),[28] and the most experienced radiologist from Rigaud (2022).[29] Sensitivity analyses were also conducted using the excluded groups from the Gemici (2020),[26] Eom (2018),[27] and Singh (2016) studies.[28]

Description of the evidence

The literature searches identified 1032 citations, and 215 full-text reports were selected for eligibility assessment. Of these, 5 reports were unavailable, of the remaining 184 were excluded. We attempted to contact the corresponding authors of 28 reports where it was unclear whether the reports included eligible populations (n=25),[30-54] whether they included eligible mammography (n=2)[55, 56] or whether they were a secondary report of the same participant sample of another included study (n=1).[57] The email addresses of five authors of 6 reports were invalid. We were unable to identify alternative email addresses for these authors. We were also unable to locate an email address for 1 further study where the email address of the corresponding author was not provided in the written report.[50] Of the remaining 21 reports, we received replies from 5 authors, and we were subsequently able to include t studies.[33, 56] We did not include the 19 studies where we received no reply from the study authors.[30, 31, 34, 35, 38-47, 51, 53, 55-57]

We also excluded 11 studies[58-68] that used mammograms from databases of digitised film mammography, specifically the Digital Database of Screening Mammography (DDSM)[69] and the Mammographic Image Analysis Society (MIAS) databases.[70]

In total, 26 reports were included in this systematic review. Details of the screening process and the lists of included and excluded studies with the main reasons for exclusion are presented in Appendix 2.

Characteristics of the included studies

The characteristics of the included studies are detailed in Table 4, Appendix 3. The included studies were conducted in Europe (n=9: Sweden [n=5],[56, 71-74] the Netherlands [n=2],[75, 76] France [n=1],[77] Norway [n=1][78]); the USA (n=4);[29, 33, 79, 80] the Republic of Korea (n=4);[17, 27, 81, 82] Peru (n=2);[83, 84] and 1 study each was conducted in Argentina;[85] Australia;[86] Brazil;[87] India;[28] Saudia Arabia;[88] Turkey;[26] and the UK and USA.[89]

Eight studies evaluated Volpara software,[17, 26-28, 73, 75, 80, 88] two studies evaluated Quantra,[78, 86] two evaluated Volpara and Quantra in the same study,[76, 81] three

studies evaluated Volpara and other automated algorithms (Cumulus Hand Delineation and ImageJ software,[89] EfficientNetB0 deep learning software[29] and AI-CAD Lunit INSIGHT MMG).[82] The remaining 11 studies[33, 56, 71, 72, 74, 77, 79, 83-85, 87] evaluated various other individual automated software. The studies varied in reporting their units of analysis by the number of participants (total across studies, n=29,784) and the number of mammograms (total across studies, n=16,194). The youngest reported mean age of the participants was 48.8 years[28] and the oldest was 58.8 years.[75] Only 1 study reported the ethnicity of the participants.[82] None of the studies reported details of the socioeconomic status or transgender characteristics of the participants. Five studies[56, 71-74] obtained mammograms from the Mälmo Breast Tomosynthesis Screening Trial (MBTST).[90] It is unclear whether any participant overlap exists between these studies; however, each study evaluates different automated software. Therefore, we believe these studies are not problematic in terms of double counting participants. Similarly, while it is unclear whether there is participant overlap between two studies conducted in Peru, both studies also evaluated different automated software, reducing concerns about duplication. Four studies explicitly reported evaluating raw (for processing/pre-processed) images;[71-73, 83] however, most studies did not specify whether they evaluated raw or processed (for presentation) images.

Quality assessment

Overall, the included studies were of moderate quality even though many items of the ReBIP checklist items were rated as unclear due to insufficient reporting in the full-text publications. All studies clearly defined the intervention and comparison, included valid and reliable assessment of important outcomes and included radiologists with breast imaging experience. Half (50%) of the studies were judged to be representative of the target population, with the remainder judged as unclear, and the comparison groups of all studies were judged to be similar in terms of their demographic and clinical features. Eleven (42.3%) studies were deemed to have conducted blinded outcome assessments. The authors of 11 studies either developed the automated software or had associations with the manufacturers of the automated software.[56, 71-76, 79, 83, 84, 87] The results of the study-level quality assessment are provided in Table 5, Appendix 3.

Overall findings

Full-length results tables are provided in Tables 6 to 8 in Appendix 4.

In some studies, more than one agreement statistic was used, and across all of the related studies, there are three general types of agreement reported. These are kappa statistics, correlation coefficients and percentage agreement. Nearly 70% of the reported methods are kappa with over 80% either linear or quadratic weighted Cohen's kappa. All studies included participants with a full range of breast density categories, so low prevalence in either dense or non-dense classifications was not an issue. Even in studies using a four-level density agreement - where categories A and D had fewer participants - there were sufficient participants in categories B and C to avoid the need for a prevalence-adjusted Kappa. We are, therefore, happy that the studies are using appropriate methods to measure agreement for ordinal rating categories between different reviewers. Where studies had more than two reviewers, Fleis Kappa was used to account for multiple raters.

Volpara software versus manual measurement

The results of the 13 studies[17, 26-29, 73, 75, 76, 80-82, 88, 89] that compared Volpara software with manual density measurement using BI-RADs are summarised in Table 2. The studies evaluated 15 versions of Volpara (Version 1.4.2 [n=1];[26] Version 1.4.5 [n=1];[28] Version 1.5.0 [n=3];[75, 76, 89] Version 1.5.1 [n=3];[17, 26, 80] Version 1.5.11 [n=1];[73] Version 1.5.12 [n=1];[27] Version 1.5.2 [n=1];[80] Version 1.5.5.1 [n=1];[88] Version 3.1 [n=1];[81] Version 3.4.1 [n=2])[29, 82] compared against BI-RADS 4th (n=7 studies)[17, 26, 28, 73, 75, 80, 89] and BI-RADS 5th (n=5 studies)[76, 80-82, 88] editions. Two studies did not report the BI-RADS edition.[27, 29] Portnow and colleagues (2022) evaluated BI-RADS 4th and 5th editions in their study.[80] The different versions of Volpara and BI-RADS editions were considered as suitably similar for combining in our meta-analysis comparing the proportions of density categories between Volpara and manual measurement. The studies evaluated mammograms obtained using Hologic (n=5);[26, 29, 75, 76, 81] GE Healthcare (n=4),[17, 27, 80, 82] Siemens AG (n=1)[73] and Phillips (n=1)[28] systems. Two studies did not report the mammography system.[88, 89]

Discussion of Volpara findings

Concordance ranged from Kappa -0.40 to 0.83. Most studies (53.8%)[17, 27, 28, 75, 76, 80, 89] showed substantial agreement between Volpara and manual measurement with BI-RADS, both for categorising mammograms into the four density and non-dense/dense categories, although one of these studies by Alomaim and colleagues (2020) showed only moderate agreement for mammograms that did not contain image distractors.[89] The study found that there was a statistically significant difference in the average time required to make density assessments by manual and automated methods, with manual assessment taking an average of 14 seconds (range: 7 to 33 seconds) and Volpara taking approximately 57 seconds (range not reported) per patient ($p < 0.001$). This study also reported that the average time for the manual subjective categorisation for mammograms with image distractors was 12 seconds (range: 7–23 seconds), and 16 seconds (range: 9–33 seconds) for those without any distractors, however, the time difference was not significant ($p > 0.05$). The study by Eom and colleagues (2018) found almost perfect agreement between Volpara version 1.5.12 and visual measurement by expert radiologists for classifying mammograms into dense and non-dense categories, although this reduced to substantial agreement for measuring the four density categories and the agreement between Volpara and general radiologists was substantial for both the 4-way and 2-way density classifications in this study.[27] Four studies (30.8%) showed moderate agreement between Volpara and BI-RADS although, of these, the study by Aloufi and colleagues (2022)[88] found only fair agreement for categorising mammograms into the four density categories compared with moderate agreement for the dense/non-dense categories. One study by Riguad and colleagues (2022)[29] showed only fair agreement and the study by Gemici and colleagues (2020)[26] showed poor agreement. The version of Volpara software, BI-RADS edition and the type of mammography system used in the studies were not consistently associated with the strength of agreement between the automated and visual density measurements.

The meta-analysis comparing the density categorisation of Volpara and manual measurement is shown in Figure 1. The meta-analysis shows a slightly higher categorisation as dense than non-dense from Volpara in comparison to manual classification. In Gemici (2020),[26] two versions of Volpara were used and the most

recent used in the meta-analysis below. In Eom (2018),[27] and Singh (2016)[28] there were two control groups and so breast imaging experts were the chosen control group for Eom (2018) and observer 1 was used as the control group for Singh (2016). The sensitivity analyses where the alternative groups from these three studies are used are all consistent with overall differences of 0.03 (95% CI-0.04, 0.10), 0.03 (95% CI -0.03, 0.09), and 0.04 (95% CI -0.03, 0.11) in comparison to the overall difference in the forest plot of 0.04 (-0.03, 0.11). There are no dominant or small studies in the meta-analysis with weights between 7.38% and 9.53% but the I^2 statistic value of 97.92% indicates considerable heterogeneity between the studies.

Figure 1 Meta-analysis of Volpara versus manual measurement of mammographic breast density

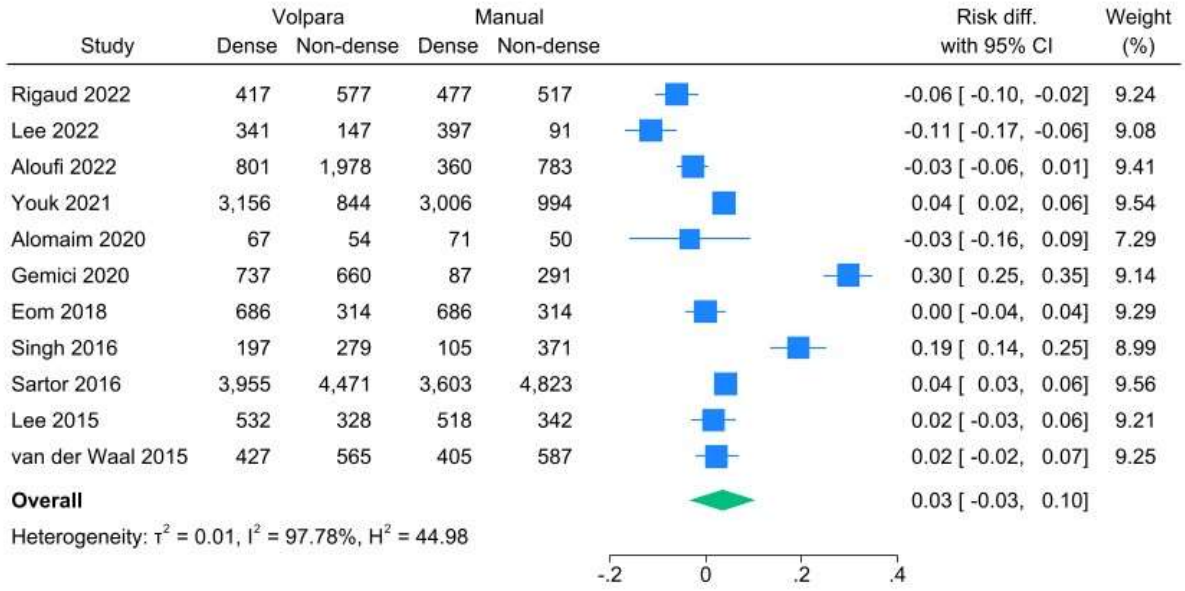


Table 2 Summary results of the studies evaluating Volpara versus manual measurement of mammographic breast density

Study ID	Sample size used in the analysis	Type of mammography	Automated software version and BI-RADS edition	Agreement value (95% CI)	Agreement interpretation
Gemici 2020[91]	379 mammograms	FFDM Selenia, Hologic	Volpara version 1.4.2 (n=1399 mammograms) Volpara version 1.5.1 (n=1399 mammograms) BI-RADS 4th edition (n=379 mammograms) Two radiologists with 5 and 8 years, experience	V 1.4.2 vs BI-RADS 4: κ -0.41 (NR) V1.5.1 vs BI-RADS 4: κ -0.40 (NR)	Poor agreement Poor agreement
Holland 2016[75]	500 participants; 1000 mammograms	Digital mammography Lorad Selenia, Hologic	Volpara version 1.5.0 BI-RADS 4th edition Three radiologists (R1, R2 and R3) with ≥8 years of experience in breast imaging and a PhD student (R4) with a medical degree and 2 years'experience	*4-way density* Experienced radiologists: κ 0.73 to 0.78; PhD student: κ 0.77 *Non-dense vs. dense* Experienced radiologists: κ 0.63 to 0.70; PhD student: κ 0.71	Substantial agreement Substantial agreement
Rigaud 2022[29]	995 participants	FFDM Selenia Dimensions, Hologic	Volpara version 3.4.1 BI-RADS (edition NR) 7 radiologists with 5-22 years' experience	κ 0.34 (range 0.17 to 0.48); Average percent agreement: 56% (range 43 to 64%)	Fair agreement
	992 mammograms	FFDM	Volpara version 1.5.0	κ 0.80 (0.77, 0.82) *Proportion agreement* 65.4%	Substantial agreement

van der Waal 2015[76]		Selenia system, Hologic	BI-RADS 5th edition 3 experienced radiologists	*Prediction of dense category* AUC: 0.948 (0.935, 0.960) *Accuracy at 8.0% cut-off* Sensitivity: 84%; Specificity 91%	
Youk 2021[81]	4000 participants	FFDM Lorad Selenia, Hologic	Volpara version 3.1 BI-RADS 5th edition Three radiologists with 7, 10 and 14 years' experience	κ 0.48 (0.46, 0.50)	Moderate agreement
Eom 2018[27]	1000 participants	FFDM Senographe DS, GE Healthcare	Volpara version 1.5.12 BI-RADS (edition NR) Two breast-imaging experts with >5 years'experience BI-RADS (edition NR) Two general radiologists with <5 years of experience	Expert radiologists: κ 0.77 (0.75, 0.80); General radiologists: κ 0.71 (0.68, 0.74) *Non-dense vs. dense* Expert radiologists: κ 0.83 (0.80, 0.87) General radiologists: κ 0.73 (0.68, 0.77)	Substantial agreement Almost perfect agreement Substantial agreement
Lee 2015[17]	860 participants	FFDM Senographe DS; GE Healthcare	Volpara version 1.5.1 BI-RADS 4th edition One radiologist with 6 years' experience	κ 0.80 (0.77, 0.83) ρ 0.86 (NR) P<0.0001	Substantial agreement
Lee 2022[82]	488 participants 488 mammograms	FFDM Senographe Pristina, GE Healthcare	Volpara version 3.4.1 BI-RADS 5th edition Three radiologists with 2-, 10- and 25-years' experience	*4-way density:* κ 0.50 (0.45, 0.56) *Non-dense vs. dense:* κ 0.56 (0.48, 0.65)	Moderate agreement Moderate agreement
Portnow 2022[80]	200 mammograms	FFDM Senograph ES and Senograph	Volpara version 1.5.1 BI-RADS 4th edition	*Non-dense vs dense* V 1.5.1 vs. BI-RADS 4: κ 0.68 to 0.83	Substantial agreement

		DS, GE Healthcare	Six radiologists with 23-30 years' experience Volpara version 1.5.2 BI-RADS 5th edition Six radiologists with 23-30 years' experience	V 1.5.2 vs. BI-RADS 5: κ 0.76 to 0.85	Substantial agreement
Sartor 2016[73]	8426 participants; 8426 mammograms	Digital mammography, Mammomat Inspiration, Siemens AG	Volpara version 1.5.11 BI-RADS 4th edition five breast radiologists with >10 years' experience	κ 0.55 (0.53, 0.56)	Moderate agreement
Singh 2016[28]	476 participants	FFDM MicroDose SI, Philips	Volpara version 1.4.5 BI-RADS 4th edition Observer 1 (1/2 radiologists with 5–10 years' experience) BI-RADS 4th edition Observer 2 (1/2 radiologists with 5–10 years' experience)	Observer 1: ρ 0.73 Observer 2: ρ 0.73	Strong agreement Strong agreement
Alomaim 2020[89]	With distractors: 92 mammograms Without distractors: 158 mammograms	Digital mammography (machine NR)	Volpara version 1.5.0 (n=122 mammograms) BI-RADS 4th edition (n=122 mammograms) 25 USA and 24 UK radiologists with >8 years' experience (UK 12% ≤1 years' experience)	κ 0.66, p<0.001 *With distractors*: κ 0.67 (NR) p<0.001 *Without distractors*: κ 0.52 (NR) p<0.001	Substantial agreement Substantial agreement Moderate agreement

