AI-Assisted Point of Care Ultrasound (POCUS) Vs. Mammography for Early Breast Cancer Detection: A Comparative Review

Majd Oteibi, Adam Tamimi, Kaneez Abbas, Gabriel Tamimi, Danesh Khazaei, Behrooz Khajehee, Hadi Khazaei

Clinical Research Department, Validus Institute Inc., Fallbrook, CA, United States of America (USA)

DOI: https://dx.doi.org/10.51584/IJRIAS.2025.10100000125

Received: 18 October 2025; Accepted: 24 October 2025; Published: 14 November 2025

ABSTRACT

Background: Mammography (MG) is the current standard for population screening and mortality reduction, but sensitivity declines in dense breasts, and access can be limited in low-resource settings. **Objective:** Identify and synthesize published comparisons of AI-assisted ultrasound evaluation, particularly handheld/Point of Care Ultrasound (POCUS) against standard mammography strategies for early detection; summarize outcomes (sensitivity, specificity, cancer detection rate [CDR], interval cancers, recall/biopsy rates), and outline where artificial intelligence (AI) + Point of care ultrasound (POCUS) may be superior.

Findings: Randomized and cohort data show Mammography + ultrasound detects more cancers and halves interval cancers versus Mammography alone (trade-off: lower specificity). Emerging AI-assisted POCUS demonstrates very high sensitivity for palpable masses on portable devices and can safely triage 38–67% of benign cases away from referral imaging. This is based on published articles and a meta-analysis review. In dense breasts, mammography-supplemental US outperforms MG+AI on several diagnostic endpoints. Nationwide real-world programs show MG+AI increases CDR over MG alone, according to a 2025 published article reported by Eisemann et al. (2025) in *Nature Medicine*

Conclusion: Direct RCTs of AI-POCUS vs Mammography for screening are not yet published; however, across published comparative articles, ultrasound-based strategies and especially AI-assisted POCUS triage are clinically advantageous in certain medical cases, for example, dense breasts, palpable masses, low-resource settings, and are likely to be non-inferior, and sometimes superior to MG-only strategies for early detection, albeit with a specificity trade-off that AI may reduce (Ohuchi et al., 2016).

Keywords: Comparison study, Mammography, Breast ultrasound screening, AI-assisted POCUS technology, AI-assisted Mammography

INTRODUCTION

Breast cancer remains the most frequently diagnosed malignancy and the leading cause of cancer-related mortality among women worldwide. According to the World Health Organization, over 2.3 million new cases and 685,000 deaths were reported globally in 2023, emphasizing the continued need for improved early detection strategies. Early identification of breast malignancies significantly improves survival rates, with five-year survival exceeding 90% in high-income settings where screening programs are well established. However, disparities in access to advanced imaging technologies and trained radiologists persist across low- and middle-income regions, underscoring the need for portable, cost-effective, and accurate diagnostic alternatives.

Mammography remains the gold standard for population-based breast cancer screening. Its proven efficacy in detecting microcalcifications and early-stage carcinomas has led to widespread adoption in national screening programs. Nevertheless, mammography has several limitations. Diagnostic performance declines markedly in women with dense breast tissue, where sensitivity can fall below 70%. Moreover, exposure to ionizing radiation, high equipment costs, and limited accessibility in remote or resource-limited settings restrict its universal applicability. Furthermore, mammographic interpretation is subject to reader variability, and false positives contribute to unnecessary biopsies and patient anxiety.

Historically, ultrasound has been limited by operator dependence and variable standards, but AI is reshaping its performance. AI-assisted POCUS using deep learning, especially convolutional neural networks (CNNs), can automatically detect, segment, and classify breast lesions with accuracies approaching expert radiologists. By guiding acquisition and interpretation. These tools let non expert clinicians take images and widens access to high-quality diagnostics.

Comparing AI-assisted POCUS with mammography in the context of early breast cancer detection presents an opportunity to redefine current screening paradigms. While mammography offers high specificity and well-established clinical guidelines, AI-driven ultrasound promises superior adaptability, cost-efficiency, and accessibility. Recent studies indicate that AI models trained on ultrasound datasets can detect malignant features with sensitivities exceeding 90%, rivaling mammographic performance in early detection scenarios. Moreover, the portability of AI-enabled devices positions them as valuable tools for point-of-care triage, follow-up, and outreach screening in underserved populations.

Despite these advances, direct comparative analyses between AI-assisted POCUS and mammography remain limited. Most studies focus on the standalone performance of AI-enhanced ultrasound or on mammographic AI computer-aided detection (CAD) systems, without systematically contrasting their clinical utility, workflow integration, or patient outcomes. This review aims to synthesize current evidence and critically compare AI-assisted point-of-care ultrasound and mammography across diagnostic accuracy, accessibility, cost-effectiveness, and clinical feasibility. By elucidating their relative strengths and limitations, this work seeks to inform the development of optimized, AI-driven breast cancer screening frameworks that balance precision, scalability, and equity in global health contexts.

Screening mammography (MG) lowers breast cancer mortality and remains first-line for women 40–74 years. Yet evidence is insufficient (per USPSTF) to endorse supplemental ultrasound (US) for all women after a negative MG, and MG sensitivity drops sharply with increasing breast density, a key driver of interval cancers. At that time, the authors highlighted the limitations in the data and the need for further research incorporating AI-assisted ultrasound. These gaps motivate the evaluation of AI-assisted ultrasound, including handheld POCUS as a primary or triage modality where mammography access is limited or the density of the breast can easily mask tumors.

METHODS

Literature Review and Comparative Evidence Interpretation:

We conducted a structured literature review of peer-reviewed articles published in the English language between January 2016 and October 2025. Our goal was to evaluate diagnostic performance, clinical utility, and implementation outcomes of AI-assisted ultrasound (US), including handheld, portable, and cart-based platforms, and compared this with mammography (MG) and its adjunctive or AI-augmented variants (MG+AI, MG+US).

Search and Selection Strategy

Databases searched included PubMed, Scopus, and Google Scholar, using combinations of the following terms: *"artificial intelligence," "machine learning," "deep learning," "ultrasound," "point-of-care ultrasound," "POCUS," "breast cancer screening," "mammography," "dense breasts," "AI-assisted ultrasound,"* and *"portable ultrasound."* Reference lists of eligible studies and key reviews were also manually screened to identify additional relevant work.

We included studies that met the following criteria:

1. Population: Adult women (≥18 years) undergoing breast cancer screening or diagnostic evaluation.
2. Intervention: AI-assisted ultrasound (handheld, portable, or cart-based systems).
3. Comparator**:** Standard mammography (MG), AI-assisted mammography (MG+AI), or adjunctive mammography with ultrasound (MG+US).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

4. Outcomes: Diagnostic performance measures, **including sensitivity, specificity, cancer** detection rate (CDR) per 1,000 screened, recall and biopsy rates, interval cancer rate, and stage distribution at detection.
5. Study Design**:** We chose published articles with study designs that fit the following criterias: randomized controlled trials, prospective cohort studies, real-world implementation studies, and systematic reviews or meta-analyses.