Aloufi 2022[88]	1022 participants	NR FFDM was used in the SNBCSP during the time mammography was conducted during the study (from 2012 to 2018)[92]	Volpara version 1.5.5.1 BI-RADS 5th edition 11 radiologists (experience NR)	*4-way density* κ (0.35 (0.29, 0.39) *Non-dense vs.dense* κ 0.53 (0.47, 0.60)	Fair agreement Moderate agreement

Aktiengesellschaft (AG); AUC, area under the curve; FFDM, full field digital mammography; GE, General Electric, κ, Kappa statistic; NR, not reported; ρ; Spearman's rank correlation coefficient; SE, standard error; SNBCSP, Saudi National Breast Cancer Screening Programme; V, Volpara; vs, versus

Quantra software versus manual measurement

The results of the four studies that compared Quantra software with manual measurement are summarised in Table 3. All studies used Selenia, Hologic mammography systems and evaluated Quantra versions 1.3 (n=1),[76] 2.0 (n=2)[78, 86] and 2.1.1 (n=1).[81]

Discussion of the Quantra findings

Concordance ranged from Kappa 0.54 to 0.84. Two studies by van der Waal and colleagues (2015)[76] and Ekpo and colleagues (2016)[86] showed excellent or almost perfect agreement between Quantra and BI-RADS density measurements, although this reduced to substantial agreement for the classification of the four density categories in the latter study. The study conducted by Osteras and colleagues (2016)[78] showed substantial agreement, while the study by Youk and colleagues (2021)[81] showed moderate agreement between Quantra and BI-RADS measurements. As with the evaluations of Volpara software, the different Quantra versions and BI-RADS editions were not consistently associated with the strength of agreement between the different methods of density measurement. We were unable to analyse the proportions of mammograms classified by the different density categories because data for both Quantra and BI-RADs measurements were only available for the study conducted by Youk and colleagues (2021).[81]

Table 3 summary results of the studies evaluating Quantra versus manual measurement of mammographic breast density

Study ID	Sample size used in the analysis	Type of mammography	Automated software version and BI-RADS edition	Agreement value (95% CI)	Agreement interpretation
Ekpo 2016[86]	292 participants (majority report)	NR Selenia Dimensions, Hologic	Quantra version 2.0 BI-RADS 4th edition 3 RANZCR-certified breast radiologists (majority report – consensus of 2/3)	Majority report *4-way density* κ 0.79 (0.75, 0.84) *Non-dense vs.dense* κ 0.84 (0.79, 0.87) Sensitivity: 91.3%; Specificity 83.6% AUC: 0.89 (0.82, 0.91)	Substantial agreement Almost perfect agreement
Osteras 2016[78]	537 mammograms	FFDM Selenia Dimensions, Hologic	Quantra version 2.0 BI-RADS 4th edition five radiologists with 1 to 34 years' experience	κ 0.73 (0.67, 0.79)	Substantial agreement
van der Waal 2015[76]	992 mammograms	FFDM Selenia system, Hologic	Quantra version 1.3 BI-RADS 5th edition 3 experienced radiologists	*Prediction of dense category* AUC: 0.98 (0.94, 0.96) Accuracy at 13.8% cut-off: Sensitivity: 82%; Specificity: 92%)	Excellent agreement
Youk 2021[81]	4000 participants	FFDM Lorad Selenia, Hologic	Quantra version 2.1.1 BI-RADS 5th edition Three radiologists with 7-, 10- and 14-years' experience	κ 0.54 (0.52, 0.56)	Moderate agreement

AUC, area under the curve; FFDM, full field digital mammography; κ, Kappa statistic; NR, not reported; ρ; Spearman's rank correlation coefficient; RANZCR, Royal Australian and New Zealand College of Radiology; SE, standard error; vs, versus

‘Other’ automated software versus manual measurement

Fourteen studies compared 16 individual automated software with manual BI-RADS density measurement (BI-RADS 3rd edition, n=1;[84] BI-RADS 4th edition n=4;[72, 79, 87, 89] BI-RADS 5th edition n=7;[33, 56, 71, 74, 77, 82, 85] BI-RADS edition was not reported n=2).[29, 83] Details of the results of the automated software are presented in Table 8, Appendix 4.

Discussion of the ‘other’ automated software findings

Concordance ranged from Kappa 0.38 to 0.86. Two automated algorithms evaluated by Alomaim and colleagues (2020)[89] and Lehman and colleagues (2019)[33] showed almost perfect agreement with manual measurements using BI-RADS 4th and 5th editions, respectively. The study by Alomaim also found that automated assessment of breast density took significantly longer on average per patient than manual assessment: 14 seconds (range: 7–33 seconds) for manual assessment compared with 320 seconds (range: 198–453 seconds) or Cumulus hand delineation (HD) and 131 seconds (range: 83–168 seconds) for Image J software ($p < 0.001$).[89] Eight (50%) studies showed substantial agreement with manual measurement. Of these, the study by Le Boulc’h and colleagues (2020)[77] demonstrated substantial agreement for DenSeeMammo software compared with both senior and junior radiologists’ density measurements using BI-RADS 5th edition. Six (37.5%) studies showed moderate agreement, although the study by Lee and colleagues (2022)[82] showed only fair agreement for the dense/non-dense classification.

We did not attempt to pool the proportions data for the manual and automated density classifications due to the heterogeneity among the different automated software.

Discussion of overall findings

This evidence synthesis included 26 reports that evaluated the concordance between automated and manual measurement of breast density for 2D digital mammography. To our knowledge, our study is the most comprehensive and up-to-date overview of the evidence for the agreement between automated and manual measurement of breast density in the general screening population.

Our findings show that, overall, there is good concordance between automated and manual measurement of breast density. Nevertheless, there is considerable variation both between automated technologies and within different versions of automated software. Robust conclusions are difficult to draw due to the small number of studies evaluating similar versions of automated software using comparable BI-RADS editions. The largest body of evidence for one type of automated software came from studies evaluating Volpara. This is unsurprising given the widespread use of Volpara in clinical practice. Our meta-analysis indicates that Volpara is more likely to categorise mammograms as dense compared with manual measurement, although the difference was not statistically significant.

Most studies reported substantial agreement between automated and manual measurements of breast density, suggesting that automated software holds promise for its use in clinical practice. However, firm conclusions cannot be made due to the heterogeneous nature of the data.

Three studies examined the impact of radiologists' experience levels on agreement.[27, 75, 77] Their findings indicate that the agreement between manual and automated density measurements is consistent regardless of whether the measurements were performed by senior/experienced radiologists or junior/general radiologists. This suggests that radiologists' experience does not influence the level of agreement between manual and automated density measurements.

One study indicated that agreement between Volpara and manual density assessment is greater for mammograms that contain image distractors.[89] The authors note that this finding was unexpected and indicates the need for further

studies to explore the impact of the image quality of mammograms on automated and manual breast density measurement. The same study found that automated density assessment with Volpara may take longer than manual assessment, although the results were derived from a small study that evaluated 250 mammograms and it is, therefore, difficult to draw firm conclusions on any differences for the time requirements of manual and automated density assessment. Nevertheless, the impact on workflow in clinical practice should be considered as part of any planned integration of automated technology in radiological services.

A 2024 multi-society paper,[93] presenting the views of Radiology Societies in the USA, Canada, Europe, Australia, and New Zealand, notes that "*exposing radiologists to an increasing number of complex interfaces is undesirable, and is liable to diminish utility and acceptance of AI tools*."[94] The authors of this paper highlight the tendency of humans to favour AI-generated decisions over those made by humans, a phenomenon known as 'automation bias'. Automation bias can lead to errors if the AI system is incorrect, a risk that may increase when radiologists are fatigued or when there is limited capacity to supervise or validate the AI output. The authors also emphasise that this issue has broader implications, particularly in scenarios where autonomous AI systems continue to learn and adapt over time. Such automated systems require ongoing monitoring to ensure their performance remains satisfactory. This might also apply to situations where new versions of automated software are released. The implementation of automated technology in radiological clinical practice would need to consider any associated training, strategic, regulatory, performance, technical, or economic considerations for the introduction of automated technology in radiological clinical practice.

While we have made every effort to ensure that the included studies are representative of the UK general breast cancer screening population, including contacting authors to clarify the composition of their study samples, confirming the eligibility of study populations has been challenging. This difficulty arises because the term 'screening' is often used interchangeably by study authors to refer to imaging for breast cancer detection in the general screening population as well as imaging for surveillance to detect recurrent or second primary breast cancer. As a result, we cannot rule out with certainty the possibility that some included studies

may have study populations with fewer than 60% general screening participants. It is also possible that some relevant studies have been excluded because we were unable to establish the composition of their populations.

Breast density appears to vary by ethnicity with Asian women, for example, typically having denser breasts than non-Asian women.[95-98] However, the generalisability of our findings for minority ethnic and underserved groups remains uncertain due to the poor reporting of participants' ethnic and socioeconomic characteristics in the included studies. Moreover, none of the studies reported data on the inclusion of individuals who do not identify as women. This lack of information could be problematic if the automated technologies were trained on datasets that exclude select groups, potentially introducing bias into their application.[93] It is, therefore, unclear whether our findings are truly representative of the broader screening population, highlighting the need for more inclusive research and transparent reporting.

Appendix 1 — Search strategy

Electronic databases

Search Terms

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions <1946 to March 29, 2024>

1 breast density/ 1628
2 ((breast or mammograph*) adj5 dens*).tw,kf. 6047
3 1 or 2 6182
4 (automat* or "semi-automat*" or "computer-assist*" or objective or quantitative).tw,kf. 3271110
5 (Cumulus or ImageJ or "DM-scan" or Densitas or LIBRA or Quantra or SXA or Volpara).tw,kf. 12130
6 Radiographic Image Interpretation, Computer-Assisted/ 16126
7 4 or 5 or 6 3289578
8 (human or manual or visual or radiologist? or reader? or subjective).tw,kf. 3861612
9 (concordance or agreement or compar* or correlat* or kappa or "Bland-Altman" or "reference standard" or "gold standard" or sensitivity or specificity or accuracy or AUC or ("area under" adj5 curve)).tw,kf. 9736279
10 3 and 7 and 8 and 9 502
11 limit 10 to yr="2014 -Current" 321

Embase <1974 to 2024 Week 13>

1 breast density/ 4090
2 ((breast or mammograph*) adj5 dens*).tw,kf. 8523
3 1 or 2 9463
4 (automat* or "semi-automat*" or "computer-assist*" or objective or quantitative).tw,kf. 4801827
5 (Cumulus or ImageJ or "DM-scan" or Densitas or LIBRA or Quantra or SXA or Volpara).tw,kf. 20331
6 exp *computer assisted diagnosis/ 334358
7 4 or 5 or 6 5070426
8 (human or manual or visual or radiologist? or reader? or subjective).tw,kf. 4788455
9 (concordance or agreement or compar* or correlat* or kappa or "Bland-Altman" or "reference standard" or "gold standard" or sensitivity or specificity or accuracy or AUC or ("area under" adj5 curve)).tw,kf. 12763430
10 3 and 7 and 8 and 9 820
11 limit 10 to yr="2014 -Current" 566
12 conference abstract.pt. 5093636
13 11 not 12 471

Cochrane Library

ID Search Hits
#1 MeSH descriptor: [Breast Density] this term only 80
#2 (breast or mammograph*) near/5 dens* 1116
#3 #1 or #2 1116
#4 automat* or (semi next automat*) or (computer next assist*) or objective or quantitative 416441
#5 Cumulus or ImageJ or "DM-scan" or Densitas or LIBRA or Quantra or SXA or Volpara 829
#6 MeSH descriptor: [Radiographic Image Interpretation, Computer-Assisted] this term only 601
#7 #4 or #5 or #6 417005
#8 human or manual or visual or radiologist? or reader? or subjective 920312
#9 concordance or agreement or compar* or correlat* or kappa or "Bland-Altman" or "reference standard" or "gold standard" or sensitivity or specificity or accuracy or AUC or ("area under" adj5 curve) 1104385
#10 #3 and #7 and #8 and #9 152
#11 2014-24 CDSR: 6; CENTRAL 81

Web of Science: SCI

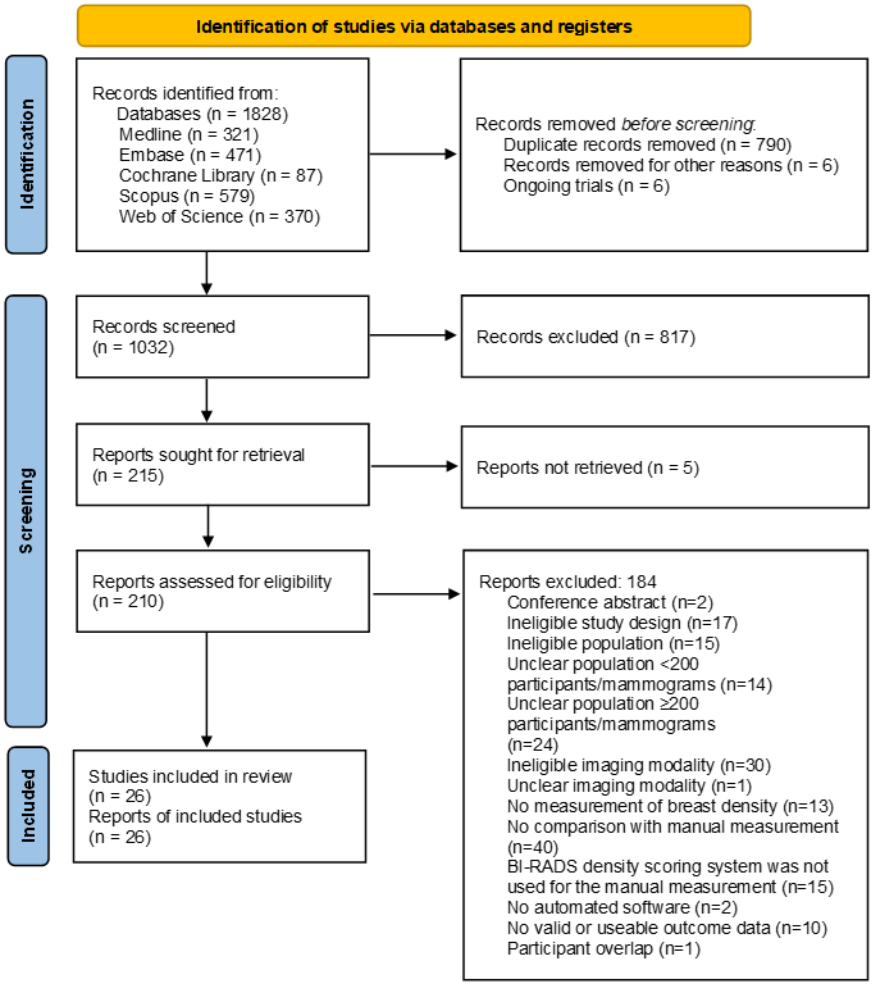
1: (breast or mammograph*) near/5 dens* (Topic) Results: 7981
2: automat* or "semi-automat*" or "computer-assist*" or objective or quantitative (Topic) 4460752
3: Cumulus or ImageJ or "DM-scan" or Densitas or LIBRA or Quantra or SXA or Volpara (Topic) 21981
4: #2 OR #34478810
5: human or manual or visual or radiologist$ or reader$ or subjective (Topic) 4993190
6: concordance or agreement or compar* or correlat* or kappa or "Bland-Altman" or "reference standard" or "gold standard" or sensitivity or specificity or accuracy or AUC or ("area under" near/5 curve) (Topic) 14542525
7: #1 AND #4 AND #5 AND#6 572
8: #1 AND #4 AND #5 AND #6 and 2024 or 2023 or 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 (Publication Years) 370

Scopus

(TITLE-ABS-KEY ((breast OR mammograph*) W/5 dens* W/5 (measur* OR assess* OR stratif* OR classif*))) AND (TITLE-ABS-KEY (automat* OR "semi-automat*" OR "computer-assist*" OR objective OR quantitative OR cumulus OR imagej OR "dm-scan" OR densitas OR libra OR quantra OR sxa OR volpara)) AND (TITLE-ABS-KEY (human OR manual OR visual OR radiologist$ OR reader$ OR subjective)) AND (TITLE-ABS-KEY (concordance OR agreement OR compar* OR correlat* OR kappa OR "bland-altman" OR "reference standard" OR "gold standard" OR sensitivity OR specificity OR accuracy OR auc OR ("area under" W/5 curve))) AND PUBYEAR > 2013 AND PUBYEAR < 2025

Appendix 2 — Included and excluded studies

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only



Publications included after review of full-text articles

Alomaim W, O'Leary D, Ryan J, Rainford L, Evanoff M, Foley S. Subjective versus quantitative methods of assessing breast density. Diagnostics. 2020;**10**(5):331.

Aloufi AS, AlNaeem A, Almousa A, et al., Breast density distribution among the Saudi screening population and correlation between radiologist visual assessment and two automated methods. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2022.

Angulo A, Ferrer J, Pinto J, Lavarello R, Guerrero J, Castaneda B, Experimental assessment of an automatic breast density classification algorithm based on principal component analysis applied to histogram data. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2015.

Ekpo EU, McEntee MF, Rickard M, et al. Quantra™ should be considered a tool for two-grade scale mammographic breast density classification. Br J Radiol. 2016;**89**(1060).

Eom HJ, Cha JH, Kang JW, Choi WJ, Kim HJ, Go E. Comparison of variability in breast density assessment by BI-RADS category according to the level of experience. Acta Radiol. 2018;**59**(5):527-32.

Fieselmann A, Fornvik D, Fornvik H, et al. Volumetric breast density measurement for personalized screening: Accuracy, reproducibility, consistency, and agreement with visual assessment. J Med Imaging (Bellingham). 2019;**6**(3):031406.

Fonseca P, Mendoza J, Wainer J, et al., Automatic breast density classification using a convolutional neural network architecture search procedure. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2015.

Fornvik D, Fornvik H, Fieselmann A, Lang K, Sartor H. Comparison between software volumetric breast density estimates in breast tomosynthesis and digital mammography images in a large public screening cohort. Eur Radiol. 2019;**29**(1):330-6.

Gemici AA, Aribal E, Ozaydin AN, et al. Comparison of qualitative and volumetric assessments of breast density and analyses of breast compression parameters and breast volume of women in bahcesehir mammography screening project. Meme Sagligi Dergisi / Journal of Breast Health. 2020;**16**(2):110-6.

Holland K, van Zelst J, den Heeten GJ, et al. Consistency of breast density categories in serial screening mammograms: A comparison between automated and human assessment. Breast. 2016;**29**:49-54.

Kaiser N, Fieselmann A, Vesal S, et al., Mammographic breast density classification using a deep neural network: Assessment based on inter-observer variability. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2019.

Le Boulc'h M, Bekhouche A, Kermarrec E, et al. Comparison of breast density assessment between human eye and automated software on digital and synthetic mammography: Impact on breast cancer risk. Diagn Interv Imaging. 2020;**101**(12):811-9.

Lee HN, Sohn YM, Han KH. Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: Analysis of clinical-radiologic factors affecting discrepancy between them. Acta Radiol. 2015;**56**(9):1061-8.

Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. Med Phys. 2018;**45**(3):1178-90.

Lee SE, Son NH, Kim MH, Kim EK. Mammographic Density Assessment by Artificial Intelligence-Based Computer-Assisted Diagnosis: A Comparison with Automated Volumetric Assessment. J Digit Imaging. 2022;**35**(2):173-9.

Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: clinical implementation. Radiology. 2019;**290**(1):52-8.

Osteras BH, Martinsen ACT, Brandal SHB, et al. Classification of fatty and dense breast parenchyma: Comparison of automatic volumetric density measurement and radiologists' classification and their inter-observer variation. Acta Radiol. 2016;**57**(10):1178-85.

Pavan ALM, Vacavant A, Trindade AP, de Pina DR. Fibroglandular Tissue Quantification in Mammography by Optimized Fuzzy C-Means with Variable Compactness. IRBM. 2017;**38**(4):228-33.

Pesce K, Tajerian M, Chico MJ, et al. Interobserver and intraobserver variability in determining breast density according to the fifth edition of the BI-RADS Atlas. Radiologia (Roma). 2020;**62**(6):481-6.

Portnow LH, Georgian-Smith D, Haider I, et al. Persistent inter-observer variability of breast density assessment using BI-RADS 5th edition guidelines. Clin Imaging. 2022;**83**:21-7.

Rigaud B, Weaver OO, Dennison JB, et al. Deep Learning Models for Automated Assessment of Breast Density Using Multiple Mammographic Image Types. Cancers (Basel). 2022;**14**(20).

Sartor H, Lang K, Rosso A, Borgquist S, Zackrisson S, Timberg P. Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification. Eur Radiol. 2016;**26**(12):4354-60.

Singh T, Sharma M, Singla V, Khandelwal N. Breast Density Estimation with Fully Automated Volumetric Method: Comparison to Radiologists' Assessment by BI-RADS Categories. Acad Radiol. 2016;**23**(1):78-83.

Timberg P, Fieselmann A, Dustler M, et al., Breast density assessment using breast tomosynthesis images. Lecture Notes in Computer Science; 2016.

Van Der Waal D, Den Heeten GJ, Pijnappel RM, et al. Comparing visually assessed BI-RADS breast density and automated volumetric breast density software: A cross-sectional study in a breast cancer screening setting. PLoS One. 2015;**10**(9):e0136667.

Youk JH, Gweon HM, Son EJ, Eun NL, Kim JA. Fully automated measurements of volumetric breast density adapted for BIRADS 5th edition: a comparison with visual assessment. Acta Radiol. 2021;**62**(9):1148-54.

Publications excluded after review of full text articles

R1 unavailable studies (n=5)

Cai S, Tang JN, Liu PZ, et al. Breast density classification based on wavelet transform. Journal of Medical Imaging and Health Informatics. 2018;8(6):1157-63.

Kuzmiak CM, Ko EY, Tuttle LA, Steed D, Zeng D, Yoon SC. Whole Breast Ultrasound: Comparison of the Visibility of Suspicious Lesions with Automated Breast Volumetric Scanning Versus Hand-Held Breast Ultrasound. Acad Radiol. 2015;22(7):870-9.

Pavusa R, Fernandes JCL, Da Silva AP, et al. Automated methodology for breast segmentation and mammographic density classification using co-occurrence and statistical and SURF descriptors. Int J Biomed Eng Technol. 2022;39(3):213-48.

Virmani J. Breast density classification using Laws' mask texture features. Int J Biomed Eng Technol. 2015;19(3):279-302.

Zhou J, Wang T, Li M, Zhao J, Shuang P, Sheng F. Evaluation of mammography calcification detection system based on deep learning. Chinese Journal of Radiology (China). 2019;53(11):968-73.

R1 excluded studies (n=184)

Conference abstracts and conference proceedings (n=2)

Bogaert E, Monten C, De Wagter C, De Neve W. Investigation of a fast CBCT protocol for supine accelerated whole breast Irradiation. Radiother Oncol. 2016;119:S434.

Kim HK, Lee ES, Lim C, et al. 201TiP Density lowering effect of ovarian function suppression in premenopausal breast cancer patients who had no density change with one year of tamoxifen treatment (DELFINO study). Ann Oncol. 2022;33:S627-S8.

Ineligible study design (n=17)

Anandarajah A, Chen Y, Stoll C, Hardi A, Jiang S, Colditz GA. Repeated measures of mammographic density and texture to evaluate prediction and risk of breast cancer: a systematic review of the methods used in the literature. Cancer Causes Control. 2023;34(11):939-48.

Bell RJ. Mammographic density and breast cancer screening. Climacteric. 2020:1-6.

Brandt KR, Scott CG, Ma L, et al. Measurements: Implications for risk prediction and supplemental screening. Radiology. 2016;279(3):710-9.

Brentnall AR, Cohn WF, Knaus WA, Yaffe MJ, Cuzick J, Harvey JA. A Case-Control Study to Add Volumetric or Clinical Mammographic Density into the Tyrer-Cuzick Breast Cancer Risk Model. Journal of breast imaging. 2019;1(2):99-106.

Eng A, Gallant Z, Shepherd J, et al. Digital mammographic density and breast cancer risk: A case-control study of six alternative density assessment methods. Breast Cancer Res. 2014;16(5).

Gardezi SJS, Elazab A, Lei B, Wang T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. J Med Internet Res. 2019;21(7):e14464.

Haji Maghsoudi O, Gastounioti A, Scott C, et al. Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment. Med Image Anal. 2021;73:102138.

Hernández A, Miranda DA, Pertuz S. Algorithms and methods for computerized analysis of mammography images in breast cancer risk assessment. Comput Methods Programs Biomed. 2021;212.

Jeffers AM, Sieh W, Lipson JA, et al. Breast cancer risk and Mammographic Density assessed with semiautomated and Fully automated Methods and Bi-raDs. Radiology. 2017;282(2):348-55.

Kerlikowske K, Ma L, Scott CG, et al. Combining quantitative and qualitative breast density measures to assess breast cancer risk. Breast Cancer Res. 2017;19.

Lamb LR, Lehman CD, Gastounioti A, Conant EF, Bahl M. Artificial Intelligence (AI) for Screening Mammography, from the AJR Special Series on AI Applications. American Journal of Roentgenology. 2022;219(3):369-81.

Mina LM, Isa NAM. A review of computer-aided detection and diagnosis of breast cancer in digital mammography. Journal of Medical Sciences (Faisalabad). 2015;15(3):110-21.

Pérez-Benito FJ, Signol F, Pérez-Cortés JC, et al. Global parenchymal texture features based on histograms of oriented gradients improve cancer development risk estimation from healthy breasts. Comput Methods Programs Biomed. 2019;177:123-32.

Tina Shih YC, Dong W, Xu Y, Etzioni R, Shen Y. Incorporating baseline breast density when screening women at average risk for breast cancer: A cost-effectiveness analysis. Ann Intern Med. 2021;174(5):602-12.

Winkel RR, von Euler-Chelpin M, Nielsen M, et al. Inter-observer agreement according to three methods of evaluating mammographic density and parenchymal

pattern in a case control study: Impact on relative risk of breast cancer. BMC Cancer. 2015;15(1):274.

Winkel RR, von Euler-Chelpin M, Nielsen M, et al. Mammographic density and structural features can individually and jointly contribute to breast cancer risk assessment in mammography screening: A case-control study. BMC Cancer. 2016;16(1).

Winkler NS, Raza S, Mackesy M, Birdwell RL. Breast density: clinical implications and assessment methods. Radiographics : a review publication of the Radiological Society of North America, Inc. 2015;35(2):316-24.

Ineligible population (n=15)

Castillo-García M, Chevalier M, Garayoa J, Rodriguez-Ruiz A, García-Pinto D, Valverde J. Automated Breast Density Computation in Digital Mammography and Digital Breast Tomosynthesis: Influence on Mean Glandular Dose and BIRADS Density Categorization. Acad Radiol. 2017;24(7):802-10.

Chiarelli AM, Prummel MV, Muradali D, et al. Digital versus screen-film mammography: impact of mammographic density and hormone therapy on breast cancer detection. Breast Cancer Res Treat. 2015;154(2):377-87.

Jari I, Ursaru M, Gheorghe L, Naum AG, Negru D. Computerized calculation of breast density: our experience from Arcadia Medical Imaging Center. Rev Med Chir Soc Med Nat Iasi. 2014;118(4):979-85.

Moshina N, Roman M, Sebuodegard S, Waade GG, Ursin G, Hofvind S. Comparison of subjective and fully automated methods for measuring mammographic density. Acta Radiol. 2018;59(2):154-60.

Pahwa S, Hari S, Thulkar S, Angraal S. Evaluation of breast parenchymal density with QUANTRA software. Indian Journal of Radiology and Imaging. 2015;25(4):391-6.

Perry N, Moss S, Dixon S, et al. Mammographic breast density and urbanization: Interactions with BMI, environmental, lifestyle, and other patient factors. Diagnostics. 2020;10(6):418.

Rahbar K, Gubern-Merida A, Patrie JT, Harvey JA. Automated Volumetric Mammographic Breast Density Measurements May Underestimate Percent Breast Density for High-density Breasts. Acad Radiol. 2017;24(12):1561-9.

Rahmat K, Ab Mumin N, Ramli Hamid MT, Fadzli F, Ng WL, Muhammad Gowdh NF. Evaluation of automated volumetric breast density software in comparison with visual assessments in an Asian population: A retrospective observational study. Medicine (United States). 2020;99(39):E22405.

Rangarajan K, Hari S, Thulkar S, Sharma S, Srivastava A, Parshad R. Characterization of lesions in dense breasts: Does tomosynthesis help? Indian Journal of Radiology and Imaging. 2016;26(2):210-5.

Regini E, Mariscotti G, Durando M, et al. Radiological assessment of breast density by visual classification (Bi–rads) compared to automated volumetric digital software (quantra): Implications for clinical practice. Radiologia Medica. 2014;119(10):741-9.

Rodriguez-Ruiz A, Gubern-Merida A, Imhof-Tas M, et al. One-view digital breast tomosynthesis as a stand-alone modality for breast cancer detection: do we need more? Eur Radiol. 2018;28(5):1938-48.

Sacchetto D, Morra L, Agliozzo S, et al. Mammographic density: Comparison of visual assessment with fully automatic calculation on a multivendor dataset. Eur Radiol. 2016;26(1):175-83.

Valencia-Hernandez I, Peregrina-Barreto H, Reyes-Garcia CA, Lopez-Armas GC. Density map and fuzzy classification for breast density by using BI-RADS. Comput Methods Programs Biomed. 2021;200.

Youk JH, Kim SJ, Son EJ, Gweon HM, Kim JA. Comparison of visual assessment of breast density in BI-RADS 4th and 5th editions with automated volumetric measurement. American Journal of Roentgenology. 2017;209(3):703-8.

Unclear population <200 participants/mammograms (n=14)

Badawy E, ElNaggar R, Soliman SAM, Elmesidy DS. Performance of AI-aided mammography in breast cancer diagnosis: Does breast density matter? Egyptian Journal of Radiology and Nuclear Medicine. 2023;54(1):178.

Carneiro P, Debs C, Andrade A, Patrocinio A. CLAHE Parameters Effects on the Quantitative and Visual Assessment of Dense Breast Mammograms. Ieee Latin America Transactions. 2019;17(5):851-7.

Damases CN, Brennan PC, Mello-Thoms C, McEntee MF. Mammographic Breast Density Assessment Using Automated Volumetric Software and Breast Imaging Reporting and Data System (BIRADS) Categorization by Expert Radiologists. Acad Radiol. 2016;23(1):70-7.

He W, Harvey S, Juette A, Denton ERE, Zwiggelaar R, Mammographic segmentation and density classification: A fractal inspired approach. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2016.

Machida Y, Tozaki M, Yoshida T, Saita A, Yakabe M, Nii K. Feasibility study of a breast density measurement within a direct photon-counting mammography scanner system. Japanese Journal of Radiology. 2014:561-7.

McEntee MF, Damases CN, Mammographic density measurement: A comparison of automated volumetric density measurement to BIRADS. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2014.

Oiwa M, Endo T, Suda N, et al. Can quantitative evaluation of mammographic breast density, "volumetric measurement", predict the masking risk with dense breast tissue? Investigation by comparison with subjective visual estimation by Japanese radiologists. Breast Cancer. 2019;26(3):349-58.

Oiwa M, Suda N, Morita T, et al. Validity of computed mean compressed fibroglandular tissue thickness and breast composition for stratification of masking risk in Japanese women. Breast Cancer. 2023;30(4):541-51.

Prinsloo E, Minne C, Greeff W. Measuring breast density: Comparing computer-automated breast density quantification with an observer-based method in a South African academic context. South African Journal of Radiology. 2018;22(2):1358.

Razali NF, Isa IS, Sulaiman SN, N.K AK, Osman MK. CNN-Wavelet scattering textural feature fusion for classifying breast tissue in mammograms. Biomed Signal Process Control. 2023;83:104683.

Sadad T, Khan AR, Hussain A, et al. Internet of medical things embedding deep learning with data augmentation for mammogram density classification. Microsc Res Tech. 2021;84(9):2186-94.

Saffari N, Rashwan HA, Abdel-Nasser M, et al. Fully automated breast density segmentation and classification using deep learning. Diagnostics. 2020;10(11):988.

Uyun S, Hartati S, Harjoko A, Choridah L, Comparison between automatic and semiautomatic thresholding method for mammographic density classification. Advanced Materials Research; 2014.

Wang J, Kato F, Yamashita H, et al. Automatic Estimation of Volumetric Breast Density Using Artificial Neural Network-Based Calibration of Full-Field Digital Mammography: Feasibility on Japanese Women With and Without Breast Cancer. J Digit Imaging. 2017;30(2):215-27.

Unclear population ≥200 participants/mammograms (n=24)

Balleyguier C, Arfi-Rouche J, Boyer B, et al. A new automated method to evaluate 2D mammographic breast density according to BI-RADS Atlas Fifth Edition recommendations. Eur Radiol. 2019;29(7):3830-8.

Carneiro PC, Franco MLN, De Lima Thomaz R, Patrocinio AC. Breast density pattern characterization by histogram features and texture descriptors. Revista Brasileira de Engenharia Biomedica. 2017;33(1):69-77.

Chen B, Ruth C, Zhang Y, Jing Z, Breast density assessment: Image feature extraction and density classification with machine intelligence. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2018.

Couwenberg AM, Verkooijen HM, Li J, et al. Assessment of a fully automated, high-throughput mammographic density measurement tool for use with processed digital mammograms. Cancer Causes Control. 2014;25(8):1037-43.