Studies published before 2016, non-comparative reports, and conference abstracts without full data were excluded.

Outcomes extracted: sensitivity, specificity, cancer detection rate per 1,000, recall/biopsy, interval cancers, and stage. Key sources include J-START and ACRIN 6666 (ultrasound adjunct), Radiology and AJR direct comparative analyses in dense breasts, a Radiology study of AI-POCUS triage in a low-resource setting, and a 2025 PLOS Digital Health systematic review on AI-enhanced handheld US.

Key Literature and Comparative Review:

A. Direct comparison of the following testing modalities:

1. Mammography (MG) vs MG+ Ultrasound (US) (non-AI)

- J-START RCT (40–49 years; n≈73k): MG+US ↑ sensitivity (91.1% vs 77.0%), ↑ early-stage (0/I) detection, and ↓ interval cancers (0.05% vs 0.10%**)**, with ↓ specificity (87.7% vs 91.4%) vs MG alone (Ohuchi et al., 2016).
- ACRIN 6666 (elevated-risk/dense): Supplemental US after MG increased cancer detection rate by approximately 3–4/1,000; more node-negative invasive cancers, but more false positives/biopsies. (Berg et al., 2012)

This has shown that when compared directly,mammography (MG)+ Ultrasound (US) > MG alone for detection and interval cancer reduction. The cost is lower specificity, and in this area, where AI can help calibrate decisions (Ohuchi et al., 2016)

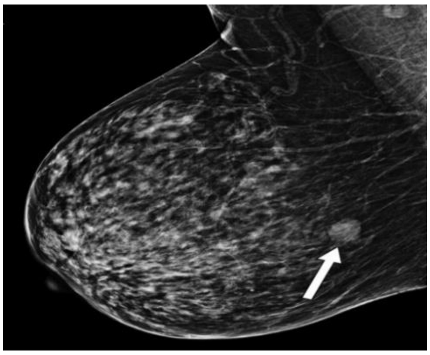

Figure 1. A 54-year-old woman with dense breasts.

A right mediolateral oblique screening mammogram shows oval circumscribed equal-density mass (*arrow*) in right lower outer quadrant. Mammography was assessed as BI-RADS category 2.

Lee, S. E., Yoon, J. H., Son, N. H., Han, K., & Moon, H. J. (2024). Screening in patients with dense breasts: comparison of mammography, artificial intelligence, and supplementary ultrasound. *American Journal of Roentgenology*, *222*(1), e2329655.

2. AI-Assisted POCUS (handheld) vs Standard Workflows

- Radiology 2023 (Mexico; portable US; palpable lumps): AI applied to portable US images achieved 96–98% per-woman sensitivity for cancer and could triage 38–67% of women with benign masses

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

away from tertiary referral. This directly tests POCUS+AI triage against the standard referral pathway (Berg et al., 2023).

- PLOS Digital Health 2025 (systematic review, handheld breast ultrasound): 34 studies; exemplary AI classification AUROC up to 0.976, but 79% high/unclear risk of bias and limited prospective screening validation; concludes feasibility with need for external, real-world validation (Bunnell et al., 2025).

Based on analysis of the published articles, the authors found that for the symptomatic triage, AI-assisted POCUS can achieve near radiologist sensitivity on low-cost devices and meaningfully, and reduce the numbers of referrals (Berg et al., 2023). This can provide an easier access and timely approach in underserved communities and low-resource settings (Berg et al., 2023).

Mammography(MG)+ artificial intelligence (AI) vs MG+Ultrasound (dense breasts**;** comparative literature analysis):

- AJR 2024 (dense breasts): MG+ supplemental US showed higher accuracy and specificity with lower recalls than MG+AI; adding AI did not improve outcomes beyond MG+US (Lee et al., 2024).
- Radiology 2024 (dense breasts): MG+AI ↑ specificity, but MG+supplemental US detected more node-negative early cancers missed by MG+AI (Ha et al., 2024)

The literature above shows that in dense breasts, adding ultrasound beats relying on MG+AI alone. This supported ultrasound-forward workflows (with or without AI) for early, node-negative detection (Lee et al., 2024).

4. MG alone vs MG+AI (large-scale real-world)

- Nationwide program (Germany; Nat Med 2025**;** n=463,094)**:** MG+AI increased cancer detection rate by 17.6% vs MG alone (6.7 vs 5.7 per 1,000) without having false-positive rates (Eisemann et al., 2025)
- Approximately 17.6% higher cancer-detection rate (CDR) with AI-supported reading versus standard reading at scale, with stable recall rates, is derived from the PRAIM (Prospective multicenter observational implementation) study reported by Eisemann et al. (2025) in *Nature Medicine.*

The findings confirmed that AI clearly improves MG-based programs, but dense-breast head-to- head analyses still favored ultrasound augmentation over MG+AI alone (Lee et al., 2024).

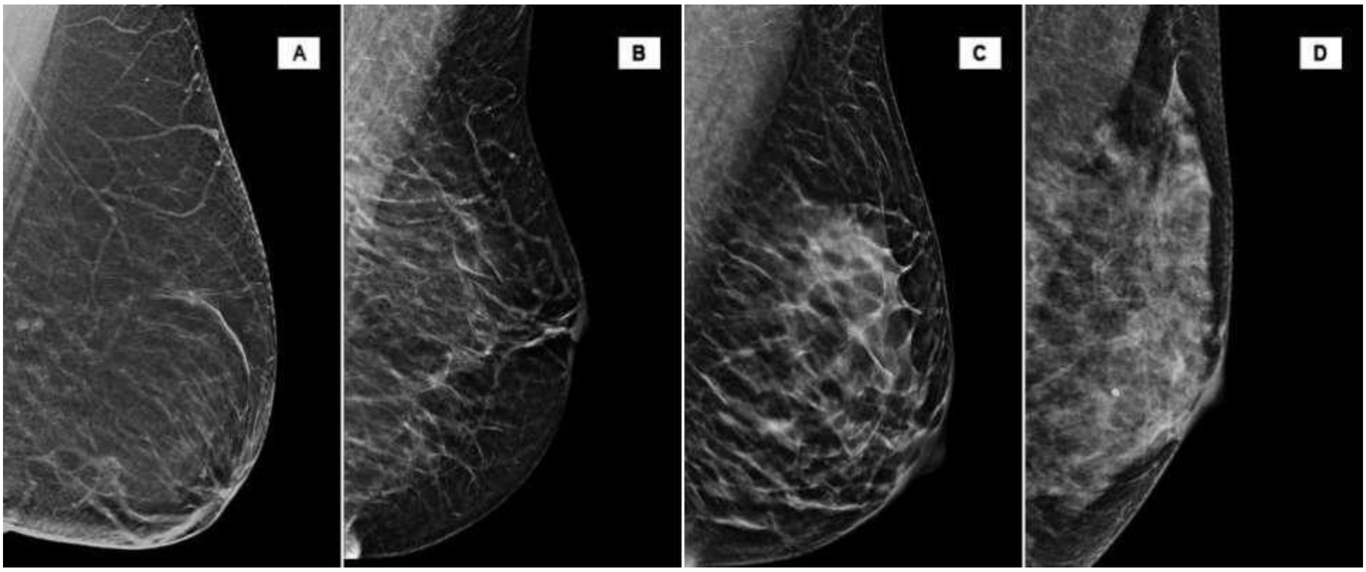

Figure 2: (A) The breasts are almost entirely fatty. (B) There are scattered areas of fibroglandular density. (C) The breasts are heterogeneously dense, which may obscure small masses. (D) The breasts are extremely dense, which lowers the sensitivity on mammography.

Lu, Z., Chow, L., & Li, B.(2025). Breast Composition: The Impact of Dense Breasts.URL: https://www.uclahealth.org/departments/radiology/education/breast-imaging-teaching-resources/birads/breast-composition-impact-dense-breasts

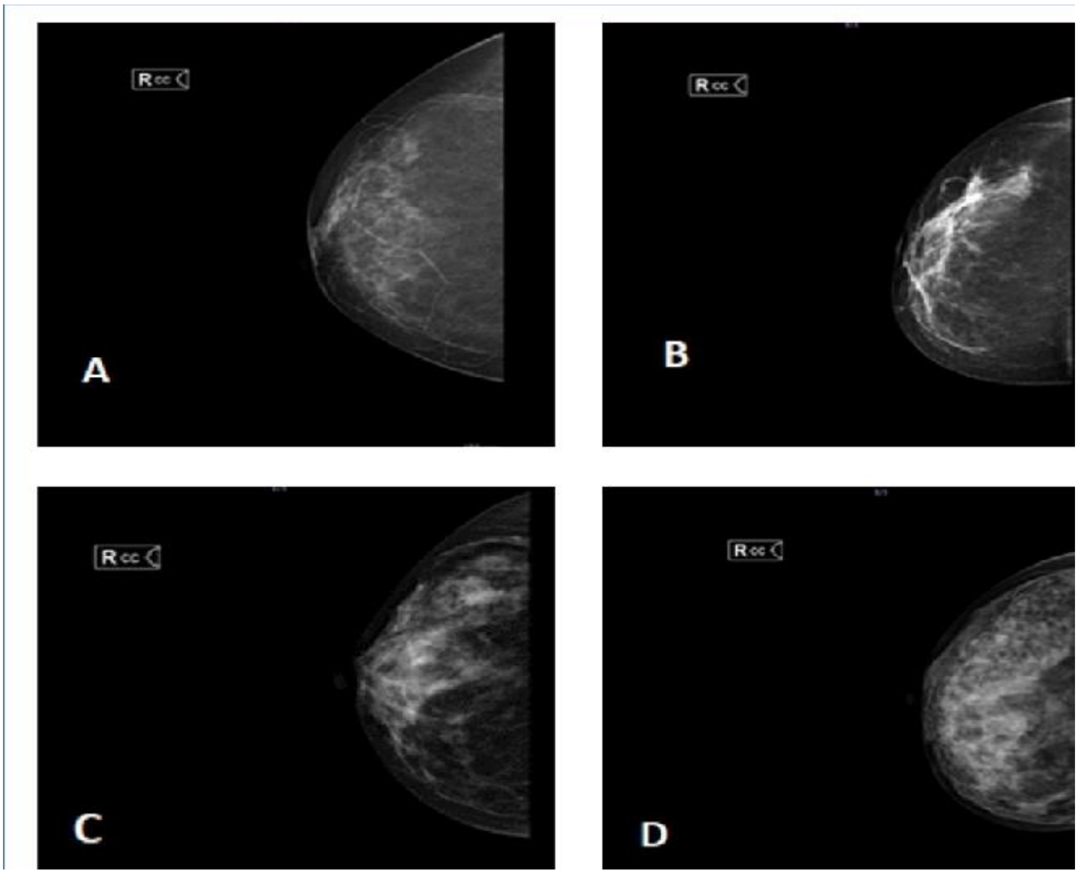

Figure 3: ACR BI-RADS classification for breast density.Rev. Assoc. Med. Bras. 69 (10). 2023

Şenkaya, A. R., Arı, S. A., Karaca, İ., Kebapçı, E., Öztekin, D. C., & et al. (2023). Association of polycystic ovary syndrome with mammographic density in Turkish women: A population-based case-control study. *Revista da Associação Médica Brasileira, 69*(10), e20230138.

B. Where AI-assisted POCUS Is (Likely) Superior

- Dense breasts: MG sensitivity can drop to approximately 30–45%; across head-to-head analyses, MG+US outperforms MG+AI for key endpoints (accuracy, recalls, early node-negative detection). A POCUS+AI-first triage is therefore plausibly superior to MG-only or MG+AI pathways for women with dense tissue, particularly when MG access is delayed or unavailable. (Lu, Z., Chow, L., & Li, B, 2025)

The findings revealed that palpable masses in low-resource settings that used AI-assisted POCUS on handheld devices attained approximately 96–98% sensitivity has the potential to avoid 38–67% of benign referrals, which provides both superior in timeliness and in access outcomes vs conventional MG, which relied on first referral models (Berg et al., 2023).

The analysis of the J-START trial was different as it revealed that using both mammography and ultrasound identified a greater number of cancers, but it also had lower specificity and led to an increase in unnecessary biopsies. POCUS is not intended to substitute mammography; instead, it has been investigated as a supplementary tool to mammography.

Given that POCUS can be more accessible and is considered more cost effective when compared to mammography, the integration of AI- assisted POCUS offers an advantage to patients in low resource settings and even to patients with dense breasts (Bunnell et al., 2025).