Dontchos BN, Yala A, Barzilay R, Xiang J, Lehman CD. External Validation of a Deep Learning Model for Predicting Mammographic Breast Density in Routine Clinical Practice. Acad Radiol. 2021;28(4):475-80.

Ji Y, Shao Z, Liu J, Hao Y, Liu P. The correlation between mammographic densities and molecular pathology in breast cancer. Cancer Biomark. 2018;22(3):523-31.

Kai C, Ishizuka S, Otsuka T, et al. Automated Estimation of Mammary Gland Content Ratio Using Regression Deep Convolutional Neural Network and the Effectiveness in Clinical Practice as Explainable Artificial Intelligence. Cancers (Basel). 2023;15(10).

Khan N, Wang K, Chan A, Highnam R, Automatic BI-RADS classification of mammograms. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2016.

Ko SY, Kim EK, Kim MJ, Moon HJ. Mammographic density estimation with automated volumetric breast density measurement. Korean Journal of Radiology. 2014;15(3):313-21.

Li C, Xu J, Liu Q, et al. Multi-View Mammographic Density Classification by Dilated and Attention-Guided Residual Learning. IEEE/ACM Trans Comput Biol Bioinform. 2021;18(3):1003-13.

Li S, Wei J, Chan HP, et al. Computer-aided assessment of breast density: comparison of supervised deep learning and feature-based statistical learning. Phys Med Biol. 2018;63(2):025005.

Lin X, Wu S, Li L, et al. Automatic mammographic breast density classification in Chinese women: clinical validation of a deep learning model. Acta Radiol. 2023;64(5):1823-30.

Ma X, Fisher C, Wei J, et al., Multi-path deep learning model for automated mammographic density categorization. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2019.

Matthews TP, Singh S, Mombourquette B, et al. A multisite study of a breast density deep learning model for full-field digital mammography and synthetic mammography. Radiology: Artificial Intelligence. 2021;3(1):e200015.

Nykanen A, Okuma H, Sutela A, Masarwah A, Vanninen R, Sudah M. The mammographic breast density distribution of Finnish women with breast cancer and comparison of breast density reporting using the 4th and 5th editions of the Breast Imaging-Reporting and Data System. Eur J Radiol. 2021;137:109585.

Oshima A, Shinohara N, Kamiya N, Investigation of the effect of image resolution on automatic classification of mammary gland density in mammography images using deep learning. Proceedings of SPIE - The International Society for Optical Engineering; 2019.

Singh N, Joshi P, Singh DK, Narayan S, Gupta A. Volumetric breast density evaluation using fully automated Volpara software, its comparison with BIRADS density types and correlation with the risk of malignancy. Egyptian Journal of Radiology and Nuclear Medicine. 2022;53(1).

Testagrose C, Gupta V, Erdal BS, et al., Impact of concatenation of digital craniocaudal mammography images on a deep-learning breast-density classifier using Inception-V3 and ViT. IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2022, Las Vegas, NV; 2022

Tsuchida T, Negishi T, Takahashi Y, Nishimura R. Dense-breast classification using image similarity. Radiological Physics and Technology. 2020;13(2):177-86.

Vállez N, Bueno G, Déniz O, et al. Breast density classification to reduce false positives in CADe systems. Comput Methods Programs Biomed. 2014;113(2):569-84.

Watanabe AT, Retson T, Wang J, Mantey R, Chim C, Karimabadi H. Mammographic Breast Density Model Using Semi-Supervised Learning Reduces Inter-/Intra-Reader Variability. Diagnostics. 2023;13(16):2694.

Wei J, Chan HP, Helvie MA, Roubidoux MA, Zhou C, Hadjiiski L, Radiomic modeling of BI-RADS density categories. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2017.

Yamamuro M, Asai Y, Hashimoto N, et al. Utility of U-Net for the objective segmentation of the fibroglandular tissue region on clinical digital mammograms. Biomedical Physics and Engineering Express. 2022;8(4):045016.

Youk JH, Gweon HM, Son EJ, Kim JA. Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: A comparison with visual assessment. American Journal of Roentgenology. 2016;206(5):1056-62.

Ineligible imaging modality (n=30)

Ahmadinejad N, Rasoulighasemlouei S, Rostamzadeh N, Arian A, Mohajeri A, Miratashi Yazdi SN. Our experience using synthesized mammography vs full field digital mammography in population-based screening. European Journal of Radiology Open. 2023;10:100475.

Alshafeiy TI, Wadih A, Nicholson BT, et al. Comparison between digital and synthetic 2D mammograms in breast density interpretation. American Journal of Roentgenology. 2017;209(1):W36-W41.

Bandeira Diniz JO, Bandeira Diniz PH, Azevedo Valente TL, Corrêa Silva A, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Methods Programs Biomed. 2018;156:191-207.

Choi EJ, Choi H, Byon JH, Youk JH. Analysis of background echotexture on automated breast ultrasound using BI-RADS and modified classification: Association with clinical features and mammographic density. Journal of clinical ultrasound : JCU. 2023;51(4):687-95.

Chugh S, Goyal S, Pandey A, Joshi S. Morphological and Otsu's Technique Based Mammography Mass Detection and Deep Neural Network Classifier Based Prediction. Traitement Du Signal. 2022;39(4):1283-94.

Ekpo EU, Mello-Thoms C, Rickard M, Brennan PC, McEntee MF. Breast density (BD) assessment with digital breast tomosynthesis (DBT): Agreement between Quantra™ and 5th edition BI-RADS®. Breast. 2016;30:185-90.

Fischer U, Luftner-Nagel S, Baum F, Marten-Engelke K, Wienbeck S. The Value of Quality-Assured Magnetic Resonance Imaging of the Breast for the Early Detection of Breast Cancer in Asymptomatic Women. J Comput Assist Tomogr. 2018;42(1):1-5.

Gamdonkar Z, Tay K, Ryder W, Brennan PC, Mello-Thoms C, IDensity: An automatic Gabor filter-based algorithm for breast density assessment. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2015.

Hamid MTR, Mumin NA, Hamid SA, Rahmat K. Application of Artificial Intelligence (AI) System in Opportunistic Screening and Diagnostic Population in a Middle-income Nation. Current medical imaging. 2024.

Johansson H, Von Tiedemann M, Cederström B, Breast density classification based on volumetric glandularity measured by spectral mammography. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2014.

Kaliyaperumal V, Selvarajan S. Automated characterization of mammographic density for early detection of breast cancer risk. International Journal of Simulation: Systems, Science and Technology. 2014;15(1):56-63.

Khanani S, Xiao L, Jensen MR, et al. Comparison of breast density assessments between synthesized C-View™ & intelligent 2D™ mammography. Br J Radiol. 2022;95(1134).

Lee ZY, Goh YLE, Lai C. Classification of mammographic breast density and its correlation with BI-RADS in elder women using machine learning approach. Journal of Medical Imaging and Radiation Sciences. 2022;53(1):28-34.

Lewin J, Schoenherr S, Seebass M, et al. PACS-integrated machine learning breast density classifier: clinical validation. Clin Imaging. 2023;101:200-5.

Magni V, Interlenghi M, Cozzi A, et al. Development and Validation of an AI-driven Mammographic Breast Density Classification Tool Based on Radiologist Consensus. Radiology: Artificial Intelligence. 2022;4(2):e210199.

Molloi S, Ding H, Feig S. Breast density evaluation using spectral mammography, radiologist reader assessment, and segmentation techniques: a retrospective study based on left and right breast comparison. Acad Radiol. 2015;22(8):1052-9.

Mumin NA, Rahmat K, Fadzli F, et al. Diagnostic Efficacy of Synthesized 2D Digital Breast Tomosynthesis in Multi-ethnic Malaysian Population. Sci Rep. 2019;9(1):1459.

Natesan R, Wiskin J, Lee S, Malik BH. Quantitative assessment of breast density: Transmission ultrasound is comparable to mammography with tomosynthesis. Cancer Prevention Research. 2019;12(12):871-6.

Osteras BH, Martinsen ACT, Gullien R, Skaane P. Digital mammography versus breast tomosynthesis: Impact of breast density on diagnostic performance in population-based screening. Radiology. 2019;293(1):60-8.

Pawar SD, Sharma KK, Sapate SG, et al. Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection. Frontiers in public health. 2022;10:885212.

Sexauer R, Hejduk P, Borkowski K, et al. Diagnostic accuracy of automated ACR BI-RADS breast density classification using deep convolutional neural networks. Eur Radiol. 2023;33(7):4589-96.

Sharma V, Singh S. CFS-SMO based classification of breast density using multiple texture models. Med Biol Eng Comput. 2014;52(6):521-9.

Simon S, Lavanya R, Vijayan D, PSO based density classifier for mammograms. IFMBE Proceedings; 2017.

Testagrose CT, Gupta V, Erdal BS, et al., Breast-Density Semantic Segmentation with Probability Scaling for BI-RADS Assessment using DeepLabV3. ACM-BCB 2023 - 14th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics; 2023.

Tiryaki VM, Kaplanoğlu V. Deep Learning-Based Multi-Label Tissue Segmentation and Density Assessment from Mammograms. IRBM. 2022;43(6):538-48.

Tittmann J, Csanadi M, Agh T, Szeles G, Voko Z, Kallai A. Development of a breast cancer screening protocol to use automated breast ultrasound in a local setting. Frontiers in public health. 2022;10:1071317.

Trivizakis E, Ioannidis GS, Melissianos VD, et al. A novel deep learning architecture outperforming 'off-the-shelf' transfer learning and feature-based methods in the automated assessment of mammographic breast density. Oncol Rep. 2019;42(5):2009-15.

Tunissen SAM, Motta A, Mauter F, et al., End-to-End Mammographic Breast Density Quantification with Deep Learning: Preliminary Study on Simulated Mammograms. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2023.

Wiskin J, Malik B, Natesan R, Lenox M. Quantitative assessment of breast density using transmission ultrasound tomography. Med Phys. 2019;46(6):2610-20.

Yi PH, Lin A, Wei J, et al. Deep-Learning-Based Semantic Labeling for 2D Mammography and Comparison of Complexity for Machine Learning Tasks. J Digit Imaging. 2019;32(4):565-70.

Unclear imaging modality (n=1)

Ekpo EU, Hogg P, Wasike E, McEntee MF. A self-directed learning intervention for radiographers rating mammographic breast density. Radiography. 2017;23(4):337-42.

No measurement of breast density (n=13)

Ain QUI, Xue B, Al-Sahaf H, Zhang M, A Genetic Programming Approach to Automatically Construct Informative Attributes for Mammographic Density Classification. IEEE International Conference on Data Mining Workshops, ICDMW; 2022.

Alonzo-Proulx O, Mainprize J, Hussein H, Jong R, Yaffe M, Local detectability maps as a tool for predicting masking probability and mammographic performance. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2016.

Arslan ZF, Altunkeser A, Aksoy N, Korez MK, Omeroglu E. Which Modality Should Be Integrated to Increase the Diagnostic Efficiency of BI-RADS 0, 3, and 4 Lesions? Ultrasonography or Digital Breast Tomosynthesis? Iranian Journal of Radiology. 2021;18(4).

Hanis TM, Ruhaiyem NIR, Arifin WN, et al. Developing a Supplementary Diagnostic Tool for Breast Cancer Risk Estimation Using Ensemble Transfer Learning. Diagnostics. 2023;13(10):1780.

Lacson R, Goodrich ME, Harris K, Brawarsky P, Haas JS. Assessing Inaccuracies in Automated Information Extraction of Breast Imaging Findings. J Digit Imaging. 2017;30(2):228-33.

Mackenzie A, Thomson EL, Mitchell M, et al. Virtual clinical trial to compare cancer detection using combinations of 2D mammography, digital breast tomosynthesis and synthetic 2D imaging. Eur Radiol. 2022;32(2):806-14.

Menegatti Pavan AL, Vacavant A, Petean Trindade A, Quini CC, Rodrigues De Pina D, Enhancement of breast periphery region in digital mammography. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2018.

Pawar SD, Sharma KK, Sapate SG, Yadav GY. Segmentation of pectoral muscle from digital mammograms with depth-first search algorithm towards breast density classification. Biocybernetics and Biomedical Engineering. 2021;41(3):1224-41.

Saffari N, Rashwan HA, Herrera B, Romani S, Arenas M, Puig D, On Improving Breast Density Segmentation Using Conditional Generative Adversarial Networks. Frontiers in Artificial Intelligence and Applications; 2018.

Salim M, Dembrower K, Eklund M, Smith K, Strand F. Differences and similarities in false interpretations by AI CAD and radiologists in screening mammography. Br J Radiol. 2023;96(1151):20230210.

Shermis RB, Redfern RE, Burns J, Kudrolli H. Molecular Breast Imaging in Breast Cancer Screening and Problem Solving. Radiographics : a review publication of the Radiological Society of North America, Inc. 2017;37(5):1309-606.

Uematsu T, Nakashima K, Harada TL, Nasu H, Igarashi T. Artificial intelligence computer-aided detection enhances synthesized mammograms: comparison with original digital mammograms alone and in combination with tomosynthesis images in an experimental setting. Breast Cancer. 2023;30(1):46-55.

Verburg E, Van Gils CH, Van Der Velden BHM, et al. Deep Learning for Automated Triaging of 4581 Breast MRI Examinations from the DENSE Trial. Radiology. 2022;302(1):29-36.

No comparison with manual measurement (n=40)

Al-Qershi O, Nguyen TL, Elliott MS, et al. AutoCumulus: An Automated Mammographic Density Measure Created Using Artificial Intelligence. medRxiv. 2024.

Alabdulali N, Dris AB, Alqahtani F, Othman AB. A Hybrid Approach of Texture Feature and Gradient Orientation for Computer Aided Diagnosis System Based on Breast Density Classification. Journal of Computer Science. 2020;16(10):1491-500.

Alexeeff SE, Odo NU, McBride R, et al. Reproductive Factors and Mammographic Density: Associations among 24,840 Women and Comparison of Studies Using Digitized Film-Screen Mammography and Full-Field Digital Mammography. Am J Epidemiol. 2019;188(6):1144-54.

Alonzo-Proulx O, Mawdsley GE, Patrie JT, Yaffe MJ, Harvey JA. Reliability of automated breast density measurements. Radiology. 2015;275(2):366-76.

Araujo ALC, Soares HB, Carvalho DF, Mendonca RM, Oliveira AG. Design and clinical validation of a software program for automated measurement of mammographic breast density. BMC Med Inform Decis Mak. 2020;20(1):45.

Bell RJ, Evans J, Fox J, Pridmore V. Using an automated measure of breast density to explore the association between ethnicity and mammographic density in Australian women. J Med Imaging Radiat Oncol. 2019;63(2):183-9.

Chang K, Beers AL, Brink L, et al. Multi-Institutional Assessment and Crowdsourcing Evaluation of Deep Learning for Automated Classification of Breast Density. Journal of the American College of Radiology. 2020;17(12):1653-62.

Cheddad A, Czene K, Eriksson M, et al. Area and volumetric density estimation in processed full-field digital mammograms for risk assessment of breast cancer. PLoS One. 2014;9(10).

Conant EF, Keller BM, Pantalone L, Gastounioti A, McDonald ES, Kontos D. Agreement between Breast Percentage Density estimations from standard-Dose versus synthetic Digital Mammograms: Results from a Large Screening Cohort Using Automated Measures. Radiology. 2017;283(3):673-80.

Damases CN, Brennan PC, McEntee MF. Mammographic density measurements are not affected by mammography system. Journal of Medical Imaging. 2015;2(1):15501.

Deng J, Ma Y, Li DA, Zhao J, Liu Y, Zhang H. Classification of breast density categories based on SE-Attention neural networks. Comput Methods Programs Biomed. 2020;193:105489.

Devi SS, Vidivelli S, Classification of breast tissue density in digital mammograms. Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017; 2017.

Dontchos BN, Cavallo-Hom K, Lamb LR, et al. Impact of a Deep Learning Model for Predicting Mammographic Breast Density in Routine Clinical Practice. Journal of the American College of Radiology. 2022;19(9):1021-30.

Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis--a systematic review. The British journal of radiology. 2014;87(1043):20140460.

Fowler EEE, Smallwood A, Miltich C, Drukteinis J, Sellers TA, Heine J. Generalized breast density metrics. Phys Med Biol. 2019;64(1).

Fowler EEE, Vachon CM, Scott CG, Sellers TA, Heine JJ. Automated percentage of breast density measurements for full-field digital mammography applications. Acad Radiol. 2014;21(8):958-70.

García E, Diaz O, Martí R, et al. Local breast density assessment using reacquired mammographic images. Eur J Radiol. 2017;93:121-7.

Gastounioti A, Hsieh MK, Pantalone L, Conant EF, Kontos D, Volumetric versus area-based density assessment: Comparisons using automated quantitative measurements from a large screening cohort. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2018.