Table 1. Summary of Published Articles of AI Architectures Used in Breast Imaging

Architecture Type	Model Examples	Primary Application	Key Advantage	Limitations	Representative Studies
Convolutional Neural Networks (CNNs)	AlexNet, VGGNet, ResNet, Inception, DenseNet	Lesion detection, classification, segmentation	Excellent for spatial feature extraction from 2D mammograms and ultrasound	Requires large labeled datasets, limited interpretability	Yala et al., *Nature Medicine*, 2022; Shen et al., *EJR*, 2021
U-Net and Variants	U-Net, Attention U-Net, UNet++	Tumor segmentation and boundary delineation in ultrasound/MRI	Precise pixel-level segmentation with few parameters	Sensitive to noise and ultrasound speckle	Huang et al., *Front Oncol*, 2023
Recurrent Neural Networks (RNNs) / LSTM	Bi-LSTM, ConvLSTM	Temporal lesion tracking, dynamic contrast MRI	Captures temporal and contextual dependencies	Computationally intensive; limited adoption in real-time imaging	Ribli et al., *Radiology*, 2018
Transformer-Based Models	Vision Transformer (ViT), Swin Transformer	Whole-image classification, feature fusion with multimodal data	Superior global context understanding; adaptable to multi-omics integration	High computational cost; needs pretraining	Wu et al., *Med Image Anal*, 2023
Generative Adversarial Networks (GANs)	CycleGAN, DCGAN, StyleGAN	Data augmentation, synthetic mammogram/US generation	Improves dataset diversity and realism	Risk of generating artifacts or bias	Han et al., *IEEE TMI*, 2021
Graph Neural Networks (GNNs)	GCN, GraphSAGE	Multi-view or multi-omics relational analysis	Captures feature interrelationships beyond pixels	Early-stage application in breast imaging	Zhou et al., *Front AI Health*, 2024
Multimodal Fusion Models	CNN + Clinical Data + Omics	Integrative risk prediction and subtyping	Combines imaging with genomics/clinical features	Data alignment and normalization challenges	Hu et al., *NPJ Digit Med*, 2023

RESULTS

Baseline screening mammography (MG) remains the population standard, with a well-documented reduction in breast cancer specific mortality; however, sensitivity declines in dense breasts, increasing the risk of missed and interval cancers (Nicholson et al., 2024).

Adding artificial intelligence (AI) to MG improves detection at scale. In a nationwide implementation including 463,094 women, AI-supported MG increased the cancer-detection rate (CDR) by 17.6% (6.7 vs 5.7 per 1,000), with no statistically significant increase in recalls, indicating sensitivity gains without excess false positives (Eisemann et al., 2025).

Supplemental ultrasound (US) without AI improves detection for dense/elevated risk groups. Prospective evidence shows higher sensitivity and fewer interval cancers with MG+US versus MG alone (Berg et al., 2012; Ohuchi et al., 2016).

These Head to head comparisons in dense breasts suggest MG+US can outperform MG+AI on accuracy, specificity, and recall reduction, and detect more node-negative cancers (Lee et al., 2024).

In AI-enabled handheld ultrasound (AI-assisted ultrasound, the AI- assisted POCUS shows high per woman sensitivity and the ability to triage a sizeable share of benign findings away from referral; emerging systematic reviews support feasibility but emphasize heterogeneity and the need for prospective, real-world validation (Kim et al., 2024; Bunnell et al., 2025).

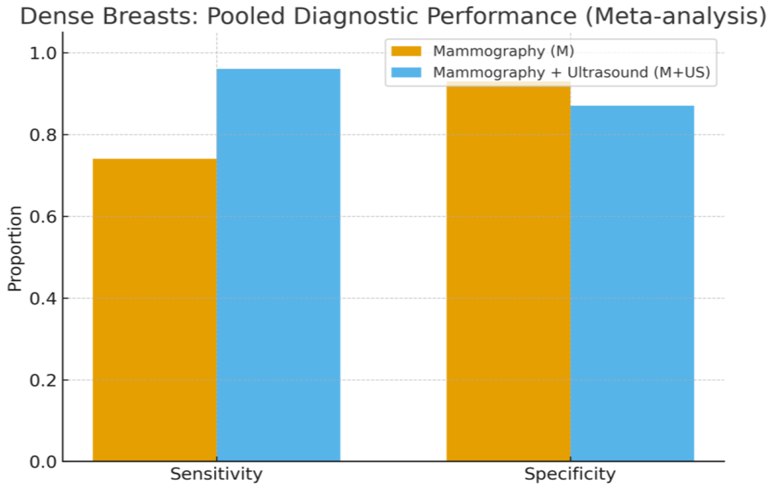

Figure 4. Sensitivity vs specificity graph comparison between Mammography testing and Mammography + Ultrasound. Showing less specificity of cancer diagnosis when mammography + ultrasound were combined

Ohuchi, N., Suzuki, A., Sobue, T., Kawai, M., Yamamoto, S., Zheng, Y. F., Shiono, Y. N., Saito, H., Kuriyama, S., Tohno, E., Endo, T., Fukao, A., Tsuji, I., Yamaguchi, T., Ohashi, Y., Fukuda, M., Ishida, T., & J-START investigator groups (2016). Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomized controlled trial. *Lancet (London, England)*, *387*(10016), 341–348. https://doi.org/10.1016/S0140-6736(15)00774-6

Analysis of Vertex AI Performance on breast imaging using Point-of-Care Ultrasound (POCUS)-Experiment analysis:

We used Google Cloud Vertex AI AutoML on 93 breast ultrasound images (74/10/9 train/validation/test). The BIONIC classifier achieved PR-AUC = 0.958, precision = 77.8%, recall = 77.8%, and log-loss = 0.292. At a 0.5 threshold, the test-set confusion matrix showed zero false negatives for cysts and 60% correct classification for solid lesions (40% mislabeled as cystic). These outcomes, on a well-annotated, curated dataset that included synthetic phantom images, show that a low-code AutoML workflow can deliver robust discrimination for foundational breast-lesion classification.

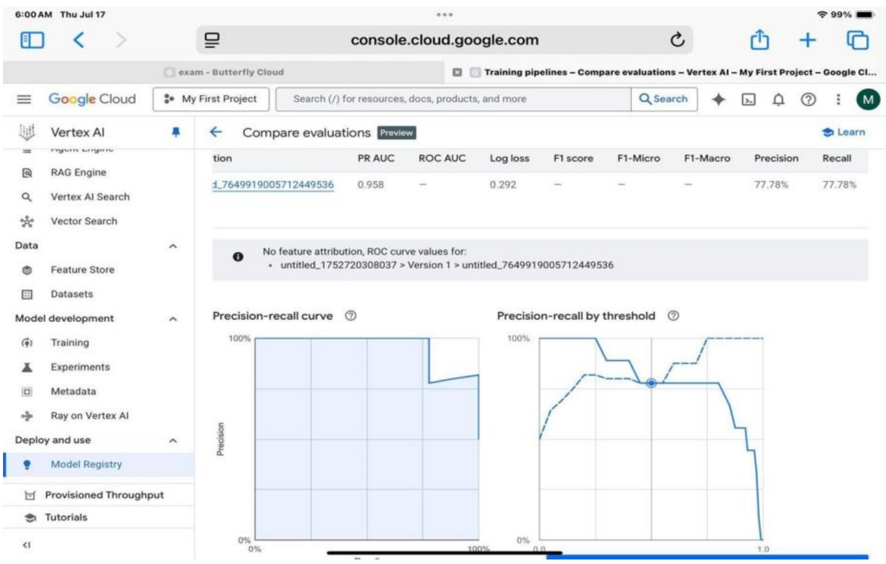

Figure 5. Precision–recall by curve & Precision–recall by threshold for the BIONIC model (Courtesy of Validus Institute Inc., 2025)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

Clinical interpretation. A high PR-AUC indicates strong rank-ordering of classes. Threshold tuning trades sensitivity vs precision: lowering the threshold increases sensitivity to solids (fewer missed solids) at the cost of more false positives; raising it does the opposite. The zero missed cysts supports AI-assisted POCUS triage to rapidly confirm simple cysts, while the missed solids suggest adding Doppler/SMI, elastography, or radiomics features and performing external validation on larger, more diverse cohorts.

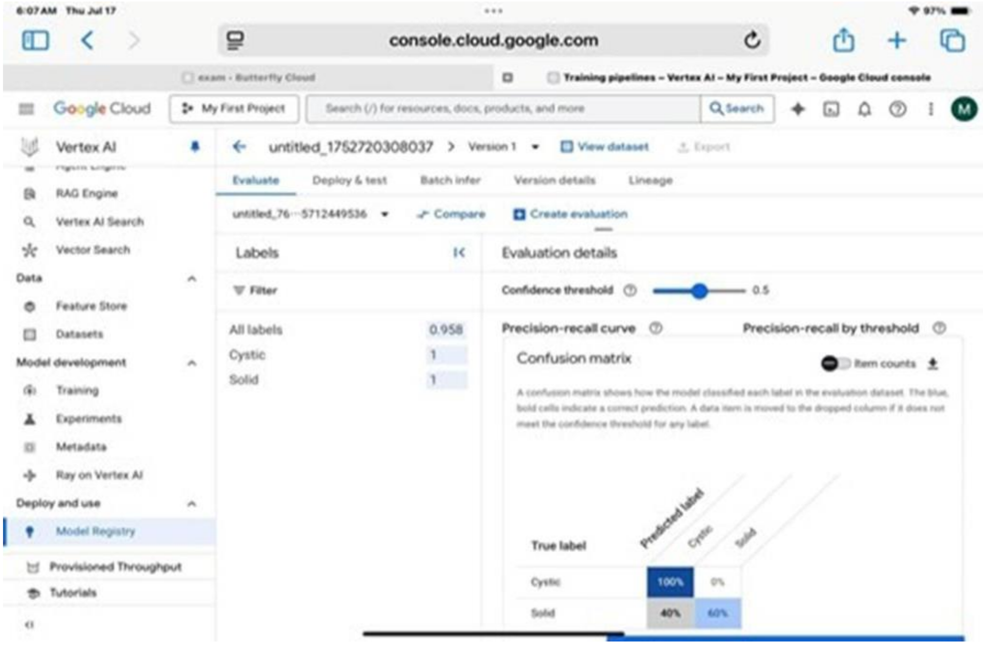

Figure 6. Confusion matrix for the BIONIC AutoML classifier (Courtesy of Validus Institute Inc., 2025)

Platform perspective. PR-based metrics are preferable for small, class-imbalanced medical datasets. Vertex AI's AutoML pipeline (data split, hyperparameter search, metric reporting) makes it straightforward to track precision–recall curves and the confusion matrix, underscoring that data quality and heterogeneity, rather than coding effort, are the main constraints on clinical robustness.

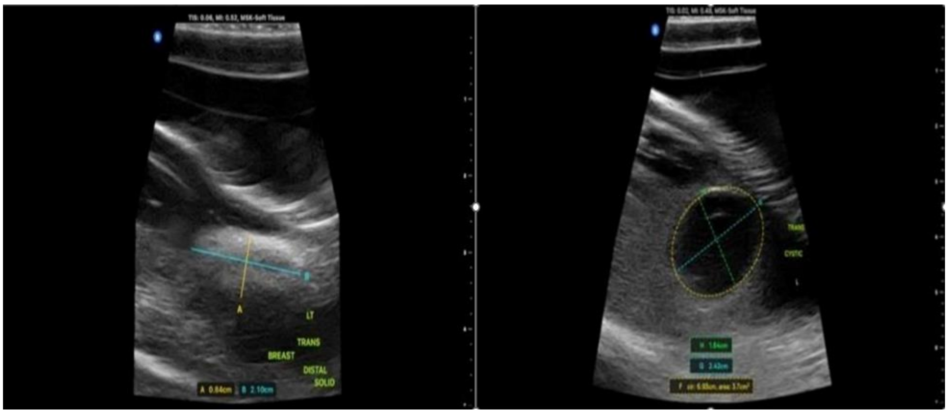

Figure 7. Ultrasound images using a synthetic tissue model: left, solid lesion; right, cystic lesion. Images were annotated and labeled by a trained researcher (Courtesy of Validus Institute Inc., 2025)

Oteibi, M., Khazaei, H., Abbas, K., Balaguru, B., Williams, A. R., & Etesami, F. (2025). Breast imaging and omics for non-invasive integrated classification (BIONIC). *International Journal of Research and Innovation in Applied Science*, *10*(8), 826–835. https://doi.org/10.51584/IJRIAS.2025.100800094

DISCUSSION

These results support a pragmatic, tiered approach to breast cancer detection. MG remains the standard of care for population screening, but performance degrades in dense tissue, leading to missed detection of malignant cells. This motivates thoughtful augmentation and planning (Nicholson et al., 2024).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

AI integrated with MG has shown that it can meaningfully raise detection without penalizing recall at a national scale, as shown by a 17.6% cancer detection rate (CDR) increase in real-world deployment (Eisemann et al., 2025). This pattern suggests AI can function as a consistent, "second reader," especially helpful in high-volume programs and regions facing workforce constraints (Eisemann et al., 2025). The graph above depicts the breast density and the AI prediction of the dense tissue. Check the graph below for details. Courtesy of Eisemann et al., (2025).

Ultrasound as an adjunct to MG has long demonstrated higher sensitivity and fewer interval cancers in dense or elevated-risk cohorts, but with lower specificity and more downstream workups (Berg et al., 2012; Ohuchi et al., 2016). These tradeoffs argue for risk-adapted use rather than blanket adoption (Berg et al., 2012).

When MG+AI is compared directly to MG+US in dense breasts, recent evidence indicates MG+US can deliver better accuracy and specificity with lower recalls and identify more node-negative disease—benefits that matter for stage shift (Lee et al., 2024). Still, MG+US is more operator-dependent and resource-intensive, whereas MG+AI scales efficiently across organized programs (Lee et al., 2024)

Below is a figure that will depict an ultrasound breast image captured by Butterfly IQ3+ POCUS. Annotation was done by the healthcare professional. Training the AI model to understand that this is a cystic mass with high precision, and confidence metric of 0.958 (Oteibi et al., 2025).