Gastounioti A, Kasi CD, Scott CG, et al. Evaluation of LIBRA software for fully automated mammographic density assessment in breast cancer risk prediction. Radiology. 2020;296(1):24-31.

Giorgi Rossi P, Djuric O, Helin V, et al. Validation of a new fully automated software for 2D digital mammographic breast density evaluation in predicting breast cancer risk. Sci Rep. 2021;11(1):19884.

Habel LA, Alexeeff SE, Achacoso N, et al. Examination of fully automated mammographic density measures using LIBRA and breast cancer risk in a cohort of 21,000 non-Hispanic white women. Breast Cancer Res. 2023;25(1):92.

Kang E, Lee EJ, Jang M, et al. Reliability of computer-assisted breast density estimation: Comparison of interactive thresholding, semiautomated, and fully automated methods. American Journal of Roentgenology. 2016;207(1):126-34.

Keller BM, Chen J, Daye D, Conant EF, Kontos D. Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: Comparison of fully automated area and volumetric density measures in a case-control study with digital mammography. Breast Cancer Res. 2015;17(1).

Kim Y, Hong BW, Kim SJ, Kim JH. A population-based tissue probability map-driven level set method for fully automated mammographic density estimations. Med Phys. 2014;41(7):071905.

Larroza A, Perez-Benito FJ, Perez-Cortes JC, et al. Breast Dense Tissue Segmentation with Noisy Labels: A Hybrid Threshold-Based and Mask-Based Approach. Diagnostics. 2022;12(8):1822.

Llobet R, Pollan M, Anton J, et al. Semi-automated and fully automated mammographic density measurement and breast cancer risk prediction. Comput Methods Programs Biomed. 2014;116(2):105-15.

Lopez-Almazan H, Javier Perez-Benito F, Larroza A, et al. A deep learning framework to classify breast density with noisy labels regularization. Comput Methods Programs Biomed. 2022;221:106885.

Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. Med Phys. 2018;45(1):314-21.

Moshina N, Sebuødegård S, Lee CI, et al. Automated volumetric analysis of mammographic density in a screening setting: Worse outcomes for women with dense breasts. Radiology. 2018;288(2):343-52.

Mustra M, Stajduhar A. Segmentation Masks for the Mini-Mammographic Image Analysis Society (mini-MIAS) Database. Ieee Consumer Electronics Magazine. 2020;9(5):28-33.

Perez-Benito FJ, Signol F, Perez-Cortes JC, et al. A deep learning system to obtain the optimal parameters for a threshold-based breast and dense tissue segmentation. Comput Methods Programs Biomed. 2020;195:105668.

Ren K, Harkness E, Boggis C, et al., Mammographic density and breast cancer characteristics. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2014.

Richard-Davis G, Whittemore B, Disher A, et al. Evaluation of Quantra Hologic Volumetric Computerized Breast Density Software in Comparison With Manual

Interpretation in a Diverse Population. Breast Cancer: Basic and Clinical Research. 2018;12.

Saikiran P, Ramzan R, Nandish S, Kamineni PD, John AM. Mammographic breast density assessed with fully automated method and its risk for breast cancer. J Clin Imaging Sci. 2019;9(1).

Schmachtenberg C, Hammann-Kloss S, Bick U, Engelken F. Intraindividual Comparison of Two Methods of Volumetric Breast Composition Assessment. Acad Radiol. 2015;22(4):447-52.

Sohn G, Lee JW, Park SW, et al. Reliability of the percent density in digital mammography with a semi-automated thresholding method. J Breast Cancer. 2014;17(2):174-9.

Sovio U, Li J, Aitken Z, et al. Comparison of fully and semi-automated area-based methods for measuring mammographic density and predicting breast cancer risk. Br J Cancer. 2014;110(7):1908-16.

Verma P, Dumka A, Bhardwaj A, Kestwal MC. Classifying Breast Density in Mammographic Images Using Wavelet-Based and Fine-Tuned Sensory Neural Networks. International Journal of Image and Graphics. 2021;21(5).

Wimmer M, Sluiter G, Major D, et al. Multi-Task Fusion for Improving Mammography Screening Data Classification. IEEE Trans Med Imaging. 2022;41(4):937-50.

Zhao W, Wang R, Qi Y, et al. BASCNet: Bilateral adaptive spatial and channel attention network for breast density classification in the mammogram. Biomed Signal Process Control. 2021;70.

BI-RADS density scoring system was not used for the manual measurement (n=15)

Beattie L, Harkness E, Bydder M, et al., Factors affecting agreement between breast density assessment using volumetric methods and visual analogue scales. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2014.

Benitez CS, Pertuz S, Arponen O, Laaperi AL, Rinta-Kiikka I, Automatic Dense Tissue Segmentation in Digital Mammography Images Based on Fully Convolutional Network and Intensity-Based Clustering. 2022 IEEE Colombian Conference on Applications of Computational Intelligence, ColCACI 2022 - Proceedings; 2022.

Damases CN, Mello-Thoms C, McEntee MF. Inter-observer variability in mammographic density assessment using Royal Australian and New Zealand College of Radiologists (RANZCR) synoptic scales. J Med Imaging Radiat Oncol. 2016;60(3):329-36.

Duffy SW, Morrish OWE, Allgood PC, et al. Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. Eur J Cancer. 2018;88:48-56.

Evans DG, Astley S, Stavrinos P, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and family history clinics: a dual cohort study. 2016.

Gudhe NR, Behravan H, Sudah M, et al. Area-based breast percentage density estimation in mammograms using weight-adaptive multitask learning. Sci Rep. 2022;12(1):12060.

Ionescu GV, Fergie M, Berks M, et al., Using a convolutional neural network to predict readers' estimates of mammographic density for breast cancer risk assessment. Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2018.

Ionescu GV, Fergie M, Berks M, et al. Prediction of reader estimates of mammographic density using convolutional neural networks. Journal of Medical Imaging. 2019;6(3):031405.

Morrish OWE, Tucker L, Black R, Willsher P, Duffy SW, Gilbert FJ. Mammographic breast density: Comparison of methods for quantitative evaluation. Radiology. 2015;275(2):356-65.

Ryan F, Roman KLL, Gerboles BZ, et al. Unsupervised domain adaptation for the segmentation of breast tissue in mammography images. Comput Methods Programs Biomed. 2021;211:106368.

Squires S, Harkness E, Gareth Evans D, Astley SM. Automatic assessment of mammographic density using a deep transfer learning method. Journal of Medical Imaging. 2023;10(2):024502.

Squires S, Harkness EF, Evans DG, Astley SM. The effect of variable labels on deep learning models trained to predict breast density. Biomedical Physics and Engineering Express. 2023;9(3):035030.

Wong Sik Hee JR, Harkness EF, Gadde S, et al., Does the prediction of breast cancer improve using a combination of mammographic density measures compared to individual measures alone? Progress in Biomedical Optics and Imaging - Proceedings of SPIE; 2017.

Woolcott CG, Conroy SM, Nagata C, et al. Methods for assessing and representing mammographic density: An analysis of 4 case-control studies. Am J Epidemiol. 2014;179(2):236-44.

Yeo I, Akwo J, Ekpo E. Automated mammographic density measurement using QuantraTM: Comparison with the Royal Australian and New Zealand College of Radiology synoptic scale. Journal of Medical Imaging. 2020;7(3):035501.

No automated software (n=2)

Angelo MF, Carneiro PC, Granado TC, Patrocinio AC, Influence of contrast enhancement to breast density classification by using sigmoid function. IFMBE Proceedings; 2015.

Krammer J, Zolotarev S, Hillman I, et al. Evaluation of a new image reconstruction method for digital breast tomosynthesis: Effects on the visibility of breast lesions and breast density. Br J Radiol. 2019;92(1103):20190345.

No valid or useable outcome data (n=10)

Aloufi AS, Alnaeem AN, Almousa AS, et al. Mammographic breast density and breast cancer risk in the Saudi population: a case-control study using visual and automated methods. Br J Radiol. 2022;95(1134):A16.

Arzehgar A, Khalilzadeh MM, Varshoei F. Assessment and Classification of Mass Lesions Based on Expert Knowledge Using Mammographic Analysis. Curr Med Imaging Rev. 2019;15(2):199-208.

Astley SM, Harkness EF, Sergeant JC, et al. A comparison of five methods of measuring mammographic density: A case-control study. Breast Cancer Res. 2018;20(1):10.

Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep Learning in Mammography Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. Invest Radiol. 2017;52(7):434-40.

Brand JS, Czene K, Shepherd JA, et al. Automated measurement of volumetric mammographic density: A tool for widespread breast cancer risk assessment. Cancer Epidemiology Biomarkers and Prevention. 2014;23(9):1764-72.

Chakraborty D, Palit S, Bhattacharya U, Deep Classification of Mammographic Breast Density: DCBARNet. International Conference Image and Vision Computing New Zealand; 2023.

Martinez Gomez I, Casals El Busto M, Anton Guirao J, Ruiz Perales F, Llobet Azpitarte R. Semiautomatic estimation of breast density with DM-Scan software. Radiologia (Roma). 2014;56(5):429-34.

Oliver A, Tortajada M, Llado X, et al. Breast Density Analysis Using an Automatic Density Segmentation Algorithm. J Digit Imaging. 2015;28(5):604-12.

Vilmun BM, Vejborg I, Lynge E, et al. Impact of adding breast density to breast cancer risk models: A systematic review. Eur J Radiol. 2020;127:109019.

Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology. 2019;292(1):60-6.

Participant overlap (n=1)

Osteras BH, Martinsen ACT, Brandal SHB, et al. BI-RADS density classification from areometric and volumetric automatic breast density measurements. Acad Radiol. 2016;23(4):468-78.

Appendix 3 — Summary and appraisal of individual studies

Data Extraction

Table 4 Characteristics of the included studies

Study ID	Authors associated with automated software development or declared relationship with software developer	Date when mammograms were taken	Name of centre(s) or database(s) providing mammography images	Participant demographics	Population and eligibility criteria
Alomaim 2020[89] **UK and USA**	No	NR	NR	Age: NR Ethnicity: NR	The cases were collected as part of previous research from 18 centres in a national breast screening programme Exclusion criteria were not reported
Aloufi 2022[88] **Saudi Arabia**	No	2012 to 2018	Saudi National Breast Cancer Screening Programme (SNBCSP)	Age, years, median (IQR): 50 (45-55) Ethnicity: NR	All mammograms were obtained from women attending the Saudi National Breast Cancer Screening Programme (SNBCSP). Participants younger than screening age (40 years) and older than 75 years. Thirteen (0.4%) participants were excluded due to an error in the Volpara output indicating a

					mosaic image or because they had an extra mammographic image for any screening view
Anguloa 2015[83] **Peru**	Yes, the authors were involved in developing the automated software	NR	Three medical centres in Lima	Age, years, mean, (SD), range: 56.7 (9.5), 31 to 86 Ethnicity: NR	All participants underwent routine breast cancer screening. Participants with detected pathologies were excluded
Fonseca 2015[84] **Peru**	Yes. The fifth author, Joseph Pinto, is an author on the two studies cited for the development of the algorithm	NR	Two medical centres in Lima	Age, years, mean, (SD): 56.7 (9.5) Ethnicity: NR	All participants underwent routine breast cancer screening Exclusion criteria were not reported
Ekpo 2016[86] **Australia**	No	March to July 2014	NR	Age: NR Ethnicity: NR	All participants underwent screening mammography Exclusion criteria were not reported

Eom 2018[27] **Republic of Korea**	No	January to June 2016	Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul,	Age, years, mean, range: 52.9, 30 to 89 Ethnicity: NR	Mammograms are reported as a set of screening mammograms Participants with a history of breast surgery, augmentation, or foreign body injections were excluded
Fieselmann 2019[71] **Sweden**	Yes. Kristina Lång, Hanna Sartor, and Sophia Zackrisson received grants and speaking fees from Siemens Healthcare GmbH. Andreas Fieselmann, Steffen Kappler, Thomas Mertelmeier, and Ludwig Ritschl are employed by Siemens Healthcare GmbH. Siemens Healthcare GmbH developed the automated software	NR	Mälmo Breast Tomosynthesis Screening Trial (MBTST)	Age, years, mean, (SD), range: 57 (9), 39 to 75 Ethnicity: NR	All mammograms were obtained from the MBTST. The MBTST invited every third women aged 40–74 years living in Malmö who were invited to mammography screening as part of the Sweden national breast cancer screening programme. The MBTST excluded participants if they were pregnant or could not speak Swedish or English. No study specific exclusion criteria are reported
Fornvik 2019[72]	Yes. The authors of the manuscript declared relationships with the following companies: Siemens	NR	Mälmo Breast Tomosynthesis Screening Trial (MBTST)	Age, years, mean, range: 57, 39 to 75	All mammograms were obtained from the MBTST.

Sweden	Healthcare. Siemens provided the mammography/tomosynthesis equipment and VBD software prototype. The sponsors (Siemens) of the study had no role in the design and performance of the study, data analyses, or data interpretation. One of the co-authors, AF PhD, Siemens Healthcare, provided expertise on the software prototype. HS and KL received speaker's fees and travel grants for talking at Siemens seminars			Ethnicity: NR	Out of 14,848 participants in the MBTST, 4735 women lacked saved raw data, and 129 participants lacked BI-RADS density scores. Seventy-three participants with implants were excluded and two participants were excluded due to technical error in image handling.
Kaiser 2019[56] **Sweden**	Yes. The authors developed the automated software.	NR	Mälmo Breast Tomosynthesis Screening Trial (MBTST)	Age: NR Ethnicity: NR	All mammograms were obtained from the MBTST. Images with no density classification were excluded.
Sartor 2016[73] **Sweden**	Volpara sponsored the study but had no role in the design and performance of the study, data analysis, or data interpretation	8th February 2012 to 11th March 2014	Mälmo Breast Tomosynthesis Screening Trial (MBTST)	Age, years, mean, range: 58, 40 to 76 Ethnicity: NR	All mammograms were obtained from the MBTST. Examinations from participants with breast cancer with at least 10 months of follow up

					and examinations with previously known breast implants were excluded
Timberg 2016[74] **Sweden**	Yes. The authors developed the automated algorithm	NR Participants in the MBTST were recruited between 27th January 2010 and 13th February 2015	Mälmo Breast Tomosynthesis Screening Trial (MBTST)	Age, years, mean, (SD): 57 (4) Ethnicity: NR	A random sample of mammograms were taken from the MBTST
Gemici 2020[26] **Turkey**	No	May 2014 to May 2015	NR	Age, years, mean, (SD): 52.4 (8.3) Ethnicity: NR	All mammograms were described as screening mammograms of women who participated in a community based mammographic screening programme for the first time. Healthy women in the 40-69 age group who participated in the population based mammographic screening programme were eligible
Holland 2016[75]	K Holland, J van Zelst and M Imhof-Tas have nothing to disclose. GJ den Heeten is founder of Sigma screening, a spin company of the Academic Medical Centre	2003 to 2012	NR	Age, years, mean, (SD): 58.8 (6.7) Ethnicity: NR	All participants were recruited from the Dutch breast cancer screening programme.

The Netherlands	Amsterdam. RM Mann is connected to the Speaker Bureau of the Bayer AG. CH van Gils is principal investigator of a trial that is financially supported by Bayer AG and for which Volpara Health Technologies has provided Volpara software. N Karssemeijer is shareholder of Volpara Health Technologies, consultant and shareholder of QView Medical Inc, and director and shareholder of ScreenPoint Medical BV.				Participants with a prior and current exam were randomly selected from the Dutch breast cancer screening programme.
Le Boulc'h 2020[77] **France**	No	1st January 2019 to 28th February 2019	Department of Radiology, Tenon Hospital, AP-HP Sorbonne University	Age, years, mean, (SD), range: 55.6 (8.5), 40 to 74 Ethnicity: NR	All participants had no personal history of breast cancer and underwent routine mammography. Exclusions included high risk of breast cancer, including BRCA mutations serous ovarian cancer) or a personal history of radiotherapy for lymphoma, personal history of breast cancer and high-risk breast lesions, breast implants, incomplete risk factor data, mammographic artefacts or technical problems.

Lee 2022[82] **Republic of Korea**	No	March to May 2020	Yongin Severance Hospital	Age, years, mean, (SD): 56.3 (10.9) Ethnicity, %: Asian 100%	Screening: 91% Diagnostic: 9% Patients with a history of partial mastectomy and patients without available density data were excluded
Lee 2018[79] **USA**	RM Nishikawa received royalties from and has a research contract with Hologic, Inc. The authors developed the automated algorithm	2007 to 2013	University of Pittsburgh Medical Center	Age: NR Ethnicity: NR	The authors report that the datasets are of "screening mammograms" Exclusion criteria were not reported.
Lee 2015[17] **Republic of Korea**	No	February 2011 to September 2012	NR	Age, years, mean, (SD), range: 54.7 (10.2), 26 to 89 Ethnicity: NR	Screening: 88.4% Diagnostic: 11.6% Exclusion criteria were not reported.
Lehman 2019[33] **USA**	The authors declared no relevant relationships related to the present article. The authors' institutions have submitted a patent for the density algorithm.	January 2018 to May 2018	NR	Age, years, mean, range: 57.5, 28 to 92 Ethnicity: NR	The authors confirmed by email dated 16th October 2024 that ≥80% of the study sample were from the general screening population with no previous history of breast cancer.