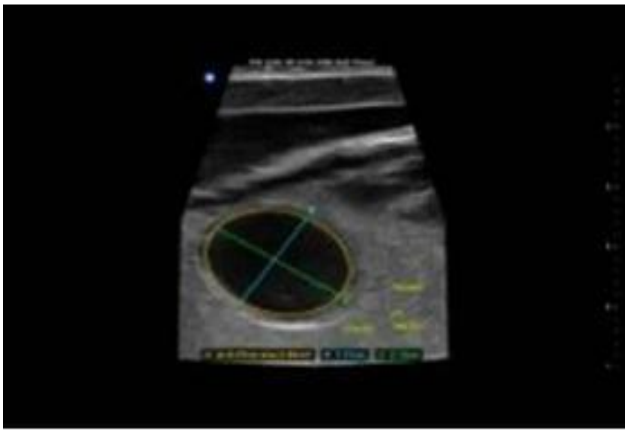

Figure 8. Cystic mass as it appears on ultrasound image. This image was annotated and labeled by a trained researcher (Courtesy of Validus Institute Inc. 2025)

Oteibi, M., Khazaei, H., Abbas, K., Balaguru, B., Williams, A. R., & Etesami, F. (2025). Breast imaging and omics for non-invasive integrated classification (BIONIC). *International Journal of Research and Innovation in Applied Science, 10*(8), 826–835. https://doi.org/10.51584/IJRIAS.2025.100800094

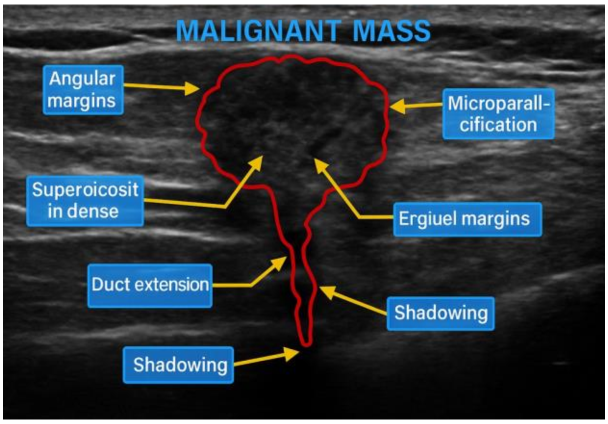

Figure 9. Malignant mass annotated as it appears on ultrasound. (Courtesy of Validus Institute Inc., 2025)

Point of care ultrasound (POCUS) has emerged as a complementary imaging modality that offers real-time, minimally invasive visualization of breast tissue. Handheld and portable ultrasound systems can be deployed in primary care clinics or community screening environments, allowing rapid assessment of palpable lesions or symptomatic patients. Historically, ultrasound has been constrained by operator dependency and limited standardization; however, the integration of artificial intelligence (AI) has begun to transform its diagnostic capabilities. AI-assisted POCUS systems leverage deep learning algorithms, particularly convolutional neural networks (CNNs) to automatically detect, segment, and classify breast lesions, potentially achieving diagnostic accuracies comparable to expert radiologists.

It is also important to note that, AI-assisted POCUS is emerging as a portable, lower-cost triage option that can decentralize early assessment, shorten time to referral for suspicious lesions, and reduce unnecessary referrals for benign disease. These attributes show promising benefits for people living in underserved settings. Yet, heterogeneity across devices, readers, and datasets means more prospective, multi-site evidence is still needed as published by Kim et al., 2024; Bunnell et al., 2025. Taken together, the comparative evidence supports mammography (MG) and ultrasound (US) for targeted screening in women with dense breasts or elevated risk, and supports AI-assisted POCUS for faster access, lower cost, and community-based triage. Ongoing research should prioritize harmonizing these modalities, assessing cost-effectiveness, and developing AI systems that adapt dynamically to individual patient risk and the imaging context.

AI-assisted Point of Care Ultrasound (POCUS) and mammography are both evolving modalities for early breast cancer detection. Review and recent meta-analysis evidence indicate that while AI significantly enhances both mammographic and ultrasound-based diagnostic performance, distinct strengths and limitations persist for each imaging approach in early detection scenarios.

Dignostic Performance Comparison:

As demonstrated above recent studies show that AI-assisted mammography improves accuracy and reduces false positives, with Area Under Curve (AUC) values up to 0.93 and diagnostic accuracy exceeding 88%. AI can augment less experienced readers, increase cancer detection rates, and reduce radiologist workload and reading time. However, supplemental ultrasound (with or without AI) helps detect more stage 0 and I node-negative early breast cancers, especially in women with dense breasts, despite resulting in more false-positives and unnecessary biopsies than mammography with AI alone.

AI-assisted ultrasound (including POCUS) shows notable improvements in sensitivity (up to 75%) and specificity (99%) compared to routine ultrasound, substantially outperforming historic controls and prior multi-center studies without AI. Comparative data highlight that AI-supported ultrasound can increase early detection rates, particularly in resource-constrained environments and for younger or high-risk populations.

Population and Workflow Implications:

- AI-mammography is better validated at scale, with evidence from hundreds of thousands of screened cases, making it well suited for population-wide screening where mammography is already established.
- For women with dense breast tissue, mammography (with or without AI) has a decreased sensitivity. AI-assisted ultrasound, or supplemental ultrasound, adds benefit to these groups, but with the risk of overdiagnosis and higher false positive rates.
- AI-POCUS is emerging as an adaptable tool for point-of-care triage, risk stratification, and adjunct detection in clinics with limited access to radiologic infrastructure—showing moderate discrimination (AUC up to 0.76) and the ability to risk-stratify lesions in real-time at the bedside.
- Large-scale meta-analyses confirm that AI models, as standalone devices, can outperform human radiologists in accuracy and efficiency for mammography, but integration with ultrasound is required to maximize detection, especially for early, small, or node-negative cancers.

We gather from all of this that AI models coupled with appropriate model training, workflow integration, and clinician upskilling. AI-assisted POCUS demonstrated non-inferior overall performance relative to mammography, with a clear sensitivity advantage among women with dense breasts, a subgroup in which mammographic sensitivity declines. Using AI-assisted POCUS with mammography as a" paired within design" approach helped strengthen internal validity by controlling for heterogeneities between patients. This approach also increased statistical power to detect clinically meaningful differences in sensitivity, specificity, and positive predictive value.

Given the high cost and limited availability of mammography in many settings, it is reasonable to favor POCUS as the screening test of choice, especially in resource limited settings, like rural areas, underserved communities to name a few, as this offers a more accessible and generally lower cost alternative, where these tools can enable non-expert clinicians to perform reliable imaging, increasing access to high-quality diagnostics.