					There were no exclusion criteria.
Osteras 2016[78] **Norway**	No	22nd November 2010 to 31st December 2011	Oslo Tomosynthesis Screening Trial (part of The Norwegian Breast Cancer Screening Programme [NBCSP])	Age: NR Ethnicity: NR	Mammograms were obtained from The Breast Cancer Screening Programme in Oslo, which is part of the NBCSP Participants with pacemaker or implants were excluded
Pavan 2017[87] **Brazil**	Yes. The authors developed the automated algorithm	2013 to 2015	Botucatu Medical School	Age: NR Ethnicity: NR	All participants underwent screening mammography Eligible participants were aged ≥18 years, had no previous history of breast cancer or breast surgery and had a BI-RADS assessment of either 1 or 2 (negative or benign finding, respectively).
Pesce 2020[85] **Argentina**	No	All were conducted in February 2019	Breast Diagnosis and Intervention Section of the Diagnostic Imaging Department at a tertiary hospital	Age, years, range: 40 to 90 Ethnicity: NR	All participants were asymptomatic at the time of screening and did not have a personal history of breast cancer Focalised and magnified incidents as well as mammograms for patients with a personal history of breast surgery (including

					breast implants) or gigantomastia were excluded
Portnow 2022[80] **USA**	No	September 2012 to January 2013	NR	Age: NR Ethnicity: NR	The authors report that the datasets are of "screening mammograms" Exclusion criteria were not reported.
Rigaud 2022[29] **USA**	Unclear if the authors developed the pre-trained EfficientNetB0 DL model	2017 to 2021	NR	Age, years, range: 25 to 80	All participants were enrolled in a prospective breast cancer screening cohort at a single academic institution. Eligible participants had no history of breast cancer, no history of treatment for any invasive cancer within the last five years, 25–80 years of age, and no breastfeeding within the last six months
Singh 2016[28] **India**	No	March 2013 to April 2014	NR	Age, years, mean, (SD), range: 48.8 (7.07), 36 to 76 Ethnicity: NR	All mammograms were obtained from asymptomatic patients with no with breast symptoms, and no previous history of any breast surgery.

					Participants aged less than 35 years, with breast symptoms, and with previous history of any breast surgery were excluded
van der Waal 2015[76] **The Netherlands**	GJ den Heeten is a former member of the medical advisory board of Mātakina (owners of Volpara), founder of sigmascreening (spin-off company Academic Medical Centre Amsterdam), and consultant of Philips ICT	During 2013	Nijmegen screening unit	Age: NR Ethnicity: NR	All mammograms were taken from participants in the Netherlands breast cancer screening programme. Retrieval dates were chosen at random.
Ref ID 245 Youk 2021[81] **Republic of Korea**	No	January 2016 to July 2016	Department of Radiology, Gangnam Severance Hospital	Age, years, mean, (SD): 52.7 (10.2) Ethnicity: NR	(95%) screening and (5%) diagnostic examinations. 1271 (31.8%) women had images of a unilateral breast due to prior breast cancer surgery on the contralateral breast. Examinations of breast conserving surgery or mastectomy for cancer treatment had been performed; examinations of augmented breasts; and examinations from which the software failed to obtain volumetric density data were excluded

IQR, interquartile range; MBTS, Mälmo Breast Tomosynthesis Screening Trial; NBCSP, Norwegian Breast Cancer Screening Programme; not reported; SD, standard deviation; SNBCSP, Saudi National Breast Cancer Screening Programme

Appraisal for quality and risk of bias

Table 5 Quality assessment of the included studies

Study ID	1.Were participants a representative sample selected from a relevant patient population	2. Were the inclusion / exclusion criteria of participants clearly described?	3.Were participants entering the study at a similar point in their disease progression?	4.Was selection of patients consecutive?	5. Was data collection undertaken prospectively?	6. Were the groups comparable on demographic characteristics and clinical features?	7. Was the intervention and comparison clearly defined?	8. Was the intervention undertaken by someone experienced at performing the procedure?	9. Were the staff, place, and facilities where the patients were treated appropriate for performing the procedure?	10. Were any of the important outcomes considered i.e. on clinical effectiveness, cost-effectiveness, or learning curves?	11. Were objective (valid and reliable) outcome measures used, including satisfaction scale?	12. Was the assessment of main outcomes blind?
Alomaim 2020[89]	Unclear	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Aloufi 2022[88]	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Angulo 2015[83]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Eom 2018[27]	Unclear	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ekpo 2016[99]	Unclear	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fieselmann 2019[71]	Yes	Partial	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Fonseca 2015[84]	Yes	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fornvik 2019[72]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Gemici 2020[26]	Unclear	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Holland 2016[75]	Unclear	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Kaiser 2019[56]	Yes	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear

Le Boulc'h 2020[77]	Unclear	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lee 2018[79]	Unclear	No	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Lee 2016[17]	Unclear	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lee 2022[82]	Unclear	No	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lehman 2019[33]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Osteras 2016[78]	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Ye	Yes	Yes	Yes	Unclear
Pavan 2017[87]	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Pesce 2020[85]	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Portnow 2022[80]	Unclear	No	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Rigaud 2022[29]	Yes	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Sartor 2016[73]	Yes	Partial	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Singh 2016[28]	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Timberg 2016[74]	Yes	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
van der Waal 2015[76]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Youk 2021[81]	Unclear	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

MBTST, Mälmo Breast Tomosynthesis Screening Trial

Appendix 4 – Full-length results tables

Table 6 Full-length results of the studies evaluating Volpara versus manual measurement of mammographic breast density

Study ID	Sample size used in the analysis	Type of mammography	Automated software version and BI-RADS edition	Density categories, n (%) [a]						Agreement value (95% CI)	Agreement interpretation
				A	B	C	D	Non-dense	Dense		
Gemici 2020[26]	379 mammograms	FFDM Selenia, Hologic	Volpara version 1.4.2 (n=1399 mammograms)	n, NR, 12.7%	n, NR, 39.3%	n, NR, 34.1%	n, NR, 13.9%	n, NR 52% [b]	n, NR 48% [b]	V 1.4.2 vs BI-RADS 4: κ -0.41 (NR)	Poor agreement
			Volpara version 1.5.1 (n=1399 mammograms)	n, NR, 1.2%	n, NR, 46%	n, NR, 36.8%	n, NR, 15.9%	n, NR 47.2% [b]	n, NR 52.7 [b]		
			BI-RADS 4th edition (n=379 mammograms) Two radiologists with 5 and 8 years, experience	n, NR, 25.9%	n, NR, 50.9%	n, NR, 19.8%	n, NR, 3.4%	n, NR 76.8% [b]	n, NR 23.2 [b]	V1.5.1 vs BI-RADS 4: κ -0.40 (NR)	Poor agreement
Holland 2016[75]	500 participants; 1000 mammograms	Digital mammography Lorad Selenia, Hologic	Volpara version 1.5.0	NR	NR	NR	NR	NR	NR	*4-way density* Experienced radiologists: κ 0.73 to 0.78; PhD student: κ 0.77 *Non-dense vs. dense* Experienced radiologists: κ 0.63 to 0.70; PhD student: κ 0.71	Substantial agreement Substantial agreement
			BI-RADS 4th edition Three radiologists (R1, R2 and R3) with ≥8 years of experience in breast imaging and a PhD student (R4) with a medical degree and 2 years'experience	NR	NR	NR	NR	NR	NR		

Rigaud 2022[29]	995 participants	FFDM Selenia Dimensions, Hologic	Volpara version 3.4.1	n, NR, 13%	n, NR, 45%	n, NR, 30%	n, NR, 12%	n, NR 58% [b]	n, NR 42% [b]	κ 0.34 (range 0.17 to 0.48); Average percent agreement: 56% (range 43 to 64%)	Fair agreement
			BI-RADS (edition NR) 7 radiologists with 5-22 years' experience	n, NR, 5-39%	n, NR, 30-44%	n, NR, 21-56%	n, NR, 2-7%	n, NR 5-40% [b]	n, NR 2-56% [b]		
van der Waal 2015[76]	992 mammograms	FFDM Selenia system, Hologic	Volpara version 1.5.0	261 (26.3)	304 (30.6)	305 (30.7)	122 (12.3)	565 (57) [b]	427 (43) [b]	κ 0.80 (0.77, 0.82) *Proportion agreement* 65.4% *Prediction of dense category* AUC: 0.948 (0.935, 0.960) *Accuracy at 8.0% cut-off* Sensitivity: 84%; Specificity 91%	Substantial agreement
			BI-RADS 5th edition 3 experienced radiologists	177 (17.8)	410 (41.3)	294 (29.6)	111 (11.2)	587 (59.2) [b]	405 (40.8) [b]		
Youk 2021[81]	4000 participants	FFDM Lorad Selenia, Hologic	Volpara version 3.1	6 (0.2)	838 (21.0)	1542 (38.6)	1614 (40.4)	844 (21.1) [b]	3156 (78.9) [b]	κ 0.48 (0.46, 0.50)	Moderate agreement
			BI-RADS 5th edition Three radiologists with 7, 10 and 14 years' experience	201 (5.0)	793 (19.8)	2418 (60.5)	588 (14.7)	994 (24.9) [b]	3006 (75.1) [b]		
Eom 2018[27]	1000 participants	FFDM Senographe DS, GE Healthcare	Volpara version 1.5.12	67 (6.7)	247 (24.7)	391 (39.1)	295 (29.5)	314 (31.4)	686 (68.6)	Expert radiologists: κ 0.77 (0.75, 0.80); General radiologists: κ 0.71 (0.68, 0.74) *Non-dense vs. dense* Expert radiologists:	Substantial agreement Substantial agreement Almost perfect agreement
			BI-RADS (edition NR) Two breast-imaging experts with >5 years'experience	67 (6.7)	247 (24.7)	476 (47.6)	210 (21.0)	314 (31.4)	686 (68.6)		
			BI-RADS (edition NR) Two general radiologists with <5 years of experience	25 (2.5)	209 (20.9)	381 (38.1)	385 (38.5)	234 (23.4)	766 (76.6)		

										κ 0.83 (0.80, 0.87) General radiologists: κ 0.73 (0.68, 0.77)	Substantial agreement
Lee 2015[17]	860 participants	FFDM Senographe DS; GE Healthcare	Volpara version 1.5.1	73 (8.5)	255 (29.7)	337 (39.2)	195 (22.7)	328 (38.1) [b]	532 (61.9) [b]	κ 0.80 (0.77, 0.83) ρ 0.86 (NR) P<0.0001	Substantial agreement
			BI-RADS 4th edition One radiologist with 6 years' experience	68 (7.9)	274 (31.9)	307 (35.7)	211 (24.5)	342 (39.8) [b]	518 (60.2) [b]		
Lee 2022[82]	488 participants 488 mammograms	FFDM Senographe Pristina, GE Healthcare	Volpara version 3.4.1	16 (3.3)	131 (26.8)	199 (40.8)	142 (29.1)	147 (30.1) [b]	341 (69.9) [b]	*4-way density:* κ 0.50 (0.45, 0.56) *Non-dense vs. dense:* κ 0.56 (0.48, 0.65)	Moderate agreement Moderate agreement
			BI-RADS 5th edition Three radiologists with 2, 10, and 25 years' experience	3 (0.6)	88 (18)	334 (68.5)	63 (12.9)	91 (18.6) [b]	397 (81.4) [b]		
Portnow 2022[80]	200 mammograms	FFDM Senograph ES and Senograph DS, GE Healthcare	Volpara version 1.5.1	NR	NR	NR	NR	NR	NR	*Non-dense vs dense* V 1.5.1 vs. BI-RADS 4: κ 0.68 to 0.83 V 1.5.2 vs. BI-RADS 5: κ 0.76 to 0.85	Substantial agreement Substantial agreement
			BI-RADS 4th edition Six radiologists with 23-30 years' experience	NR	NR	NR	NR	n, NR, 52%	n, NR, 48%		
			Volpara version 1.5.2	NR	NR	NR	NR	NR	NR		
			BI-RADS 5th edition Six radiologists with 23-30 years' experience	NR	NR	NR	NR	n, NR, 43%	n, NR, 57%		
Sartor 2016[73]	8426 participants; 8426 mammograms	Digital mammography, Mammomat Inspiration, Siemens AG	Volpara version 1.5.11	1763 (20.9)	2708 (32.1)	2650 (31.5)	1305 (15.5)	4471 (53.1) [b]	3955 (46.9) [b]	κ 0.55 (0.53, 0.56)	Moderate agreement
			BI-RADS 4th edition five breast radiologists with >10 years' experience	1378 (16.4)	3445 (40.9)	2967 (35.2)	636 (7.5)	4823 (57.2) [b]	3603 (42.8) [b]		
Singh 2016[28]	476 participants	FFDM	Volpara version 1.4.5	148 (31.1)	131 (27.5)	164 (34.5)	33 (6.9)	279 (58.6) [b]	197 (41.4) [b]	Observer 1: ρ 0.73	Strong agreement

		MicroDose SI, Philips	BI-RADS 4th edition Observer 1 (1/2 radiologists with 5–10 years' experience)	159 (33.4)	212 (44.5)	98 (20.6)	7 (1.5)	371 (77.9) [b]	105 (22.1) [b]	Observer 2: ρ 0.73	Strong agreement
			BI-RADS 4th edition Observer 2 (1/2 radiologists with 5–10 years' experience)	163 (34.2)	220 (46.2)	90 (18.9)	3 (0.6)	383 (80.5) [b]	93 (19.5) [b]		
Alomaim 2020[89]	With distractors: 92 mammograms Without distractors: 158 mammograms	Digital mammography (machine NR)	Volpara version 1.5.0 (n=122 mammograms)	n, NR, 16%	n, NR, 29%	n, NR, 41%	n, NR, 14%	n, NR, 45% [b]	n, NR, 55% [b]	κ 0.66, p<0.001 *With distractors*: κ 0.67 (NR) p<0.001; *Without distractors*: κ 0.52 (NR) p<0.001	Substantial agreement Substantial agreement Moderate agreement
			BI-RADS 4th edition (n=122 mammograms) 25 USA and 24 UK radiologists with >8 years' experience (UK 12% ≤1 years' experience)	n, NR, 10%	n, NR, 31%	n, NR, 31%	n, NR, 28%	n, NR, 41% [b]	n, NR, 59% [b]		
Aloufi 2022[88]	1022 participants	NR FFDM was used in the SNBCSP during the time mammography was conducted during the study (from 2012 to 2018)[92]	Volpara version 1.5.5.1	n, NR, 20.6%	n, NR, 50.5%	n, NR, 24.4%	n, NR, 4.4%	n, NR, 71.1% [b]	n, NR, 28.8% [b]	*4-way density* κ (0.35 (0.29, 0.39) *Non-dense vs.dense* κ 0.53 (0.47, 0.60)	Fair agreement Moderate agreement
			BI-RADS 5th edition 11 radiologists (experience NR)	n, NR, 17.4%	n, NR, 51.1%	n, NR, 28.7%	n, NR, 2.8%	n, NR, 68.5% [b]	n, NR, 31.5% [b]		

a. A, almost entirely fat; B, scattered fibroglanduar densities; C, heterogeneously dense; D, extremely dense; b. values were manually calculated by summing A+B and C+D
Aktiengesellschaft (AG); AUC, area under the curve; FFDM, full field digital mammography; GE, General Electric, κ, Kappa statistic; NR, not reported; ρ; Spearman's rank correlation coefficient; SE, standard error; SNBCSP, Saudi National Breast Cancer Screening Programme; V, Volpara; vs, versus

Table7 Full-length results of the studies evaluating Quantra versus manual measurement of mammographic breast density