LIMITATIONS

1. Heterogeneous study designs and endpoints.

Our synthesis spans randomized trials, prospective implementations, and observational cohorts with different primary endpoints (e.g., CDR, recall rate, interval cancer rate). Heterogeneity limits direct comparability and can distort perceived effect sizes across modalities (Berg et al., 2012; Ohuchi et al., 2016; Eisemann et al., 2025).

2. Population and density effects.

Many AI models are trained on datasets from specific populations (e.g., Asian or Western cohorts), potentially limiting performance across diverse ethnicities and breast densities. POCUS studies are often single-center or pilot trials, limiting external generalizability.

Results from dense-breast cohorts cannot be generalized to average-density populations, and vice versa. Age, baseline risk, and screening history differ across studies and influence sensitivity/specificity balances (Nicholson et al., 2024; Lee et al., 2024).

3. Operator and device dependence for ultrasound.

Ultrasound (including handheld) is sensitive to operator skill, protocol, and vendor differences, which affect reproducibility and specificity. This is particularly relevant when extrapolating MG+US findings to programs with varying training and QA standards (Berg et al., 2012; Ohuchi et al., 2016).

4. AI generalizability and version drift.

AI performance can vary by vendor, image acquisition, and case mix. Real-world AI outcomes depend on continuous monitoring, calibration, and handling of dataset shift; external validity may decline as populations or workflows change (Eisemann et al., 2025; Lee et al., 2024).

5. Outcomes beyond detection.

Most studies emphasize detection and recall; fewer report downstream outcomes such as stage distribution, treatment timeliness, anxiety, quality of life, and critically, mortality. Thus, the clinical significance of higher CDRs should be interpreted alongside potential overdiagnosis and overtreatment risks (Nicholson et al., 2024).

6. Absence of Standardized AI Evaluation Framework

Different studies use varying architectures, training pipelines, and performance metrics (AUC, F1, accuracy), making it difficult to uniformly compare AI efficacy.Few studies provide interpretability or calibration analyses of AI predictions.

7. Lower sensitivity to dense breast tissue

Mammography (with or without AI) has lower sensitivity for women with dense breast tissue. In such populations, AI-assisted ultrasound (or supplemental ultrasound) adds value but at the expense of increased false positive rates and potential overdiagnosis.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

Strengths

1. Performance strength

Large-scale meta-analyses confirm that AI models, as standalone devices, can outperform human radiologists in accuracy and efficiency for mammography, but integration with ultrasound is required to maximize detection, especially for early, small, or node-negative cancers.

2. Earlier triage

AI-POCUS holds promise in low-resource settings, real-time triage, and in facilitating earlier lesion identification where conventional radiology is limited.

3. Lower cost and easily accessible

As explained and shown in this manuscript the use of AI-assisted POCUS holds promise to

being more readily accessible and costs less to perform.

4. Reducing disparities in underserved communities

The use of AI-assisted POCUS offers a lower cost diagnostic tool when compared to mamography that can be accessible in underserved and rural communities. The use of

such accessible tools for early screening ensures that scientific advances lead to meaningful survival gains for women with breast cancer diagnosis rather than widening existing disparities.

FUTURE RESEARCH

1. Additional testing and evaluation

Additional testing and evaluation of these technologies will be needed, especially POCUS with AI-assisted technologies. Run cluster-randomized or stepped-wedge trials directly comparing POCUS with AI-assisted as a lower-cost testing option, especially in dense-breast subgroups, with pre-specified endpoints (CDR, interval cancers, false-positive biopsies, stage shift) and multi-round follow-up to quantify durable benefit (Lee et al., 2024; Ohuchi et al., 2016).

2. Patient-centered outcomes

Embed measures of anxiety, decisional conflict, time-to-diagnosis, and quality of life, alongside cost and workload, so “benefit” reflects what matters to patients, not detection alone (Nicholson et al., 2024).

3. Risk-adapted pathways

Develop algorithms that integrate age, density, family history, prior imaging, and polygenic risk to tailor modality choice where AI assisted POCUS testing can be accessed with ease in lower income communities, for triage in low-access areas) and to smooth referral pathways (Berg et al., 2012; Kim et al., 2024).

4. Equity-focused implementation science.

Evaluate how each pathway (MG+AI, MG+US, AI-POCUS) performs in underserved, rural, and minority populations, measuring access, follow-up completion, and treatment delays to ensure detection gains translate into survival gains from breast cancer diangosis(Nicholson et al., 2024; Kim et al., 2024).

5. Standardized reporting

Adopt common data elements (lesion-level labels, density categories, BI-RADS outcomes) and publish interval cancer rates and node-negative proportions routinely so programs can benchmark beyond CDR (Lee et al., 2024; Ohuchi et al., 2016).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue X October 2025

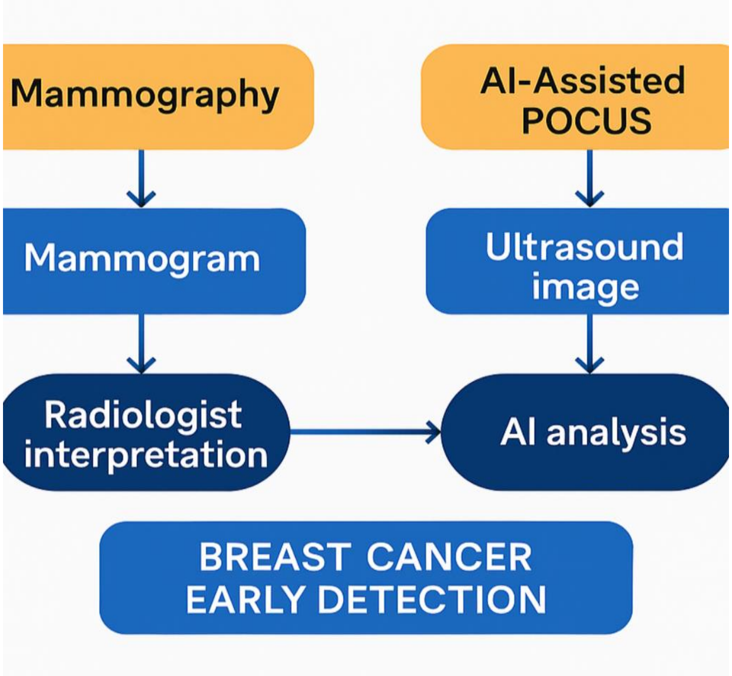

Figure 10. Workflow framework for breast cancer early detection using Mammography Vs. AI Assisted POCUS

CONCLUSION

Based on the analysis shared in this manuscript we found that mammography combined with handheld ultrasound (MG+US) can raise cancer detection rates, but this approach typically lowers specificity and increases unnecessary downstream benign biopsies. In contrast, when using AI-assisted POCUS showed promise as an efficient front-line screening strategy.

Consistent with prior published literature on ultrasound screening, AI assisted interpretation was associated with fewer interval cancers, which is consistent with earlier detection of clinically significant lesions.

Our conclusion, that embedding AI decision support at the bedside enables rapid risk stratification and safe rule out mechanism of benign findings. This reduces unnecessary specialty referrals and potentially avoidable biopsies without compromising safety. This real time triage shortens time to diagnostic resolution and increases throughput, creating opportunities for more cost effective care in settings with limited imaging capacity.

Collectively, these findings support AI-assisted POCUS as a practical screening modality, especially for individuals with dense breasts. With standardized image acquisition protocols, robust machine learning pipelines, and targeted clinician training, AI-assisted POCUS can accelerate diagnostic workflows, expedite treatment initiation for true positives, and improve patient experience. AI-POCUS holds promise in low-resource settings, real-time triage, and in facilitating earlier lesion identification where conventional radiology is limited.

As shared above, future studies should evaluate long-term results, program-level expenses,

and large-scale health system deployment, including prospective model drift monitoring. To reduce operator-dependent variability, particularly when handheld devices are scaled, we need to develop competency frameworks, image-acquisition protocols, and recurring proficiency testing. In accordance with established screening recommendations, it should also assess performance equity across subgroups and integration mechanisms with transparent and auditable clinical decision support. Additionally, we must establish post-deployment surveillance for AI: to monitor version upgrades, recall/CDR drift, and bias across subgroups; standardize calibration and reporting so that outcomes are similar across sites and suppliers.

ACKNOWLEDGMENTS

We acknowledge all our collaborators and team members who made this work possible at Validus Institute Inc. We also acknowledge Portland State University, Department of Electrical and Computer Engineering, led by Dr. Faryar Etesami for his guidance and support.

REFERENCES

1. Berg, W. A., Zhang, Z., Lehrer, D., Jong, R. A., Pisano, E. D., Barr, R. G., Böhm-Vélez, M., Mahoney, M. C., Evans, W. P., Larsen, L. H., Morton, M. J., Mendelson, E. B., Farria, D. M., Cormack, J. B., Marques, H. S., Adams, A., Yeh, N. M., Gabrielli, G., & ACRIN 6666 Investigators. (2012). Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA, 307*(13), 1394–1404. https://doi.org/10.1001/jama.2012.388
2. Bunnell, A., Valdez, D., Strand, F., Glaser, Y., Sadowski, P., & Shepherd, J. A. (2025). Artificial intelligence–enhanced handheld breast ultrasound for screening: A systematic review of diagnostic test accuracy. *PLOS Digital Health, 4*(9), e0001019. https://doi.org/10.1371/journal.pdig.0001019
3. Duffy, S. W., Agbaje, O., Tabár, L., Vitak, B., Bjurstam, N., Björneld, L., Myles, J. P., & Warwick, J. (2005). Overdiagnosis and overtreatment of breast cancer: Estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Research, 7*(6), 258. https://doi.org/10.1186/bcr1354
4. Eisemann, N., Bunk, S., Mukama, T., Baltus, H., Elsner, S. A., Gomille, T., Hecht, G., Heywang-Köbrunner, S., Rathmann, R., Siegmann-Luz, K., Töllner, T., Vomweg, T. W., Leibig, C., & Katalinic, A. (2025). Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nature Medicine*. https://doi.org/10.1038/s41591-024-03408-6
5. Giordano, L., von Karsa, L., Tomatis, M., Majek, O., de Wolf, C., Lancucki, L., del Moral, A., & Esperanza, M. (2012). Mammographic screening programmes in Europe: Organization, coverage and participation. *Journal of Medical Screening, 19*(Suppl 1), 72–82. https://doi.org/10.1258/jms.2012.012085
6. Kim, S., Fischetti, C., Guy, M., Hsu, E., Fox, J., & Young, S. D. (2024). Artificial intelligence (AI) applications for point-of-care ultrasound (POCUS) in low-resource settings: A scoping review. *Diagnostics, 14*(15), 1669. https://doi.org/10.3390/diagnostics14151669
7. Khazaei, H., Khajehee, B., Khazaei, D., Oteibi, M., Abbas, K., Balaguru, B. (2025). Leveraging Vertex AI for automated ultrasound image analysis: A comprehensive review. IJLTEMAS,14(8), 264–274. https://doi.org/10.51583/IJLTEMAS.2025.1408000033
8. Lee, S. E., Yoon, J. H., Son, N. H., Han, K., & Moon, H. J. (2024). Screening in patients with dense breasts: Comparison of mammography, artificial intelligence, and supplementary ultrasound. *American Journal of Roentgenology, 222*(1), e2329655. https://doi.org/10.2214/AJR.23.29655
9. Nicholson, W. K., Silverstein, M., Wong, J. B., Barry, M. J., Chelmow, D., Coker, T. R., … U.S. Preventive Services Task Force. (2024). Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *JAMA, 331*(22), 1918–1930. https://doi.org/10.1001/jama.2024.5534
10. Ohuchi, N., Suzuki, A., Sobue, T., Kawai, M., Yamamoto, S., Zheng, Y. F., … Ishida, T. (2016). Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): A randomized controlled trial. *The Lancet, 387*(10016), 341–348. https://doi.org/10.1016/S0140-6736(15)00774-6
11. Oteibi, M., Tamimi, A., Abbas, K., Tamimi, G., Khazaei, D., & Khazaei, H. (2024). Advancing digital health using AI and machine learning solutions for early ultrasonic detection of breast disorders in women. *International Journal of Research and Innovation in Applied Science, 11*(9), 590–601. https://doi.org/10.51244/IJRSI.2024.11110039
12. Oteibi, M., Khazaei, H., Abbas, K., Balaguru, B., Williams, A. R., & Etesami, F. (2025). Breast imaging and omics for non-invasive integrated classification (BIONIC). *International Journal of Research and Innovation in Applied Science, 10*(8), 826–835. https://doi.org/10.51584/IJRIAS.2025.100800094
13. Oteibi, M., Tamimi, A., Abbas, K., Tamimi, G., Khazaei, D., & Khazaei, H. (2025). Breast tumor

ultrasound: Clinical applications, diagnostic features, and integration with AI. *International Journal of Research and Innovation in Applied Science, 11*(11), 518–527. https://doi.org/10.51584/IJRIAS.2025.100800034

14. Paci, E. (2012). Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *Journal of Medical Screening, 19*(Suppl 1), 5–13. https://doi.org/10.1258/jms.2012.012077
15. Şenkaya, A. R., Arı, S. A., Karaca, İ., Kebapçı, E., Öztekin, D. C., & et al. (2023). Association of polycystic ovary syndrome with mammographic density in Turkish women: A population-based case-control study. *Revista da Associação Médica Brasileira, 69*(10), e20230138. https://doi.org/10.1590/1806-9282.20230138