Study ID	Sample size used in the analysis	Type of mammography	Automated software version and BI-RADS edition	Density categories, n (%) [a]						Agreement value (95% CI)	Agreement interpretation
				A	B	C	D	Non-dense	Dense		
Ekpo 2016[86]	292 participants (majority report)	NR Selenia Dimensions, Hologic	Quantra version 2.0	Unable to extract [b]	Unable to extract [b]	Unable to extract [b]	Unable to extract [b]	Unable to extract [b]	Unable to extract [b]	Majority report *4-way density* κ 0.79 (0.75, 0.84);	Substantial agreement
			BI-RADS 4th edition 3 RANZCR-certified breast radiologists (majority report – consensus of 2/3)	28 (9.6)	103 (35.3)	79 (27.1)	82 (28.1)	131 (44.9) [c]	161 (55.1) [c]	*Non-dense vs.dense* κ 0.84 (0.79, 0.87); Sensitivity: 91.3%; Specificity 83.6%; AUC: 0.89 (0.82, 0.91)	Almost perfect agreement
Osteras 2016[78]	537 mammograms	FFDM Selenia Dimensions, Hologic	Quantra version 2.0	NR	NR	NR	NR	NR	NR	κ 0.73 (0.67, 0.79)	Substantial agreement
			BI-RADS 4th edition five radiologists with 1 to 34 years' experience	73 (13.6)	220 (41.0)	188 (35.0)	56 (10.4)	293 (54.6) [c]	244 (45.4) [c]		
van der Waal 2015[76]	992 mammograms	FFDM Selenia system, Hologic	Quantra version 1.3	NR	NR	NR	NR	NR	NR	*Prediction of dense category* AUC: 0.98 (0.94, 0.96)	Excellent agreement

			BI-RADS 5th edition 3 experienced radiologists	177 (17.8)	410 (41.3)	294 (29.6)	111 (11.2)	587 (59.2) [c]	405 (40.8) [c]	Accuracy at 13.8% cut-off: Sensitivity: 82%; Specificity: 92%)	
Youk 2021[81]	4000 participants	FFDM Lorad Selenia, Hologic	Quantra version 2.1.1	216 (5.4)	1622 (40.6)	1725 (43.1)	437 (10.9)	1838 (46) [c]	2162 (54) [c]	κ 0.54 (0.52, 0.56)	Moderate agreement
			BI-RADS 5th edition Three radiologists with 7, 10 and 14 years' experience	201 (5.0)	793 (19.8)	2418 (60.5)	588 (14.7)	994 (24.9) [c]	3006 (75.1) [c]		

a. A, almost entirely fat; B, scattered fibroglanduar densities; C, heterogeneously dense; D, extremely dense; b. data were reported in graph format. Data extraction of graph data using Webplot digitizer version 4 was attempted but was unreliable and data are, therefore, not presented; c. values were manually calculated by summing A+B and C+D
AUC, area under the curve; FFDM, full field digital mammography; κ, Kappa statistic; NR, not reported; ρ; Spearman's rank correlation coefficient; RANZCR, Royal Australian and New Zealand College of Radiology; SE, standard error; vs, versus

Table 8 Full-length results of the studies evaluating individual automated software versus manual measurement of mammographic breast density

Study ID	Sample size used in the analysis	Type of mammography	Automated software version and BI-RADS edition	Density categories, n (%) [a]						Agreement value (95% CI)	Agreement interpretation
				A	B	C	D	Non-dense	Dense		
Le Boulc'h 2020[77]	311 participants	FFDM, Selenia Dimensions, Hologic	DenSeeMammo®	25 (8)	114 (37)	151 (49)	21 (7)	139 (44.7)	172 (55.3)	DenSee Mammo vs. BIRADS 5, senior radiologist κ 0.79 (0.74, 0.84)	Substantial agreement
			BI-RADS 5th edition, junior radiologist with 1 year's experience	33 (11)	104 (33)	159 (51)	15 (5)	137 (44.1)	174 (55.9)	DenSee Mammo vs. BI-RADS 5, junior radiologist κ 0.76 (0.71, 0.82)	Substantial agreement
			BI-RADS 5th edition, senior radiologist with 5 years' experience	28 (9)	122 (39)	151 (49)	10 (3)	150 (48.2)	161 (51.8)		
Lee 2018[79]	182 mammograms	FFDM, Model NR, Hologic	Deep Learning algorithm	NR	NR	NR	NR	NR	NR	DL vs BIRADS 4 (averaged CC-MLO views) r 0.85 (0.80, 0.89)	Highly correlated
			LIBRA	NR	NR	NR	NR	NR	NR		
			BI-RADS 4th edition Number and experience of the radiologists NR	30 (16.5)	44 (24.2)	56 (30.8)	52 (28.5)	74 (40.7)	108 (59.3)	LIBRA vs BIRADS 4 (averaged CC-MLO views) r 0.69 (0.55, 0.77)	Moderately correlated
Rigaud 2022[29]	995 participants		EfficientNetB0 deep learning	NR	NR	NR	NR	NR	NR		

		FFDM, Selenia Dimensions, Hologic	BI-RADS (edition NR) 7 radiologists with 5-22 years' experience	n, NR, 5-39%	n, NR, 30-44%	n, NR, 21-56%	n, NR, 2-7%	n, NR, 5-44%	n, NR, 2-56%	<u>For processing images</u> *4-way density* Without clinical history κ 0.61; With clinical history κ 0.61 *Non-dense vs. dense* Without clinical history κ 0.70; With clinical history κ 0.71 <u>For presentation images</u> *4-way density* Without clinical history κ 0.66; With clinical history κ 0.66 *Non-dense vs. dense* Without clinical history κ 0.75; With clinical history κ 0.75	Substantial agreement Substantial agreement Substantial agreement Substantial agreement
Fonseca 2015[84]	1157 participants	Digital mammography, Selenia	HT-L3 convolutional network with a support vector machines classifier	NR	NR	NR	NR	NR	NR	κ 0.58	Moderate agreement

		Dimensions, Hologic	BI-RADS 3rd edition seven radiologists 5 to 25 years' experience	NR	NR	NR	NR	NR	NR	Average accuracy 73.05%	
Anguloa 2015[83]	1050 [b] mammograms	Digital mammography, Selenia Dimensions, Hologic (16% mammograms); Mammomat 3000, Siemens Medical (84% mammograms)	Principal component analysis	NR	NR	NR	NR	NR	NR	κ 0.44	Moderate agreement
			BI-RADS (edition NR) Eight expert radiologists with varying experience (mean 8.9 years (SD 6.2 years) experience)	263 (25)	530 (50.5)	198 (18.9)	59 (5.6)	793 (75.5)	257 (24.5)		
Fieslemann 2019[71]	600 mammograms	FFDM, Mammomat Inspiration, Siemens Healthcare	Insight BD	132 (22)	231 (38.5)	180 (30)	57 (9.5)	363 (60.5)	237 (39.5)	4- way density κ 0.67 Percent agreement 69.5% Non-dense vs. dense κ 0.76 Percent agreement 88.5%	Substantial agreement Substantial agreement
			BI-RADS 5th edition 32 experienced radiologists	97 (16.2)	271 (45.2)	188 (31.3)	44 (7.3)	368 (61.3)	232 (38.7)		
Fornvik 2019[72]	9909 participants	Digital mammography Mammomat Inspiration, Siemens Healthcare	Automated software (n=9204)	1832 (18.5)	3385 (34.2)	3080 (31.1)	1612 (16.3)	5217 (52.7)	4692 (47.3)	κ 0.55 Percent agreement 57.1%	Moderate agreement
			BI-RADS 4th edition Five radiologists with >10 years' experience (n=	1475 (14.9)	3880 (39.2)	3647 (36.8)	907 (9.2)	5355 (54)	4554 (46)		
Timberg 2016[74]	348 participants	FFDM, Mammomat Inspiration, Siemens Healthcare	Automated volumetric breast density analysis	NR	NR	NR	NR	NR	NR	κ 0.73 Percent agreement 70%	Substantial agreement
			BI-RADS 5th edition Four radiologists	NR	NR	NR	NR	NR	NR		

Lee 2022[82]	488 participants 488 mammograms	FFDM, Senographe Pristina, GE Healthcare	AI-CAD mean, Lunit INSIGHT MMG, version 1.1.4.3	22 (4.5)	218 (44.7)	222 (45.4)	26 (5.3)	240 (49.2)	248 (50.8)	AI-CAD mean *4-way density* κ 0.41 (0.36, 0.47) p=0.002 *Non-dense vs. dense* κ 0.38 (0.32, 0.45) p<0.001 AI-CAD max *4-way density* κ 0.52 (0.46, 0.58) p=0.67 *Non-dense vs. dense* κ 0.51 (0.43, 0.58) p=0.037	Moderate agreement Fair agreement Moderate agreement Moderate agreement
			AI-CAD max, Lunit INSIGHT MMG, version 1.1.4.3	13 (2.7)	184 (37.7)	247 (50.6)	44 (9)	197 (40.4)	291 (59.6)		
			BI-RADS 5th edition Three radiologists with 2-, 10-, and 25-years' experience	3 (0.6)	88 (18)	334 (68.5)	63 (12.9)	91 (18.6)	397 (81.4)		
Pavan 2017[87]	30 mammograms	Digital mammography, Senographe 600T, GE Healthcare	Optimizing fuzzy C-means with variable compactness (FCMVC)	NR	NR	NR	NR	NR	NR	*Non-dense vs. dense* FCMVC skewness [c] vs. BI-RADS 4 κ 0.65 FCMVC FTP [d] vs. BI-RADS 4 κ 0.47 ρ 0.62, p <0.001	Substantial agreement Moderate agreement Strongly correlated
			BI-RADS 4th edition One experienced radiologist	4 (13.3)	13 (43.3)	8 (26.7)	5 (16.7)	17 (56.7)	13 (43.3)		
Alomaim 2020[89]	250 mammograms	Digital mammography, machine NR	Cumulus Hand Delineation (HD)	NR	NR	NR	NR	NR	NR	Cumulus HD vs. BI-RADS 4 κ 0.78, p <0.001 Image J vs. BI-RADS 4	Substantial agreement
			ImageJ	NR	NR	NR	NR	NR	NR		
			BI-RADS 4th edition (n=122 mammograms)	Image set A: 14%; Image	Image set A 36%; Image	Image set A 36%; Image	Image set A 14%; Image	Image set A: 50%; Image	Image set A: 50%; Image		

			25 USA and 24 UK radiologists with >8 years' experience (UK 12% ≤1 years' experience)	set B 25%; Image set C 22%; Image set D 17%; Image set E 22%	set B 17%; Image set C 22%; Image set D 28%; Image set E 25%	set B 39%; Image set C 39% Image set D 39%; Image set E 36%	set B 19%; Image set C 17%; Image set D 17%; Image set E 17%	set B 42%; Image set C 44%; Image set D 45%; Image set E 47%	set B 58%; Image set C 56%; Image set D 56%; Image set E 53%	κ 0.86, p <0.001	Almost perfect agreement
Lehman 2019[33]	10763 mammograms	Digital mammography, machine NR	Deep learning model	646 (6)	5597 (52)	4305 (40)	215 (2)	6243	4520	κ 0.85 (0.84, 0.86)	Almost perfect agreement
			BI-RADS 5th edition Eight breast imagers with 2–23 years' experience	NR	NR	NR	NR	6031 [e]	4732 [e]	*4-way density* Percent agreement 90% (90, 91) *Non-dense vs. dense* Percent agreement 94% (94, 95)	
Pesce 2020[85]	451 mammograms	Digital mammography, machine NR	AMULET Innovality	NR	NR	NR	NR	NR	NR	*4-way density* κ 0.46 (0.39, 0.52)	Moderate agreement
			BI-RADS 5th edition Six physicians with an average of nine years (range 2 to 18) years' experience	NR	NR	NR	NR	NR	NR	*Non-dense vs. dense* κ 0.51 (0.43, 0.59)	Moderate agreement
Ref 1501 Kaiser 2019[56]	600 mammograms	2D mammography, machine NR	Convolutional neural network (CNN)	NR	NR	NR	NR	NR	NR	κ 0.75 Percent agreement 88%	Substantial agreement
			BI-RADS 5th edition 32 radiologists (experience NR)	NR	NR	NR	NR	NR	NR		

a. A, almost entirely fat; B, scattered fibroglanduar densities; C, heterogeneously dense; D, extremely dense; b. Table 2 in the Anguloa 2015[83] paper indicates n=1057 mammograms but the BIRADS data sum to 1050 mammograms; c. The authors based skewness values from an equalized histogram image. "Fat" group has greater skewness values than "dense" group; d. The authors assigned a different weight to clusters to calculate the percentage of fibroglandular tissue. The algorithm estimated the fibroglandular tissue percentage (FTP) in mammograms. FTP was computed by dividing the fibroglandular tissue area by total breast area; e. values were manually calculated by summing A+B and C+D

CC; craniocaudal; CNN, convolutional neural network; DL, deep learning; FCMVC, Optimizing fuzzy C-means with variable compactness; FFDM, full field digital mammography; FTP, fibroglandular tissue percentage; GE, General Electric; κ, Kappa statistic; MLO, mediolateral oblique; NR, not reported; ρ; Spearman's rank correlation coefficient; SD, standard deviation; SE, standard error; vs, versus

Appendix 5 – UK NSC reporting checklist for evidence summaries

All items on the UK NSC Reporting Checklist for Evidence Summaries have been addressed in this report. A summary of the checklist, along with the page or pages where each item can be found in this report, is presented in Table 9.

Table 9 – UK NSC reporting checklist for evidence summaries

Section	Item	Page no.
Title and summaries		
Title Sheet	Identify the review as a UK NSC Evidence summary	1
Plain English summary	Plain English description of the executive summary.	5
Executive summary	Structured overview of the whole report. To include: the purpose/aim of the review; background; previous recommendations; findings and gaps in the evidence; recommendations on the screening that can or cannot be made on the basis of the review	7
Introduction and Approach		

Section	Item	Page no.
Background and objectives	Background – Current policy context and rationale for the current review – for example, reference to details of previous reviews, basis for current recommendation, recommendations made, gaps identified, drivers for new reviews Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary, criteria they address, and number of studies included per question, description of the overall results of the literature search. Method – briefly outline the rapid review methods used.	11-13
Eligibility for inclusion in the review	State all criteria for inclusion and exclusion of studies to the review clearly(PICO, dates, language, study type, publication type, publication status etc.) To be decided a priori	15-18
Appraisal for quality/ risk of bias tool	Details of tool/ checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	19
Search strategy and study selection		
Databases/ sources searched	Give details of all databases searched (including platform/ interface and coverage dates) and date of final search.	15-16

Section	Item	Page no.
Search strategy and results	Present the full search strategy for at least one database(usually a version of Medline), including limits and search filters if used. Provide details of the total number of (results from each database searched), number of duplicates removed, and the final number of unique records to consider for inclusion.	36-38, 39
Study selection	State the process for selecting studies – inclusion and exclusion criteria, number of studies screened by title/abstract and full text, number of reviewers, any cross checking carried out.	16-17
Study level reporting of results (for each key question)		
Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.). Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available. For each study, present the results of any assessment of quality/risk of bias.	26-29, 31, 59-68, 69-70, 71-82
Additional analyses	Describe additional analyses (for example, sensitivity, specificity, PPV, etc.) carried out by the reviewer. [Remove if not performed]	19, 24-25
Question level synthesis		
Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and inclusion in the review, with summary reasons for exclusion	21, 38, 42-58

Section	Item	Page no.
Combining and presenting the findings	Provide a balanced discussion of the body of evidence which avoids over reliance on one study or set of studies. Consideration of four compartments should inform the reviewer's judgement on whether the criterion is "met", "not met" or "uncertain": quantity; quality; applicability and consistency.	33-35
Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion. Summarise the main findings including the quality/ risk of bias issues for each question. Have the criteria addressed been "met", "not met" or "uncertain"?	33-35
Review Summary		
Conclusions and implications for policy	Do findings indicate whether screening should be recommended? IS further work warranted? Are there gaps in the evidence highlighted by the review?	33-35
Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	34-35

References

1. Cancer Research UK. Cancer statistics for the UK. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk (Accessed January 2024).

2. Duffy SW, Tabár L, Yen AM, et al. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. Cancer. 2020;**126**(13):2971-9.

3. Lauby-Secretan B, Loomis D, Straif K. Breast cancer screening: viewpoint of the IARC Working Group. N Engl J Med. 2015;**373**(15):1479.

4. Njor S, Nyström L, Moss S, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen. 2012;**19 Suppl 1**:33-41.

5. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;**380**(9855):1778-86.

6. Edmonds CE, O'Brien SR, Conant EF. Mammographic breast density: current assessment methods, clinical implications, and future directions. Semin Ultrasound CT MR. 2023;**44**(1):35-45.

7. Patterson J, Stinton C, Alkhudairy L, et al. UK NSC Evidence Review. Additional Breast Screening with Ultrasound in Women with Dense Breasts Final 27 02 2019 2019. 2019. Available from: https://view-health-screening-recommendations.service.gov.uk/document/465/download. (Accessed January 2024)

8. American College of Radiology. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System 2013. 2013. Available from: https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads (Accessed January 2024).

9. Sickles E, D'Orsi CJ, Bassett L, Morris E. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. 5th ed. Reston, VA: American College of Radiology; 2013.

10. Sprague BL, Kerlikowske K, Bowles EJA, et al. Trends in Clinical Breast Density Assessment from the Breast Cancer Surveillance Consortium. J Natl Cancer Inst. 2019;**111**(6):629-32.

11. Weigel S, Heindel W, Heidrich J, Hense HW, Heidinger O. Digital mammography screening: sensitivity of the programme dependent on breast density. Eur Radiol. 2017;**27**(7):2744-51.

12. Radhakrishna S, Agarwal S, Parikh PM, et al. Role of magnetic resonance imaging in breast cancer management. South Asian J Cancer. 2018;**7**(2):69-71.

13. Clinical Trials.Gov. Breast Screening - Risk Adaptive Imaging for Density (BRAID) Protocol ID NCT04097366. 2019. Available from: https://clinicaltrials.gov/study/NCT04097366?term=BRAID&checkSpell=false&rank=2#study-record-dates (Accessed January 2024).

14. Higgins J, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023) 2024 December 2024]. Available from: www.training.cochrane.org/handbook.)

15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;**372**:n71.

16. Gweon HM, Youk JH, Kim JA, Son EJ. Radiologist assessment of breast density by BI-RADS categories versus fully automated volumetric assessment. AJR Am J Roentgenol. 2013;**201**(3):692-7.

17. Lee HN, Sohn YM, Han KH. Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: analysis of clinical-radiologic factors affecting discrepancy between them. Acta Radiol. 2015;**56**(9):1061-8.

18. Patterson J, Stinton C, Alkhudairy L, et al. Additional screening with ultrasound after negative mammography screening in women with dense breasts: a systematic review. London: National Screening Committee; 2019.

19. American College of Radiology. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. 3rd ed. Reston, VA: American College of Radiology; 1998.

20. D'Orsi C, Bassett L, Berg W. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. 4th ed. Reston, VA: American College of Radiology; 2003.

21. Trial Forge. PRO EDI participant characteristics table 22/3/2024. 2024. Available from: https://www.trialforge.org/trial-diversity/pro-edi-improving-how-equity-diversity-and-inclusion-is-handled-in-evidence-synthesis/ (Accessed October 2024).

22. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking systematic reviews in health care. University of York. 2009. Available

from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. (Accessed March 2024)

23. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol. 1998;**51**(12):1235-41.

24. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;**52**(6):377-84.

25. Jackson R, Ameratunga S, Broad J, et al. The GATE frame: critical appraisal with pictures. Evid Based Med. 2006;**11**(2):35-8.

26. Gemici AA, Aribal E, Ozaydin AN, et al. Comparison of qualitative and volumetric assessments of breast density and analyses of breast compression parameters and breast volume of women in Bahcesehir mammography screening project. Meme Sagligi Dergisi / Journal of Breast Health. 2020;**16**(2):110-6.

27. Eom HJ, Cha JH, Kang JW, Choi WJ, Kim HJ, Go E. Comparison of variability in breast density assessment by BI-RADS category according to the level of experience. Acta Radiol. 2018;**59**(5):527-32.

28. Singh T, Sharma M, Singla V, Khandelwal N. Breast density estimation with fully automated volumetric nethod: comparison to radiologistsa Assessment by BI-RADS categories. Acad Radiol. 2016;**23**(1):78-83.

29. Rigaud B, Weaver OO, Dennison JB, et al. Deep learning models for automated assessment of breast density using multiple mammographic image types. Cancers (Basel). 2022;**14**(20).

30. Balleyguier C, Arfi-Rouche J, Boyer B, et al. A new automated method to evaluate 2D mammographic breast density according to BI-RADS Atlas Fifth Edition recommendations. Eur Radiol. 2019;**29**(7):3830-8.

31. Carneiro PC, Franco MLN, De Lima Thomaz R, Patrocinio AC. Breast density pattern characterization by histogram features and texture descriptors. Revista Brasileira de Engenharia Biomedica. 2017;**33**(1):69-77.

32. Couwenberg AM, Verkooijen HM, Li J, et al. Assessment of a fully automated, high-throughput mammographic density measurement tool for use with processed digital mammograms. Cancer Causes Control. 2014;**25**(8):1037-43.

33. Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: clinical implementation. Radiology. 2019;**290**(1):52-8.

34. Lin X, Wu S, Li L, et al. Automatic mammographic breast density classification in Chinese women: clinical validation of a deep learning model. Acta Radiol. 2023;**64**(5):1823-30.

35. Matthews TP, Singh S, Mombourquette B, et al. A multisite study of a breast density deep learning model for full-field digital mammography and synthetic mammography. Radiol Artif Intell. 2021;**3**(1):e200015.

36. Li S, Wei J, Chan HP, et al. Computer-aided assessment of breast density: comparison of supervised deep learning and feature-based statistical learning. Phys Med Biol. 2018;**63**(2):025005.

37. Wei J, Chan HP, Helvie MA, Roubidoux MA, Zhou C, Hadjiiski L. Radiomic modeling of BI-RADS density categories. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2017;**10134**:101340P.

38. Kai C, Ishizuka S, Otsuka T, et al. Automated estimation of mammary gland content ratio using regression deep convolutional neural network and the effectiveness in clinical practice as explainable artificial intelligence. Cancers (Basel). 2023;**15**(10).

39. Li C, Xu J, Liu Q, et al. Multi-view mammographic density classification by dilated and attention-guided residual learning. IEEE/ACM Trans Comput Biol Bioinform. 2021;**18**(3):1003-13.

40. Khan N, Wang K, Chan A, Highnam R. Automatic BI-RADS classification of mammograms. Lecture Notes in Computer Science. 2016;**9431**:475-87.

41. Watanabe AT, Retson T, Wang J, Mantey R, Chim C, Karimabadi H. Mammographic breast density model using semi-supervised learning reduces inter-/intra-reader variability. Diagnostics. 2023;**13**(16):2694.

42. Yamamuro M, Asai Y, Hashimoto N, et al. Utility of U-Net for the objective segmentation of the fibroglandular tissue region on clinical digital mammograms. Biomed Phys Eng Express. 2022;**8**(4):045016.

43. Youk JH, Gweon HM, Son EJ, Kim JA. Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: A comparison with visual assessment. AJR Am J Roentgenol. 2016;**206**(5):1056-62.

44. Singh N, Joshi P, Singh DK, Narayan S, Gupta A. Volumetric breast density evaluation using fully automated Volpara software, its comparison with BIRADS density types and correlation with the risk of malignancy. Egyptian Journal of Radiology and Nuclear Medicine. 2022;**53**(1).

45. Vállez N, Bueno G, Déniz O, et al. Breast density classification to reduce false positives in CADe systems. Comput Methods Programs Biomed. 2014;**113**(2):569-84.

46. Ji Y, Shao Z, Liu J, Hao Y, Liu P. The correlation between mammographic densities and molecular pathology in breast cancer. Cancer Biomark. 2018;**22**(3):523-31.

47. Nykanen A, Okuma H, Sutela A, Masarwah A, Vanninen R, Sudah M. The mammographic breast density distribution of Finnish women with breast cancer and comparison of breast density reporting using the 4th and 5th editions of the Breast Imaging-Reporting and Data System. Eur J Radiol. 2021;**137**:109585.

48. Dontchos BN, Yala A, Barzilay R, Xiang J, Lehman CD. External Validation of a Deep Learning Model for Predicting Mammographic Breast Density in Routine Clinical Practice. Acad Radiol. 2021;**28**(4):475-80.

49. Oshima A, Shinohara N, Kamiya N. Investigation of the effect of image resolution on automatic classification of mammary gland density in mammography images using deep learning. Proceedings of SPIE - The International Society for Optical Engineering. 2019;**11050**:1105018.

50. Ma X, Fisher C, Wei J, et al. Multi-path deep learning model for automated mammographic density categorization. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2019;**10950**:109502E.

51. Chen B, Ruth C, Zhang Y, Jing Z. Breast density assessment: image feature extraction and density classification with machine intelligence. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2018;**10573**:105735F.

52. Ko SY, Kim EK, Kim MJ, Moon HJ. Mammographic density estimation with automated volumetric breast density measurement. Korean J Radiol. 2014;**15**(3):313-21.

53. Testagrose C, Gupta V, Erdal BS, et al., Impact of concatenation of digital craniocaudal mammography images on a deep-learning breast-density classifier using Inception-V3 and ViT. IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2022, Las Vegas, NV; 2022.

54. Tsuchida T, Negishi T, Takahashi Y, Nishimura R. Dense-breast classification using image similarity. Radiol Phys Technol. 2020;**13**(2):177-86.

55. Ekpo EU, Hogg P, Wasike E, McEntee MF. A self-directed learning intervention for radiographers rating mammographic breast density. Radiography. 2017;**23**(4):337-42.

56. Kaiser N, Fieselmann A, Vesal S, et al. Mammographic breast density classification using a deep neural network: assessment based on inter-observer variability. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2019;**10952**:109520O.

57. Osteras BH, Martinsen ACT, Brandal SHB, et al. BI-RADS density classification from areometric and volumetric automatic breast density measurements. Acad Radiol. 2016;**23**(4):468-78.

58. Pawar SD, Sharma KK, Sapate SG, et al. Multichannel DenseNet architecture for classification of mammographic breast density for breast cancer detection. Front Public Health. 2022;**10**:885212.

59. Chugh S, Goyal S, Pandey A, Joshi S. Morphological and Otsu's technique based mammography mass detection and deep neural network classifier based prediction. Traitement Du Signal. 2022;**39**(4):1283-94.

60. Bandeira Diniz JO, Bandeira Diniz PH, Azevedo Valente TL, Corrêa Silva A, de Paiva AC, Gattass M. Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Methods Programs Biomed. 2018;**156**:191-207.

61. Gamdonkar Z, Tay K, Ryder W, Brennan PC, Mello-Thoms C. IDensity: An automatic Gabor filter-based algorithm for breast density assessment. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2015;**9416**:941607.

62. Kaliyaperumal V, Selvarajan S. Automated characterization of mammographic density for early detection of breast cancer risk. International Journal of Simulation: Systems, Science and Technology. 2014;**15**(1):56-63.

63. Yi PH, Lin A, Wei J, et al. Deep-learning-based semantic labeling for 2D mammography and comparison of complexity for machine learning tasks. J Digit Imaging. 2019;**32**(4):565-70.

64. Tiryaki VM, Kaplanoğlu V. Deep learning-based multi-label tissue segmentation and density assessment from mammograms. IRBM. 2022;**43**(6):538-48.

65. Trivizakis E, Ioannidis GS, Melissianos VD, et al. A novel deep learning architecture outperforming ‘off-the-shelf’ transfer learning and feature-based methods in the automated assessment of mammographic breast density. Oncol Rep. 2019;**42**(5):2009-15.

66. Simon S, Lavanya R, Vijayan D, PSO based density classifier for mammograms. IFMBE Proceedings; 2017.

67. Sharma V, Singh S. CFS-SMO based classification of breast density using multiple texture models. Med Biol Eng Comput. 2014;**52**(6):521-9.

68. Lee ZY, Goh YLE, Lai C. Classification of mammographic breast density and its correlation with BI-RADS in elder women using machine learning approach. J Med Imaging Radiat Sci. 2022;**53**(1):28-34.

69. Heath M, Bowyer K, Kopans DB, Moore RH, Kegelmeyer P. The Digital Database for Screening Mammography. In: Yaffe M, editor. Proceedings of the Fifth International Workshop on Digital Mammography. Madison, WI: Medical Physics Publishing; 2001.

70. Suckling J. The mammographic images analysis society digital mammogram database. Excerpta Medica International Congress Series. 1994;**1069**:375-8.

71. Fieselmann A, Fornvik D, Fornvik H, et al. Volumetric breast density measurement for personalized screening: accuracy, reproducibility, consistency, and agreement with visual assessment. J Med Imaging (Bellingham). 2019;**6**(3):031406.

72. Fornvik D, Fornvik H, Fieselmann A, Lang K, Sartor H. Comparison between software volumetric breast density estimates in breast tomosynthesis and digital mammography images in a large public screening cohort. Eur Radiol. 2019;**29**(1):330-6.

73. Sartor H, Lang K, Rosso A, Borgquist S, Zackrisson S, Timberg P. Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification. Eur Radiol. 2016;**26**(12):4354-60.

74. Timberg P, Fieselmann A, Dustler M, et al., Breast density assessment using breast tomosynthesis images. Lecture Notes in Computer Science; 2016.

75. Holland K, van Zelst J, den Heeten GJ, et al. Consistency of breast density categories in serial screening mammograms: A comparison between automated and human assessment. Breast. 2016;**29**:49-54.

76. Van Der Waal D, Den Heeten GJ, Pijnappel RM, et al. Comparing visually assessed BI-RADS breast density and automated volumetric breast density software: A cross-sectional study in a breast cancer screening setting. PLoS One. 2015;**10**(9):e0136667.

77. Le Boulc'h M, Bekhouche A, Kermarrec E, et al. Comparison of breast density assessment between human eye and automated software on digital and synthetic mammography: Impact on breast cancer risk. Diagn Interv Imaging. 2020;**101**(12):811-9.

78. Osteras BH, Martinsen ACT, Brandal SHB, et al. Classification of fatty and dense breast parenchyma: Comparison of automatic volumetric density measurement and radiologists' classification and their inter-observer variation. Acta Radiol. 2016;**57**(10):1178-85.

79. Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. Med Phys. 2018;**45**(3):1178-90.

80. Portnow LH, Georgian-Smith D, Haider I, et al. Persistent inter-observer variability of breast density assessment using BI-RADS 5th edition guidelines. Clin Imaging. 2022;**83**:21-7.

81. Youk JH, Gweon HM, Son EJ, Eun NL, Kim JA. Fully automated measurements of volumetric breast density adapted for BIRADS 5th edition: a comparison with visual assessment. Acta Radiol. 2021;**62**(9):1148-54.

82. Lee SE, Son NH, Kim MH, Kim EK. Mammographic density assessment by artificial intelligence-based computer-assisted diagnosis: a comparison with automated volumetric assessment. J Digit Imaging. 2022;**35**(2):173-9.

83. Angulo A, Ferrer J, Pinto J, Lavarello R, Guerrero J, Castaneda B. Experimental assessment of an automatic breast density classification algorithm based on principal component analysis applied to histogram data. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2015;**9287**:92870E.

84. Fonseca P, Mendoza J, Wainer J, et al. Automatic breast density classification using a convolutional neural network architecture search procedure. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2015;**9414**:941428.

85. Pesce K, Tajerian M, Chico MJ, et al. Interobserver and intraobserver variability in determining breast density according to the fifth edition of the BI-RADS Atlas. Radiologia (Roma). 2020;**62**(6):481-6.

86. Ekpo EU, McEntee MF, Rickard M, et al. Quantra™ should be considered a tool for two-grade scale mammographic breast density classification. Br J Radiol. 2016;**89**(1060).

87. Pavan ALM, Vacavant A, Trindade AP, de Pina DR. Fibroglandular tissue quantification in mammography by optimized fuzzy C-Means with variable compactness. IRBM. 2017;**38**(4):228-33.

88. Aloufi AS, AlNaeem A, Almousa A, et al. Breast density distribution among the Saudi screening population and correlation between radiologist visual assessment and two automated methods. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. 2022;**12035**:120350L.

89. Alomaim W, O'Leary D, Ryan J, Rainford L, Evanoff M, Foley S. Subjective versus quantitative methods of assessing breast density. Diagnostics. 2020;**10**(5):331.

90. Zackrisson S, Lang K, Rosso A, et al. One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol. 2018;**19**(11):1493-503.

91. Ali KAS, Fateh SM. Mammographic breast density status in women aged more than 40 years in Sulaimaniyah, Iraq: a cross-sectional study. J Int Med Res. 2022;**50**(12).

92. Aloufi AS, Alnaeem AN, Almousa AS, et al. Mammographic breast density and breast cancer risk in the Saudi population: a case-control study using visual and automated methods. Br J Radiol. 2022;**95**(1134):A16.

93. Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. Insights Imaging. 2024;**15**(1):16.

94. Mongan J, Kohli M. Artificial intelligence and human life: five lessons for radiology from the 737 MAX disasters. Radiol Artif Intell. 2020;**2**(2):e190111.

95. Uematsu T. Rethinking screening mammography in Japan: next-generation breast cancer screening through breast awareness and supplemental ultrasonography. Breast cancer (Tokyo, Japan). 2024;**31**(1):24-30.

96. Mizukoshi MM, Hossain SZ, Poulos A. Mammographic breast density of Japanese women living in Australia: implications for breast screening policy. Asian Pac J Cancer Prev. 2019;**20**(9):2811-7.

97. Kerlikowske K, Bissell MCS, Sprague BL, et al. Impact of BMI on prevalence of dense breasts by race and ethnicity. Cancer Epidemiol Biomarkers Prev. 2023;**32**(11):1524-30.

98. del Carmen MG, Halpern EF, Kopans DB, et al. Mammographic breast density and race. AJR Am J Roentgenol. 2007;**188**(4):1147-50.

99. Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis - a systematic review. Br J Radiol. 2014;**87**(1043):20140460.